



**Università
degli Studi
di Ferrara**

Ph.D. Course

in

Evolutionary Biology and Ecology

In cooperation with:

Università degli Studi di Parma

Università degli Studi di Firenze

Cycle XXXIII

Coordinator Prof. Guido Barbujani

**The green deal challenge:
exploiting biotic interactions from bacterial strains to
communities**

Scientific Sector (SDS): BIO/18

Candidate

Dott. Fagorzi Camilla

Supervisor

Prof. Mengoni Alessio

Co-supervisor

Prof. diCenzo George

Years 2017/2020

PhD dissertation

**The green deal challenge: exploiting
biotic interactions from bacterial strains
to communities**

Camilla Fagorzi

Supervisor: Professor Alessio Mengoni

Co-supervisor: Professor George DiCenzo

July 2021

Camilla Fagorzi

The green deal challenge:

Exploiting biotic interactions

from bacterial strains to communities

PhD dissertation, July 2021

Supervisor: Prof. Alessio Mengoni

Co-supervisor: Ass. Prof. George DiCenzo

University of Florence

Department of Biology

Via Madonna del Piano, 6

Sesto Fiorentino, 50019, Florence

Preface

Once upon a time in the rhizosphere, two organisms met. One was a bacterium, a small being but full of inventive and ready to face many new situations. The other one was a plant, a legume, able to adapt to many complicated situations, but always starving for nitrogen to grow strong. Little is known on their first date; they talked a lot and at a certain moment they touched each other. Sometimes they quarrelled and had to compromise, but now they are still together, the one eating sugar-candies on the sofa, and the other drinking nitrogen in the garden. Meanwhile, in the soil, thousands of other bacteria are living their existences, starting other symbiotic relationships with plants and animals, being harmful pathogens to conquer the world, mutating their lifestyle to face environmental changes, and evolving in a complex messy world.

This thesis tells this story with scientific coldness but describes it with all the passion that researchers dedicates to their work.

Three chapters reporting my PhD journey include the main investigated topics (first chapter), and two side (but scientifically and technically connected) researches (second and third chapter). All of them were essential for my formation and constitute a whole broad project aimed at analysing the interactions between bacteria and the environment (as nutrients, biotic community, symbiosis).

The first chapter focuses on the symbiosis between legumes and rhizobia, ranging from the most recent discoveries on rhizobia to improve agricultural practices in harsh soils to the choice of the best partnership (the rhizobium and its “beloved” legume). This part dissects the many evolutions of the partnership along the *Sinorhizobium/Ensifer* genus and its pangenome and going deeper with one of the model species for plant-rhizobium symbiosis, *Sinorhizobium meliloti*, by looking at the many ways (transcriptional responses) the “scent” of the host plant can be perceived. The second chapter looks at the small in the large, the study of the microbiota and its interaction with the environment, analysing the many signatures of microbial life in nature. The third chapter talks about sugar-candies and how complex is their use. Here, a study on the diauxie is presented, showing the many metabolic networks a heterotrophic marine bacterium, *Pseudoalteromonas haloplanktis*, can play with when the utilization of different carbon sources is the challenge.

In conclusion, this thesis reports a journey into the complexity of small simple things, the bacteria, and on their ways to cope with nutrients, signals and trophic interactions. Enjoy reading it as I enjoyed writing the stories reported here.

Contents

Chapter 1	1
1.1 An ecological problem.....	1
1.1.1 The nitrogen cycle.....	1
1.2 We are children of the Haber-Bosch synthesis.....	3
1.2.1 Air and water pollution.....	5
1.3 An ecological solution.....	7
1.3.1 Biological nitrogen fixation.....	7
1.3.2 Associative and Symbiotic Nitrogen Fixation.....	8
1.3.3 Rhizobia and legumes: the partnership.....	9
1.4 An overview on the nodulation process and symbiosis establishment.....	11
1.4.1 More than nitrogen fixation: plant growth promoting rhizobacteria as a multipurpose tool to improve bio-sustainable agriculture.....	16
1.4.2 A symbiosis model: <i>Sinorhizobium meliloti</i> and <i>Medicago sativa</i>	19
1.4.3 Not only symbionts.....	21
1.5 Biotechnological approaches to harness the potential of rhizobia.....	22
Chapter 2	24
2.1 The small in the large: exploring environmental bacterial communities.....	24
Chapter 3	26
3.1 Going complex: a short journey into bacteria metabolic complexity.....	26
Concluding remarks	28
Bibliography	30
Publications	41
List of Publications	42
Overall activities	44

Chapter 1

1.1 An ecological problem

1.1.1 The nitrogen cycle

Nitrogen is the most abundant element in our planet's atmosphere. Nearly 78% of the atmosphere is made up of nitrogen gas (N_2). We are talking about an essential and often limiting element for living organisms, eukaryotes and prokaryotes. It is a major component of amino acids, the key building blocks of proteins. It is also present in other important biomolecules, such as ATP and nucleic acids. Plant growth and crop production is limited by the availability of this element. The variety of processes in which nitrogen is involved requires its transformation from one chemical form to another. The nitrogen cycle includes all those transformations that nitrogen undergo moving between the atmosphere, the soil and living organisms. The term "fixation" refers to the process that leads to the conversion of nitrogen to a form that can be used in biological processes. Nitrogen fixation is a multipath routing process, whose result comes from both natural phenomena, as biological labour, lightning and volcanic activity and anthropological intervention, mainly due to the development of Haber-Bosh synthesis and Ostwald oxidation process (Bartholomew & Farrauto 2010; Vitousek *et al.* 2013).

Lightning and volcanic activity makes ~1% ammonia of the net nitrogen fixed per year, contributing to a minor part of the total nitrates generated (Guo *et al.* 2019; Igarashi & Seefeldt 2003). Atmospheric nitrogen is converted into ammonia and nitrate (NO_3^-) that enter soil with rainfalls.

Biological nitrogen fixation (BNF) converts di-nitrogen (N_2) into plant-usable form (NH_4^+ primarily). Nitrogen gas (N_2) diffuses from the atmosphere into the soil where species of bacteria convert it to ammonia (NH_3). The nitrogen molecule (N_2) is composed of two nitrogen atoms joined by a triple covalent bond, thus making the molecule highly inert and nonreactive. To break down a N_2 molecule 16 ATP molecules are required and further 12 ATP molecules are needed for assimilation and transport (Roley *et al.* 2019). 28 ATP molecules represent a high cost for organisms exerting BNF. These organisms obtain the required energy by oxidizing organic molecules. Non-photosynthetic free-living microorganisms must obtain these molecules from other organisms, while photosynthetic ones, such as cyanobacteria, use sugars produced by photosynthesis. Associative and symbiotic nitrogen-fixing microorganisms obtain these compounds from their host plants (Wagner 2011). Symbiotic plant partners, harbouring nitrogen fixing bacteria in their nodules, must provide 12 g of glucose to benefit 1 g of nitrogen (Rensing 2006). Nevertheless, BNF is less energetically expensive than the Haber-Bosh process that, to produce the same amount of nitrogen, requires temperatures of 400–500 °C and a pressure of ~200–250 bars (Gilchrist & Benjamin 2017).

Nitrogen cycle is completed by denitrifying bacteria that use nitrate (NO_3^-) instead of oxygen when obtaining energy, releasing nitrogen gas (N_2) to the atmosphere (Figure 1).

In the cycle of nitrogen, an important environmental impact is due by the industrial production of nitrogen compounds to fertilize crops. Increased nitrogen inputs into the soil have led to the so-called "green-revolution", with lots more food being produced to feed more people. However, the excess of nitrogen with

respect to plant demand leaches from the soil into waterways, contributing to eutrophication. Lakes, rivers and sea areas receive excess nutrients that stimulate excessive algae or plant growth and thus deplete the availability of oxygen in the water and causing other organisms' death. Nitrogen compounds, if balanced in the environment, support life and do not affect animals survival. Problems occur only when the cycle is not balanced. Moreover, the high energy demand of industrial synthesis of ammonia is still mostly supplemented by fossil fuel, then increasing indirectly the atmospheric CO₂ increase.

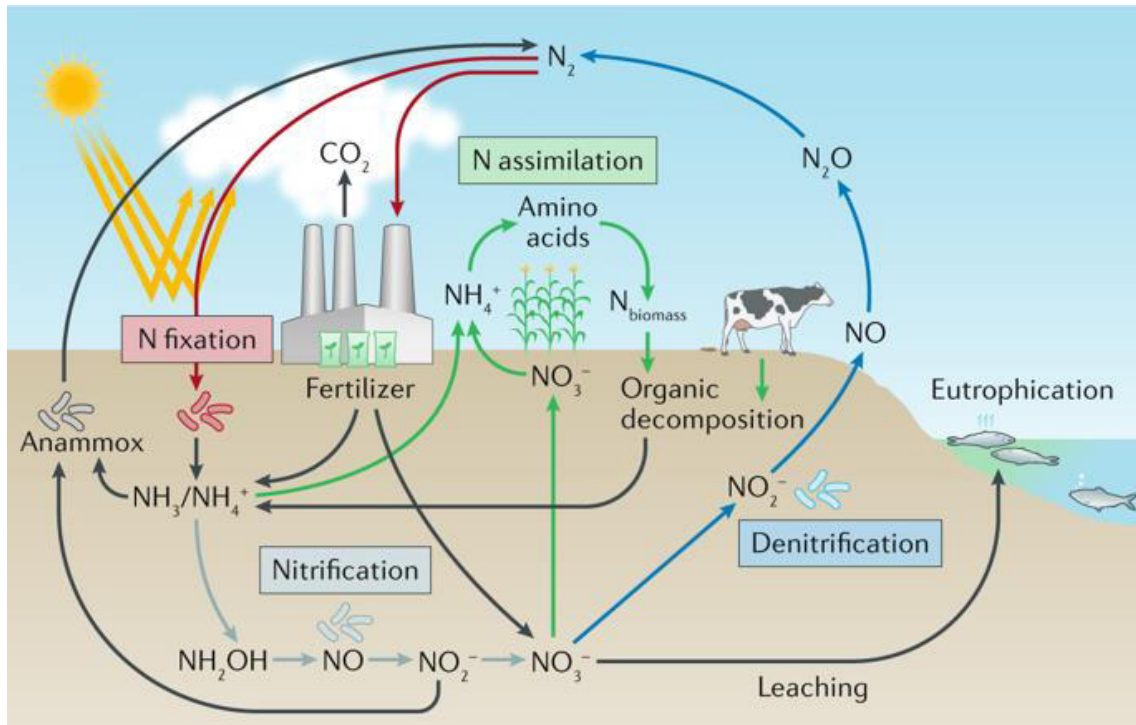


Figure 1. A schematic view of the nitrogen cycle (source: Lehnert et al. 2018)

1.2 We are children of the Haber-Bosch synthesis

Conventional agriculture has depended upon the Haber-Bosch synthesis process to produce the commercial fertilizers needed to grow most of the world's hybrid and inbred crops. Under high temperature and very high pressure, hydrogen and nitrogen (from thin air) are combined to produce ammonia (Figure 2). This conversion is typically conducted at pressures above 10 MPa (100 bar) and between 400 and 500 °C, as the gases (nitrogen and hydrogen) are passed over four beds of catalyst, with cooling between each pass (Gamble 2019).

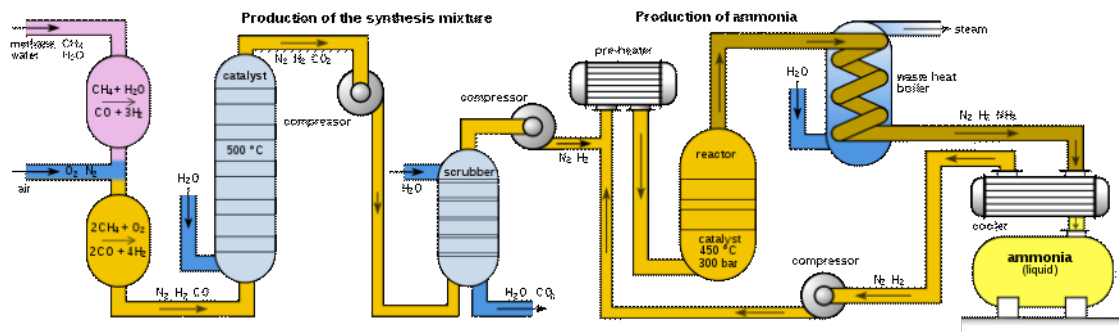


Figure 2. A schematic view of the Haber-Bosch process (source: *Wikipedia.org*)

A century after its invention, the process is still applied all over the world to produce more than 100 million tons of artificial nitrogen fertilizers per year (Figure 3 and 4). 1% of the world's energy supply is used for it (Smith 2002).

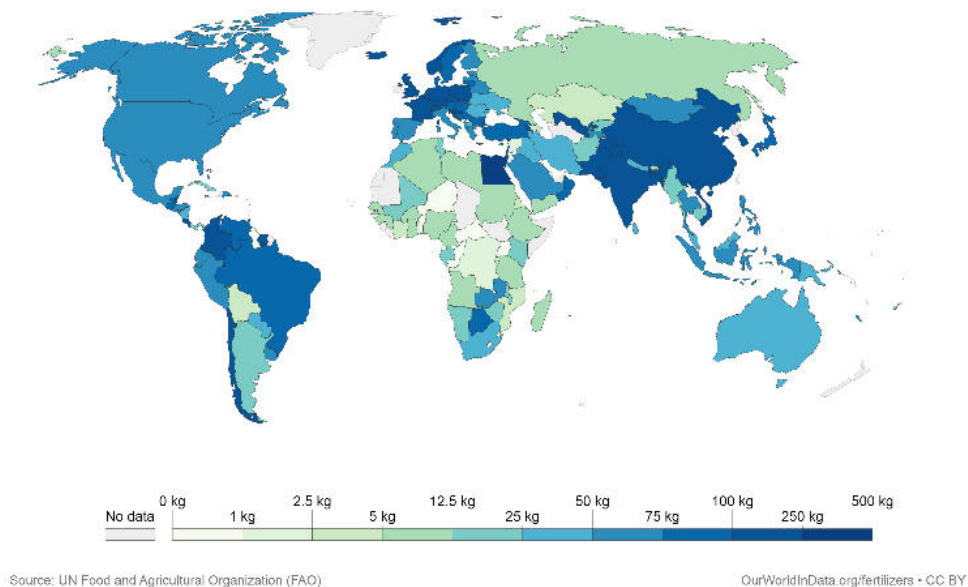


Figure 3. Nitrogen fertilizers used per hectare of cropland, 2017. Application of nitrogen fertilizers, measured in kilograms of total nutrients per hectare of cropland (source: *OurWorldData.org/fertilizers*).

Given that nitrogen is such a key element in all the living organisms, Vaclav Smil calculated how many people in the world population are dependent on its synthesis (Smil & Streatfeild 2002). He derived his calculations based on the use of nitrogen crop, livestock and human protein balances. Since nitrogen fertilizers provided around half of the nutrients in a harvested crop, he estimated that it supplies 40 percent of dietary proteins in the mid-1990s. In 2004, it sustained roughly 2 out of 5 people (Fryzuk 2004). In 2015, it already sustained nearly 1 out of 2; soon it will sustain 2 out of 3 (Figure 5). The Haber process served as the "detonator of the population explosion", enabling the global population to increase from 1.6 billion in 1900 to 7.8 billion by June 2021. Billions of people would never have existed without it; our dependence will only increase as the global count moves to ten billion people (Erisman *et al.* 2008).

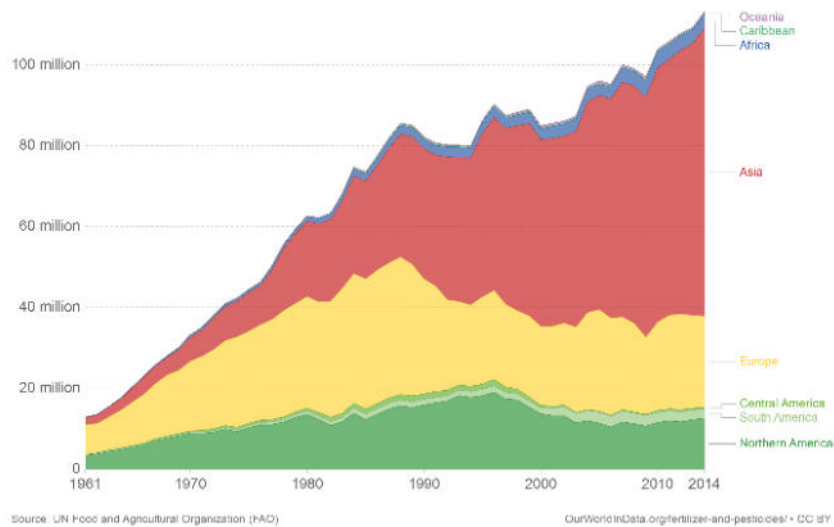


Figure 4. Nitrogen fertilizers production from 1961 to 2014. Global nitrogen fertilizers production, measured in tonnes of nitrogen produced per year (source: OurWorldData.org/fertilizers-and-pesticides).

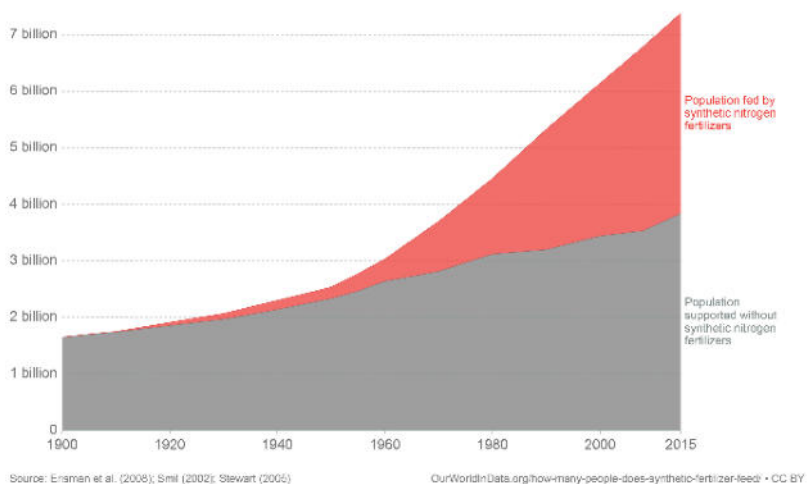


Figure 5. World population supported by synthetic nitrogen fertilizers. Estimate of the share of the global population which could be supported with and without the production of the synthetic nitrogen fertilizers for food production (source: OurWorldData.org/how-many-people-does-synthetic-fertilizer-feed).

Notably, there have been a number of additional contributors to productivity gains in agriculture throughout the 20th and 21st centuries, including crop breeding, irrigation, mechanization, and farm management techniques. Furthermore, the global food system is complex and geographically highly unequal: high-income countries have moved beyond the stage of aiming to meet basic nutritional requirements from food production, and now dedicate a large share of agricultural production to meat production (which is a much less efficient nitrogen converter).

Addressing this question partly relies on retrospective guesswork about whether, in the absence of synthetic nitrogen fertilizers, we would have managed to supply nitrogen via other methods. If Fritz Haber or Carl Bosch hadn't developed a method for transforming inert atmospheric nitrogen into reactive nitrogen plants could utilize, what is the likelihood that other solutions would have filled the gap?

A potential nitrogen source is that of organic wastes — nitrogen is supplied in many organic farming systems today in the form of animal manure. The issue is that these existed in limited supply: recycling nutrients, by definition, means you have a limited supply. Another solution would have been to greatly increase the production of nitrogen-fixing legume crops. Leguminous crops possess a unique ability to transform atmospheric nitrogen into reactive nitrogen in the soil. Growing these crops can therefore increase soil nitrogen sources over time. Unfortunately, despite the many nutritional and environmental benefits of legumes, they form only a small component of most peoples' diets. Additionally, they tend to be lower-yielding relative to cereals and other staple crops.

1.2.1 Air and water pollution

Physio-chemical process for di-nitrogen transformation comes with many consequences, including using fossil fuels for the energy needed during the process, the resulting carbon dioxide emissions and pollution from their combustion, and therefore adverse effects on human health (Vitousek *et al.* 1997).

The nitrogen cycle has been affected by the overuse of chemical fertilizers, leading to air, surface water and groundwater pollution. Reactions of nitric oxide with excited oxygen contribute to the destruction of ozone in the stratosphere (where these molecules serve to screen out dangerous ultraviolet light). Aquatic ecosystems are undergoing eutrophication due to the transfer of nitrogen fertilizer to freshwater (Figure 6). The evident "greening" of water column, triggered by the proliferation of microorganisms, especially algae, caused the decreasing of dissolved oxygen levels in bottom waters. This depletion results in a massive death of aquatic organisms and is responsible for the formation on the so called "dead zones", areas where aquatic life can hardly be detected (Diaz & Rosenberg 2008).

By the end of the 20th century, oxygen depletion of marine systems had become a major worldwide environmental problem, affecting more than 245,000 square kilometres of coastal regions with more than 400 systems that had developed hypoxia (Diaz & Rosenberg 2008; Kirchman 2021).

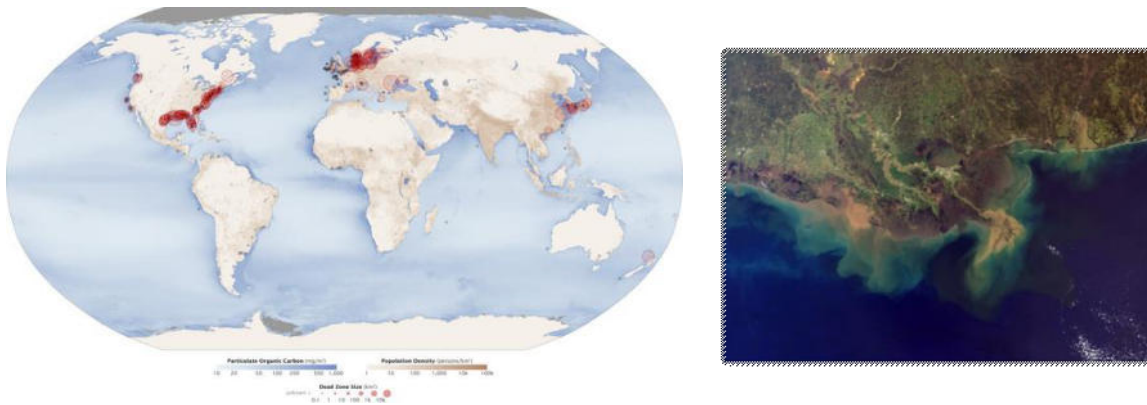


Figure 6. Coastal regions affected by eutrophication (highlighted in red on the map) (left); True-color image of Mississippi River sediment deposition into the Gulf of Mexico. Courtesy of NASA. (right)

Further expansion of dead zones will depend on how climate change affects water-column stratification and how nutrient runoff touches organic-matter production. From a prediction obtained from general circulation models, climate change alone will deplete oceanic oxygen both by increasing stratification and warming and changing rainfall patterns, so as to enhance discharges of fresh water and agricultural nutrients to coastal ecosystems (“Climate change 2013: The physical science basis-conclusions” 2013).

Industrialization weights on aquatic pollution. In pre-industrialized times, most coastal and offshore ecosystems never became hypoxic except in natural upwellings. For example, the shallow northwest continental shelf of the Black Sea, after a drastic reduction of fishing activities because of the lack of commercial species, hypoxia that was rising on the coast disappeared and ecosystems recovered. However, nutrient inputs increased again, and agriculture practices resumed with a consequent return to hypoxic conditions (Laurence 2006). The strategy to reduce dead zones will be to keep fertilizers on the land and out of the sea. For agricultural systems in general, methods need to be developed so that the nutrients cycle can be closed in a circle from soil to crop and back to agricultural soil (Tilman *et al.* 2001).

Can we overcome problems related to an impaired cycle of nitrogen moving towards a more sustainable agriculture?

1.3 An ecological solution

We know that nitrogen cycle should remain balanced and closed in the loop soil-crop-soil, and fertilizers should remain on the land, without filtering into the waters. Nature itself offers us the possibility to fertilize soils by exploiting the potential of biological nitrogen fixation (BNF) by harnessing the symbiosis between legumes and bacteria, the so called “symbiotic nitrogen fixation” or SNF.

1.3.1 Biological nitrogen fixation

Biological nitrogen fixation was discovered by Beijerinck at the beginning of the 20th century (Kamminga 1984). It is carried out by a specialized group of prokaryotes that utilize the enzyme nitrogenase to catalyse the conversion of atmospheric nitrogen (N₂) to ammonia (NH₃). Plants can readily assimilate ammonia to produce the above-mentioned nitrogenous biomolecules. Bacteria can carry out nitrogen fixation with different levels of autonomy, passing from organisms capable of producing ammonia from atmospheric nitrogen on their own, to organisms that need the presence of well-defined environmental factors, to others that ask to be housed within the physical structures of another organism. Prokaryotes able to perform BNF include aquatic organisms, such as cyanobacteria, free-living soil bacteria, such as *Azotobacter*, bacteria that form associative relationships with plants, such as *Azospirillum*, and most importantly, bacteria, such as rhizobia, that form symbioses with legumes (Hardy 1984; Shridhar *et al.* 2012) . These processes are summarized in Figure 1.

Species of *Azotobacter*, *Bacillus*, *Clostridium*, and *Klebsiella* represent some examples of heterotrophic bacteria that live in the soil and fix significant levels of nitrogen without the direct interaction with other organisms (free-living nitrogen fixers). These organisms find their own source of energy typically by oxidizing organic molecules released by other organisms or from organic matter decomposition. There are some free-living organisms (e.g *Thiobacillus* spp. and *Paracoccus denitrificans*) that have chemolithotrophic capabilities and can thereby utilize inorganic compounds as a source of energy (Stewart 1969).

Because nitrogenase can be inhibited by oxygen, free-living organisms behave as anaerobes or microaerophiles while fixing nitrogen. The scarcity of suitable carbon sources for these organisms makes their contribution to global nitrogen fixation rates generally considered minor. However, a study carried out in Australia on an intensive wheat rotation farming system demonstrated that free-living microorganisms contributed 20 kilograms per hectare per year to the long-term nitrogen needs of this cropping system (30-50% of the total needs (Vadakattu & Paterson 2006)).

Researches over the past decade has greatly expanded the knowledge of the diversity of organisms capable of carrying out BNF, but lots is still unknown about the rates and regulation of BNF. A recent evaluation of BNF is 58 Tg N yr⁻¹ (44 Tg N yr⁻¹ accounting for geological N), with a plausible range from 40 to 100 Tg N, estimated to be worth the equivalent of USD160–180 billion (Rajwar *et al.* 2013). Modern BNF is likely to be less than our pre-industrial estimation (40–100 Tg N yr⁻¹, with a preferred single estimate near 60 Tg N yr⁻¹) owing to land conversion, and perhaps to downregulation of BNF caused by increased deposition of anthropogenic reactive N. Harnessing BNF on ecosystem and regional scales is a key to understanding, predicting and managing the consequences of multiple components of anthropogenic global change (Finzi *et al.* 2015; Vitousek *et al.* 2013).

1.3.2 Associative and Symbiotic Nitrogen Fixation

Free-living fixation and associative nitrogen fixation are hard to distinguish from one another in the rhizosphere (Roley *et al.* 2019). When talking of bacteria capable to perform nitrogen fixation in association, but not in symbiosis, with other organisms, an example can be represented by species of *Azospirillum*, able to form close associations with several members of the Poaceae, including agronomically important cereal crops (such as rice, wheat, corn, oats, and barley). These bacteria fix appreciable amounts of nitrogen within the rhizosphere of the host plants. The level of nitrogen fixation is determined by several factors, including soil temperature (*Azospirillum* species thrive in more temperate and/or tropical environments), the ability of the host plant to provide a rhizosphere environment low in oxygen pressure, the availability of host photosynthates for the bacteria, the competitiveness of the bacteria, and the efficiency of nitrogenase (Reynders & Vlassak 1982).

The close association between microorganisms and plants, which can result in a symbiotic partnership, arises from the need for nutrients for microorganisms. The plant provides sugars from photosynthesis and in exchange for their carbon sources, microbes provides ammonia for its host.

By far, the most important nitrogen-fixing symbiotic associations are the relationships between legumes and rhizobia. Alfalfa, beans, clover, cowpeas, lupines, peanut, soybean, and vetches represent the most relevant legumes used in agriculture, with soybean that tops the rankings of the most cultivated legume (it is grown on 50% of the global area devoted to legumes and represent 68% of the total global legume production) (Figure 7) (Vance 2001).

Rhizobium-legume symbioses provide more than half of the world's biologically fixed nitrogen, and it was reported that rhizobial nitrogen fixation introduces 40–48 million tonnes of nitrogen into agricultural systems each year (Herridge *et al.* 2008; Smill & Streatfeild 2002).

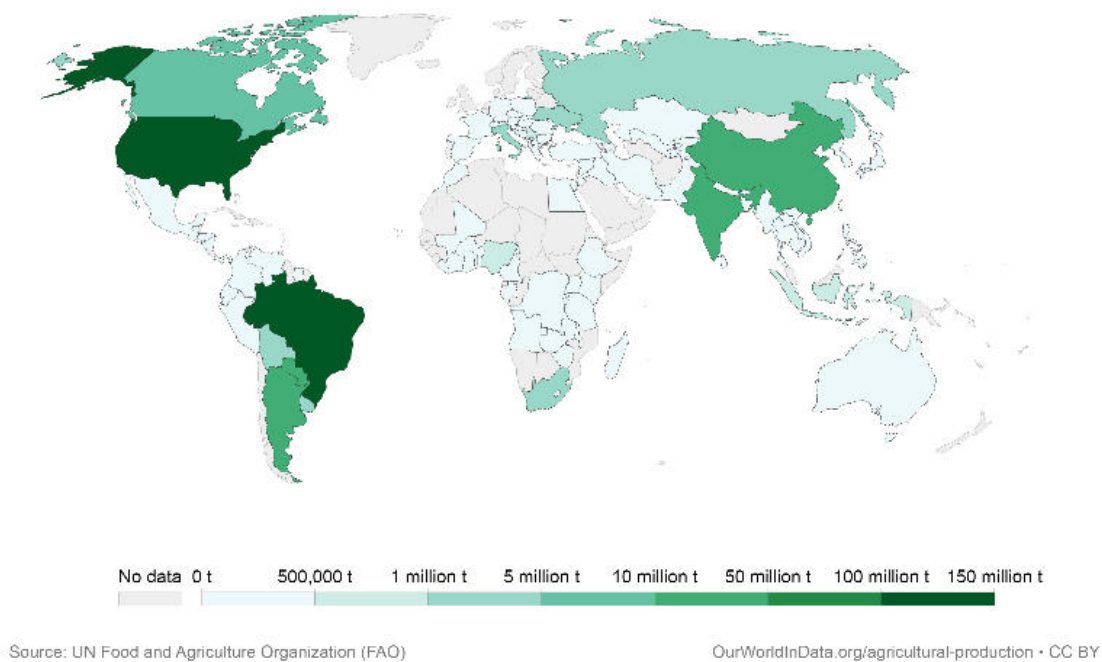


Figure 7. Soybean production in the world. Production is measured in tonnes (source: OurWorldData.org/agricultural-production).

1.3.3 Rhizobia and legumes: the partnership

The term “rhizobia” indicates a group of bacteria able to infect and colonize plant roots (*rhiza*, as root in Greek) and performing BNF. Rhizobia live as free bacteria in soil or plants (as commensal endophytes). When compatible leguminous plants are present rhizobia may form symbiotic nitrogen fixing associations with them. Rhizobia are a polyphyletic group of *Proteobacteria*, with all rhizobia identified to date belonging to two proteobacterial classes: *Alphaproteobacteria* and *Betaproteobacteria*. In the *Alphaproteobacteria*, rhizobial strains (alpha-rhizobia) are present in the genera *Sinorhizobium* (syn. *Ensifer*), *Rhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Methylobacterium*, *Devosia*, *Ochrobactrum*, *Aminobacter*, *Microvirga*, *Shinella*, and *Phyllobacterium*. In the *Betaproteobacteria*, rhizobia (beta-rhizobia) are present within the genera *Paraburkholderia*, *Cupriavidus*, and *Trinickia* (Checcucci *et al.* 2019; Estrada-de los Santos *et al.* 2019). While the beta-rhizobia are mainly found in association with tropical legumes, the alpha-rhizobia appear to be more widespread and make symbiosis with tropical to temperate legumes including pasture, tree, and grain legumes. The alpha-rhizobia have received more research attention than the beta-rhizobia; among the alpha-rhizobia, *Sinorhizobium* (syn. *Ensifer*) is likely the most studied genus, followed by the genera *Rhizobium* and *Bradyrhizobium* (Sprent *et al.* 2017).

Fabaceae (Leguminosae) is the third largest family of Angiosperms and includes more than 20,000 species including trees, shrubs, lianas and herbs, with different climatic typologies (Schmid *et al.* 2006). The establishment of this group in different environments (from coastal areas, to mountains, in rain forests and deserts, in equatorial areas and near the poles) is related to different adaptation strategies and chemical evolution. A part of the family’s evolutionary success is linked to their association with rhizobia. It is thanks to this partnership that legumes can colonize nitrogen-poor environments and store large amount of nitrogen compounds in their seeds (Benjamim *et al.* 2020). This symbiotic relationship has not always been such: according to the phylogenetic analysis based on conserved genes in different genera of rhizobia, there are evidence supporting the arise of rhizobia before their host plants, as free-living organisms, probably predators (Turner & Young 2000).

The relationship between plant and rhizobia changed probably 55 to 50 million years ago, when an important climate change, with a strong increase of temperature, favoured the carbon mobilization from sea sediments. Since nitrogen fixation requires a large amount of carbon, it was hypothesised that a driving force could have been an excess of carbon dioxide coupled with a deficit of combined nitrogen (Bowen *et al.* 2004; Sprent 2007). The co-evolution model of symbiotic association can be suggested also by the correspondence between both rhizobial and plant genes in the two phylogenesis (Dobert *et al.* 1994). According to the model proposed by Werner *et al.*, a “predisposition”, a common leguminous plant symbiotic trait, could have represented the starting point of nodulation process (Werner *et al.* 2014). From here, the accumulation of mutations could give birth to a stable nitrogen fixer bacterium or, on the other side, this trait could disappear (Figure 8).

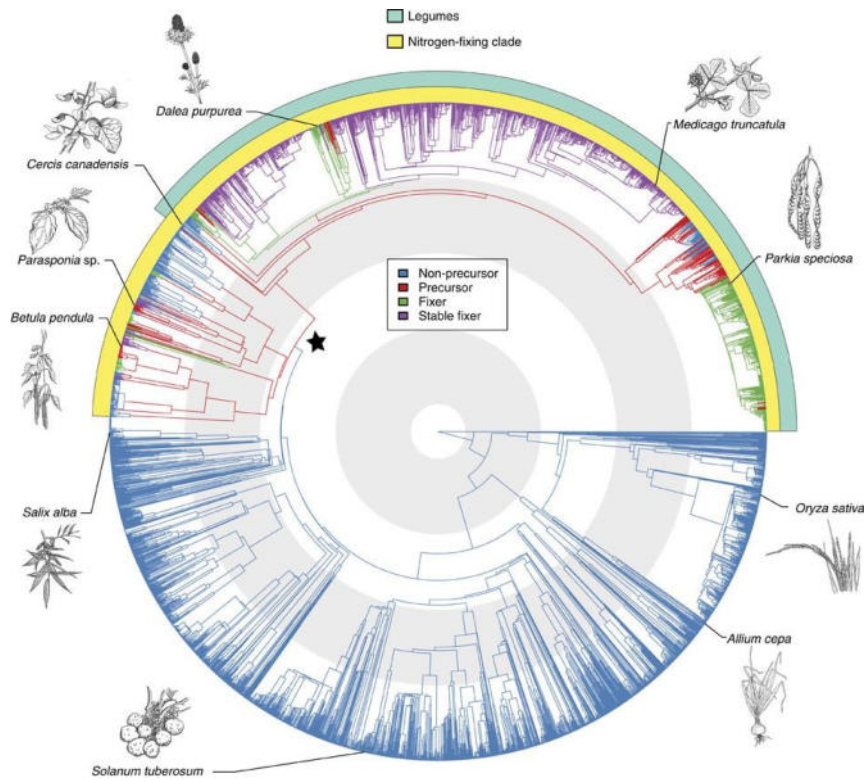


Figure 8. Angiosperm phylogeny of 3,467 species showing reconstruction of node states (Werner *et al.* 2014).

1.4 An overview on the nodulation process and symbiosis establishment

Nodulation and nitrogen fixation abilities are spread in a wide variety of α - and β - proteobacteria and leguminous plants (Chen *et al.* 2003; Zhao *et al.* 2021). In this thesis, our attention will be focused on the model of symbiosis between *Sinorhizobium meliloti* and *Medicago sativa*. *M. sativa*, commonly known as alfalfa, has an importance as a forage crop in many countries, because of its high proteins and vitamins content. Its relevance is also related to the possibility of this plant to be used as biofuel and to improve soil quality, by restoring nitrogen levels after previous depletion due to crops cultivation (Pini *et al.* 2012). Furthermore, *M. sativa* represents a source of compounds (e.g. saponins and flavonoids) with a potential in agriculture and pharmaceutical industries (Rafińska *et al.* 2017). Originally spread among temperate regions, the plant is known to be tolerant to a wide temperature range, although global warming reduces its productivity (Jacob *et al.* 2020).

Lots of details on the interaction and crosstalk between the rhizobium and the leguminous plant, still have to be elucidated. To date, with the information in our possession, it is possible to reconstruct 5 fundamental stages that run through the sequential steps of the symbiosis between the rhizobium and the legume, from the first approach to the exit from the nodule.

Mutual recognition

The symbiotic process starts from the mutual recognition between plant and bacteria present in the rhizosphere. The carbohydrates-rich area near plant roots attracts many microorganisms by signaling mechanisms in which root exudates play a fundamental role. Legumes secrete signals represented by flavonoid compounds (2-phenyl-1,4-benzopyrone derivatives, as for example luteolin) (Barreca *et al.* 2020). The rhizobial flavonoid receptor protein is NodD. The ligand binding activates the production and the secretion of the Nod factor, a lipo-chito-oligosaccharide molecule consisting of a chitin backbone, four to five N-acetylglucosamine units in length, with a lipid attached to the nonreducing end and host-specific modifications on the backbone (Miao *et al.* 2018). The specificity of host-symbiont specificity is due to the chemical structures of flavonoids and Nod factors acting in this first part of the recognition (Ankati & Podile 2019; van Dam & Bouwmeester 2016).

The last decade has changed the paradigm that the capacity to instigate nodule formation is solely determined by a canonical set of rhizobial nodulation (*nod*) genes. To date, three nodulation strategies have been identified in rhizobia: i) Nod, ii) type three secretion system (T3SS), and iii) non-Nod/non-T3SS strategies (Okazaki *et al.* 2013).

- i. In the Nod strategy, strain-specific lipo-chitoooligosaccharides called Nod factors (NFs) are produced under the control of nodulation genes (*nod/nol/noe* genes collectively referred to as *nod* genes). Their core structure is determined by the common *nodABC* genes. NFs are perceived by plant receptors (LysM-receptor like kinases) that activate the common symbiotic signaling pathway (CSSP).
- ii. In the T3SS strategy, T3SS effectors, whose nature and function are under investigation, activate CSSP components by bypassing NF recognition.
- iii. The mechanism of the third nodulation strategy is still unknown, but it involves neither nod nor T3SS functions. Nod-independent and T3SS-independent nodulation occurs via CSSP activation, since most CSSP components have been identified in *Aeschynomene evenia* - that is only nodulated via this strategy - some of which have been shown to be involved in symbiosis in this legume (Masson-Boivin & Sachs 2018).

Initiation of nodule formation (root hair curling and cortical cell division)

The Nod factor triggers the expression of plant developmental cascade for nodule formation, starting with root hair curling. Root hair curling is accompanied by cortical cells division. Together, these steps favor the invasion of the plant root cells and the formation of the infection thread by bacteria. After complete entry of rhizobia into plant cells the bacteria are found within vesicles called symbiosomes and here they differentiate into a specialized form, the bacteroid. This specialization, in terminal differentiation, is characterized by morphological changes (i.e. acquisition of an irregular form) and metabolic downregulation for many classical functions (as DNA replication) in favor of the upregulation of genes linked to the nitrogen fixation process (Wu & Wang 2019). At the same time bacteroid develops, infected root cells enlarge themselves and produce a swell on root surface up to the formation of a new organ: the root nodule (Gage 2004) (Figure 9).

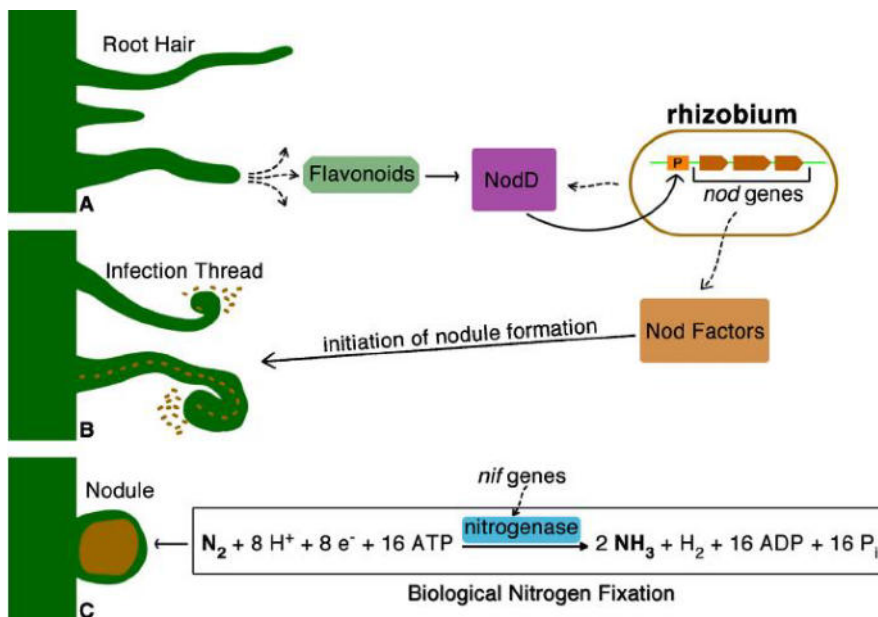


Figure 9. Schematic overview of the nodulation process and biological nitrogen fixation (source: Laranjo *et al.* 2014)

Nodule development

Two predominant nodule types exist in legumes. Determinate nodules, with a transient meristem, are spherical since cell divisions occur only during a short period of time. The final growth of the nodule is achieved by cell enlargement. In contrast, indeterminate nodules, with a persistent meristem, are cylindrical as they grow during the whole nodule life by increasing the number and the size of cells. Determinate nodules are generally formed in tropical legumes such as *Phaseolus*, *Vigna* and *Glycine*, whereas indeterminate nodules are found in temperate legumes such as *Medicago*, *Trifolium* or *Pisum* (Kazmierczak *et al.* 2020). Alfalfa produces nodules that display an indeterminate developmental program: a meristem at the tip of the growing nodule continually produces plant cells that are infected by bacteria, followed by differentiation of both the plant cells and the endosymbiont. As a result, all stages of development are present in a distal-to-proximal gradient within a single nodule (Wang *et al.* 2012) (Figure 10).

Beside the nodule types, the bacteroids possess also three different morphotypes named U, S and E. The U-morphotype (where U stands for Undifferentiated) corresponds to bacteroids with minor morphological modifications compared to the free-living bacteria and the ability to reverse from the symbiotic lifestyle to the free-living state. The S- and E-morphotypes correspond to large Spherical and large Elongated bacteroids,

respectively. These bacteroids cannot return to free-living life and are termed “terminally differentiated” (Kondorosi *et al.* 2013; Oono *et al.* 2010). This terminal differentiation is associated with the increase of permeability of the bacteroid membrane and of DNA content compared to free-living bacteria (Mergaert *et al.* 2006). The bacteroid differentiation is driven by the plant host via the production of a class of peptides called nodule-specific cysteine-rich peptides (NCRs) (Van De Velde *et al.* 2010). At the functional level, NCRs have intracellular bacteroid targets, including the cell division protein FtsZ and the chaperone protein GroEL, suggesting that NCRs interact with the bacterial metabolism to induce the terminal differentiation (Farkas *et al.* 2014; Lamouche *et al.* 2019). However, the way these targets are affected by NCRs is still unknown

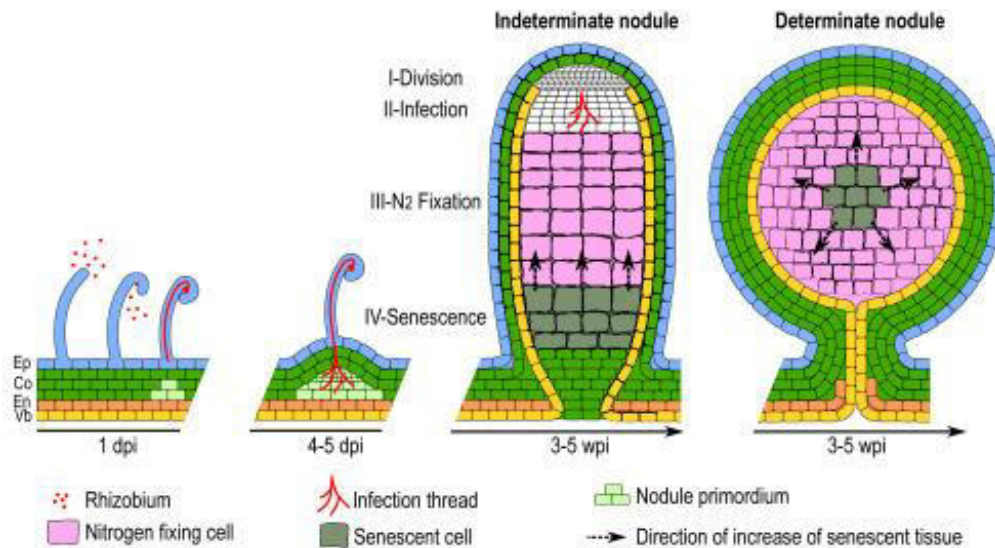


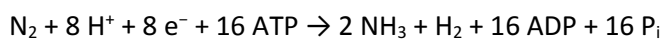
Figure 10. Schematic representation of the establishment and of the development of the legume-rhizobium symbiosis, and of the structure of mature indeterminate and determinate nodules. En, endodermis; Co, cortex; dpi, day post-inoculation; Ep, epidermis; Vb, vascular bundle; wpi, week post-inoculation (source: Kazmierczak *et al.* 2020)

Initiation of nitrogen fixation

Inside the nodule microaerobic environment, the bacteroid expresses the oxygen-sensitive enzyme nitrogenase, that catalyzes the conversion of atmospheric nitrogen to ammonia.

The nitrogenase complex consists of two protein components: the Fe-protein and the MoFe-protein. The nitrogen fixation process consists in a step by step protein association and dissociation and ATP hydrolysis for electrons delivery.

The reaction nitrogenase catalyzes is the following (Hoffman *et al.* 2014):



Energy required for the reaction is provided by the host plant as photosynthates (i.e. carbohydrates, sucrose) transported to the nodule via phloem. In the nodule, sucrose is cleaved by the sucrose synthase to UDP-glucose and fructose and the hydrolyzed compounds are metabolized by glycolytic enzymes to produce phosphoenolpyruvate (PEP), which is finally converted to malate (Ledermann *et al.* 2021). The dicarboxylates are transported into the bacteroid *via* dicarboxylic acid transport (Dct) system, and metabolized thanks to the tricarboxylic acid (TCA) cycle. Ammonium is the primary stable product of the reaction and it may diffuse through bacterial cytoplasm into plant cells or be incorporated directly into amino acids which are exchanged with organic acids (Lodwig *et al.* 2003; Lodwig & Poole 2003) (Figure 11). An experimentally based

computational model of metabolic interaction between the bacteroid and host legume has recently been developed, helping to clarify the relative importance of such metabolic exchanges over the plant growth (diCenzo *et al.* 2020).

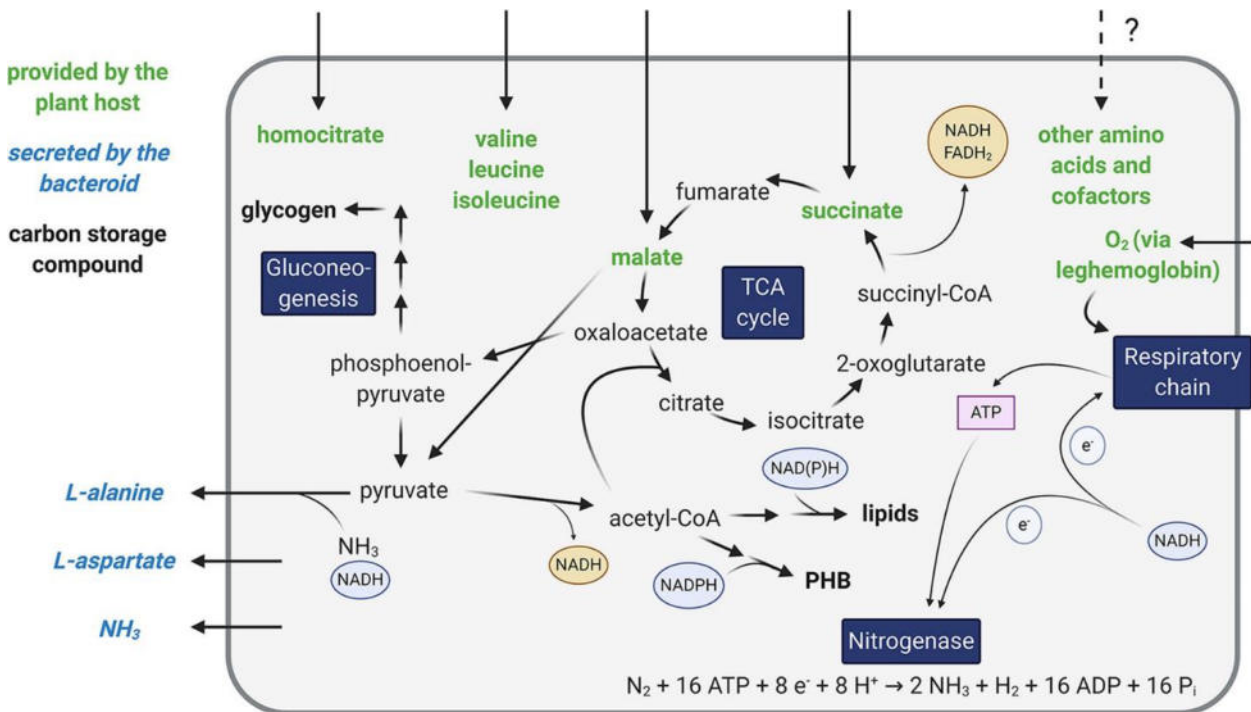


Figure 11. Central metabolic and transport reactions in bacteroids. Bacteroid metabolism is closely interdependent with the plant host because of bidirectional nutrient exchanges between the symbiotic partners (from Ledermann *et al.* 2021)

In the cytosol of the plant cells of the nodule, nitrogen is mainly assimilated into glutamine and glutamate thanks to the action of glutamine and glutamate synthetase. Fixed nitrogen can be transferred to asparagine, glutamine or purine derivatives depending on the species of legume (Vance 2000).

In this metabolic pathway, oxygen represents a limiting factor since it can inactivate the nitrogenase activity. Nevertheless, oxygen is fundamental for aerobic nitrogen fixing bacteria to obtain the large amount of ATP required. For these reasons, a system to regulate the oxygen concentration in nodules has evolved, with a major component being leghemoglobin. Leghemoglobin is a protein that underlines the strength of the linkage between rhizobia and plant. The heme group of the protein is synthesized by the microorganism and the globin is provided by the plant. By binding free oxygen with high affinity, leghemoglobin both facilitates the respiration in the bacteroid, and, at the same time, prevents the inactivation of nitrogenase by maintaining a low free-oxygen concentration (Li *et al.* 2018). Mutation or silencing of leghemoglobin biosynthesis results in inefficient symbioses (Ledermann *et al.* 2021). In oxygen-rich environments nitrogenase can also be sequestered in differentiated cells with morphological and biochemical characteristics that limit exposure of the enzyme to oxygen, protecting it from inactivation (Gage 2004).

Destiny of determinate and indeterminate nodules

Nodule senescence represents a strong agronomical interest through the possible improvement of N₂ fixation efficiency under adverse environmental conditions to maintain or even increase the yield of legume crops. Although the lifespan of nodules in most legumes is 10–12 weeks, their ability to fix N starts to decline 3–5 weeks post-inoculation (Puppo *et al.* 2005). As nodules and leaves are tightly connected in a reciprocal sink-

to-source relationship for carbon and nitrogen, any defect in photosynthetic activity will indeed impair the metabolic functions of the nodule associated with nitrogen assimilation and consequently provoke nodule senescence. Phenotypically, nodule senescence is characterized by the colour shift of tissues from pink (evidence of the presence of leghemoglobin and active nitrogenase) to green (determined by the degradation of leghemoglobin) (Rojonen 1970). Metabolically, senescence is represented by intense proteolytic activity responsible for high levels of degradations of proteins (Pladys & Vance 1993). Developmental senescence (also known as natural senescence) occurs because of organ aging, but exposure to various environmental changes also induces nodule senescence, e.g. when plants are submitted to drought, cadmium stress, photosynthesis inhibitors, prolonged darkness, defoliation, or treatments with, nitrate or abscisic acid. Alternately, senescence can result from a deficient recognition between the two partners. Depending on the type of triggering factor, the senescence process is rapid and implements variable responses, but a number of common typical features are found in most nodule senescence processes (Kazmierczak *et al.* 2020). In nodules of the determinate type, senescence develops radially, starting from the centre of the nodule and then progressively extending to the peripheral tissues in a few weeks. In nodules of indeterminate type, the N₂ fixation zone is constantly renewed and the senescence zone (zone IV) grows as the nodule grows; it increases from the root proximal zone toward the apical zone of the nodule. Because this zonation is not strictly defined, a senescence inter-zone is sometimes described between zone III and zone IV (Montiel *et al.* 2016; Puppo *et al.* 2005).

While from the point of view of plant physiology, the senescence of the nodule has been well explored, little is known about the release of symbiotic bacteria into the soil. Bacteroids in the determined nodules are not terminally differentiated, so when this kind of nodule dies, the bacteroids released in the soil are able to de-differentiate to the free-living form, enriching the soil of the substances stored during their symbiotic period. On the contrary, bacteroids in indeterminate nodules are metabolically and irreversibly different from free-living rhizobia. There terminally differentiated organisms die when the nodule they inhabit dies, but undifferentiated forms maintain the possibility to be released from the nodule, giving rise to a new rhizobial population in the rhizosphere (Denison 2000).

Host sanction

After nodule organogenesis is underway, legumes can respond to nodules that house ineffective rhizobia (strains that do not fix nitrogen) and reduce bacterial within-nodule growth rates, a trait known as “sanctions” (Regus *et al.* 2015; Sachs *et al.* 2010). But how legumes detect and sanction ineffective strains is mostly unknown. Models predict that legumes target ineffective rhizobia at the level of the whole nodule, hence that sanctions are modulated dependent upon the total amount of fixed nitrogen provided by individual nodules whether they are clonal or contain mixed strains infection. But growing evidence suggests that hosts can sanction ineffective rhizobia even when individual nodules are co-infected by a mix of effective and ineffective strains (West *et al.* 2002b). Furthermore, legumes can sanction ineffective strains at a cell-level, with the host plant inducing programmed cell death of cells that house ineffective rhizobia (Masson-Boivin & Sachs 2018; Regus *et al.* 2017).

For rhizobia, nodulation offers substantial fitness benefits that appear unachievable in the highly competitive soil and rhizosphere environments outside of the host. Conversely, plant hosts benefit from nodulation only under certain conditions, such as when soils are nitrogen poor and contain compatible, nitrogen-fixing rhizobia. Among legume taxa, there is wide variation in their specificity for restricting nodulation, often dependent on the production of host-specific flavonoids and host root receptors that recognize specific rhizobia (Kawaharada *et al.* 2017; Liu & Murray 2016; Sachs *et al.* 2018).

Plant costs to nodulation (carbon, C) are predicted to be a linear function of the number of nodules formed (Nod#) with a slope of m (cost per nodule):

$$f_C = m \cdot \text{Nod\#}$$

Plant benefits from nodulation (nitrogen, N) are predicted to be a negative exponential function:

$$f_N = \alpha (1 - e^{-B \cdot \text{Nod\#}})$$

with diminishing returns that reach an asymptote at α and diminish at a rate corresponding to B (Figure 12, a). Net benefits of nodulation can be calculated by subtracting the cost from the benefit functions. The net benefit function for nodulation is unimodal, increasing with the formation of nodules (zone of cooperation) until the optimal number of nodules is reached $(N-C)^{\text{max}}$, and above which additional nodules reduce the host benefit (zone of conflict). If too few or too many nodules are formed, the host does not acquire the net minimal benefit to set seed (i.e. $< (N-C)^{\text{min}}$) (Figure 12, b). Host fitness (i.e. seed set) varies with the number of nodules formed (Figure 12, c).

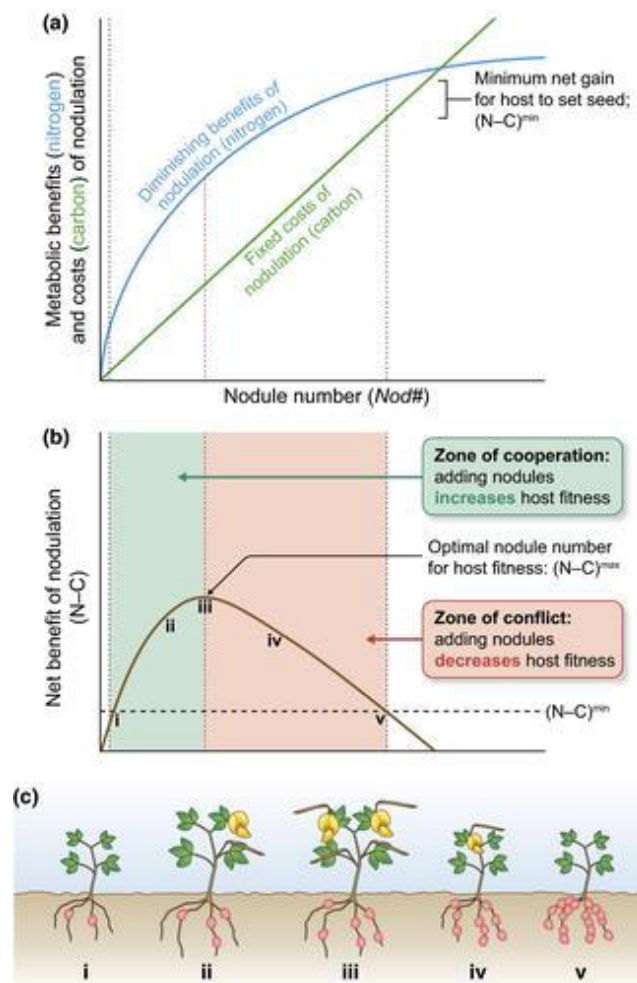


Figure 12. Legume and rhizobia conflict over nodule number. (a) Costs vs benefits of nodulation (b) Net benefits of nodulation (c) Host fitness (i.e. growth, seed set) (Source: Sachs et al., 2018).

Legumes can also induce nodule senescence when the costs of symbiosis become too great, as in case of dark stress or when nodules are infected by rhizobia that fix negligible nitrogen (i.e. Fix⁻) (Berrabah *et al.* 2015; Vauclare *et al.* 2010). A key step during nodule senescence is the neutralization of the peribacteroid space surrounding the otherwise acidic environment of the symbiosome, whose homeostasis is required to import host resources (Pierre *et al.* 2013).

1.4.1 More than nitrogen fixation: plant growth promoting rhizobacteria as a multipurpose tool to improve bio-sustainable agriculture

In general, microorganisms have become popular models for addressing sociobiological questions related to exchange of nutrients and communication between different organisms (Checcucci *et al.* 2017b). To study of mutualistic interactions, microbial systems represents an optimal model, due to their readiness of manipulation, their short generation times, and the simple methods for tracking of resources (West *et al.* 2002a).

It is becoming increasingly clear that the development, growth, and health (in one word, all functions) of macroorganisms are influenced by the complex microbial communities they host that shape their ecology and evolution (Foster *et al.* 2017; Selosse *et al.* 2014). Biology is indeed undergoing a paradigm shift, where individual phenotypes are seen as a result of complex interactions resulting from the combined expression of the host and associated microbial genomes, leading to the popularization of notions of the holobiont and the hologenome (Bordenstein & Theis 2015; Simon *et al.* 2019). In a global picture of the soil, plants and their related microbiota can be considered holobionts, complex systems ruled by interdependent and composite interactions (Sánchez-Cañizares *et al.* 2017). Plants influence the composition of their rhizosphere microbiota through the production of root exudates, contributing to the positive selection of plant-growth promoting (PGP) and beneficial bacteria (Edwards *et al.* 2018; Hartmann *et al.* 2009). In the past three decades, eco-sustainable agronomic practices have been employed to replace chemical fertilizers and pesticide-based agriculture by exploiting beneficial microorganisms as biofertilizers (Alori *et al.* 2017; Hartmann *et al.* 2009). The high economical relevance of the symbiotic interaction has promoted the production of commercial rhizobia inoculants (already widely used in agriculture), since this microbial interplay can provide a direct benefit for the host plant. Inoculation of leguminous crops with rhizobia strains is known to considerably increase crop yield, mediated largely through symbiotic nitrogen fixation (Checcucci *et al.* 2017b). However, the function of these bacteria in the rhizosphere can be more than their role in the nitrogen cycle: their role can be also those of plant-growth promoters, by helping the plant surviving to stresses.

The initial step in recognition of a single microbe occurs, in the plant, at the cellular level. Biotic and abiotic stresses detected by the plant are transformed into signals and thus transmitted to downstream intracellular and intercellular signalling networks by receptor-like kinases. Subsequently, reactive oxygen species and calcium ions (Ca^{2+}) transmit these stimuli to downstream signalling components in the local and systemic immune system in plant cells. The transferred transducers activate kinase cascades such as mitogen-activated protein kinases and accumulate plant hormones, regulators of plant stress response, such as abscisic acid, salicylic acid (SA), jasmonic acid (JA), and ethylene (Xu & Huang 2017). Survival of plants in the natural environment depends on their interaction with a complex and dynamic community of microbes. Root metabolism and root exudates are the main physiological factors used by plants to control microbial growth (Chaparro *et al.* 2014). Root exudate composition (e.g. sugars, organic acids, secondary metabolites, mucus-like polymers) is controlled by the plant genotype, developmental stage, seasonal effects and stress level (Phillips *et al.* 2008; Sasse *et al.* 2018). In a study conducted on *Arabidopsis* plants, different microbial recruitment in the rhizosphere occurred if present mutations in the plant impacting on the biosynthesis of JA and SA, indicating that these hormones modulate microbial community composition directly or indirectly through root exudates (Carvalhais *et al.* 2015; Lebeis *et al.* 2015). This suggest that the crosstalk between plant and bacteria and consequently and rhizospheric recruitment are finely regulated.

Although the image of rhizobia is linked to the formation of nodules in leguminous plants and their ability to help survival in case of nutrients depletion, their role in the rhizosphere can be extended thanks to their ability to more generally increase host plant performance under harsh conditions (as for instance drought,

osmotic stress, heavy-metals-induced oxidative stress) (Bellabarba *et al.* 2019; Bruning & Rozema 2013). Problems as soil nutrients deficiency, drought, and soil contamination are becoming a worldwide problem for crop productivity. In case of heavy-metals, phytoremediation is a low-impact, low-cost plant-based technique to remediate polluted soils (Chibuike & Obiora 2014; Fagorzi *et al.* 2018). Depending on the plant species used, it can be described as phytoextraction (accumulation of toxic compounds in the plant tissues) or phytostabilization (adsorption through the roots and conversion into harmless compounds). In legumes, associated rhizobia can promote chemical transformation and the chelation of heavy-metal compounds (Kong & Glick 2017). This is due to quite general plant growth promoting (PGP) features of some rhizobial strains, such as indoleacetic acid and siderophore production, phosphate and zinc solubilization, and the synthesis of 1-aminocyclopropane-1-carboxylate (ACC) deaminase. However, rhizobial strains with heavy-metal resistance phenotypes have been also isolated and characterized (Chaintreuil *et al.* 2007; Mahieu *et al.* 2011; Mengoni *et al.* 2010; Porter *et al.* 2017). The effect of rhizobia with such PGP abilities is more evident under stressful conditions, since they alleviate plant stress (for instance through the reduction of ethylene production) and increase nutrition (increasing root growth and fixing nitrogen) (Checcucci *et al.* 2017a; Li *et al.* 2014b; Lu *et al.* 2017; Wani *et al.* 2007).

Therefore, plant growth and agricultural yield are related not only to the plant genotype and the soil conditions, but, especially for legumes, their associated and symbiotic microbiota also plays important roles. As such, the selection of rhizobial strains resistant to harsh conditions, as for example water deficiency, and capable of alleviating metal phytotoxicity could be a crucial strategy to improve the yield of legumes in arid or in metal-contaminated soils.

To summarize key findings in this area and promote the selection of *ad hoc* rhizobial strains, the following review papers were presented.

- Bellabarba, A., Fagorzi, C., diCenzo, G. C., Pini, F., Viti, C., & Checcucci, A. (2019). Deciphering the symbiotic plant microbiome: translating the most recent discoveries on rhizobia for the improvement of agricultural practices in metal-contaminated and high saline lands. *Agronomy*, 9(9), 529.
- Fagorzi, C.; Checcucci, A.; diCenzo, G.; Debiec-Andrzejewska, K.; Dziewit, L.; Pini, F.; Mengoni, A. Harnessing rhizobia to improve heavy-metal phytoremediation by legumes. *Genes* 2018, 9, 542.

1.4.2 A symbiosis model: *Sinorhizobium meliloti* and *Medicago sativa*

Sinorhizobium meliloti is able to infect roots of *Medicago* species (*Medicago sativa* and its close relative *Medicago truncatula*) and develop a symbiosis. *S. meliloti* has been chosen by researchers as a laboratory model organism, interesting for its high genetic and phenotypic variability. Consequently, its association with *Medicago sativa* (alfalfa), a model organism for plant nodulation, gave birth to a worldwide model system for symbiosis and nitrogen fixation studies.

In the recent years, complete genomes of rhizobia species have sparked a great interest and proof of this is the fact that the GenBank database comprises around 250 sequenced strains of *Sinorhizobium meliloti*. As several bacterial species, *S. meliloti* presents a high genomic variability among different strains and is a classical example of a species with an open pangenome (Galardini *et al.* 2013). The term pangenome refers to the sum of shared (core) and unshared (dispensable) genome fractions. Core genes include housekeeping and mostly essential genes, while the dispensable fraction comprises more specialized, unique and accessory genes for specific functions, more related with the environmental adaptation and survival in particular ecological niches. In bacteria with multipartite genome structure core genes are classically located mostly on the chromosome, while dispensable genes are more likely harbored by plasmids or secondary replicons (diCenzo & Finan 2017). The case of *S. meliloti* is not far from this generalization. Indeed, the multipartite genome of our model bacterium comprises three replicons with distinctive structural and functional features. In the specific case of *S. meliloti* 1021, the genome is divided into a chromosome (3.65 Mb), a megaplasmid (pSymA, 1.35Mb) and a chromid (pSymB, 1.68 Mb), for a total length of 6.7 Mb (Galardini *et al.* 2011; Galibert *et al.* 2001). The pSymA megaplasmid harbors genes related to the symbiosis and to the process of nitrogen fixation. It is the most variable and recently evolved replicon, likely acquired through horizontal gene transfer (HGT) (diCenzo *et al.* 2018). The pSymB chromid mainly contains genes related to carbon sources transport and metabolism. Experimental depletion of the accessory replicons of *S. meliloti* indicated that genes present on the chromosome are sufficient to allow the bacteria to survive and replicate, as free-living organism, in the soil and suggested that the role of pSymA and pSymB is linked to the ability to establish an effective symbiosis with an host plant (in the case of the megaplasmid) and to growth in the rhizosphere (in the case of the chromid) (diCenzo *et al.* 2014). More recently, the minimal bacterial gene set necessary and sufficient for symbiosis was identified by removal of the pSymA replicon and following recovery of the symbiotic-nitrogen-fixing ability with 62 kb of pSymA, corresponding to 58 genes reorganized into three discrete *nod*, *nif*, and *fix* modules (Geddes *et al.* 2021). Insertion sequence (IS) elements and phage sequences compose 2.2% of the *S. meliloti* genome, but their distribution varies with a higher abundance on pSymA, especially near symbiotic genes, providing additional evidence that symbiotic regions are prone to DNA rearrangements (Galibert *et al.* 2001). In *S. meliloti*, the establishment of symbiosis with *M. sativa* follows the main steps described in paragraph 1.4. Transcriptomics analyses recurrently show that most genes on the symbiosis plasmid of diverse rhizobia are specifically induced during nodulation and nitrogen fixation, but not under free-living conditions lacking a compatible host or its symbiotic signal molecules (Jiao *et al.* 2016). Nodule formation and infection are determined via multiple systems of specificity. Host control over nodulation can be subtle, such as when legumes form fewer nodules with rhizobial strains that tend to provide less benefit to the host or nodulation can be blocked completely (Hahn & Studer 1986; Masson-Boivin & Sachs 2018; Sachs *et al.* 2010).

The interest in plant root exudates and their influence on the plant–soil microbiome in shaping nutrients cycle has largely increased in recent years. Root exudates that specifically inhibit soil nitrification have been identified in important crop species, (e.g. rice, wheat, and sorghum) while others have been shown to stimulate root nodulation and N fixation, even in neighboring plants. Can host plants reliably discriminate between beneficial and harmful symbionts during initial colonization? Is there a reason for symbionts to

“signal honestly”? Is there a perfect coupling between the two partners to obtain the best result in terms on mutual benefit? In fact, plant-rhizobia relationship is not exclusive, as multiple, genetically different rhizobial strains can colonize the same plant (Kiers et al., 2006 Carelli et al., 2000). Plants have evolved mechanisms to discriminate among different rhizobial strains, in terms of symbiotic efficiency, with mechanisms generally termed as “plant sanctions” (Denison 2000; Kiers *et al.* 2003; Westhoek *et al.* 2021). A mixed rhizobial population inside the same root nodule has been observed in the *Sinorhizobium-Medicago* symbiosis (Checcucci *et al.* 2016) and recently a series of experiments, based on select and resequence approach (Burghardt et al., 2018), genome-wide association analysis (Bellabarba *et al.* 2020) and RNA-Seq (Burghardt *et al.* 2017) have reported found that the selection that the host plant exerts on the rhizobial symbiotic population depends on competition among strains and that the bacterial fitness in hosts was associated with variation in genes involved in bacterial motility, host recognition and N fixation.

Host plant fitness and rhizobium fitness have been shown to be correlated. Plant fitness (and yield) is a quantitative trait, linked to genetic and environmental contributions (Friesen 2012). Consequently, we can view the genetic contribution as due to both host and symbiont genetics and the interaction between host genes and symbiont genes (G x G interaction). Indeed, G x G interactions clearly influence the fitness advantage of the symbiosis (Burghardt *et al.* 2017; Heath 2008). In 2020, Kang et al., analysed the gene expression of *M. sativa* during the interaction with different *S. meliloti* strains. Results highlighted a differential expression of plant genes (genes encoding nodulins, NCR peptides and proteins belonging to the NBS-LRR family) depending on the bacterial strain tested (Kang *et al.* 2020). Looking at the rhizobial side of the partnership, we sequenced the transcriptome of several *S. meliloti* strains incubated in a medium with root exudates of different alfalfa varieties. Results put in evidence an intricate network of genes whose expression is influenced by the genotype of both symbiotic partners, suggesting the presence of a large number of determinants involved in G x G interactions which can have an impact on the outcome of the rhizobium–legume symbiosis (Cangioli *et al.* 2021; Fagorzi *et al.* 2021).

Synthetic symbiotic communities can be rationally designed harnessing bacterial and plant -omics technologies’ outcomes to sustainably improve leguminous plant yield. In the following papers, the knowledge, challenges and novel findings in legume-rhizobium interactions have been reported. In particular, a review paper presents the rational of tailored selection of symbionts and host plant partnership, while a research paper was dedicated to dissecting G x G interactions during early plant recognition by the model rhizobium *S. meliloti*.

- Cangioli, L., Checcucci, A., Mengoni, A., & Fagorzi, C. (2021). Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations. *Biological Communications*, 66(1), 47-54.
- Fagorzi, C., Bacci, G., Huang, R., Cangioli, L., Checcucci, A., Fini, M., ... & Mengoni, A. (2021). Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis. *mSystems*, 6(1), e00974-20.

1.4.3 Not only symbionts

How did symbiotic rhizobia evolve? Several studies have shown that an effective symbiosis cannot be obtained by the transfer of symbiotic plasmids from rhizobia of the genera *Rhizobium* or *Sinorhizobium* to closely related non-rhizobia from the genera *Agrobacterium* or *Sinorhizobium* (diCenzo *et al.* 2019). Other studies on beta-rhizobia have elegantly deciphered the genetic requirements to transform a pathogenic species into a symbiotic one with experiments of symbiotic plasmid transfer followed by experimental evolution (Capela *et al.* 2017; de Moura *et al.* 2020; Marchetti *et al.* 2014). However, these experiments were done forcing evolution in laboratory conditions. Few information is on the contrary present on the evolutionary changes occurred along strains phylogeny (Pini *et al.* 2011).

The bacterial genus *Sinorhizobium* (syn. *Ensifer*) contains ecologically important nitrogen fixing symbionts of leguminous plants, as well as nonsymbiotic species. However, the evolutionary dynamics of symbiotic nitrogen fixation within this genus are unclear, and it remains an open question of whether the gain of classical symbiotic N₂-fixation genes is sufficient to allow a bacterium to fix nitrogen. In general, the primary genes required for SNF (i.e., the *nod*, *nif*, and *fix* genes) are located within mobile genetic elements that include symbiotic islands and symbiotic (mega)-plasmids simplifying their spread through horizontal gene transfer (Checcucci *et al.* 2019; Geddes *et al.* 2020; Pérez Carrascal *et al.* 2016).

Hence, the genus *Ensifer* provides an ideal model to further explore the differentiation of symbiotic bacteria from nonsymbionts. This genus comprises rhizobia such as *Sinorhizobim meliloti* and *Sinorhizobium fredii*, as well as nonrhizobia like *Ensifer morelense* and *Ensifer adhaerens*, and many members have been extensively studied producing an abundant set of experimental and genomic data (diCenzo *et al.* 2017).

Taxonomy of the genus *Ensifer/Sinorhizobium* has been widely discussed in the last 30 years. The genus *Ensifer* was proposed to describe a predator bacterium, *Ensifer adhaerens*, in 1982 (Casida 1982). Then, 6 years after, the genus *Sinorhizobium* arose, when *Rhizobium fredii* was reclassified as *Sinorhizobium fredii* (Chen *et al.* 1988). In 2002, the Subcommittee on the Taxonomy of *Agrobacterium* and *Rhizobium* gave the priority to the name *Ensifer*, while still endorsing the conservation of the name *Sinorhizobium* (Lindstrom & Martinez-Romero 2002).

Consequently, the exploration of the phylogeny of the genus *Ensifer/Sinorhizobium* can allow scientist to provide new clues on the evolution of SNF and on possible genomic definition of a bacterial genus. Moreover, nonsymbiotic *Ensifer* strains may harbour interesting traits that can be useful candidates for microbe and plant-based biotechnologies (see the previous paragraph 1.4.1). The three research papers presented below report novel findings on SNF evolution in the genus *Ensifer/Sinorhizobium*, propose a taxonomic revision of this genus as well as indicate some guidelines for genus delineation in *Rhizobiaceae*, and an investigation on a novel strain of this genus with interesting features for bioremediation.

- Fagorzi, C., Ilie, A., Decorosi, F., Cangioli, L., Viti, C., Mengoni, A., & diCenzo, G. C. (2020). Symbiotic and Nonsymbiotic Members of the Genus *Ensifer* (syn. *Sinorhizobium*) Are Separated into Two Clades Based on Comparative Genomics and High-Throughput Phenotyping. *Genome biology and evolution*, 12(12), 2521-2534.
- DiCenzo, G. C., Debiec, K., Krzysztoforski, J., Uhrynowski, W., Mengoni, A., Fagorzi, C., ... & Drewniak, L. (2018). Genomic and biotechnological characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14. *Genes*, 9(8), 379.
- Kuzmanovic N., Fagorzi C., Mengoni A., Lassalle F., diCenzo G. C., (in preparation) Taxonomy of *Rhizobiaceae* revisited: proposal of a new framework for genus delimitation.

1.5 Biotechnological approaches to harness the potential of rhizobia

Symbiotic and free-living nitrogen-fixing organisms have the potential to reduce our overall dependence on human-made forms of N fertilizer while providing an efficient way of producing protein-rich foods. This biological process plays a critical role in sustainable agriculture and there is a considerable economic incentive to explore ways to increase the efficiency of biological nitrogen fixation. Biotechnological (molecular genetics, genetic engineering, etc.) approaches have been used to improve the symbiotic efficiency and nitrogen fixation rate. In nature, the spontaneous generation of genetic variation in bacteria can occur through i) small local changes in the nucleotide sequence of the genome, ii) intragenomic reshuffling of segments of genomic sequences and iii) by acquisition of DNA sequences from another organism. Nitrogen fixation ability can be improved through genetic manipulation of the micro-symbiont and/or of the plant.

The physiological development of the plant system is significantly affected by microbial communities. Selected members of the microbial community can exert advantageous roles while few may be unfavourable to the plant growth. The significance of microbial communities related to rhizosphere has been widely recognized. To improve the plant growth and development, it is beneficial to know the microbial structure present in the rhizospheric microbiome. This can ameliorate the present situation of sustainable growth of agro-ecosystem related to soil microbiome by enhancing the final yields (Jha & Kumar 2021). A sustainable strategy to improve crop production and protection is represented by the possibility to engineer the microbiome.

Due to reduced sequencing costs, a big number of plant microbiomes have been unravelled employing metagenomics studies. Two main type of sequencing are commonly employed in plant microbiome studies: i) amplicon sequencing of universal genetic markers, like 16S rRNA gene for bacteria and archaea and internal transcribed spacer (ITS) region for eukaryotic organisms and ii) shotgun sequencing of the entire genetic material of an environmental sample (Sergaki *et al.* 2018). Sequencing a whole metagenome can give a clear idea of not just the composition, but also the functions of the microbiome. Indeed, besides the microbiome structure, also its functions change in different plant genotypes and growth conditions. The challenge, here, is the assembly of the sequenced reads into a high quality metagenome assembly, especially for sample with a high degree of heterogeneity (Levy *et al.* 2018; Singh *et al.* 2020).

The -omics approaches for the study of plant-microbe interactions and most specifically, in rhizobium-legume symbiosis, finds its position in a potentially recursively used pipeline, a DBTL (Design–Build–Test–Learn) cycle. The design is represented by the inoculation of the plant with microorganisms to observe their phenotypic changes and to collect datasets generated with -omics approaches (build) that can be integrated, thanks to system biology, identifying candidate genes. These functions can be introduced in a system, genes can undergo gene editing (test and learn) and results are thus evaluated to consider further improvement (Jensen & Keasling 2018; Petzold *et al.* 2015) (Figure 13).

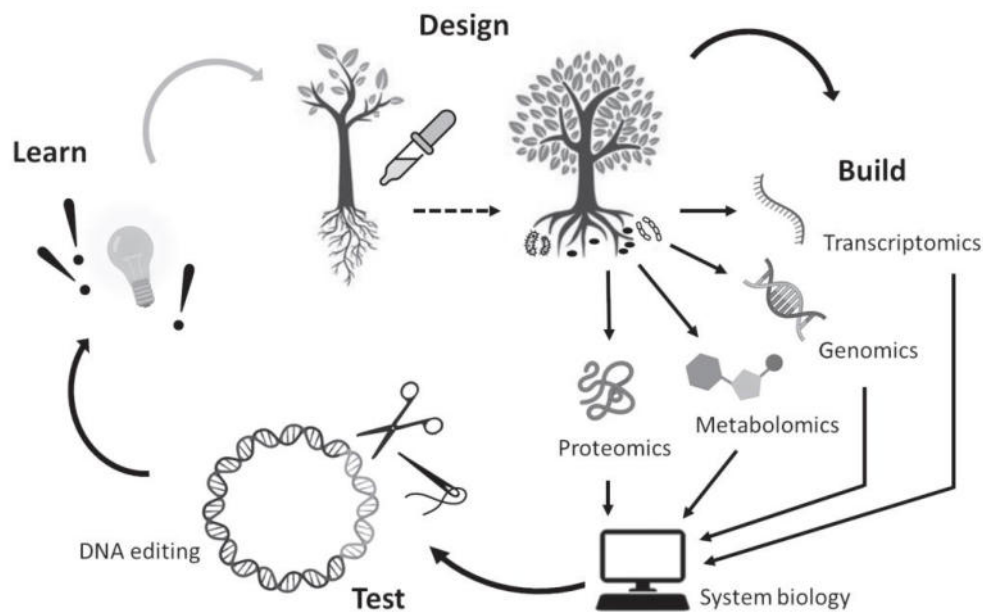


Figure 13. The potentially recursive pipeline of a DBTL cycle (source: Fagorzi, 2020)

As an example, rhizobia adapted to drastic environmental conditions, such as arid climate, may harbour specific genes expressed under these conditions, which can be transferred in other bacteria, thus improving the nitrogen fixation rate in arid soils. The exploitation of this system is especially important for the subsistence of farmers in developing countries, who typically face the difficulty of cultivating a few hectares of land with low yields. Next to point-like modification of the bacterial genomes, large-scale manipulation of rhizobia can also represent a viable way to harness the potential of specific strains (Lau *et al.* 2017; Smith *et al.* 2003). The ability to predict the phenotypic outcomes of large-scale genome modification requires a precise understanding of the genetic and regulatory interactions between each gene or gene product in the genome (Checcucci *et al.* 2017b). The multipartite genome of rhizobia lends itself optimally to large manipulations as secondary replicons might act as plug-and-play functional modules, potentially allowing the recipient strain to gain new abilities (Young 2016). The following paper which I contributed along my PhD research program explores the feasibility of large-scale manipulation within the genus *Sinorhizobium*. The *cis*-hybrid strain created derives from *cis*-genic manipulation and contains genetic material from the pangenome pool of two members of the same species. Biotechnology and synthetic biology are here environmental-friendly methods to improve the potential of bacterial species of interest in agricultural and environmental microbiology.

- Checcucci, A., diCenzo, G. C., Ghini, V., Bazzicalupo, M., Becker, A., Decorosi, F., ... & Mengoni, A. (2018). Creation and characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *ACS synthetic biology*, 7(10), 2365-2378.

Chapter 2

2.1 The small in the large: exploring environmental bacterial communities

Environmental microbiology can be defined as the study of microbes, their functions, and interactions in all habitats on Earth (and beyond). This is an expanding research area that covers air, soil and aquatic systems, astrobiology, biogeochemical cycles, plant-soil interactions, extreme environments, and spatial, temporal and perturbation studies (Watts 2019). The papers presented in this chapter are aimed at exploring three different environments sharing the feature of being “extreme”, that is non-conventional for aerobic, mesophilic microbes.

In ecology, the term “extreme” commonly refers to unfavourable environmental factors that depress the ability of organisms to function (Li *et al.* 2014a). The relevance of organisms inhabiting these environment lies in their ability to rapidly evolve to face the adverse conditions in which they live. Adaptative features of these extremophiles permit them to survive under such extreme and hostile environmental conditions (Yadav *et al.* 2015). Bacterial communities from natural extreme environments represent not only a gene reservoir for potential biotechnological applications (e.g., the discovery of the *Taq* DNA polymerase, isolated from the thermophilic bacteria *Thermus aquaticus*), but also offer the opportunity to discover novel metabolic mechanisms that could inspire innovations in such diverse areas as synthetic biology and research into human survival in space . Microorganisms adapted to extreme living conditions can provide research ideas for applications in bioremediation of polluted sites deemed too unbearable for most living organisms, they can represent sources for novel therapeutics in medicine and potentially, alternative processes for biofuel or energy production (Mengoni *et al.* 2010). Many extreme environments offer relatively accessible proxies for the harsh environments found beyond Earth (Tighe *et al.* 2017). In the last couple of decades microbiome research has received tremendous attention and it has become evident that microbiota associated with higher organisms has highly important functions supporting health, growth, and well-being of their hosts (Sessitsch *et al.* 2019).

Following the aim to provide novel insights into microbiota of soils and associated with plants from “extreme” environments, I explored the bacterial communities impacted by volcanic gases on Vulcano Islands, and contributed to two papers on the root microbiota of the common reed, *Phragmites australis*, a plant used in phytodepuration and on the biofilm covering an iron-contaminated river.

The diversity of bacterial communities of Vulcano Island (Aeolian Islands, Sicily, Italy) was investigated by sampling two areas characterised by intense exhalative activity: La Fossa Crater and Levante Bay. Differences between the two areas comprise temperature and composition of the fumarolic fluids. Pioneer plants were found both on the volcanic edifice and on the bay, in high salinity environments. For the different sampling areas, both soil bacterial communities and gases were analysed, resulting in a geo-microbiological study of a volcanic environment. Data obtained on the composition of the bacterial communities and different emission profiles suggest a possible correlation between the two parameters. The presence of similar microbial communities inhabiting sites with different emission profiles might be explained on the basis of possible sharing of metabolic abilities related to gas composition. Moreover, a possible correlation between some gases (i.e. decane, argon, i-octane and undecane) and microbial communities emerged.

- Fagorzi, C., Del Duca, S., Venturi, S., Chiellini, C., Bacci, G., Fani, R., & Tassi, F. (2019). Bacterial Communities from Extreme Environments: Vulcano Island. *Diversity*, 11(8), 140.

Epilithic biofilm of the Acquarossa river (Viterbo, Italy) are particularly interesting in terms of spatial structuring, with red and black biofilms present on the rocks very close, without blending together. This peculiarity has raised questions about the biotic and abiotic phenomena that might avoid mixing of the two biofilms. The goal of this study was to characterise the bacterial communities inhabiting this environment, both through culture dependent and independent approaches. Furthermore, a phenotypic characterization of the main bacterial genera cultivated was assessed and physio-chemical parameters of water and biofilms were measured. Differences in taxonomic composition of black and red epilithon highlights the domination of *Acinetobacter* spp. in the first case, and iron-oxidizing bacteria in the second one. No differences between the two epilithon were identified if observing the culturable bacterial fraction and heavy-metal resistance patterns. On the other side, antibiotic resistance pattern as well as antagonistic interactions between the dominant bacterial genera seems to affect the global structuring of the two biofilms. This study highlights the competition for different niches and selection that a bacterial populations can undergo at small scale.

- Chiellini, C., Miceli, E., Bacci, G., Fagorzi, C., Coppini, E., Fibbi, D., Bianconi, G., Mengoni, A., Canganella, F., Fani, R. (2018) Spatial structuring of bacterial communities in epilithic biofilms in the Acquarossa river (Italy). *FEMS Microbiology Ecology*

As deeply discussed along this thesis with particular reference to the legume-rhizobium symbiosis, the role of plant-associated microorganisms is frequently crucial for the development of the host itself. In many cases endophytes exert beneficial effects on plants, by enhancing the uptake of nutrients, promoting the plant growth, preventing pathogen infections, accelerating seedling emergence and inducing tolerance to pollution and environmental stresses. The focus of the following work is the ability of plant-associated bacteria to increase the ability of plants to detoxify polluted environments. Here, we characterized the cultivable bacterial community associated to the roots of *Phragmites australis* in the constructed wetlands (CW) of Calice (Prato, Italy), managed by G.I.D.A SpA. One of the goals of the company is to verify the action of the CW in the tertiary treatment of landfill leachate. Bacterial communities were studied for 22 months, before, during and after the activation of the plant, evaluating in this way the wastewater effect in shaping the composition of the microbiome. Bacteria isolated from roots were tested for their ability to grow in the presence of synthetic wastewater (SWW) along with their resistance against a panel of antibiotics commonly used to treat infections in humans. The cultivable bacterial community existing before the activation of the plant underwent fluctuations in terms of taxonomic composition. As expected, the influx of wastewater exerted a selective pressure on the resident bacterial community, selecting and/or making bacterial strains progressively more resistant to SWW and antibiotics. The selection of strains resistant to SWW and antibiotics and able to increase the phytodepuration properties of *P. australis* can improve the yield of the CW. Such strains should be tested for their plant-growth promotion activity and plant resistance to pollution to be considered as possible bio-inoculants for CW.

- Vassallo, A., Miceli, E., Fagorzi, C., Castronovo, L. M., Del Duca, S., Chioccioli, S., ... & Fani, R. (2020). Temporal Evolution of Bacterial Endophytes Associated to the Roots of *Phragmites australis* Exploited in Phytodepuration of Wastewater. *Frontiers in microbiology*, 11, 1652.

Chapter 3

3.1 Going complex: a short journey into bacteria metabolic complexity

Model organisms: if it is true for Escherichia coli, is it also true for the elephant?

In the eighteenth and nineteenth centuries, biologists approached biological phenomena observationally, looking for common patterns that underlie the diversity, complexity and development of organisms. In the second half of the nineteenth century, an increasing craving for experimentation shaped rudimentary knowledge of metabolism, the mechanisms of embryogenesis, animal and plant physiology, conditioned behaviour and photosynthesis. The discovery of biological diversity created in scientists the need to appreciate the evolutionary mechanisms and led biologists to engage in studies of a wide variety of organisms. The work strongly supported generalizations, but it also contributed to making biology an organism-oriented science. The result, perhaps unexpected, of this process was the expansion of the study of phylogeny, of distinct modes of development and behaviour and of biological complexity. Mendel's work on inheritance in peas and the advent of classical genetics shaped our concept of life. Genetic research on several organisms led to an appreciation of the generality of the rules of inheritance. The number of organisms deeply studied to the geneticist soon diminished as the experimental role of a few models became more prominent (Davis 2004).

Escherichia coli, *Salmonella* and their phages opened the doors to molecular biology. Research on the spontaneous mutation of *E. coli* to bacteriophage resistance initiated a disciplined study of mutation and bacterial inheritance, conjugation and transduction, analysis of metabolic pathways and gene exchange (Davis 2010; K S 1962; Lederberg & Tatum 1946). It became clear that virtually any phenotypic or behavioural attribute of bacteria could be studied with a genetic approach and *E. coli* became a supermodel (Davis 2004).

As the genome sequences of various species became available, the science of comparative genomics was born, together with the first approaches to systems biology. At the same time, genetic engineering created a link between model organisms to lesser-known forms. Comparative genomics enriched the field of molecular evolution and diversity became an observable both within and among species. The genes of one organism can potentially be compared, both in sequence and in biological function, to those of any other. As the databases grow, the appreciation of both the unity and the diversity of living things deepens (Davis 2004).

As an example, bacterial metabolic adaptation strategies (e.g. patterns including simultaneous carbon consumption, diauxic growth and bistable growth) have been studied in model organisms (e.g., *Escherichia coli* and *Lactococcus lactis*), grown on defined media containing simple mixtures of carbohydrates. However, natural conditions are often very complex, and bacteria need to dynamically modulate specific degradation pathways according to the type and the concentration of external nutrients. The marine environment represents a paradigmatic example of the challenges encountered by microorganisms. To investigate metabolic adaptation strategies in these mutable conditions, a model is represented by *Pseudoalteromonas haloplanktis* TAC125, a heterotrophic marine bacterium isolated from Antarctica. This bacterium shows a somewhat classical diauxic shape of the growth curve. However, this hypothetical diauxie is not related to simple/single carbon source, but to a complex mix of amino acids, which may likely be related to the complex

nutritional conditions encountered in the environment. It is then interesting to understand and model this diauxie, which emphasize that *E. coli* is a starting paradigm, but cannot clearly represent the overall microbial metabolic versatility. The research paper presented below was a systems-biology investigation, where metabolomic experiments and genome-scale constraint based metabolic modelling allowed to clarify the life under complex nutritional environments. Theoretical modelling indicates that this metabolic phenotype, combining diauxie and co-utilization, is compatible with a tight regulation that allows the modulation of assimilatory pathways. Future perspectives include the investigation on the molecular mechanisms allowing the implementation of this mixed feeding strategy and the regulatory circuits responsible for the switching among the available carbon sources.

- Perrin, E., Ghini, V., Giovannini, M., Di Patti, F., Cardazzo, B., Carraro, L., ... & Fondi, M. (2020). Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nature communications*, 11(1), 1-16.

Concluding remarks

Sustainability from different perspectives

In a world where the term “sustainability” is overused, with a mood that looks a lot like a fashion (hopefully not temporary), it is necessary to distinguish the sentences aimed at capturing likes from seriously applicable purposes and strategies.

The Technical Advisory Committee of Consultative Group on International Agricultural Research defined sustainability as “the successful management of resources for agriculture to satisfy changing human needs while maintaining or enhancing the quality of the environment and conserving resources”. Similarly, the World Commission on Environment and Development talked about sustainability enhancing human well-being to more equitably meet the needs of both current and future generations. In a 1987 document it is reported “*Environment* is where we all live; and *development* is what we all do in attempting to improve our lot within that abode. The two are inseparable....Humanity has the ability to make development sustainable: to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs” (Butlin 1989; Clark & Harley 2020).

Sustainable agriculture fulfils the needs of the present without compromising the needs of the future. In other words, it can be described as agriculture that is managed toward greater resource efficiency and conservation while maintaining an environment favourable for evolution of all species. In agriculture, stocks include soil, water, non-renewable energy resources and environmental quality. One of the driving forces behind agricultural sustainability is effective management of N in the environment. Industrially produced N fertilizer depletes non-renewable energy resources and poses human and environmental hazards. Moreover, judicious management of N inputs into cropping systems is a prerequisite for sustainable agricultural practice. A very successful way for manipulation of N inputs that results in farming practices that are economically viable and environmentally prudent is the use of biologically fixed nitrogen. Symbiotic systems can be a major source of N in most cropping systems (Vance 1997).

What is important to keep in mind is the attention that the sustainability theories reserve to future generations. This means that sustainability is not only a matter of environment (considered as nature preservation) but also wellness and economy.

A transformation from a linear to a circular economy might result in assessed integrated impacts - possible societal benefits. “What-if-scenarios” have been applied into a constructed interactive input–output model with a number of changes linked to a circular economy. The model has then been applied to eight European economies showing close to 70% reductions in CO₂-emissions while offering new and additional employment and improving the trade balance of fossil-fuels-importing countries (Brandão *et al.* 2020). Most policies dedicated to reducing agricultural nitrogen pollution focus on changing farmer behaviour and enhancing the circular nitrogen economy. However, farmers are just one of several actors in the agri-food chain. The activities of other actors — from fertilizer manufacturers to wastewater treatment companies to composting industries — can also impact nitrogen losses at the farm level and beyond (Kanter *et al.* 2020). According to a study carried out by Gerten *et al.*, the present food system could provide a balanced diet (2,355 kcal per capita per day) for 3.4 billion people only. Transformation towards more sustainable production and consumption patterns could support 10.2 billion people within the planetary boundaries analysed. Key

prerequisites are spatially redistributed cropland, improved water–nutrient management, food waste reduction and dietary changes (Gerten *et al.* 2020) (Figure 14).

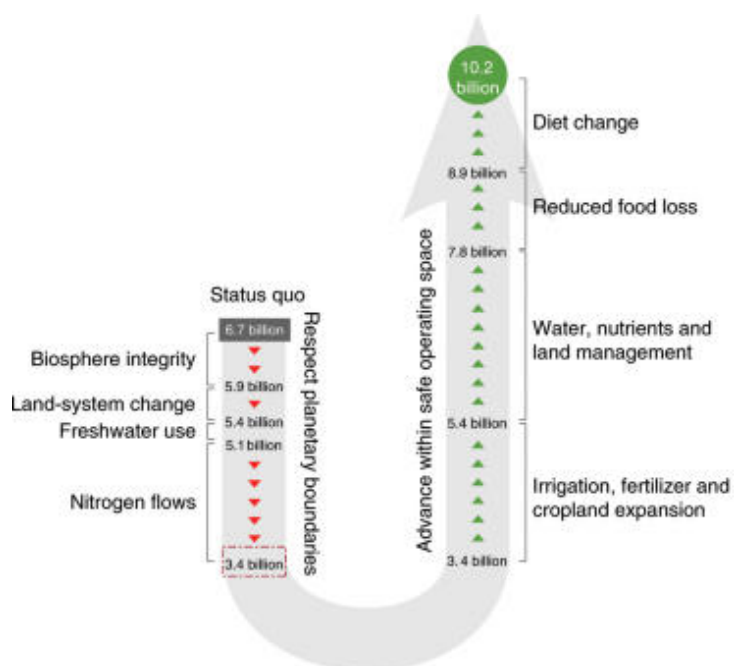


Figure 14. Simulated technological–cultural ‘U-turn’ towards increasing global food supply within four planetary boundaries. Global population that can be provided with a global average net food supply of $2,355 \text{ kcal cap}^{-1} \text{ d}^{-1}$ (including sufficient protein content) when respecting the different planetary boundaries given unchanged current practices (left side) and when making use of opportunities of agricultural land expansion, management and sociocultural changes (right side) (modified from Gerten *et al.* 2020).

From our point of view, the idea of exploiting legume-rhizobium symbiosis to improve the sustainability of agriculture represents a realistic starting point to - as mentioned a few lines above - fulfil the needs of the present without compromising the needs of the future. The experience of farmers, scientific research and innovative methodologies paved the way to the use of rhizobia as bioinoculants in field. Nonetheless, a more holistic understanding is still needed to better understand the intermicrobial interactions within the microbiota of plants and to better define the functional relevance of the microbial networks for holobiont fitness (Hacquard & Schadt 2015). Prokaryotic and eukaryotic microbes have evolved a myriad of cooperative and competitive interaction mechanisms that shape and likely stabilize microbial assemblages on plant tissues. However, most of the data are derived from one-to-one interaction studies, and only few incorporate complex microbial communities in controlled laboratory conditions to reconstitute the plant microbiota and to understand the role of intermicrobial interactions. Such experiments will shed new light on the fundamental principles that govern the assembly of complex microbial communities and the maintenance of host-microbial homeostasis. It is important to consider microbe-microbe interactions to accept or reject the hologenome theory, which postulates that selection can operate on horizontally acquired plant microbiota members. According to this concept, it is likely that microbes that tightly associate with plants also evolve community level microbe-microbe interaction strategies that allow them to persist within the plant holobiont (Bardgett *et al.* 2014; Hassani *et al.* 2018).

Final thoughts

The complexity of small things is sometimes surprising, especially when looking at the positive effects a bacterium can bring to the growth of host (i.e. a plant). The idea that the presence in the soil of bacterial strains able to ameliorate the production of a crop is intriguing; the possibility to use specific bacterial strains as if they are “precision energy supplements” for specific plant species gives the idea of how powerful and rich of novelty research in the field of sustainable agriculture is becoming (Figure 15).

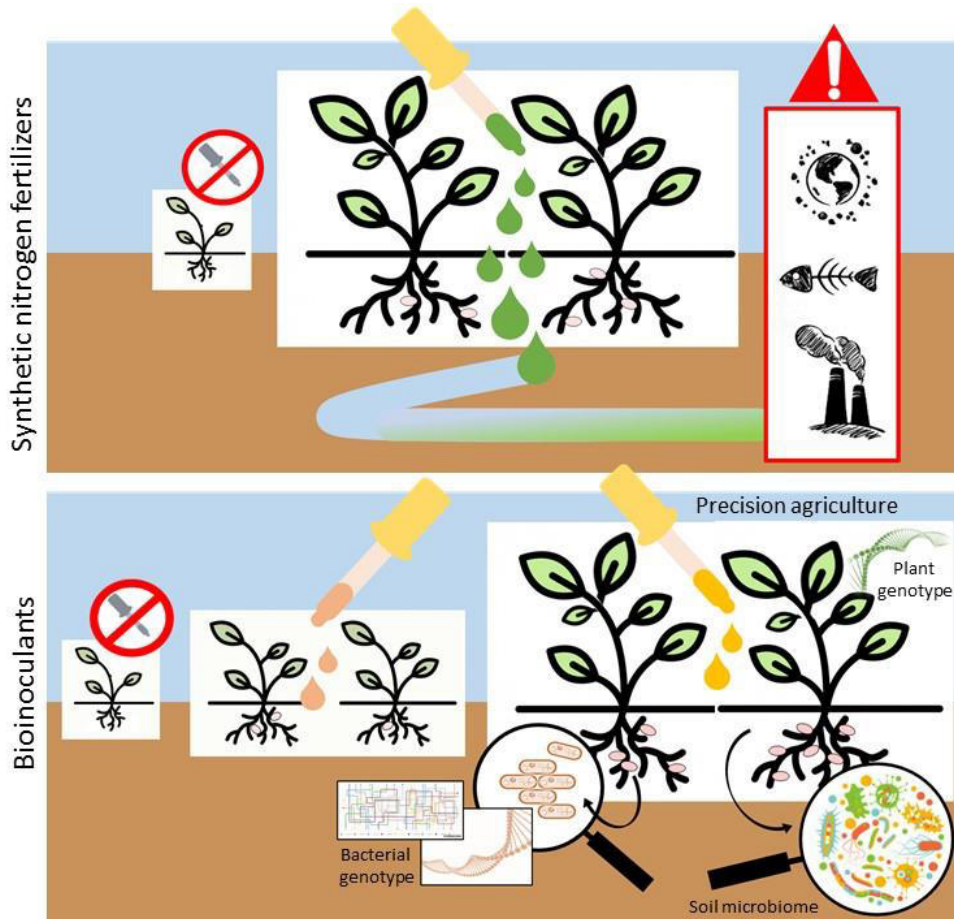


Figure 15. The use of synthetic nitrogen fertilizers increases the crop yield at the high cost of serious and lasting environmental damages. Bioinoculants are currently used as a “green” alternative to synthetic products, with good results in terms of yield of the field and safety. The future is moving towards precision agriculture, with specific strains, capable to enhance at high levels the growth of well-defined plant species and able to survive in a well-defined environment, where physio-chemical conditions and microbiome context are known.

A challenge is to make people aware that the use of bacterial inocula in agriculture is useful for limiting the use of synthetic fertilizers. The challenge in the challenge is to understand the rules (mechanisms) of the plant-association game and use our biotechnological skills to safely combine multiple traits of different bacterial strains to obtain more efficient inoculants. Omics technologies are providing huge amounts of data that help us to deeply understand the finest mechanisms of bacterial metabolism and the precise composition of microbial communities present in the most disparate environments. A lot of work has still to be carried out to answer the many questions on our small, simple but indeed complex (micro)organisms. We have a lot of inventiveness and many resources to be brought into play.

Bibliography

- Alori, E. T., Dare, M. O., & Babalola, O. O. (2017). Microbial Inoculants for Soil Quality and Plant Health. doi:10.1007/978-3-319-48006-0_9
- Ankati, S., & Podile, A. R. (2019). Metabolites in the root exudates of groundnut change during interaction with plant growth promoting rhizobacteria in a strain-specific manner. *Journal of Plant Physiology*. doi:10.1016/j.jplph.2019.153057
- Bardgett, R. D., Mommer, L., & De Vries, F. T. (2014). Going underground: Root traits as drivers of ecosystem processes. *Trends in Ecology and Evolution*. doi:10.1016/j.tree.2014.10.006
- Barreca, D., Mandalari, G., Calderaro, A., ... Gattuso, G. (2020). Citrus flavones: An update on sources, biological functions, and health promoting properties. *Plants*. doi:10.3390/plants9030288
- Bartholomew, C. H., & Farrauto, R. J. (2010). *Fundamentals of Industrial Catalytic Processes: Second Edition*. *Fundamentals of Industrial Catalytic Processes: Second Edition*. doi:10.1002/9780471730071
- Bellarbarba, A., Bacci, G., Decorosi, F., ... Pini, F. (2020). Competitiveness prediction for nodule colonization in *Sinorhizobium meliloti* through combined in vitro tagged strain characterization and genome-wide association analysis. *BioRxiv*.
- Bellarbarba, A., Fagorzi, C., DiCenzo, G. C., Pini, F., Viti, C., & Checcucci, A. (2019). Deciphering the symbiotic plant microbiome: Translating the most recent discoveries on rhizobia for the improvement of agricultural practices in metal-contaminated and high saline lands. *Agronomy*, **9**(9). doi:10.3390/agronomy9090529
- Benjamim, J. K. F., Dias da Costa, K. A., & Silva Santos, A. (2020). Chemical, Botanical and Pharmacological Aspects of the Leguminosae. *Pharmacognosy Reviews*. doi:10.5530/phrev.2020.14.15
- Berrabah, F., Ratet, P., & Gourion, B. (2015). Multiple steps control immunity during the intracellular accommodation of rhizobia. *Journal of Experimental Botany*. doi:10.1093/jxb/eru545
- Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology*. doi:10.1371/journal.pbio.1002226
- Bowen, G. J., Beerling, D. J., Koch, P. L., Zachos, J. C., & Quattlebaum, T. (2004). A humid climate state during the Palaeocene/Eocene thermal maximum. *Nature*. doi:10.1038/nature03115
- Brandão, M., Lazarevic, D., & Finnveden, G. (2020). *Handbook of the Circular Economy*. *Handbook of the Circular Economy*. doi:10.4337/9781788972727
- Bruning, B., & Rozema, J. (2013). Symbiotic nitrogen fixation in legumes: Perspectives for saline agriculture. *Environmental and Experimental Botany*. doi:10.1016/j.envexpbot.2012.09.001
- Burghardt, L. T., Epstein, B., Guhlin, J., ... Tiffin, P. (n.d.). Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. doi:10.1073/pnas
- Burghardt, L. T., Epstein, B., Guhlin, J., ... Tiffin, P. (2018). Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1714246115
- Burghardt, L. T., Guhlin, J., Chun, C. L., ... Tiffin, P. (2017). Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Molecular Ecology*. doi:10.1111/mec.14285

- Butlin, J. (1989). Our common future. By World commission on environment and development. (London, Oxford University Press, 1987, pp.383 £5.95.). *Journal of International Development*. doi:10.1002/jid.3380010208
- Cangioli, L., Checcucci, A., Mengoni, A., & Fagorzi, C. (2021). Legume tasters: Symbiotic rhizobia host preference and smart inoculant formulations. *Biological Communications*. doi:10.21638/SPBU03.2021.106
- Capela, D., Marchetti, M., Clérissi, C., ... Masson-Boivin, C. (2017). Recruitment of a Lineage-Specific Virulence Regulatory Pathway Promotes Intracellular Infection by a Plant Pathogen Experimentally Evolved into a Legume Symbiont. *Molecular Biology and Evolution*. doi:10.1093/molbev/msx165
- Carvalhais, L. C., Dennis, P. G., Badri, D. V., Kidd, B. N., Vivanco, J. M., & Schenk, P. M. (2015). Linking Jasmonic acid signaling, root exudates, and rhizosphere microbiomes. *Molecular Plant-Microbe Interactions*. doi:10.1094/MPMI-01-15-0016-R
- Casida, L. E. (1982). *Ensifer adhaerens* gen. nov., sp. nov.: A bacterial predator of bacteria in soil. *International Journal of Systematic Bacteriology*. doi:10.1099/00207713-32-3-339
- Chaintreuil, C., Rigault, F., Moulin, L., ... Bailly, X. (2007). Nickel resistance determinants in bradyrhizobium strains from nodules of the endemic New Caledonia legume *Serianthes calycina*. *Applied and Environmental Microbiology*, **73**(24), 8018–22.
- Chaparro, J. M., Badri, D. V., & Vivanco, J. M. (2014). Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal*. doi:10.1038/ismej.2013.196
- Checcucci, A., Azzarello, E., Bazzicalupo, M., ... Mengoni, A. (2016). Mixed nodule infection in *Sinorhizobium meliloti*-*medicago sativa* symbiosis suggest the presence of cheating behavior. *Frontiers in Plant Science*. doi:10.3389/fpls.2016.00835
- Checcucci, A., Azzarello, E., Bazzicalupo, M., ... Mengoni, A. (2017a). Role and regulation of ACC deaminase gene in *Sinorhizobium meliloti*: Is it a symbiotic, rhizospheric or endophytic gene? *Frontiers in Genetics*. doi:10.3389/fgene.2017.00006
- Checcucci, A., DiCenzo, G. C., Bazzicalupo, M., & Mengoni, A. (2017b). Trade, diplomacy, and warfare: The Quest for elite rhizobia inoculant strains. *Frontiers in Microbiology*, **8**(NOV). doi:10.3389/fmicb.2017.02207
- Checcucci, A., diCenzo, G. C., Perrin, E., Bazzicalupo, M., & Mengoni, A. (2019). Genomic Diversity and Evolution of Rhizobia. In *Microbial Diversity in the Genomic Era*. doi:10.1016/B978-0-12-814849-5.00003-4
- Chen, W. M., Moulin, L., Bontemps, C., Vandamme, P., Béna, G., & Boivin-Masson, C. (2003). Legume Symbiotic Nitrogen Fixation by β -Proteobacteria Is Widespread in Nature. *Journal of Bacteriology*. doi:10.1128/JB.185.24.7266-7272.2003
- Chen, W. X., Yan, G. H., & Li, J. L. (1988). Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. *International Journal of Systematic Bacteriology*. doi:10.1099/00207713-38-4-392
- Chibuike, G. U., & Obiora, S. C. (2014). Heavy metal polluted soils: effect on plants and bioremediation methods. *Applied and Environmental Soil Science*, **2014**.
- Clark, W. C., & Harley, A. G. (2020). Sustainability science: Toward a synthesis. *Annual Review of Environment and Resources*. doi:10.1146/annurev-environ-012420-043621
- Climate change 2013: The physical science basis-conclusions. (2013). *Bulletin Fuer Angewandte Geologie*.

doi:10.5169/seals-391142

- Davis, R. H. (2004). The age of model organisms. *Nature Reviews Genetics*. doi:10.1038/nrg1250
- Davis, R. H. (2010). *The Microbial Models of Molecular Biology: From Genes to Genomes. The Microbial Models of Molecular Biology: From Genes to Genomes*. doi:10.1093/acprof:oso/9780195154368.001.0001
- de Moura, G. G. D., Remigi, P., Masson-Boivin, C., & Capela, D. (2020). Experimental evolution of legume symbionts: What have we learnt? *Genes*. doi:10.3390/genes11030339
- Denison, R. F. (2000). Legume Sanctions and the Evolution of Symbiotic Cooperation by Rhizobia. *The American Naturalist*. doi:10.1086/316994
- Diaz, R. J., & Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science*. doi:10.1126/science.1156401
- diCenzo, G. C., Benedict, A. B., Fondi, M., ... Griffiths, J. S. (2018). Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genetics*, **14**(4). doi:10.1371/journal.pgen.1007357
- diCenzo, G. C., & Finan, T. M. (2017). The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiology and Molecular Biology Reviews*, **81**(3).
- diCenzo, G. C., MacLean, A. M., Milunovic, B., Golding, G. B., & Finan, T. M. (2014). Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. *PLoS Genetics*. doi:10.1371/journal.pgen.1004742
- diCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A., & Fondi, M. (2020). Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nature Communications*. doi:10.1038/s41467-020-16484-2
- Dicenzo, G. C., Zamani, M., Checcucci, A., ... Mengoni, A. (2019). Multidisciplinary approaches for studying rhizobium–legume symbioses. *Canadian Journal of Microbiology*. doi:10.1139/cjm-2018-0377
- diCenzo, G. C., Zamani, M., Ludwig, H. N., & Finan, T. M. (2017). Heterologous Complementation Reveals a Specialized Activity for BacA in the *Medicago – Sinorhizobium meliloti* Symbiosis. *Molecular Plant-Microbe Interactions*. doi:10.1094/MPMI-02-17-0030-R
- Dobert, R. C., Breil, B. T., & Triplett, E. W. (1994). DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationship to those of other nodulating bacteria. *Molecular Plant-Microbe Interactions*. doi:10.1094/MPMI-7-0564
- Edwards, J. A., Santos-Medellín, C. M., Liechty, Z. S., ... Sundaresan, V. (2018). Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biology*. doi:10.1371/journal.pbio.2003862
- Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience*. doi:10.1038/ngeo325
- Estrada-de los Santos, P., Palmer, M., Steenkamp, E. T., ... Venter, S. N. (2019). *Trinickia dabaoshanensis* sp. nov., a new name for a lost species. *Archives of Microbiology*. doi:10.1007/s00203-019-01703-2
- Fagorzi, C., Bacci, G., Huang, R., ... Mengoni, A. (2021). Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-rhizobium symbiosis. *MSystems*, **6**(1). doi:10.1128/mSystems.00974-20
- Fagorzi, C., Checcucci, A., Dicenzo, G. C., ... Mengoni, A. (2018). Harnessing rhizobia to improve heavy-metal

phytoremediation by legumes. *Genes*. doi:10.3390/genes9110542

Farkas, A., Maróti, G., Dürgo, H., ... Kondorosi, É. (2014). Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1404169111

Finzi, A. C., Abramoff, R. Z., Spiller, K. S., ... Phillips, R. P. (2015). Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles. *Global Change Biology*. doi:10.1111/gcb.12816

Foster, K. R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. *Nature*. doi:10.1038/nature23292

Friesen, M. L. (2012). Widespread fitness alignment in the legume – rhizobium symbiosis. *New Phytologist*. doi:10.1111/j.1469-8137.2012.04099.x

Fryzuk, M. D. (2004). Ammonia transformed. *Nature*. doi:10.1038/427498a

Gage, D. J. (2004). Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews*. doi:10.1128/mmbr.68.2.280-300.2004

Galardini, M., Mengoni, A., Brilli, M., ... Biondi, E. G. (2011). Exploring the symbiotic pangenome of the nitrogen-fixing bacterium Sinorhizobium meliloti. *BMC Genomics*. doi:10.1186/1471-2164-12-235

Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E. G., & Mengoni, A. (2013). Replicon-dependent bacterial genome evolution: The case of Sinorhizobium meliloti. *Genome Biology and Evolution*. doi:10.1093/gbe/evt027

Galibert, F., Finan, T. M., Long, S. R., ... Batut, J. (2001). The composite genome of the legume symbiont Sinorhizobium meliloti. *Science*. doi:10.1126/science.1060966

Gamble, A. (2019). Ullmann's Encyclopedia of Industrial Chemistry. *The Charleston Advisor*. doi:10.5260/chara.20.4.46

Geddes, B. A., Kearsley, J., Morton, R., diCenzo, G. C., & Finan, T. M. (2020). The genomes of rhizobia. *Advances in Botanical Research*. doi:10.1016/bs.abr.2019.09.014

Geddes, B. A., Kearsley, J. V. S., Huang, J., ... Finan, T. M. (2021). Minimal gene set from Sinorhizobium (Ensifer) meliloti pSymA required for efficient symbiosis with Medicago. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.2018015118

Gerten, D., Heck, V., Jägermeyr, J., ... Schellnhuber, H. J. (2020). Feeding ten billion people is possible within four terrestrial planetary boundaries. *Nature Sustainability*. doi:10.1038/s41893-019-0465-1

Gilchrist, M., & Benjamin, N. (2017). From Atmospheric Nitrogen to Bioactive Nitrogen Oxides. In *Nitrite and Nitrate in Human Health and Disease*. doi:10.1007/978-3-319-46189-2_2

Guo, W., Zhang, K., Liang, Z., Zou, R., & Xu, Q. (2019). Electrochemical nitrogen fixation and utilization: Theories, advanced catalyst materials and system design. *Chemical Society Reviews*. doi:10.1039/c9cs00159j

Hacquard, S., & Schadt, C. W. (2015). Towards a holistic understanding of the beneficial interactions across the Populus microbiome. *New Phytologist*. doi:10.1111/nph.13133

Hahn, M., & Studer, D. (1986). Competitiveness of a nif⁺ Bradyrhizobium japonicum mutant against the wild-type strain. *FEMS Microbiology Letters*. doi:10.1111/j.1574-6968.1986.tb01228.x

Hardy, R. W. F. (1984). The Fundamentals of Nitrogen Fixation. J. R. Postgate. *The Quarterly Review of*

Biology. doi:10.1086/413793

- Hartmann, A., Schmid, M., van Tuinen, D., & Berg, G. (2009). Plant-driven selection of microbes. *Plant and Soil*. doi:10.1007/s11104-008-9814-y
- Hassani, M. A., Durán, P., & Hacquard, S. (2018). Microbial interactions within the plant holobiont. *Microbiome*. doi:10.1186/s40168-018-0445-0
- Heath, K. D. (2008). The coevolutionary genetics of plant-microbe interactions. In *New Phytologist*. doi:10.1111/j.1469-8137.2008.02633.x
- Herridge, D. F., Peoples, M. B., & Boddey, R. M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*. doi:10.1007/s11104-008-9668-3
- Hoffman, B. M., Lukoyanov, D., Yang, Z. Y., Dean, D. R., & Seefeldt, L. C. (2014). Mechanism of nitrogen fixation by nitrogenase: The next stage. *Chemical Reviews*. doi:10.1021/cr400641x
- Igarashi, R. Y., & Seefeldt, L. C. (2003). Nitrogen Fixation: The Mechanism of the Mo-Dependent Nitrogenase. *Critical Reviews in Biochemistry and Molecular Biology*. doi:10.1080/10409230391036766
- Jacob, V., Zhang, H., Churchill, A. C., ... Tissue, D. T. (2020). Warming reduces net carbon gain and productivity in medicago sativa l. And festuca arundinacea. *Agronomy*. doi:10.3390/agronomy10101601
- Jensen, M. K., & Keasling, J. D. (2018). *Synthetic Metabolic Pathways*. *Methods in Molecular Biology*. doi:10.1007/978-1-4939-7295-1_5
- Jha, P., & Kumar, V. (2021). Role of Metagenomics in Deciphering the Microbial Communities Associated with Rhizosphere of Economically Important Plants. doi:10.1007/978-981-15-6949-4_4
- Jiao, J., Wu, L. J., Zhang, B., ... Tian, C. F. (2016). MucR is required for transcriptional activation of conserved ion transporters to support nitrogen fixation of Sinorhizobium fredii in soybean nodules. *Molecular Plant-Microbe Interactions*. doi:10.1094/MPMI-01-16-0019-R
- K S, &NA; (1962). Sexuality and the Genetics of Bacteria. *The American Journal of the Medical Sciences*. doi:10.1097/00000441-196207000-00016
- Kamminga, H. (1984). Martinus Willem Beijerinck - His life and his work. *Trends in Biochemical Sciences*. doi:10.1016/0968-0004(84)90228-7
- Kang, W., Kang, W., Jiang, Z., ... Zhang, X. X. (2020). Plant transcriptome analysis reveals specific molecular interactions between alfalfa and its rhizobial symbionts below the species level. *BMC Plant Biology*. doi:10.1186/s12870-020-02503-3
- Kanter, D. R., Bartolini, F., Kugelberg, S., Leip, A., Oenema, O., & Uwizeye, A. (2020). Nitrogen pollution policy beyond the farm. *Nature Food*. doi:10.1038/s43016-019-0001-5
- Kawaharada, Y., Nielsen, M. W., Kelly, S., ... Stougaard, J. (2017). Differential regulation of the Epr3 receptor coordinates membrane-restricted rhizobial colonization of root nodule primordia. *Nature Communications*. doi:10.1038/ncomms14534
- Kazmierczak, T., Yang, L., Boncompagni, E., ... Brouquisse, R. (2020). Legume nodule senescence: a coordinated death mechanism between bacteria and plant cells. *Advances in Botanical Research*. doi:10.1016/bs.abr.2019.09.013
- Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume-rhizobium mutualism. *Nature*. doi:10.1038/nature01931
- Kirchman, D. L. (2021). *Dead Zones*. *Dead Zones*. doi:10.1093/oso/9780197520376.001.0001

- Kondorosi, E., Mergaert, P., & Kereszt, A. (2013). A Paradigm for endosymbiotic life: Cell differentiation of rhizobium bacteria provoked by host plant factors. *Annual Review of Microbiology*. doi:10.1146/annurev-micro-092412-155630
- Kong, Z., & Glick, B. R. (2017). The Role of Plant Growth-Promoting Bacteria in Metal Phytoremediation. In *Advances in Microbial Physiology*. doi:10.1016/bs.ampbs.2017.04.001
- Lamouche, F., Bonadé-Bottino, N., Mergaert, P., & Alunni, B. (2019). Symbiotic efficiency of spherical and elongated bacteroids in the aescynomene-bradyrhizobium symbiosis. *Frontiers in Plant Science*. doi:10.3389/fpls.2019.00377
- Laranjo, M., Alexandre, A., & Oliveira, S. (2014). Legume growth-promoting rhizobia: An overview on the Mesorhizobium genus. *Microbiological Research*. doi:10.1016/j.micres.2013.09.012
- Lau, Y. H., Stirling, F., Kuo, J., ... Silver, P. A. (2017). Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Research*. doi:10.1093/nar/gkx415
- Laurence, M. E. E. (2006). Reviving dead zones. *Scientific American*. doi:10.1038/scientificamerican1106-78
- Lebeis, S. L., Paredes, S. H., Lundberg, D. S., ... Dangl, J. L. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*. doi:10.1126/science.aaa8764
- Lederberg, J., & Tatum, E. L. (1946). Novel Genotypes in Mixed Cultures of Biochemical Mutants of Bacteria. *Cold Spring Harbor Symposia on Quantitative Biology*.
- Ledermann, R., Schulte, C. C. M., & Poole, P. S. (2021). How rhizobia adapt to the nodule environment. *Journal of Bacteriology*. doi:10.1128/JB.00539-20
- Lehnert, N., Dong, H. T., Harland, J. B., Hunt, A. P., & White, C. J. (2018). Reversing nitrogen fixation. *Nature Reviews Chemistry*. doi:10.1038/s41570-018-0041-7
- Levy, A., Conway, J. M., Dangl, J. L., & Woyke, T. (2018). Elucidating Bacterial Gene Functions in the Plant Microbiome. *Cell Host and Microbe*. doi:10.1016/j.chom.2018.09.005
- Li, S. J., Hua, Z. S., Huang, L. N., ... Shu, W. S. (2014a). Microbial communities evolve faster in extreme environments. *Scientific Reports*, **4**, 1–9.
- Li, X., Feng, H., Wen, J., Dong, J., & Wang, T. (2018). MtCAS31 aids symbiotic nitrogen fixation by protecting the leghemoglobin MtLb120-1 under drought stress in *Medicago truncatula*. *Frontiers in Plant Science*. doi:10.3389/fpls.2018.00633
- Li, Z., Ma, Z., Hao, X., Rensing, C., & Wei, G. (2014b). Genes conferring copper resistance in *Sinorhizobium meliloti* CCNWSX0020 also promote the growth of *Medicago lupulina* in copper-contaminated soil. *Applied and Environmental Microbiology*, **80**(6). doi:10.1128/AEM.03381-13
- Lindstrom, K., & Martinez-Romero, M. E. (2002). International Committee on Systematics of Prokaryotes: Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*. *International Journal of Systematic and Evolutionary Microbiology*. doi:10.1099/ijs.0.02524-0
- Liu, C. W., & Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: An update. *Plants*. doi:10.3390/plants5030033
- Lodwig, E. M., Hosie, A. H. F., Bourdès, A., ... Poole, P. S. (2003). Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature*. doi:10.1038/nature01527
- Lodwig, E., & Poole, P. (2003). Metabolism of *Rhizobium* bacteroids. *Critical Reviews in Plant Sciences*. doi:10.1080/713610850
- Lu, M., Jiao, S., Gao, E., ... Wei, G. (2017). Transcriptome response to heavy metals in *Sinorhizobium meliloti*

- CCNWSX0020 reveals new metal resistance determinants that also promote bioremediation by *Medicago lupulina* in metal contaminated soil. *Applied and Environmental Microbiology*, **83**(20). doi:10.1128/AEM.01244-17
- Mahieu, S., Frérot, H., Vidal, C., ... Cleyet-Marel, J.-C. (2011). *Anthyllis vulneraria*/Mesorhizobium metallidurans, an efficient symbiotic nitrogen fixing association able to grow in mine tailings highly contaminated by Zn, Pb and Cd. *Plant and Soil*, **342**(1–2), 405–417.
- Marchetti, M., Jauneau, A., Capela, D., ... Masson-Boivin, C. (2014). Shaping Bacterial Symbiosis With Legumes by Experimental Evolution. *Molecular Plant-Microbe Interactions*, **27**(9), 956–964.
- Masson-Boivin, C., & Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia — the roots of a success story. *Current Opinion in Plant Biology*, **44**, 7–15.
- Mengoni, A., Schat, H., & Vangronsveld, J. (2010). Plants as extreme environments? Ni-resistant bacteria and Ni-hyperaccumulators of serpentine flora. *Plant and Soil*, **331**(1). doi:10.1007/s11104-009-0242-4
- Mergaert, P., Uchiumi, T., Alunni, B., ... Kondorosi, E. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.0600912103
- Miao, J., Zhang, N., Liu, H., Wang, H., Zhong, Z., & Zhu, J. (2018). Soil commensal rhizobia promote *Rhizobium etli* nodulation efficiency through CinR-mediated quorum sensing. *Archives of Microbiology*. doi:10.1007/s00203-018-1478-2
- Montiel, J., Arthikala, M. K., Cárdenas, L., & Quinto, C. (2016). Legume NADPH oxidases have crucial roles at different stages of nodulation. *International Journal of Molecular Sciences*. doi:10.3390/ijms17050680
- Okazaki, S., Kaneko, T., Sato, S., & Saeki, K. (2013). Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1302360110
- Oono, R., Schmitt, I., Sprent, J. I., & Denison, R. F. (2010). Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytologist*. doi:10.1111/j.1469-8137.2010.03261.x
- Pérez Carrascal, O. M., VanInsberghe, D., Juárez, S., Polz, M. F., Vinuesa, P., & González, V. (2016). Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. *Environmental Microbiology*. doi:10.1111/1462-2920.13415
- Petzold, C. J., Chan, L. J. G., Nhan, M., & Adams, P. D. (2015). Analytics for metabolic engineering. *Frontiers in Bioengineering and Biotechnology*. doi:10.3389/fbioe.2015.00135
- Phillips, R. P., Erlitz, Y., Bier, R., & Bernhardt, E. S. (2008). New approach for capturing soluble root exudates in forest soils. *Functional Ecology*. doi:10.1111/j.1365-2435.2008.01495.x
- Pierre, O., Engler, G., Hopkins, J., Brau, F., Boncompagni, E., & Hérouart, D. (2013). Peribacteroid space acidification: A marker of mature bacteroid functioning in *Medicago truncatula* nodules. *Plant, Cell and Environment*. doi:10.1111/pce.12116
- Pini, F., Frascella, A., Santopolo, L., ... Mengoni, A. (2012). Exploring the plant-associated bacterial communities in *Medicago sativa* L. *BMC Microbiology*, **12**(1), 1.
- Pini, F., Galardini, M., Bazzicalupo, M., & Mengoni, A. (2011). Plant-bacteria association and symbiosis: Are there common genomic traits in alphaproteobacteria? *Genes*. doi:10.3390/genes2041017
- Pladys, D., & Vance, C. P. (1993). Proteolysis during development and senescence of effective and plant gene-controlled ineffective alfalfa nodules. *Plant Physiology*. doi:10.1104/pp.103.2.379

- Porter, S. S., Chang, P. L., Conow, C. A., Dunham, J. P., & Friesen, M. L. (2017). Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic Mesorhizobium. *ISME Journal*. doi:10.1038/ismej.2016.88
- Puppo, A., Groten, K., Bastian, F., ... Foyer, C. H. (2005). Legume nodule senescence: Roles for redox and hormone signalling in the orchestration of the natural aging process. *New Phytologist*. doi:10.1111/j.1469-8137.2004.01285.x
- Rafińska, K., Pomastowski, P., Wrona, O., Górecki, R., & Buszewski, B. (2017). Medicago sativa as a source of secondary metabolites for agriculture and pharmaceutical industry. *Phytochemistry Letters*. doi:10.1016/j.phytol.2016.12.006
- Rajwar, A., Sahgal, M., & Johri, B. N. (2013). Legume-rhizobia symbiosis and interactions in agroecosystems. In *Plant Microbe Symbiosis: Fundamentals and Advances*. doi:10.1007/978-81-322-1287-4_9
- Regus, J. U., Gano, K. A., Hollowell, A. C., Sofish, V., & Sachs, J. L. (2015). Lotus hosts delimit the mutualism-parasitism continuum of Bradyrhizobium. *Journal of Evolutionary Biology*. doi:10.1111/jeb.12579
- Regus, J. U., Quides, K. W., O'Neill, M. R., ... Sachs, J. L. (2017). Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *American Journal of Botany*. doi:10.3732/ajb.1700165
- Rensing, C. (2006). Microorganisms in Soils: Roles in Genesis and Function. *Microbe Magazine*. doi:10.1128/microbe.1.152.1
- Reynders, L., & Vlassak, K. (1982). Use of Azospirillum brasilense as biofertilizer in intensive wheat cropping. *Plant and Soil*. doi:10.1007/BF02183980
- Roley, S. S., Duncan, D. S., Liang, D., ... Philip Robertson, G. (2019). Associative nitrogen fixation (ANF) in switchgrass (Panicum virgatum) across a nitrogen input gradient. *PLoS ONE*. doi:10.1371/journal.pone.0197320
- Roponen, I. (1970). The Effect of Darkness on the Leghemoglobin Content and Amino Acid Levels in the Root Nodules of Pea Plants. *Physiologia Plantarum*. doi:10.1111/j.1399-3054.1970.tb06435.x
- Sachs, J. L., Quides, K. W., & Wendlandt, C. E. (2018). Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *New Phytologist*. doi:10.1111/nph.15222
- Sachs, J. L., Russell, J. E., Lii, Y. E., Black, K. C., Lopez, G., & Patil, A. S. (2010). Host control over infection and proliferation of a cheater symbiont. *Journal of Evolutionary Biology*. doi:10.1111/j.1420-9101.2010.02056.x
- Sánchez-Cañizares, C., Jorrín, B., Poole, P. S., & Tkacz, A. (2017). Understanding the holobiont: the interdependence of plants and their microbiome. *Current Opinion in Microbiology*. doi:10.1016/j.mib.2017.07.001
- Sasse, J., Martinoia, E., & Northen, T. (2018). Feed Your Friends: Do Plant Exudates Shape the Root Microbiome? *Trends in Plant Science*. doi:10.1016/j.tplants.2017.09.003
- Schmid, R., Lewis, G., Schrire, B., Mackinder, B., & Lock, M. (2006). Legumes of the World. *Taxon*. doi:10.2307/25065563
- Selosse, M. A., Bessis, A., & Pozo, M. J. (2014). Microbial priming of plant and animal immunity: Symbionts as developmental signals. *Trends in Microbiology*. doi:10.1016/j.tim.2014.07.003
- Sergaki, C., Lagunas, B., Lidbury, I., Gifford, M. L., & Schäfer, P. (2018). Challenges and approaches in microbiome research: from fundamental to applied. *Frontiers in Plant Science*. doi:10.3389/fpls.2018.01205

- Sessitsch, A., Pfaffenbichler, N., & Mitter, B. (2019). Microbiome Applications from Lab to Field: Facing Complexity. *Trends in Plant Science*. doi:10.1016/j.tplants.2018.12.004
- Shridhar, B. S., Author, C., Shrimant Shridhar, B., Shridhar, B. S., Author, C., & Shrimant Shridhar, B. (2012). Nitrogen Fixing Microorganisms. *International Journal of Microbiological Research*. doi:10.5829/idosi.ijmr.2012.3.1.61103
- Simon, J. C., Marchesi, J. R., Mougel, C., & Selosse, M. A. (2019). Host-microbiota interactions: From holobiont theory to analysis. *Microbiome*. doi:10.1186/s40168-019-0619-4
- Singh, A., Kumar, M., Verma, S., Choudhary, P., & Chakdar, H. (2020). Plant Microbiome: Trends and Prospects for Sustainable Agriculture. In *Plant Microbe Symbiosis*. doi:10.1007/978-3-030-36248-5_8
- Smill, V., & Streatfeild, R. A. (2002). Enriching the earth: Fritz Haber, Carl Bosch, and the transformation of world food production. *Electronic Green Journal*. doi:10.2307/3985938
- Smith, B. E. (2002). Nitrogenase reveals its inner secrets. *Science*. doi:10.1126/science.1076659
- Smith, H. O., Hutchison, C. A., Pfannkoch, C., & Venter, J. C. (2003). Generating a synthetic genome by whole genome assembly: ϕ X174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.2237126100
- Sprent, J. I. (2007). Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation: Tansley review. *New Phytologist*. doi:10.1111/j.1469-8137.2007.02015.x
- Sprent, J. I., Ardley, J., & James, E. K. (2017). Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytologist*. doi:10.1111/nph.14474
- Stewart, W. D. (1969). Biological and ecological aspects of nitrogen fixation by free-living micro-organisms. *Proceedings of the Royal Society of London. Series B. Biological Sciences*. doi:10.1098/rspb.1969.0027
- Tighe, S., Afshinnkoo, E., Rock, T. M., ... Mason, C. E. (2017). Genomic methods and microbiological technologies for profiling novel and extreme environments for the extreme microbiome project (XMP). *Journal of Biomolecular Techniques*. doi:10.7171/jbt.17-2801-004
- Tilman, D., Fargione, J., Wolff, B., ... Swackhamer, D. (2001). Forecasting agriculturally driven global environmental change. *Science*. doi:10.1126/science.1057544
- Turner, S. L., & Young, J. P. W. (2000). The glutamine synthetases of rhizobia: Phylogenetics and evolutionary implications. *Molecular Biology and Evolution*. doi:10.1093/oxfordjournals.molbev.a026311
- Vadakattu, G., & Paterson, J. (2006). Free-living bacteria lift soil nitrogen supply. *Farming Ahead*.
- van Dam, N. M., & Bouwmeester, H. J. (2016). Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. *Trends in Plant Science*. doi:10.1016/j.tplants.2016.01.008
- Van De Velde, W., Zehirov, G., Szatmari, A., ... Mergaert, P. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*. doi:10.1126/science.1184057
- Vance, C. P. (1997). Enhanced Agricultural Sustainability Through Biological Nitrogen Fixation. In *Biological Fixation of Nitrogen for Ecology and Sustainable Agriculture*. doi:10.1007/978-3-642-59112-9_36
- Vance, C. P. (2000). Amide biosynthesis in root nodules of temperate legumes. In *Prokaryotic nitrogen fixation: a model system for the analysis of a biological process*.
- Vance, C. P. (2001). Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. *PLANT PHYSIOLOGY*. doi:10.1104/pp.127.2.390
- Vauclare, P., Bligny, R., Gout, E., de Meuron, V., & Widmer, F. (2010). Metabolic and structural rearrangement



during dark-induced autophagy in soybean (*Glycine max* L.) nodules: An electron microscopy and ³¹P and ¹³C nuclear magnetic resonance study. *Planta*. doi:10.1007/s00425-010-1148-3

- Vitousek, P. M., Menge, D. N. L., Reed, S. C., & Cleveland, C. C. (2013). Biological nitrogen fixation: Rates, patterns and ecological controls in terrestrial ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi:10.1098/rstb.2013.0119
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*. doi:10.1126/science.277.5325.494
- Wang, D., Yang, S., Tang, F., & Zhu, H. (2012). Symbiosis specificity in the legume - rhizobial mutualism. *Cellular Microbiology*. doi:10.1111/j.1462-5822.2011.01736.x
- Wani, P. A., Khan, M. S., & Zaidi, A. (2007). Effect of metal tolerant plant growth promoting Bradyrhizobium sp. (vigna) on growth, symbiosis, seed yield and metal uptake by greengram plants. *Chemosphere*. doi:10.1016/j.chemosphere.2007.07.028
- Watts, J. E. M. (2019). Welcome to Environmental Microbiome. *Environmental Microbiomes*. doi:10.1186/s40793-019-0340-8
- Werner, G. D. A., Cornwell, W. K., Sprent, J. I., Kattge, J., & Kiers, E. T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nature Communications*. doi:10.1038/ncomms5087
- West, S. A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002a). Sanctions and mutualism stability: Why do rhizobia fix nitrogen? *Proceedings of the Royal Society B: Biological Sciences*. doi:10.1098/rspb.2001.1878
- West, S. A., Toby Kiers, E., Pen, I., & Denison, R. F. (2002b). Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology*. doi:10.1046/j.1420-9101.2002.00441.x
- Westhoek, A., Clark, L. J., Culbert, M., ... Turnbull, L. A. (2021). Conditional sanctioning in a legume-Rhizobium mutualism. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.2025760118
- Wu, X., & Wang, D. (2019). The defective in nitrogen fixation genes of *Medicago truncatula* reveal key features in the intracellular association with rhizobia. In *The Model Legume Medicago truncatula*. doi:10.1002/9781119409144.ch74
- Xu, W., & Huang, W. (2017). Calcium-dependent protein Kinases in phytohormone signaling pathways. *International Journal of Molecular Sciences*. doi:10.3390/ijms18112436
- Yadav, A. N., Verma, P., Kumar, M., ... Saxena, A. K. (2015). Diversity and phylogenetic profiling of niche-specific Bacilli from extreme environments of India. *Annals of Microbiology*, **65**(2), 611–629.
- Young, J. P. W. (2016). Bacteria Are Smartphones and Mobile Genes Are Apps. *Trends in Microbiology*. doi:10.1016/j.tim.2016.09.002
- Zhao, C., Ni, H., Zhao, L., Zhou, L., Borrás-Hidalgo, O., & Cui, R. (2021). High nitrogen concentration alter microbial community in *Allium fistulosum* rhizosphere. *PLOS ONE*. doi:10.1371/journal.pone.0246163

Publications

Review

Deciphering the Symbiotic Plant Microbiome: Translating the Most Recent Discoveries on Rhizobia for the Improvement of Agricultural Practices in Metal-Contaminated and High Saline Lands

Agnese Bellabarba ¹, Camilla Fagorzi ² , George C. diCenzo ³, Francesco Pini ¹, Carlo Viti ¹ 
and Alice Checcucci ^{2,4,*} 

¹ Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, 50144 Florence, Italy

² Department of Biology, University of Florence, 50019 Sesto Fiorentino, Italy

³ Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada

⁴ Department of Agricultural and Food Science (DISTAL), University of Bologna, 40127 Bologna, Italy

* Correspondence: alice.checcucci2@unibo.it

Received: 1 July 2019; Accepted: 5 September 2019; Published: 10 September 2019



Abstract: Rhizosphere and plant-associated microorganisms have been intensely studied for their beneficial effects on plant growth and health. These mainly include nitrogen-fixing bacteria (NFB) and plant-growth promoting rhizobacteria (PGPR). This beneficial fraction is involved in major functions such as plant nutrition and plant resistance to biotic and abiotic stresses, which include water deficiency and heavy-metal contamination. Consequently, crop yield emerges as the net result of the interactions between the plant genome and its associated microbiome. Here, we provide a review covering recent studies on PGP rhizobia as effective inoculants for agricultural practices in harsh soil, and we propose models for inoculant combinations and genomic manipulation strategies to improve crop yield.

Keywords: soil bioremediation; high-salinity soil; plant beneficial microbes; rhizobia; microbial inoculants; plant-growth promoting rhizobacteria (PGPR)

1. Introduction

The last ten years have witnessed a number of discoveries and an increased awareness of the importance of the microbiome for the health and the growth of host macroorganisms [1,2]. Plants and their related microbiota can be considered holobionts, complex systems ruled by interdependent and composite interactions [3–5]. Indeed, plants are colonized by an astounding number of microorganisms that can reach numbers much greater than those of plant cells. This is especially relevant for the rhizosphere, the thin layer of soil surrounding and influenced by plant roots, where it is possible to observe a staggering diversity of microorganisms; a single gram of rhizospheric soil hosts tens of thousands of distinct microbial species [6,7]. Plants influence the composition of their rhizosphere microbiota through the production of root exudates [8], which differ in space and time [9], contributing to the positive selection of plant-growth promoting (PGP) and beneficial bacteria [10]. Indeed, it has been suggested that plants may have evolved the beneficial trait of secreting specific compounds to recruit protective microorganisms in response to pathogen attacks [11,12]. Plant-based bacterial selection relies on the microorganism already present in the soils where they are grown. Therefore, crop productivity could be increased by modifying root microbiota with microbial inoculants, which may be composed of a single strain or a consortia of different PGP rhizobacteria (PGPR) [13,14].

Within the beneficial plant microbiota, rhizobia constitute one of the most studied fraction [15]. The “rhizobium” definition is based on the ability to induce the formation of root/stem nodules in leguminous plants [16]. However, rhizobia are found not only on legumes, but are also found in association with several plant species [17]. Within legume nodules, rhizobia differentiate into bacteroids and synthesize a protein complex called nitrogenase that converts atmospheric dinitrogen to ammonia (biological nitrogen fixation, BNF) [18]. The produced ammonia is then transferred to the host plant to sustain its biosynthetic pathways [16]. The establishment of the symbiotic interaction between the nitrogen fixing rhizobia and leguminous plants is highly regulated and begins with mutual recognition between the plant and the rhizobia present in the rhizosphere. To date, rhizobia have been identified in two bacterial classes, the *Alphaproteobacteria* and the *Betaproteobacteria*. In the *Alphaproteobacteria*, rhizobial strains are present in the genera *Sinorhizobium* (syn. *Ensifer*), *Rhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Methylobacterium*, *Devosia*, *Ochrobactrum*, *Aminobacter*, *Microvirga*, *Shinella*, and *Phyllobacterium*. In the *Betaproteobacteria*, rhizobia are present within strains of the genera *Paraburkholderia*, *Cupriavidus*, and *Trinickia* [19–21]. While the beta-rhizobia are mainly found in association with tropical legumes, the alpha-rhizobia appear to be more widespread and nodulate tropical to temperate legumes including pasture, tree, and grain legumes. The alpha-rhizobia have received more research attention than the beta-rhizobia; among the alpha-rhizobia, *Sinorhizobium* (syn. *Ensifer*) is likely the most studied genus, followed by the genera *Rhizobium* and *Bradyrhizobium* [22,23].

In this review, we highlight the most recent studies on PGP rhizobia isolated from, and adapted to, drought-affected and metal-contaminated soils and their possible use as effective inoculants for legumes grown in harsh agricultural soils. We cover the identification of the genetic determinants of their tolerance, as well as the mechanisms that allow rhizobia to survive and to improve host plant growth in harsh soils. Models of inoculant combinations and genomic manipulation strategies for the improvement of crop yield are discussed.

2. The Need for Rhizobial Inoculants

The demand for plant proteins for human nutrition has increased tremendously over the last fifteen years. This can be related to: (i) Demographic growth and urbanization, (ii) the limited land areas that can be used for the production of food crops while farming systems are switching to specialized but unsustainable cereal production (for market competitiveness), and (iii) decreases in animal protein production due to shortage of irrigation and/or rainfall water especially in arid areas. The demand for plant proteins can be met in part through the cultivation of protein-rich leguminous crops. Additionally, legumes can help improve soil fertility through symbiotic nitrogen fixation, and they can help protect ground water from toxicity resulting from excessive application of N-fertilizers [24].

In the past three decades, eco-sustainable agronomic practices have been employed in an attempt to replace chemical fertilizers and pesticide-based agriculture [25,26]. Therefore, the exploitation of beneficial microorganisms as biofertilizer has become of primary importance [27]. In particular, rhizobial bioformulations could partially or completely substitute mineral nitrogen fertilizers [28,29]. Rhizobium-legume symbioses provide more than half of the world’s biologically fixed nitrogen [30], and it was reported that rhizobial nitrogen fixation introduces 40–48 million tonnes of nitrogen into agricultural systems each year [31]. The impact of BNF on the global agricultural economy was estimated to be worth the equivalent of USD160–180 billion [32]. Rhizobium inoculants are already widely used in agriculture, providing one of the most cost-effective ways to boost legume performances [33,34]. However, with a few exceptions, the last fifteen years has seen only small enhancements in the production of traditionally grown grain legumes such as fava bean, chickpea, lentils, or common beans [35]. Generally, yield instability is the main constraint for increasing plant productivity. Thus, special attention must be given to the factors that reduce soil quality and decrease plant yield.

Recent works have highlighted that microbial species associated with plants (rhizobial and non-rhizobial strains, including mycorrhizal fungi) can positively influence plant tolerance to water

deficiency [36]. This is due to their PGP features such as indoleacetic acid, siderophore production, phosphate and zinc solubilization [37], and the synthesis of 1-aminocyclopropane-1-carboxylate (ACC) deaminase [38], which are more evident and easily identified in stressful conditions [39]. Nevertheless, legumes are strongly affected by water deficit. In particular, BNF appears to be more sensitive to water deficit than other physiological functions such as photosynthesis or nutrient uptake [40]. Sometimes, this results in impaired nodule development [41] or the accumulation of small, generally organic, osmolytes called compatible solutes [42].

Aside from water deficiency and soil nutrient depletion, heavy-metal contamination due to anthropic activities (agricultural and industrial practices) or the weathering of metal-enriched rocks have recently increased exponentially, becoming a worldwide problem for crop productivity [43–45]. Generally, plant-associated microbes can contribute to a plant's ability to perform phytoextraction (accumulation of toxic compounds in the plant tissues) and phytostabilization (adsorption through the roots and conversion into harmless compounds). In legumes, phytostabilization is the key process when considering the phytoremediation of contaminated soils [46,47], and their associated rhizobia can promote chemical transformation and the chelation of heavy-metal compounds [48] (Figure 1). Therefore, plant growth and agricultural yield is related not only to the plant genotype and the soil condition, but, especially for legumes, their associated microbiota also play important roles [49,50]. As such, the selection of rhizobial strains resistant to water deficiency and capable of alleviating metal phytotoxicity could be a crucial strategy to improve the yield of legumes growth in arid or in metal-contaminated soils.

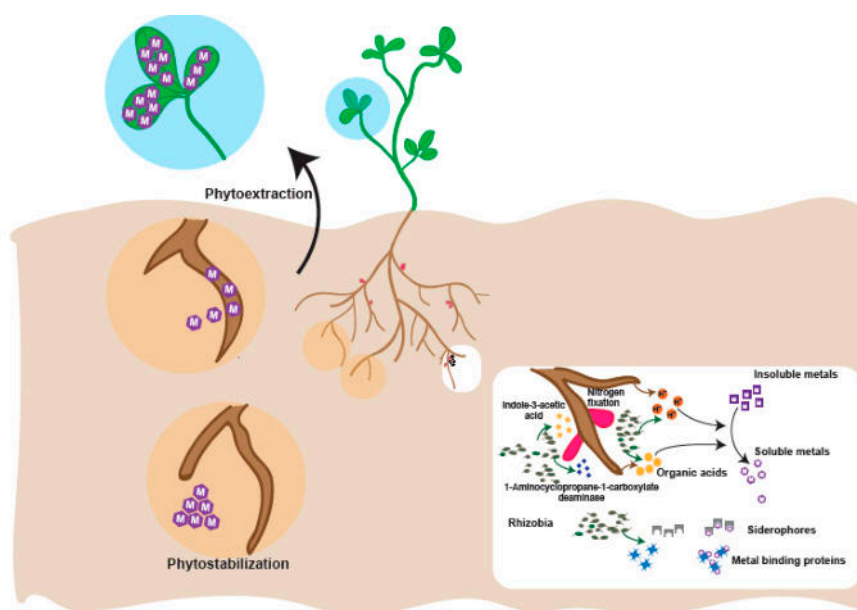


Figure 1. The mechanisms involved in bioremediation of heavy-metal contaminated soil and the contribution of PGP rhizobia.

3. Plant Growth Promoting Rhizobia in Saline and Harsh Soil

Salinity due to water deficiency is one of the largest environmental constraints for plant growth and productivity in stricken regions [51,52]. It was estimated that almost 40% of the world's lands can potentially become arid or semi-arid [52], most of which are located in the tropics and in the Mediterranean area [53,54]. The persistent increase of anthropic activities (such as poor agricultural practices) decreases soil water availability, alters the soil microbiota, and reduces the nutritive value of soils [55]. The progressive salinization of soil may cause several stresses to the plants that decrease their growth. This negatively affects crop productivity, hindering the agricultural economy of developing countries [56].

Desiccated soils can lead to alterations in plant homeostasis due to a reduction in the osmotic potential and inappropriate ionic distribution [56]. The alteration in nutrient supply and the resulting nutritional imbalance induces a loss of turgor pressure and the growth of smaller leaves [56]. Furthermore, increasing Na^+ and Cl^- concentrations commonly leads to the formation of burn-like lesions that alter leaf transpiration [57]. Continued growth interruption results in less efficient photosynthesis, respiratory variations, premature senescence, and the loss of cellular integrity leading to plant death [58,59]. Salt tolerance among legume species can fluctuate and is dependent on the chemical features of the soil, the climatic conditions, and the growth stage of the plant [53,60,61]. Tree legumes such as *Prosopis* [53,62] and *Acacia* spp. [53,63] are highly tolerant to salinity, as are the grain legumes *Glycine max* [51] and *Vicia faba* [61,64]. On the contrary, *Cicer arietinum*, *Phaseolus vulgaris*, and *Pisum sativum* are known to be extremely sensitive to salt stress [64–67]. BNF can also be strongly affected by a lack of water, influencing the mutual symbiotic interaction between the host plants and their associated rhizobia [53]. Generally, salt and osmotic stresses lead to a decrease in rhizobium nodulation, affecting the early stages of the symbiotic process [68–71]. This commonly includes poor root colonization by the bacteria and reduced curling and deformation of roots hairs [53,72,73]. Furthermore, a decrease in nitrogenase activity under drought/salinity stress is commonly attributed to a reduction in nodule respiration [67,74–76], which reduces the synthesis of cytosolic proteins such as leghemoglobin [67,76,77].

It was reported that host legumes are less tolerant to salt than their associated rhizobia [71,73,78]. For example, *Rhizobium leguminosarum* bv. *trifolii* TA-1 can tolerate up to 350 mM of NaCl in vitro [78]. The highest levels of salt tolerance seem to be associated with rhizobia isolated from woody legumes (e.g., *Prosopis*, *Acacia*, and *Leucaena*), which are tolerant of up to 850 mM of NaCl [73,79,80]. For example, Sakrouhi and collaborators [81] isolated 20 symbiotic N_2 -fixing bacteria from *Acacia tortilis* and *Acacia gummifera* that were able to grow in high salt media (400 mM). The intracellular accumulation of low-molecular-weight organic solutes, osmo-protectants, is an osmotic adaptation mechanism used by a large variety of bacterial species [53]. These compounds are acquired through de novo synthesis or uptake from the environment, and they can be accumulated to high intracellular concentrations without interfering with cellular processes [56], such as in the *S. meliloti* 102F34 strain [82–84]. The accumulation of poly-hydroxybutyrate (PHB) has also been reported as a protective measure to help rhizobia survive in high saline environments [85]. Following a decrease in osmolarity, the osmo-protectants are released by bacteria into the surrounding environment and actively recovered by plants, which are unable to synthesize them de novo [84]. The successful uptake of these compounds by the plants improve their growth under osmotic stress [56]. Additionally, intracellular trehalose accumulation by *R. leguminosuarum* seems to be involved in metabolic osmoregulation of the host plants [86,87]. The intracellular accumulation of glycine betaine was identified as one of the most frequent osmotic stress responses of rhizobia [88,89]. Several lines of evidence suggest that this process plays a role in maintaining bacteroid nitrogenase activity in *Medicago sativa* nodules [53,90–92].

Nowadays, the massive use of fertilizers to offset the effect of soil nutrient loss (also due to salinization) on crop productivity seems to be the preferred solution. This choice has progressively contributed to the deterioration of the soils that were already compromised by intense agricultural practices. An alternative is to make use of bacterial inoculants adapted to these harsh conditions. Studying the bacterial communities associated with plants growing in saline soils, and the underlying mechanism of their effectiveness, can be a good starting point for the use of microbial inoculant in agricultural practices to reduce saline stress [56,93]. To begin addressing this, the genomes of several rhizobia that nodulate plants in harsh environments were sequenced to identify stress-adaption genes. Examples include: *Rhizobium* sp. LCM 4573 (a salt-tolerant rhizobium from Senegalese soil [94]), *R. leucaenae* (a stress-tolerant species nodulating plants in tropical acid soils [95]), *S. meliloti* AK21 (from the Aral Sea Region that experiences saline and drought conditions [96]) and *Ensifer aridi* (isolated from arid soils of diverse deserts [97]). Transcriptomic and proteomic studies have been instrumental in highlighting how rhizobia respond to environmental stresses (a detailed list is provided

elsewhere [98]). Studies with *Mesorhizobium loti* suggested heat shock results in a global downregulation of protein expression possibly to conserve energy [99], while salt stress leads to over-expression of ABC transporters and genes associated with nucleotide transport and metabolism [100]. Similarly, characterization of the acid stress response in *S. meliloti* suggested that this stress results in an elevated metabolic respiration rate [101]. However, to date only a few studies have examined the responses of differentiated bacteroids to environmental stress [102,103], although several studies have examined the responses of the plant partner [104]. Additional studies characterizing the response of bacteroids to environmental stresses would be beneficial to complement the free-living datasets.

The agricultural significance of uncovering the genetic and metabolic basis of stress resistance in rhizobia and other PGP bacteria has been emphasized from at least the early 1990s [105]. In fact, numerous studies have demonstrated that genetically modifying rhizobia can increase or decrease their symbiotic abilities in stressful environments [106]. The grain yield of common bean plants grown in drought conditions was significantly higher when inoculated with a *Rhizobium etli* strain overexpressing *otsA*, encoding a trehalose-6-phosphate [107]. Similar results were obtained for maize plants inoculated with a non-rhizobium diazotroph *Azospirillum brasilense* strain overexpressing a trehalose biosynthesis gene [108]. In contrast, soybean plants inoculated with a *Bradyrhizobium japonicum putA* mutant that is unable to catabolize proline produced fewer seeds than plants inoculated with the wild-type parental strain when grown in moderate drought conditions [109]. *S. meliloti* strains overexpressing *betS* displayed improved nitrogen fixation phenotypes during salt stress [110], while salt-sensitive *Rhizobium tropici* mutants were poor symbionts even in the absence of stress [111]. Finally, ACC-deaminases encoded by some rhizobia can reduce the overproduction of the plant gas hormone ethylene during abiotic stresses [38,112], reducing the deleterious effect of ethylene and thus improving plant growth [113].

The inoculation of plants with microbial communities has also been shown to improve plant tolerance to environmental stresses. The co-inoculation of soybean (*Glycine max*) with *Chryseobacterium balustinum* Aur9 and *Ensifer (Sinorhizobium) fredii* SMH12 led to increased symbiotic performance under saline conditions (25 mM NaCl) [114]. In the same study, co-inoculation of common bean (*Phaseolus vulgaris* L.) with *R. tropici* CIAT899 and *C. balustinum* Aur9 enhanced bean growth in both saline (25 mM NaCl) and control conditions compared to single strain inoculation [114]. Moreover, co-inoculation of *Rhizobium phaseoli* M6 and M9, *Pseudomonas syringae* Mk1, *Pseudomonas fluorescens* Mk20, and *Pseudomonas fluorescens* biotype G Mk25 strains decreased the effects of salinity stress in bean, enhancing its nodulation process in vitro and in fields conditions [115,116].

4. Plant Growth Promoting Rhizobia in Heavy Metal Contaminated Soil

Anthropogenic activities, such as the use of fertilizers and pesticides in agricultural soils, the production of sewage sludge waste, and industrial and mining activities, are responsible for the accumulation of toxic heavy metals in the food chain [117]. Low concentrations of metals such as zinc (Zn), copper (Cu), iron (Fe), nickel (Ni), manganese (Mn), molybdenum (Mo) and cobalt (Co) are necessary for the metabolism of all organisms [118]. However, high concentrations of these metals, as well as the long term persistence in the soil of elements such as cadmium (Cd), lead (Pb), and arsenic (As), negatively affect the composition of microbial communities [119], the dynamics of the rhizosphere niche [120], and the growth, the biomass, and the photosynthesis of plants [121]. Plant species used for the remediation of heavy metal polluted sites represent an environment-friendly, aesthetically appealing, and cost-effective solution. Legumes may be ideal species for bioremediation as surveys of plant species surviving in long-term metal-contaminated environments have shown legumes to account for a dominant portion of these populations [122]. The metal tolerant plant species used for bioremediation have developed several mechanisms that allow them to thrive in these contaminated environments and to accumulate high concentrations of specific metals in the aboveground tissue. Among these, both enzymatic and non-enzymatic molecular mechanisms have been described. Heavy-metal stressed plants may protect themselves from reactive oxygen species through the production of antioxidant

enzymes or scavenger compounds [123]. Recently, PGP bacteria, including rhizobia, have been shown to reduce the toxicity of plant exposure to heavy metals [124,125]. Heavy-metal contaminated soil remediation can be performed with different strategies [126]. Plants able to decrease the mobility and/or the bioavailability of metals can be used in both phytostabilization and phytoimmobilization to prevent their leaching into ground water or their entry into the food chain. Mechanisms involved in this process include adsorption by roots, and the precipitation and complexation of the metals in the root zone [127]. Phytovolatilization involves the conversion of a metal (i.e., Hg as the mercuric ion) into the volatile form and its release into the atmosphere through the stomata [128]. However, the most important phytoremediation approach for removal of metals and metalloids from contaminated soils, water, and sediments is phytoextraction [129–131]. The main contribution of rhizobia towards phytostabilization and phytoimmobilization is plant growth enhancement [132]. Bacteria that nodulate their hosts may increase metal accumulation in root nodules, while those that remain in the rhizosphere would reduce metal toxicity locally by precipitation, chelation, immobilization, and biosorption. The nodule itself has an important role in metal-resistance: Once the symbiosis is established, nodules could serve as storage areas that provide plants an extra place to stock metals and reduce the risk of direct exposure [133].

Accelerating the phytoremediation of metalliferous soils by increasing mobilization and phytoextraction of heavy metals though the metabolic activity of rhizobia is a well-known practice [37,131,134–137]. Currently, the most studied metal resistance mechanisms in microorganisms include metal exclusion, protein binding-mediated extra- and intra-cellular sequestration, enzymatic detoxification, active transport of the metal, passive tolerance, and reduction in metal sensitivity of the cellular targets [138,139]. Bacteria can also contribute to phytoremediation through the production of extracellular polymeric substances. For example, studies of the interaction between metals and extracellular polymeric substances demonstrated that biosorption can reduce heavy metal contamination of wastewater systems [140,141]. Unlike salt-tolerant bacteria, there have been numerous studies on the use of bacteria isolated from metal contaminated soil as inoculants to promote plant growth in contaminated environments [37,48]. Although not all rhizobia are intrinsically tolerant to metals, metal-tolerant strains of taxonomically diverse rhizobia have been isolated from various plants in heavy metal contaminated environments [37,142]. Metal resistance determinants provide protection for rhizobia to survive and maintain effective nodulation of legumes, allowing them to play a role in promoting plant growth. In addition, the existence of a symbiotic relationship may provide protection for the survival of rhizobia in soils with elevated metal concentrations [143,144].

Arsenic toxicity, and the oxidative damage that it produces in cells through the overproduction of reactive oxygen species, affects DNA, proteins, and lipids. This provokes chlorosis, necrosis, delays in flowering, and a reduction in yield [145]. PGP rhizobia may play a beneficial role in protecting plants from arsenic contamination. This can be accomplished by stimulating the antioxidant enzymatic activities in plants, and stabilizing heavy metals and metalloids thereby reducing their accumulation in aerial organs [146–148]. For this reason, the use of PGP rhizobia in heavy metal and metalloid contaminated soils should not only promote the growth of the plant but should also immobilize and decrease the concentration of these elements in plant organs to reduce human exposure to toxic concentrations [149,150].

The presence of heavy metals can also influence the results of inoculant treatment of crops. For example, inoculation of soybean plants with two different strains of *Bradyrhizobium*, *B. diazoefficiens* USDA110 and *Bradyrhizobium* sp. Per 3.61, was studied in the Córdoba province of Argentina where arsenic contamination of groundwater is a consistent environmental problem [151]. The results demonstrated that only *B. diazoefficiens* USDA110 could nodulate soybean at the highest tested As(V) concentrations, while *Bradyrhizobium* sp. Per 3.61 was the better symbiont in the presence of low As(V) concentrations as it limited the translocation of the metal to the legume aerial compartments [151]. Numerous studies have also examined the effect of plant inoculation with pairs of PGP bacteria [152–155]. For instance, the co-inoculation of soybean with *B. japonicum* E109 and *Azospirillum brasilense* Az39

influenced plant growth and arsenic phytostabilization in arsenic contaminated conditions [156]. Furthermore, it was observed that the indole acetic acid (IAA) produced by *A. brasilense* Az39 had a protective effect on *B. japonicum* E109 when exposed to arsenic [156]. *B. japonicum* strains can use IAA as a carbon source, which seems to serve as a signal to coordinate bacterial behaviour to enhance protection under adverse conditions [157]. The presence of high levels of lead in soil is toxic for plants, resulting in chlorosis, blackening of roots, and reduced growth [158]. A study using *Brassica juncea* showed that the inoculation of autochthonous PGP rhizobial strains can alleviate the harmful effects of lead exposure. *Sinorhizobium* sp. Pb002 was isolated from the rhizosphere of *B. juncea* grown in Pb-contaminated soil [159]. In a microcosm experiment, the presence of strain Pb002 stimulated biomass formation by *B. juncea* and increased plant survival and lead uptake [159]. Nickel and zinc are essential elements for plant growth; however, excessive amounts of these nutrients can be toxic [160,161]. This toxicity can be alleviated, at least in part, through rhizobium inoculation. For example, inoculation of green gram plants with *Bradyrhizobium* sp. (vigna) RM8 or *Rhizobium* sp. RP5 increased both seed yield and grain protein in the presence of excessive nickel or zinc [162,163]. The presence of Cd in soil can impair plant growth due to a reduction in chlorophyll content and photosynthesis [164]. Moreover, Cd alters the cell redox potential and increases the amount of reactive oxygen species in the cell, which in turn negatively impacts cell membranes and biomolecules [164,165]. *Bradyrhizobium* sp. YI-6, isolated from *G. max* nodules grown in Chinese Cd-contaminated soil, displayed an ability to increase mineral nutrient (Fe) uptake while reducing Cd accumulation [166].

Researchers have identified potential metalloid stress-adaptation genes in rhizobia and they have investigated their transcriptional responses. Putative nickel adaptation genes were identified using association mapping with 47 symbiotic *Mesorhizobium* strains isolated from either nickel-enriched serpentine soils or nearby non-serpentine soils [167]. The identified genes included several transporters, an opine dehydrogenase, and an exopolysaccharide export protein, among others [166]. Additionally, investigation of the transcriptional response of *S. meliloti* strain CCNWSX0020 upon exposure to copper or zinc stress allowed the identification of several upregulated genes, including four genes (*yedYZ*, *fixH*-like, *cusA*-like, and *cueO*) whose mutation impaired either early or late symbiotic processes [168]. As for rhizobia colonizing saline and arid environments, genetically modified strains can have different symbiotic abilities. For example, a *M. amorphae* Δ *copA* deletion mutant displayed impaired symbiotic capabilities in copper contaminated soils, whereas overexpression of a flavodoxin gene in *S. meliloti* led to a more efficient symbiosis under cadmium stress [169].

Together, the studies discussed in this review (and summarized in Table 1) highlight how the rhizobial genotype can strongly influence symbiotic effectiveness and the plant response in harsh environments. Fully elucidating the genetic and molecular bases of these phenotypes would lay a strong foundation to aid the development of improved bio-inoculants, either through genetic engineering or the rational selection of optimal wild isolates.

Table 1. Studies of plant growth improvement mediated by rhizobium-inoculants on harsh soil.

Strain	Isolation Conditions		Crop	Effect/Action Mechanism	Reference
	Site	Metal Contamination			
<i>Bradyrhizobium diazoefficiens</i> USDA110	Ref. strain	As	Soybean	Limits metalloid translocation and accumulation in edible parts of the legume	[151]
<i>Bradyrhizobium</i> sp. Per 3.61	Nodules of soybeans	As			
<i>A. brasilense</i> Az39		As	Soybean	Enhances growth of the plant and phytostabilization of As when co-inoculated	[156]
<i>B. japonicum</i> E109		As			
<i>Sinorhizobium</i> sp. Pb002	Rhizosphere of <i>Brassica juncea</i>	Cd	<i>Brassica juncea</i>	Increases plant survival and lead uptake	[159]

Table 1. Cont.

Strain	Isolation Conditions		Crop	Effect/Action Mechanism	Reference
	Site	Metal Contamination			
<i>Bradyrhizobium</i> sp. (<i>vigna</i>) RM8 <i>Rhizobium</i> sp. RP5	Nodules of greengram	Ni, Zn	Greengram	Increases the number of nodules on the plant, as well as IAA and siderophore production	[163]
<i>Bradyrhizobium</i> sp.	Nodules of Glycine max	Cd	Glycine max	Increases mineral nutrient uptake (Fe) and reduces cadmium accumulation in the plant	[166]
<i>Rhizobium leguminosarum</i> (LR-30), <i>Mesorhizobium ciceri</i> (CR.30 and CR-39) and <i>Rhizobium phaseoli</i> (MR-2)	<i>Lens culinaris</i> L., <i>Vigna radiata</i> L., <i>Cicer arietinum</i> L.		Wheat (Triticum aestivum)	Improves plant growth, biomass, and drought stress through production of catalase, IAA, and exopolysaccharides under PEG-6000 simulated drought conditions	[170]
Strain	Isolation conditions		Crop	Effect/Mechanism of action	Reference
<i>Azospirillum</i> Sp245	Surface-sterilized wheat roots of (Brazil) [171]		Lettuce (<i>L. sativa</i> L.)	Promotes aerial biomass, higher ascorbic acid content accompanied by a lower oxidation rate, better overall visual quality due to higher chlorophyll content, hue, Chroma, L and lower browning intensity	[172]
<i>Sinorhizobium medicae</i> WSM419	Nodules of <i>Medicago murex</i>		<i>Medicago truncatula</i>	Delays stress-induced leaf senescence and abscission and nutrient acquisition during drought stress	[173]
<i>Sinorhizobium meliloti</i> A2 strains	Commercial strain, Eastern Canada [174]		<i>Medicago sativa</i> cv Apica, <i>Medicago sativa</i> cv Halo	Increases shoot/root ratio, shoot water content, and the concentrations of starch and pinitol in nodules	[175]
<i>Sinorhizobium meliloti</i> Rm1521	Ottawa vicinity [176]				
<i>Rhizobium etli</i> CE3 overexpressing trehalose-6-phosphate synthase gene	CFN42 derivate [177], original isolate from <i>P. vulgaris</i> nodule		<i>P. vulgaris</i> var. Negro Jamapa	Enhances drought tolerance due to upregulation of genes involved in stress tolerance, carbon and nitrogen metabolism by trehalose	[107]
Co-inoculation			Crop	Effect/Mechanism of action	Reference
Rhizobia	PGPRs				
<i>Rhizobium phaseoli</i> M6; M9, <i>phaseoli</i> M6 and M9 and PGPR	<i>Pseudomonas syringae</i> , Mk1; <i>Pseudomonas fluorescens</i> Mk20; <i>Pseudomonas fluorescens</i> biot. G Mk25		<i>Vigna radiata</i> L.	Decreases damaging effect of salinity stress on mung bean growth	[116]

5. Development of Rhizobial Inoculants

The use of rhizobial bioinoculants began in the USA at the end of 19th century, where soil containing naturally-occurring rhizobia was mixed with seeds. Since then, rhizobium inoculation has become a common practice to improve crop production [178]. Since the first marketed rhizobium biofertilizer “Nitragin”, which was developed by Nobbe and Hiltner in 1896, rhizobial bioformulations have improved dramatically. From the second half of the 19th century, liquid inoculants formulation [179] moved initially to freeze-dried inoculant lyophilization [180], and then to gel-based products such as polyacrylamide (PER) [181], alginate (AER), or xanthan (XER) [27,182]. Over the last 30 years, a huge number of formulations have been patented and commercialized; examples include vermiculite-based Gold Coat™ *Rhizobium* inoculant [183], liquid seed applied soybean inoculant Cell-Tech® [184], liquid in-furrow inoculant LIFT, and air-dried clay powder for alfalfa Nitragin® Gold [27,184]. In 1997, the first marketing of a genetically engineered *S. meliloti* strain RMBPC-2 was approved [185].

Eventually, additives and cell protectant-based liquid inoculant formulations were developed that increased cell survival through the use of compounds such as the polymer polyvinyl pyrrolidone (PVP), carboxymethyl cellulose (CMC) [186], gum Arabic [142], sodium alginate [182], and glycerol [187]. The choice of inoculant carriers that can promote the long-lasting maintenance and protection of viable microbial cells is a global issue [24,188]. Peat is currently the most common organic carrier material for bioformulation production [189], especially in North and South America, Europe, and Australia [190],

although other materials (such as coal, bagasse, coir, dust, and perlite) are also used [191]. Supported by successful in vitro experiments, Albareda and collaborators suggested the use of broth culture media as a rhizobial carrier for soybean cultivation [191]. It was demonstrated that after 3 months of storage, liquid cultures were able to maintain more than 10⁹ cfu/mL of *Sinorhizobium fredii* SMH12 and *Bradyrhizobium japonicum* USDA110 [191]. Therefore, aqueous-, oil-, or polymer-based liquid formulations have increased enormously in recent years (for details see [190,192–194]). Additionally, the inclusion of microbial or plant secondary metabolites, such as flavonoids and phytohormones, has become a common practice in bioformulation preparation to improve the efficiency of the inoculants [25,195–197].

6. Inoculant Combinations and Phenomic Strategies for Improving Crop Yield

Screening for rhizobia with high nitrogen fixation rates is performed in many laboratories; however, the use of effectiveness as the sole criterion for rhizobia selection may not always be the most relevant criterion for field applications [75]. Indeed, in the soil the rhizobia have to overcome many different adverse conditions (pH, desiccation, nutrient deficiencies, salinity/alkalinity, extreme temperatures, toxicities) [75,198] and they have to outcompete other rhizobial strains [199,200]. It is thought that there is generally an alignment between the fitness of rhizobia and the fitness of their host plants [201]. However, rhizobia are not vertically transmitted but are instead soil bacteria that colonize plant roots. Therefore, ineffective or less effective rhizobia can become abundant and outcompete more effective strains. Moreover, a single plant can be infected by multiple strains with different nitrogen fixation efficiency [202,203]. Recent data suggested that legumes cannot discriminate between effective and ineffective strains prior to infection [204]. Instead, legumes limit the loss of resources by sanctioning individual nodules containing ineffective strains [205]. However ineffective strains may escape from plant sanctions by co-infecting nodules together with effective strains [202].

One way to overcome these competition issues is to select or create highly competitive strains [3]. Rhizobial symbiosis genes (*nod*, *nif*, and *fix*) are generally located on chromosomal mobile elements or on symbiosis plasmids [206]. Taking advantage of these features, it is possible to create hybrid strains without the insertion of exogenous DNA. For example, a hybrid strain of *S. meliloti* was recently created by moving the pSymA megaplasmid (accounting for nearly 20% of total genome content) from a donor *S. meliloti* strain to an acceptor strain [207]. Interesting, the resulting cis-hybrid strain seemed to exhibit a cultivar-specific improvement in symbiotic properties, compared to the parental strains, in controlled laboratory conditions [207]. Similarly, the transfer of symbiotic plasmids between different *R. leguminosarum* strains improved various measures of symbiotic efficiency in laboratory settings [208–211]. Therefore, genome-wide replicon-based remodeling of bacterial strains, potentially supported with a metabolic modelling framework [212], could be a powerful tool in precision agriculture by creating highly efficient strains depending of the farm/soil features [213]. This “Natural Genome Assisted Breeding” approach, based on the transfer of replicons among different strains, will also prevent the introduction of non-natural genes into the environment (Figure 2).

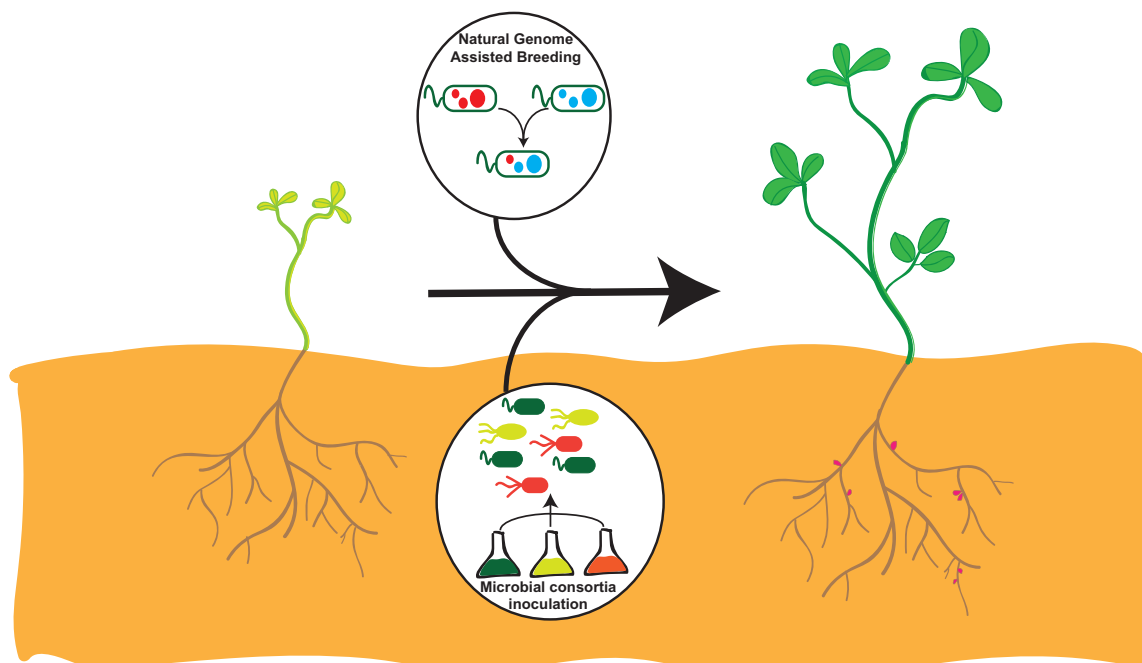


Figure 2. Different approach aimed to promote plant growth: Natural genome assisted breeding as a genomic manipulation strategy, and microbial consortia (multi-strains combination) for the improvement of plant-specific rhizobial inoculants in harsh environments.

One of the major issues related to the application of rhizobium inoculants is the inclusion of other additives (e.g., fungicides, nutrients [193], and fertilizers) that may reduce the viability or effectiveness of rhizobia. A recent trend to overcome this issue is to use microbial consortia instead of single strain biostimulants [214]. Consortia are formed by a combination of bacterial and/or fungal species to cover a broader spectrum of usage and soil conditions [214]. Microbes within a consortium are better able to handle biotic and abiotic stresses as they may work synergistically by exchanging nutrients and removing toxic compounds [215,216]. Using rhizobia in combination with other PGPR may improve their effect; for example, it has been reported that the combination of *Rhizobium* strains with *Bacillus* strains can improve root structure and increase nodule formation in bean, pigeon pea, and soybean (see [217] and references therein). Other well documented examples of mixed inoculants involving rhizobia and PGPR strains that led to improved symbiotic phenotype are: *Rhizobium* with *Bacillus subtilis* and *Bacillus megaterium*; *Rhizobium tropici* with *Chryseobacterium balustinum*, *Bacillus atrophaeus*, and *Burkholderia cepacia*; *Mesorhizobium* with *Pseudomonas*; *Mesorhizobium* in combination with *Azotobacter chroococcum*, *Pseudomonas aeruginosa*, and *Trichoderma harzianum*; and *S. meliloti* with a consortia of *Burkholderia* spp. (see [217] and references therein). Increased growth promotion could be due to a direct effect on nodulation, the production of phytohormones, or enhanced resistance to crop diseases [216,217]. Additionally, the synergistic interaction between arbuscular mycorrhiza and rhizobia for enhancing crop yield through improving nutrients uptake has also been heavily investigated [217]. Indeed, arbuscular mycorrhiza symbiosis can increase rhizobium nodulation of legumes under control [218] and saline conditions [219]. The consortium can also improve the uptake and transfer of nitrogen in a soybean/maize inter-cropping system [220].

Overall, the use of consortia composed of rhizobia and other PGPR combined with recent advances in rhizobium genomic manipulation could lead to increased inoculum efficiency in field conditions.

7. Concluding Remarks

In recent years, many studies were focused on the development and the optimization of technologies for the improvement of sustainable crop production, in particular in harsh (arid and/or metal-contaminated) environments. The studies were mainly spurred by an increasing requirement

for plant proteins, which is due to the increasing worldwide human population and the need to reduce the use of chemical fertilizers. The demand for increasing plant protein production and reducing the use of fertilizers can be accomplished, in part, through the cultivation of legumes; legumes are rich in protein and are able to improve soil fertility through BNF performed by their associated rhizobia. A huge number of studies conducted in the last thirty years have highlighted the ability of rhizobia to colonize particularly harsh soils, and to promote the growth of the leguminous plants to which they are associated. In this review, we summarized the most recent and detailed literature on plant growth promoting rhizobia isolated from (and thus, adapted to) arid and heavy metal contaminated soils, as well as their possible use as inoculants for legume-based agriculture in harsh soils. Despite the current knowledge on the topic, which ranges from genetic to molecular mechanisms, further research should be conducted on the feasibility of plant or soil specific rhizobia-based inoculations. Here, we proposed genetic manipulation strategies, which simulate natural evolution, and strain combination to optimize plant-specific rhizobial inoculants for the improvement of crop yield.

Author Contributions: A.C., conceptualization, original draft preparation, review, editing; A.B., C.F., original draft preparation, literature review sections; G.C.d., original draft preparation, review, editing; F.P., original draft preparation, figures conceptualization; C.V., review, editing. All authors contributed to critically revising the manuscript and gave final approval for publication.

Funding: This research was partially supported by project METATRACK granted by FONDAZIONE CASSA DI RISPARMIO FIRENZE. AC was partially supported by Fondazione Adriano Buzzati-Traverso.

Acknowledgments: The authors are grateful to Alessio Mengoni for critical reading of the manuscript.

Conflicts of Interest: The authors declare that the review was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Vandenkoornhuyse, P.; Quaiser, A.; Duhamel, M.; Le Van, A.; Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol.* **2015**, *206*, 1196–1206. [[CrossRef](#)] [[PubMed](#)]
2. Turnbaugh, P.J.; Ley, R.E.; Hamady, M.; Fraser-Liggett, C.M.; Knight, R.; Gordon, J.I. The human microbiome project: Exploring the microbial part of ourselves in a changing world. *Nature* **2007**, *449*, 804–810. [[CrossRef](#)] [[PubMed](#)]
3. Checcucci, A.; DiCenzo, G.C.; Bazzicalupo, M.; Mengoni, A. Trade, diplomacy, and warfare: The quest for elite rhizobia inoculant strains. *Front. Microbiol.* **2017**, *8*, 2207. [[CrossRef](#)] [[PubMed](#)]
4. Hassani, M.A.; Durán, P.; Hacquard, S. Microbial interactions within the plant holobiont. *Microbiome* **2018**, *6*, 58. [[CrossRef](#)] [[PubMed](#)]
5. Sánchez-Cañizares, C.; Jorrín, B.; Poole, P.S.; Tkacz, A. Understanding the holobiont: The interdependence of plants and their microbiome. *Curr. Opin. Microbiol.* **2017**, *38*, 188–196. [[CrossRef](#)] [[PubMed](#)]
6. Gans, J.; Wolinsky, M.; Dunbar, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **2005**, *309*, 1387–1390. [[CrossRef](#)] [[PubMed](#)]
7. Schloss, P.D.; Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput. Biol.* **2006**, *2*, e92. [[CrossRef](#)] [[PubMed](#)]
8. Hartmann, A.; Schmid, M.; van Tuinen, D.; Berg, G. Plant-driven selection of microbes. *Plant Soil* **2009**, *321*, 235–257. [[CrossRef](#)]
9. Pini, F.; East, A.K.; Appia-Ayme, C.; Tomek, J.; Karunakaran, R.; Mendoza-Suárez, M.; Edwards, A.; Terpolilli, J.J.; Roworth, J.; Downie, J.A. Bacterial biosensors for in vivo spatiotemporal mapping of root secretion. *Plant Physiol.* **2017**, *174*, 1289–1306. [[CrossRef](#)] [[PubMed](#)]
10. Edwards, J.A.; Santos-Medellín, C.M.; Liechty, Z.S.; Nguyen, B.; Lurie, E.; Eason, S.; Phillips, G.; Sundaresan, V. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol.* **2018**, *16*, e2003862. [[CrossRef](#)] [[PubMed](#)]
11. Thijs, S.; Sillen, W.; Rineau, F.; Weyens, N.; Vangronsveld, J. Towards an enhanced understanding of plant–microbiome interactions to improve phytoremediation: Engineering the metaorganism. *Front. Microbiol.* **2016**, *7*, 341. [[CrossRef](#)] [[PubMed](#)]

12. Berendsen, R.L.; Pieterse, C.M.; Bakker, P.A. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **2012**, *17*, 478–486. [[CrossRef](#)] [[PubMed](#)]
13. Mueller, U.G.; Sachs, J.L. Engineering microbiomes to improve plant and animal health. *Trends Microbiol.* **2015**, *23*, 606–617. [[CrossRef](#)] [[PubMed](#)]
14. Trivedi, P.; Schenk, P.M.; Wallenstein, M.D.; Singh, B.K. Tiny microbes, big yields: Enhancing food crop production with biological solutions. *Microb. Biotechnol.* **2017**, *10*, 999–1003. [[CrossRef](#)] [[PubMed](#)]
15. Pini, F.; Galardini, M.; Bazzicalupo, M.; Mengoni, A. Plant-bacteria association and symbiosis: Are there common genomic traits in *Alphaproteobacteria*? *Genes (Basel)* **2011**, *2*, 1017–1032. [[CrossRef](#)]
16. Brill, W.J. Genetics of N₂-fixing organisms. In *The Biology of Nitrogen Fixation*; Quispel, A., Ed.; North-Holland Pub. Co.: Amsterdam, The Netherlands, 1974.
17. Chi, F.; Shen, S.-H.; Cheng, H.-P.; Jing, Y.-X.; Yanni, Y.G.; Dazzo, F.B. Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. *Appl. Environ. Microbiol.* **2005**, *71*, 7271–7278. [[CrossRef](#)] [[PubMed](#)]
18. Vitousek, P.M.; Menge, D.N.L.; Reed, S.C.; Cleveland, C.C. Biological nitrogen fixation: Rates, patterns and ecological controls in terrestrial ecosystems. *Philos. Trans. R. Soc. B Biol. Sci.* **2013**, *368*, 20130119. [[CrossRef](#)]
19. Checcucci, A.; Perrin, E.; Bazzicalupo, M.; Mengoni, A. Genomic diversity and evolution of rhizobia. In *Microbial Diversity in the Genomic Era*; Das, S., Dash, H.R., Eds.; Elsevier Science BV: Amsterdam, The Netherlands, 2019; pp. 37–46.
20. Estrada-de Los Santos, P.; Palmer, M.; Steenkamp, E.T.; Maluk, M.; Beukes, C.; Hirsch, A.M.; James, E.K.; Venter, S.N. *Trinickia dabaoshanensis* sp. nov., a new name for a lost species. *Arch. Microbiol.* **2019**, 1–4. [[CrossRef](#)]
21. Estrada-de los Santos, P.; Palmer, M.; Chávez-Ramírez, B.; Beukes, C.; Steenkamp, E.; Briscoe, L.; Khan, N.; Maluk, M.; Lafos, M.; Humm, E.; et al. Whole genome analyses suggests that *Burkholderia* sensu lato contains two additional novel genera (*Mycetohabitans* gen. nov., and *Trinickia* gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the *Burkholderiaceae*. *Genes* **2018**, *9*, 389. [[CrossRef](#)]
22. Sprent, J.I.; Ardley, J.; James, E.K. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol.* **2017**, *215*, 40–56. [[CrossRef](#)]
23. Dwivedi, F.; Upadhyaya, K.L.; Galardini, A.; Bazzicalupo, M.; Biondi, M.; Hungria, E.G.; Kaschuk, M.; Blair, G.; Dwivedi, S.L.; Sahrawat, K.L.; et al. Advances in host plant and rhizobium genomics to enhance symbiotic nitrogen fixation in grain legumes. In *Advances in Agronomy*; Sparks, D.L., Ed.; Elsevier Science BV: Amsterdam, The Netherlands, 2015; Volume 129, pp. 1–116.
24. Galloway, J.N.; Townsend, A.R.; Erismann, J.W.; Bekunda, M.; Cai, Z.; Freney, J.R.; Martinelli, L.A.; Seitzinger, S.P.; Sutton, M.A. Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. *Science* **2008**, *320*, 889–892. [[CrossRef](#)] [[PubMed](#)]
25. Alori, E.T.; Dare, M.O.; Babalola, O.O. Microbial inoculants for soil quality and plant health. In *Sustainable Agriculture Reviews*; Lichtfouse, E., Ed.; Springer: Cham, Switzerland, 2017; Volume 22, pp. 281–307.
26. Carvajal-Muñoz, J.S.; Carmona-García, C.E. Benefits and limitations of biofertilization in agricultural practices. *Livest. Res. Rural Dev.* **2012**, *24*, 1–8.
27. Bhardwaj, D.; Ansari, M.W.; Sahoo, R.K.; Tuteja, N. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. *Microb. Cell Fact.* **2014**, *13*, 66. [[CrossRef](#)] [[PubMed](#)]
28. Arora, N.K.; Verma, M.; Mishra, J. Rhizobial bioformulations: Past, present and future. In *Rhizotrophs: Plant Growth Promotion To Bioremediation*; Mehnaz, S., Ed.; Springer: Singapore, 2017; Volume 2, pp. 69–99.
29. Tairo, E.V.; Ndakidemi, P.A. Possible benefits of rhizobial inoculation and phosphorus supplementation on nutrition, growth and economic sustainability in grain legumes. *Am. J. Res. Commun* **2013**, *1*, 532–556.
30. Smil, V. *Enriching the Earth: Fritz Haber, Carl Bosch and the Transformation of World Food Production*; The MIT Press: Cambridge, MA, USA, 2001.
31. Herridge, D.F.; Peoples, M.B.; Boddey, R.M. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **2008**, *311*, 1–18. [[CrossRef](#)]
32. Rajwar, A.; Sahgal, M.; Johri, B.N. Legume–rhizobia symbiosis and interactions in agroecosystems. In *Plant Microbe Symbiosis: Fundamentals and Advances*; Arora, N., Ed.; Springer: New Delhi, India, 2013; pp. 233–265.
33. Deshwal, V.K.; Singh, S.B.; Kumar, P.; Chubey, A. Rhizobia unique plant growth promoting rhizobacteria: A review. *Int. J. Life Sci.* **2013**, *2*, 74–86.

34. Xavier, I.J.; Holloway, G.; Leggett, M. Development of rhizobial inoculant formulations. *Crop Manag.* **2004**, *3*. [[CrossRef](#)]
35. Akibode, C.S.; Maredia, M.K. Global and Regional Trends in Production, Trade and Consumption of Food Legume Crops. 2012. Available online: <https://ispc.cgiar.org/sites/default/files/images/Legumetrendsv2.pdf> (accessed on 30 April 2019).
36. Pérez-Montaño, F.; Alías-Villegas, C.; Bellogín, R.A.; Del Cerro, P.; Espuny, M.R.; Jiménez-Guerrero, I.; López-Baena, F.J.; Ollero, F.J.; Cubo, T. Plant growth promotion in cereal and leguminous agricultural important plants: From microorganism capacities to crop production. *Microbiol. Res.* **2014**, *169*, 325–336. [[CrossRef](#)]
37. Glick, B.R. Using soil bacteria to facilitate phytoremediation. *Biotechnol. Adv.* **2010**, *28*, 367–374. [[CrossRef](#)]
38. Checcucci, A.; Azzarello, E.; Bazzicalupo, M.; Carlo, A.D.; Emiliani, G.; Mancuso, S.; Spini, G.; Viti, C.; Mengoni, A. Role and regulation of ACC deaminase gene in *Sinorhizobium meliloti*: Is it a symbiotic, rhizospheric or endophytic gene? *Front. Genet.* **2017**, *8*, 6. [[CrossRef](#)]
39. Brígido, C.; Nascimento, F.X.; Duan, J.; Glick, B.R.; Oliveira, S. Expression of an exogenous 1-aminocyclopropane-1-carboxylate deaminase gene in *Mesorhizobium* spp. reduces the negative effects of salt stress in chickpea. *FEMS Microbiol. Lett.* **2013**, *349*, 46–53. [[PubMed](#)]
40. Djekoun, A.; Planchon, C. Water status effect on dinitrogen fixation and photosynthesis in soybean. *Agron. J.* **1991**, *83*, 316–322. [[CrossRef](#)]
41. Serraj, R.; Sinclair, T.R.; Purcell, L.C. Symbiotic N₂ fixation response to drought. *J. Exp. Bot.* **1999**, *50*, 143–155. [[CrossRef](#)]
42. Elsheik, E.A.E. Effects of salt on rhizobia and bradyrhizobia: A review. *Ann. Appl. Biol.* **1998**, *132*, 507–524. [[CrossRef](#)]
43. Fagorzi, C.; Checcucci, A.; diCenzo, G.; Debiec-Andrzejewska, K.; Dziewit, L.; Pini, F.; Mengoni, A. Harnessing rhizobia to improve heavy-metal phytoremediation by legumes. *Genes* **2018**, *9*, 542. [[CrossRef](#)] [[PubMed](#)]
44. Chibuike, G.U.; Obiora, S.C. Heavy metal polluted soils: Effect on plants and bioremediation methods. *Appl. Environ. Soil Sci.* **2014**, *2014*, 752708. [[CrossRef](#)]
45. Lebrazi, S.; Fikri-Benbrahim, K. Rhizobium-legume symbioses: Heavy metal effects and principal approaches for bioremediation of contaminated soil. In *Legumes for Soil Health and Sustainable Management*; Meena, R., Das, A., Yadav, G., Lal, R., Eds.; Springer: Singapore, 2018; pp. 205–233.
46. Mahar, A.; Wang, P.; Ali, A.; Awasthi, M.K.; Lahori, A.H.; Wang, Q.; Li, R.; Zhang, Z. Challenges and opportunities in the phytoremediation of heavy metals contaminated soils: A review. *Ecotoxicol. Environ. Saf.* **2016**, *126*, 111–121. [[CrossRef](#)] [[PubMed](#)]
47. Bolan, N.S.; Park, J.H.; Robinson, B.; Naidu, R.; Huh, K.Y. Phytostabilization: A green approach to contaminant containment. *Adv. Agron.* **2011**, *112*, 145–204.
48. Kong, Z.; Glick, B.R. The role of plant growth-promoting bacteria in metal phytoremediation. In *Advances in Microbial Physiology*; Poole, R.K., Ed.; Academic Press: Cambridge, MA, USA, 2017; Volume 71, pp. 97–132.
49. Sturz, A.V.; Nowak, J. Endophytic communities of rhizobacteria and the strategies required to create yield enhancing associations with crops. *Appl. Soil Ecol.* **2000**, *15*, 183–190. [[CrossRef](#)]
50. Theis, K.R.; Dheilly, N.M.; Klassen, J.L.; Brucker, R.M.; John, F.; Bosch, T.C.G.; Cryan, J.F.; Gilbert, S.F.; Goodnight, C.J.; Lloyd, E.A.; et al. Getting the hologenome concept right: An eco—Evolutionary framework for hosts and their microbiomes. *Systems* **2016**, *29*, 00028-16. [[CrossRef](#)] [[PubMed](#)]
51. Khan, M.H.; Panda, S.K. Alterations in root lipid peroxidation and antioxidative responses in two rice cultivars under NaCl-salinity stress. *Acta Physiol. Plant.* **2008**, *30*, 81. [[CrossRef](#)]
52. Bouhmouch, I.; Souad-Mouhsine, B.; Brhada, F.; Aurag, J. Influence of host cultivars and *Rhizobium* species on the growth and symbiotic performance of *Phaseolus vulgaris* under salt stress. *J. Plant Physiol.* **2005**, *162*, 1103–1113. [[CrossRef](#)] [[PubMed](#)]
53. Zahran, H.H. *Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 968–989. [[PubMed](#)]
54. Cordovilla, M.P.; Ligeró, F.; Lluch, C. The effect of salinity on N fixation and assimilation in *Vicia faba*. *J. Exp. Bot.* **1994**, *45*, 1483–1488. [[CrossRef](#)]

55. Abbaspoor, A.; Zabihi, H.R.; Movafegh, S.; Asl, M.H.A. The efficiency of plant growth promoting rhizobacteria (PGPR) on yield and yield components of two varieties of wheat in salinity condition. *Am. Eurasian J. Sustain. Agric.* **2009**, *3*, 824–828.
56. Arora, N.K.; Tewari, S.; Singh, S.; Lal, N.; Maheshwari, D.K. PGPR for protection of plant health under saline conditions. In *Bacteria in Agrobiolgy: Stress Management*; Maheshwari, D.K., Ed.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 239–258.
57. Vijayan, K.; Chakraborti, S.P.; Ercisli, S.; Ghosh, P.D. NaCl induced morpho-biochemical and anatomical changes in mulberry (*Morus* spp.). *Plant Growth Regul.* **2008**, *56*, 61. [[CrossRef](#)]
58. Drew, M.C.; Hold, P.S.; Picchioni, G.A. Inhibition by NaCl of net CO₂ fixation and yield of cucumber. *J. Am. Soc. Hortic. Sci.* **1990**, *115*, 472–477. [[CrossRef](#)]
59. Cheeseman, J.M. Mechanisms of salinity tolerance in plants. *Plant Physiol.* **1988**, *87*, 57. [[CrossRef](#)]
60. Gupta, B.; Huang, B. Mechanism of salinity tolerance in plants: Physiological, biochemical, and molecular characterization. *Int. J. Genom.* **2014**, *2014*, 701596. [[CrossRef](#)]
61. Cordovilla, M.P.; Ocana, A.; Ligerio, F.; Lluch, C. Growth stage response to salinity in symbiosis *Vicia faba*-*Rhizobium leguminosarum* bv. *viciae*. *Plant Physiol.* **1995**, *14*, 105–111.
62. Fagg, C.W.; Stewart, J.L. The value of *Acacia* and *Prosopis* in arid and semi-arid environments. *J. Arid Environ.* **1994**, *27*, 3–25. [[CrossRef](#)]
63. Zhang, X.; Harper, R.; Karsisto, M.; Lindström, K. Diversity of *Rhizobium* bacteria isolated from the root nodules of leguminous trees. *Int. J. Syst. Evol. Microbiol.* **1991**, *41*, 104–113. [[CrossRef](#)]
64. Wahab, A.M.A.; Zahran, H.H. Effects of salt stress on nitrogenase activity and growth of four legumes. *Biol. Plant.* **1981**, *23*, 16. [[CrossRef](#)]
65. Bernstein, L.; Francois, L.E.; Clark, R.A. Interactive effects of salinity and fertility on yields of grains and vegetables 1. *Agron. J.* **1974**, *66*, 412–421. [[CrossRef](#)]
66. Maas, E.V.; Hoffman, G.J. Crop salt tolerance—current assessment. *J. Irrig. Drain. Div.* **1977**, *103*, 115–134.
67. Delgado, M.J.; Ligerio, F.; Lluch, C. Effects of salt stress on growth and nitrogen fixation by pea, faba-bean, common bean and soybean plants. *Soil Biol. Biochem.* **1994**, *26*, 371–376. [[CrossRef](#)]
68. Sulieman, S.; Tran, L.-S. Asparagine: An amide of particular distinction in the regulation of symbiotic nitrogen fixation of legumes. *Crit. Rev. Biotechnol.* **2013**, *33*, 309–327. [[CrossRef](#)] [[PubMed](#)]
69. Ballen, K.G.; Graham, P.H. The role of acid pH in symbiosis between plants and soil organisms. In *Handbook of Plant Growth pH as the Master Variable*; Rengel, Z., Ed.; Marcel Dekker: New York, NY, USA, 2002; pp. 383–404.
70. Qureshi, M.I.; Muneer, S.; Bashir, H.; Ahmad, J.; Iqbal, M. Nodule physiology and proteomics of stressed legumes. In *Advances in Botanical Research*; Kader, J.C., Delseny, M., Eds.; Elsevier: Amsterdam, The Netherlands, 2010; Volume 56, pp. 1–48. ISBN 0065-2296.
71. Zahran, H.H. Conditions for successful *Rhizobium*-legume symbiosis in saline environments. *Biol. Fertil. Soils* **1991**, *12*, 73–80. [[CrossRef](#)]
72. Tu, J.C. Effect of salinity on *Rhizobium*-root-hair interaction, nodulation and growth of soybean. *Can. J. Plant Sci.* **1981**, *61*, 231–239. [[CrossRef](#)]
73. Zahran, H.H.; Sprent, J.I. Effects of sodium chloride and polyethylene glycol on root-hair infection and nodulation of *Vicia faba* L. plants by *Rhizobium leguminosarum*. *Planta* **1986**, *167*, 303–309. [[CrossRef](#)]
74. Ikeda, J.; Kobayashi, M.; Takahashi, E. Salt stress increases the respiratory cost of nitrogen fixation. *Soil Sci. Plant Nutr.* **1992**, *38*, 51–56. [[CrossRef](#)]
75. Walsh, K.B. Physiology of the legume nodule and its response to stress. *Soil Biol. Biochem.* **1995**, *27*, 637–655. [[CrossRef](#)]
76. Hungria, M.; Vargas, M.A.T. Environmental factors affecting N₂ fixation in grain legumes in the tropics, with an emphasis on Brazil. *Field Crop. Res.* **2000**, *65*, 151–164. [[CrossRef](#)]
77. Delgado, M.J.; Garrido, J.M.; Ligerio, F.; Lluch, C. Nitrogen fixation and carbon metabolism by nodules and bacteroids of pea plants under sodium chloride stress. *Physiol. Plant.* **1993**, *89*, 824–829. [[CrossRef](#)]
78. Zahran, H.H. Rhizobia from wild legumes: Diversity, taxonomy, ecology, nitrogen fixation and biotechnology. *J. Biotechnol.* **2001**, *91*, 143–153. [[CrossRef](#)]
79. Lal, B.; Khanna, S. Selection of salt-tolerant *Rhizobium* isolates of *Acacia nilotica*. *World J. Microbiol. Biotechnol.* **1994**, *10*, 637–639. [[CrossRef](#)] [[PubMed](#)]

80. Zahran, H.H.; Räsänen, L.A.; Karsisto, M.; Lindström, K. Alteration of lipopolysaccharide and protein profiles in SDS-PAGE of rhizobia by osmotic and heat stress. *World J. Microbiol. Biotechnol.* **1994**, *10*, 100–105. [[CrossRef](#)]
81. Sakrouhi, I.; Belfquih, M.; Sbabou, L.; Moulin, P.; Bena, G.; Filali-Maltouf, A.; Le Quéré, A. Recovery of symbiotic nitrogen fixing acacia rhizobia from Merzouga desert sand dunes in South East Morocco—Identification of a probable new species of *Ensifer* adapted to stressed environments. *Syst. Appl. Microbiol.* **2016**, *39*, 122–131. [[CrossRef](#)]
82. Gouffi, K.; Pica, N.; Pichereau, V.; Blanco, C. Disaccharides as a new class of nonaccumulated osmoprotectants for *Sinorhizobium meliloti*. *Appl. Environ. Microbiol.* **1999**, *65*, 1491–1500.
83. Gouffi, K.; Blanco, C. Is the accumulation of osmoprotectant the unique mechanism involved in bacterial osmoprotection? *Int. J. Food Microbiol.* **2000**, *55*, 171–174. [[CrossRef](#)]
84. Talibart, R.; Jebbar, M.; Gouesbet, G.; Himdi-Kabbab, S.; Wroblewski, H.; Blanco, C.; Bernard, T. Osmoadaptation in rhizobia: Ectoine-induced salt tolerance. *J. Bacteriol.* **1994**, *176*, 5210–5217. [[CrossRef](#)] [[PubMed](#)]
85. Arora, N.K.; Singhal, V.; Maheshwari, D.K. Salinity-induced accumulation of poly- β -hydroxybutyrate in rhizobia indicating its role in cell protection. *World J. Microbiol. Biotechnol.* **2006**, *22*, 603–606. [[CrossRef](#)]
86. Bree6dveld, M.W.; Zevenhuizen, L.P.T.M.; Zehnder, A.J.B. Osmotically-regulated trehalose accumulation and cyclic β -(1, 2)-glucan excretion by *Rhizobium leguminosarum* biovar *trifolii* TA-1. *Arch. Microbiol.* **1991**, *156*, 501–506. [[CrossRef](#)]
87. Ghittoni, N.E.; Bueno, M.A. Changes in the cellular content of trehalose in four peanut rhizobia strains cultured under hypersalinity. *Symbiosis* **1996**, *20*, 117–127.
88. Le Rudulier, D.; Bernard, T. Salt tolerance in *Rhizobium*: A possible role for betaines. *FEMS Microbiol. Rev.* **1986**, *2*, 67–72. [[CrossRef](#)]
89. Smith, L.T.; Pocard, J.-A.; Bernard, T.; Le Rudulier, D. Osmotic control of glycine betaine biosynthesis and degradation in *Rhizobium meliloti*. *J. Bacteriol.* **1988**, *170*, 3142–3149. [[CrossRef](#)] [[PubMed](#)]
90. Elsheikh, E.A.E.; Wood, M. Salt effects on survival and multiplication of chickpea and soybean rhizobia. *Soil Biol. Biochem.* **1990**, *22*, 343–347. [[CrossRef](#)]
91. Fougere, F.; Le Rudulier, D. Uptake of glycine betaine and its analogues by bacteroids of *Rhizobium meliloti*. *Microbiology* **1990**, *136*, 157–163. [[CrossRef](#)]
92. Hickey, E.W.; Hirshfield, I.N. Low-pH-induced effects on patterns of protein synthesis and on internal pH in *Escherichia coli* and *Salmonella typhimurium*. *Appl. Environ. Microbiol.* **1990**, *56*, 1038–1045.
93. Shaharoon, B.; Arshad, M.; Zahir, Z.A.; Khalid, A. Performance of *Pseudomonas* spp. containing ACC-deaminase for improving growth and yield of maize (*Zea mays* L.) in the presence of nitrogenous fertilizer. *Soil Biol. Biochem.* **2006**, *38*, 2971–2975. [[CrossRef](#)]
94. Diagne, N.; Swanson, E.; Pesce, C.; Fall, F.; Diouf, F.; Bakhoun, N.; Fall, D.; Faye, M.N.; Oshone, R.; Simpson, S. Permanent draft genome sequence of *Rhizobium* sp. strain LCM 4573, a salt-tolerant, nitrogen-fixing bacterium isolated from Senegalese soils. *Genome Announc.* **2017**, *5*, e00285-17. [[CrossRef](#)] [[PubMed](#)]
95. Ormeño-Orrillo, E.; Gomes, D.F.; del Cerro, P.; Vasconcelos, A.T.R.; Canchaya, C.; Almeida, L.G.P.; Mercante, F.M.; Ollero, F.J.; Megías, M.; Hungria, M. Genome of *Rhizobium leucaenae* strains CFN 299 T and CPAO 29.8: Searching for genes related to a successful symbiotic performance under stressful conditions. *BMC Genom.* **2016**, *17*, 534.
96. Molina-Sánchez, M.D.; López-Contreras, J.A.; Toro, N.; Fernández-López, M. Genomic characterization of *Sinorhizobium meliloti* AK21, a wild isolate from the Aral Sea region. *Springerplus* **2015**, *4*, 259. [[CrossRef](#)] [[PubMed](#)]
97. Le Quéré, A.; Tak, N.; Gehlot, H.S.; Lavire, C.; Meyer, T.; Chapulliot, D.; Rath, S.; Sakrouhi, I.; Rocha, G.; Rohmer, M. Genomic characterization of *Ensifer aridi*, a proposed new species of nitrogen-fixing rhizobium recovered from Asian, African and American deserts. *BMC Genom.* **2017**, *18*, 85.
98. DiCenzo, G.C.; Zamani, M.; Checucci, A.; Fondi, M.; Griffiths, J.S.; Finan, T.M.; Mengoni, A. Multidisciplinary approaches for studying *Rhizobium*–legume symbioses. *Can. J. Microbiol.* **2018**, *65*, 1–33. [[CrossRef](#)]
99. Alexandre, A.N.A.; Laranjo, M.; Oliveira, S. Global transcriptional response to heat shock of the legume symbiont *Mesorhizobium loti* MAFF303099 comprises extensive gene downregulation. *DNA Res.* **2013**, *21*, 195–206. [[CrossRef](#)]

100. Laranjo, M.; Alexandre, A.; Oliveira, S. Global transcriptional response to salt shock of the plant microsymbiont *Mesorhizobium loti* MAFF303099. *Res. Microbiol.* **2017**, *168*, 55–63. [[CrossRef](#)]
101. Draghi, W.O.; Del Papa, M.F.; Hellweg, C.; Watt, S.A.; Watt, T.F.; Barsch, A.; Lozano, M.J.; Lagares Jr, A.; Salas, M.E.; López, J.L. A consolidated analysis of the physiologic and molecular responses induced under acid stress in the legume-symbiont model-soil bacterium *Sinorhizobium meliloti*. *Sci. Rep.* **2016**, *6*, 29278. [[CrossRef](#)]
102. Irar, S.; González, E.M.; Arrese-Igor, C.; Marino, D. A proteomic approach reveals new actors of nodule response to drought in split-root grown pea plants. *Physiol. Plant.* **2014**, *152*, 634–645. [[CrossRef](#)]
103. Larrainzar, E.; Wienkoop, S.; Scherling, C.; Kempa, S.; Ladrera, R.; Arrese-Igor, C.; Weckwerth, W.; González, E.M. Carbon metabolism and bacteroid functioning are involved in the regulation of nitrogen fixation in *Medicago truncatula* under drought and recovery. *Mol. Plant-Microbe Interact.* **2009**, *22*, 1565–1576. [[CrossRef](#)]
104. Larrainzar, E.; Wienkoop, S. A proteomic view on the role of legume symbiotic interactions. *Front. Plant Sci.* **2017**, *8*, 1267. [[CrossRef](#)] [[PubMed](#)]
105. Graham, P.H. Stress tolerance in *Rhizobium* and *Bradyrhizobium*, and nodulation under adverse soil conditions. *Can. J. Microbiol.* **1992**, *38*, 475–484. [[CrossRef](#)]
106. Da-Silva, J.R.; Alexandre, A.; Brígido, C.; Oliveira, S. Can stress response genes be used to improve the symbiotic performance of rhizobia. *AIMS Microbiol.* **2017**, *3*, 365–382. [[CrossRef](#)] [[PubMed](#)]
107. Suárez, R.; Wong, A.; Ramírez, M.; Barraza, A.; Orozco, M.C.; Cevallos, M.A.; Lara, M.; Hernández, G.; Iturriaga, G. Improvement of drought tolerance and grain yield in common bean by overexpressing trehalose-6-phosphate synthase in rhizobia. *Mol. Plant-Microbe Interact.* **2008**, *21*, 958–966. [[CrossRef](#)] [[PubMed](#)]
108. Rodríguez-Valera, F.; Martín-Cuadrado, A.B.; Rodríguez-Brito, B.; Pašić, L.; Thingstad, T.F.; Rohwer, F.; Mira, A. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **2009**, *7*, 828–836. [[CrossRef](#)] [[PubMed](#)]
109. Straub, P.F.; Shearer, G.; Reynolds, P.H.S.; Sawyer, S.A.; Kohl, D.H. Effect of disabling bacteroid proline catabolism on the response of soybeans to repeated drought stress. *J. Exp. Bot.* **1997**, *48*, 1299–1307. [[CrossRef](#)]
110. Boscari, A.; Van de Sype, G.; Le Rudulier, D.; Mandon, K. Overexpression of BetS, a *Sinorhizobium meliloti* high-affinity betaine transporter, in bacteroids from *Medicago sativa* nodules sustains nitrogen fixation during early salt stress adaptation. *Mol. Plant-Microbe Interact.* **2006**, *19*, 896–903. [[CrossRef](#)]
111. Nogales, J.; Campos, R.; BenAbdelkhalik, H.; Olivares, J.; Lluch, C.; Sanjuan, J. *Rhizobium tropici* genes involved in free-living salt tolerance are required for the establishment of efficient nitrogen-fixing symbiosis with *Phaseolus vulgaris*. *Mol. Plant-Microbe Interact.* **2002**, *15*, 225–232. [[CrossRef](#)]
112. Vaishnav, A.; Varma, A.; Tuteja, N.; Choudhary, D.K. PGPR-mediated amelioration of crops under salt stress. In *Plant-Microbe Interaction: An Approach to Sustainable Agriculture*; Choudhary, D.K., Varma, A., Tuteja, N., Eds.; Springer: Singapore, 2016; pp. 205–226.
113. Glick, B.R. Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *FEMS Microbiol. Lett.* **2005**, *251*, 1–7. [[CrossRef](#)]
114. Estevez, J.; Dardanelli, M.S.; Megías, M.; Rodríguez-Navarro, D.N. Symbiotic performance of common bean and soybean co-inoculated with rhizobia and *Chryseobacterium balustinum* Aur9 under moderate saline conditions. *Symbiosis* **2009**, *49*, 29–36. [[CrossRef](#)]
115. Aamir, M.; Aslam, A.; Khan, M.Y.; Jamshaid, M.U.; Ahmad, M.; Asghar, H.N.; Zahir, Z.A. Co-inoculation with *Rhizobium* and plant growth promoting rhizobacteria (PGPR) for inducing salinity tolerance in mung bean under field condition of semi-arid climate. *Asian J. Agric. Biol.* **2013**, *1*, 17–22.
116. Ahmad, M.; Zahir, Z.A.; Asghar, H.N.; Asghar, M. Inducing salt tolerance in mung bean through coinoculation with rhizobia and plant-growth-promoting rhizobacteria containing 1-aminocyclopropane-1-carboxylate deaminase. *Can. J. Microbiol.* **2011**, *57*, 578–589. [[CrossRef](#)] [[PubMed](#)]
117. Etesami, H. Bacterial mediated alleviation of heavy metal stress and decreased accumulation of metals in plant tissues: Mechanisms and future prospects. *Ecotoxicol. Environ. Saf.* **2018**, *147*, 175–191. [[CrossRef](#)] [[PubMed](#)]
118. Noll, M.R. Trace elements in terrestrial environments: Biogeochemistry, bioavailability, and risks of metals. *J. Environ. Qual.* **2003**, *32*, 374. [[CrossRef](#)]

119. Ahemad, M.; Khan, M.S. Effect of fungicides on plant growth promoting activities of phosphate solubilizing *Pseudomonas putida* isolated from mustard (*Brassica campestris*) rhizosphere. *Chemosphere* **2012**, *86*, 945–950. [[CrossRef](#)] [[PubMed](#)]
120. Pajuelo, E.; Rodríguez-Llorente, I.D.; Dary, M.; Palomares, A.J. Toxic effects of arsenic on *Sinorhizobium-Medicago sativa* symbiotic interaction. *Environ. Pollut.* **2008**, *154*, 203–211. [[CrossRef](#)] [[PubMed](#)]
121. Nagajyoti, P.C.; Lee, K.D.; Sreekanth, T.V.M. Heavy metals, occurrence and toxicity for plants: A review. *Environ. Chem. Lett.* **2010**, *8*, 199–216. [[CrossRef](#)]
122. Del Rio, M.; Font, R.; Almela, C.; Velez, D.; Montoro, R.; De Haro Bailon, A. Heavy metals and arsenic uptake by wild vegetation in the Guadiamar river area after the toxic spill of the Aznalcollar mine. *J. Biotechnol.* **2002**, *98*, 125–137. [[CrossRef](#)]
123. Adejumo, S.A. Mechanisms of lead and chromium hyperaccumulation and tolerance in plant. *Environtropica* **2019**, *15*, 11–29.
124. Ma, Y.; Oliveira, R.S.; Freitas, H.; Zhang, C. Biochemical and molecular mechanisms of plant-microbe-metal interactions: Relevance for phytoremediation. *Front. Plant Sci.* **2016**, *7*, 918. [[CrossRef](#)]
125. Zubair, M.; Shakir, M.; Ali, Q.; Rani, N.; Fatima, N.; Farooq, S.; Shafiq, S.; Kanwal, N.; Ali, F.; Nasir, I.A. Rhizobacteria and phytoremediation of heavy metals. *Environ. Technol. Rev.* **2016**, *5*, 112–119. [[CrossRef](#)]
126. Sarwar, N.; Imran, M.; Shaheen, M.R.; Ishaque, W.; Kamran, M.A.; Matloob, A.; Rehman, A.; Hussain, S. Phytoremediation strategies for soils contaminated with heavy metals: Modifications and future perspectives. *Chemosphere* **2017**, *171*, 710–721. [[CrossRef](#)] [[PubMed](#)]
127. Erakhrumen, A.A. Phytoremediation: An environmentally sound technology for pollution prevention, control and remediation in developing countries. *Educ. Res. Rev.* **2007**, *2*, 151–156.
128. Ghosh, M.; Singh, S.P. A review on phytoremediation of heavy metals and utilization of its by products. *Asian J. Energy Env.* **2005**, *6*, 18.
129. He, Z.L.; Yang, X.E.; Stoffella, P.J. Trace elements in agroecosystems and impacts on the environment. *J. Trace Elem. Med. Biol.* **2005**, *19*, 125–140. [[CrossRef](#)]
130. Seth, C.S. A review on mechanisms of plant tolerance and role of transgenic plants in environmental clean-up. *Bot. Rev.* **2012**, *78*, 32–62. [[CrossRef](#)]
131. Ali, H.; Khan, E.; Sajad, M.A. Phytoremediation of heavy metals—Concepts and applications. *Chemosphere* **2013**, *91*, 869–881. [[CrossRef](#)]
132. Glick, B.R. The enhancement of plant growth by free-living bacteria. *Can. J. Microbiol.* **1995**, *41*, 109–117. [[CrossRef](#)]
133. Hao, X.; Taghavi, S.; Xie, P.; Orbach, M.J.; Alwathnani, H.A.; Rensing, C.; Wei, G. Phytoremediation of Heavy and Transition Metals Aided by LegumeRhizobia Symbiosis, *Int. J. Phytoremediation* **2014**, *16*, 179–202. [[CrossRef](#)]
134. Rajkumar, M.; Sandhya, S.; Prasad, M.N.; Freitas, H. Perspectives of plant-associated microbes in heavy metal phytoremediation. *Biotechnol. Adv.* **2012**, *30*, 1562–1574. [[CrossRef](#)]
135. Sessitsch, A.; Kuffner, M.; Kidd, P.; Vangronsveld, J.; Wenzel, W.W.; Fallmann, K.; Puschenreiter, M. The role of plant-associated bacteria in the mobilization and phytoextraction of trace elements in contaminated soils. *Soil Biol. Biochem.* **2013**, *60*, 182–194. [[CrossRef](#)] [[PubMed](#)]
136. Sharma, R.K.; Archana, G. Cadmium minimization in food crops by cadmium resistant plant growth promoting rhizobacteria. *Appl. Soil Ecol.* **2016**, *107*, 66–78. [[CrossRef](#)]
137. Ullah, A.; Heng, S.; Munis, M.F.H.; Fahad, S.; Yang, X. Phytoremediation of heavy metals assisted by plant growth promoting (PGP) bacteria: A review. *Environ. Exp. Bot.* **2015**, *117*, 28–40. [[CrossRef](#)]
138. Wheaton, G.; Counts, J.; Mukherjee, A.; Kruh, J.; Kelly, R. The confluence of heavy metal biooxidation and heavy metal resistance: Implications for bioleaching by extreme thermoacidophiles. *Minerals* **2015**, *5*, 397–451. [[CrossRef](#)]
139. Bruins, M.R.; Kapil, S.; Oehme, F.W. Microbial resistance to metals in the environment. *Ecotoxicol. Environ. Saf.* **2000**, *45*, 198–207. [[CrossRef](#)] [[PubMed](#)]
140. Atieno, M.; Lesueur, D. Opportunities for improved legume inoculants: Enhanced stress tolerance of rhizobia and benefits to agroecosystems. *Symbiosis* **2018**, *77*, 191–205. [[CrossRef](#)]
141. Pal, A.; Paul, A.K. Microbial extracellular polymeric substances: Central elements in heavy metal bioremediation. *Indian J. Microbiol.* **2008**, *48*, 49–64. [[CrossRef](#)] [[PubMed](#)]

142. Giller, K.E.; Witter, E.; McGrath, S.P. Heavy metals and soil microbes. *Soil Biol. Biochem.* **2009**, *41*, 2031–2037. [[CrossRef](#)]
143. Martensson, A.; Witter, E. Influence of various soil amendments on nitrogen-fixing soil microorganisms in a long-term field experiment, with special reference to sewage sludge. *Soil Biol. Biochem.* **1990**, *22*, 977–982. [[CrossRef](#)]
144. Obbard, J.P.; Jones, K.C. The effect of heavy metals on dinitrogen fixation by *Rhizobium*-white clover in a range of long-term sewage sludge amended and metal-contaminated soils. *Environ. Pollut.* **1993**, *79*, 105–112. [[CrossRef](#)]
145. Gulz, P.A.; Gupta, S.-K.; Schulin, R. Arsenic accumulation of common plants from contaminated soils. *Plant Soil* **2005**, *272*, 337–347. [[CrossRef](#)]
146. Fatnassi, I.C.; Chiboub, M.; Saadani, O.; Jebara, M.; Jebara, S.H. Phytostabilization of moderate copper contaminated soils using co-inoculation of *Vicia faba* with plant growth promoting bacteria. *J. Basic Microbiol.* **2015**, *55*, 303–311. [[CrossRef](#)] [[PubMed](#)]
147. Delgadillo, J.; Lafuente, A.; Doukkali, B.; Redondo-Gómez, S.; Mateos-Naranjo, E.; Caviedes, M.A.; Pajuelo, E.; Rodríguez-Llorente, I.D. Improving legume nodulation and Cu rhizostabilization using a genetically modified rhizobia. *Environ. Technol.* **2015**, *36*, 1237–1245. [[CrossRef](#)] [[PubMed](#)]
148. Pajuelo, E.; Pérez-Palacios, P.; Romero-Aguilar, A.; Delgadillo, J.; Doukkali, B.; Rodríguez-Llorente, I.D.; Caviedes, M.A. Improving legume–rhizobium symbiosis for copper phytostabilization through genetic manipulation of both symbionts. In *Biological Nitrogen Fixation and Beneficial Plant-Microbe Interaction*; González-Andrés, F., James, F., Eds.; Springer: Cham, Switzerland, 2016; pp. 183–193.
149. Drewniak, L.; Sklodowska, A. Arsenic-transforming microbes and their role in biomining processes. *Environ. Sci. Pollut. Res.* **2013**, *20*, 7728–7739. [[CrossRef](#)] [[PubMed](#)]
150. Pinter, M.I.F.; Salomon, M.V.; Berli, F.; Gil, R.; Bottini, R.; Piccoli, P. Plant growth promoting rhizobacteria alleviate stress by AsIII in grapevine. *Agric. Ecosyst. Environ.* **2018**, *267*, 100–108. [[CrossRef](#)]
151. Bianucci, E.; Godoy, A.; Furlan, A.; Peralta, J.M.; Hernández, L.E.; Carpena-Ruiz, R.O.; Castro, S. Arsenic toxicity in soybean alleviated by a symbiotic species of *Bradyrhizobium*. *Symbiosis* **2018**, *74*, 167–176. [[CrossRef](#)]
152. Fatnassi, I.C.; Chiboub, M.; Saadani, O.; Jebara, M.; Jebara, S.H. Impact of dual inoculation with *Rhizobium* and PGPR on growth and antioxidant status of *Vicia faba* L. under copper stress. *C. R. Biol.* **2015**, *338*, 241–254. [[CrossRef](#)] [[PubMed](#)]
153. Dary, M.; Chamber-Pérez, M.A.; Palomares, A.J.; Pajuelo, E. “In situ” phytostabilisation of heavy metal polluted soils using *Lupinus luteus* inoculated with metal resistant plant-growth promoting rhizobacteria. *J. Hazard. Mater.* **2010**, *177*, 323–330. [[CrossRef](#)]
154. Kamaludeen, S.P.B.; Ramasamy, K. Rhizoremediation of metals: Harnessing microbial communities. *Indian J. Microbiol.* **2008**, *48*, 80–88. [[CrossRef](#)]
155. Singh, N.K.; Rai, U.N.; Tewari, A.; Singh, M. Metal accumulation and growth response in *Vigna radiata* L. inoculated with chromate tolerant rhizobacteria and grown on tannery sludge amended soil. *Bull. Environ. Contam. Toxicol.* **2010**, *84*, 118–124. [[CrossRef](#)]
156. Armendariz, A.L.; Talano, M.A.; Nicotra, M.F.O.; Escudero, L.; Bresler, M.L.; Porporatto, C.; Agostini, E. Impact of double inoculation with *Bradyrhizobium japonicum* E109 and *Azospirillum brasilense* Az39 on soybean plants grown under arsenic stress. *Plant Physiol. Biochem.* **2019**, *138*, 26–35. [[CrossRef](#)]
157. Spaepen, S.; Vanderleyden, J.; Remans, R. Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol. Rev.* **2007**, *31*, 425–448. [[CrossRef](#)] [[PubMed](#)]
158. Sharma, P.; Dubey, R.S. Lead toxicity in plants. *Braz. J. Plant Physiol.* **2005**, *17*, 35–52. [[CrossRef](#)]
159. Di Gregorio, S.; Barbaferri, M.; Lampis, S.; Sanangelantoni, A.M.; Tassi, E.; Vallini, G. Combined application of Triton X-100 and *Sinorhizobium* sp. Pb002 inoculum for the improvement of lead phytoextraction by *Brassica juncea* in EDTA amended soil. *Chemosphere* **2006**, *63*, 293–299. [[CrossRef](#)] [[PubMed](#)]
160. Zhao, F.J.; Lombi, E.; McGrath, S.P. Assessing the potential for zinc and cadmium phytoremediation with the hyperaccumulator *Thlaspi caerulescens*. *Plant Soil* **2003**, *249*, 37–43. [[CrossRef](#)]
161. Seregin, I.V.; Kozhevnikova, A.D. Physiological role of nickel and its toxic effects on higher plants. *Russ. J. Plant Physiol.* **2006**, *53*, 257–277. [[CrossRef](#)]
162. Wani, P.A.; Khan, M.S.; Zaidi, A. Effect of metal-tolerant plant growth-promoting *Rhizobium* on the performance of pea grown in metal-amended soil. *Arch. Environ. Contam. Toxicol.* **2008**, *55*, 33–42. [[CrossRef](#)] [[PubMed](#)]

163. Wani, P.A.; Khan, M.S.; Zaidi, A. Effect of metal tolerant plant growth promoting *Bradyrhizobium* sp. (vigna) on growth, symbiosis, seed yield and metal uptake by greengram plants. *Chemosphere* **2007**, *70*, 36–45. [[CrossRef](#)]
164. Hasanuzzaman, M.; Prasad, M.N.V.; Fujita, M. *Cadmium Toxicity and Tolerance in Plants: From Physiology to Remediation*; Academic Press: Cambridge, MA, USA, 2018; ISBN 0128148659.
165. Yu, R.; Tang, Y.; Liu, C.; Du, X.; Miao, C.; Shi, G. Comparative transcriptomic analysis reveals the roles of ROS scavenging genes in response to cadmium in two pak choi cultivars. *Sci. Rep.* **2017**, *7*, 9217. [[CrossRef](#)]
166. Guo, J.; Chi, J. Effect of Cd-tolerant plant growth-promoting *Rhizobium* on plant growth and Cd uptake by *Lolium multiflorum* Lam. and *Glycine max* (L.) Merr. in Cd-contaminated soil. *Plant Soil* **2014**, *375*, 205–214. [[CrossRef](#)]
167. Porter, S.S.; Chang, P.L.; Conow, C.A.; Dunham, J.P.; Friesen, M.L. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J.* **2017**, *11*, 248–262. [[CrossRef](#)] [[PubMed](#)]
168. Lu, M.; Jiao, S.; Gao, E.; Song, X.; Li, Z.; Hao, X.; Rensing, C.; Wei, G.; Lu, C.M. Transcriptome response to heavy metals in *Sinorhizobium meliloti* CCNWSX0020 reveals new metal resistance determinants that also promote bioremediation by *Medicago lupulina* in metal-contaminated soil. *Appl. Environ. Microbiol.* **2017**, *83*, 1244–1261. [[CrossRef](#)] [[PubMed](#)]
169. Hao, X.; Xie, P.; Zhu, Y.-G.; Taghavi, S.; Wei, G.; Rensing, C. Copper tolerance mechanisms of *Mesorhizobium amorphae* and its role in aiding phytostabilization by *Robinia pseudoacacia* in copper contaminated soil. *Environ. Sci. Technol.* **2015**, *49*, 2328–2340. [[CrossRef](#)] [[PubMed](#)]
170. Hussain, M.B.; Zahir, Z.A.; Asghar, H.N.; Asgher, M. Can catalase and exopolysaccharides producing rhizobia ameliorate drought stress in wheat? *Int. J. Agric. Biol.* **2014**, *16*, 3–13.
171. Baldani, V.L.D.; Alvarez, M.A.B.; Baldani, J.I.; Döbereiner, J. Establishment of inoculated *Azospirillum* spp. in the rhizosphere and in roots of field grown wheat and sorghum. *Plant Soil* **1986**, *90*, 35–46. [[CrossRef](#)]
172. Fasciglione, G.; Casanovas, E.M.; Quillehauquy, V.; Yommi, A.K.; Goñi, M.G.; Roura, S.I.; Barassi, C.A. *Azospirillum* inoculation effects on growth, product quality and storage life of lettuce plants grown under salt stress. *Sci. Hortic. Amst.* **2015**, *195*, 154–162. [[CrossRef](#)]
173. Staudinger, C.; Mehmeti-Tershani, V.; Gil-Quintana, E.; Gonzalez, E.M.; Hofhansl, F.; Bachmann, G.; Wienkoop, S. Evidence for a rhizobia-induced drought stress response strategy in *Medicago truncatula*. *J. Proteom.* **2016**, *136*, 202–213. [[CrossRef](#)]
174. Bordeleau, L.M.; Antoun, H.; Lachance, R.-A. Effets des souches de *Rhizobium meliloti* et des coupes successives de la luzerne (*Medicago sativa*) sur la fixation symbiotique d'azote. *Can. J. Plant Sci.* **1977**, *57*, 433–439. [[CrossRef](#)]
175. Bertrand, A.; Dhont, C.; Bipfubusa, M.; Chalifour, F.-P.; Drouin, P.; Beauchamp, C.J. Improving salt stress responses of the symbiosis in alfalfa using salt-tolerant cultivar and rhizobial strain. *Appl. Soil Ecol.* **2015**, *87*, 108–117. [[CrossRef](#)]
176. Bromfield, E.S.P.; Wheatcroft, R.; Barran, L.R. Medium for direct isolation of *Rhizobium meliloti* from soils. *Soil Biol. Biochem.* **1994**, *26*, 423–428. [[CrossRef](#)]
177. Noel, K.D.; Vandenbosch, K.A.; Kulpaca, B. Mutations in *Rhizobium phaseoli* that lead to arrested development of infection threads. *J. Bacteriol.* **1986**, *168*, 1392–1401. [[CrossRef](#)]
178. Smith, R.S. Legume inoculant formulation and application. *Can. J. Microbiol.* **1992**, *38*, 485–492. [[CrossRef](#)]
179. Nobbe, F.; Hiltner, L. Inoculation of the soil for cultivating leguminous plants. *US Pat.* **1896**, *570*, 813.
180. Brockwell, J. Inoculation methods for field experimenters and farmers. In *Nitrogen Fixation in Legumes*; Vincent, J.M., Ed.; Academic Press: New York, NY, USA, 1982; pp. 211–227.
181. Dommergues, Y.R.; Diem, H.G.; Divies, C. Polyacrylamide-entrapped *Rhizobium* as an inoculant for legumes. *Appl. Environ. Microbiol.* **1979**, *37*, 779–781. [[PubMed](#)]
182. Jung, G.; Mugnier, J.; Diem, H.G.; Dommergues, Y.R. Polymer-entrapped *Rhizobium* as an inoculant for legumes. *Plant Soil* **1982**, *65*, 219–231. [[CrossRef](#)]
183. Paau, A.S.; Graham, L.L.; Bennett, M. Progress in formulation research for PGPR and biocontrol inoculants. *Bull. OILB SROP* **1991**, *14*, 8. Available online: <http://agris.fao.org/agris-search/search.do?recordID=FR9202869> (accessed on 30 April 2019).

184. Smith, R.S. Inoculant formulations and applications to meet changing needs. In *Nitrogen Fixation: Fundamentals and Applications*; Tikhonovich, I.A., Provorov, N.A., Romanov, V.I., Newton, W.E., Eds.; Springer: Dordrecht, The Netherlands, 1995; Volume 27, pp. 653–657.
185. Singleton, P.; Keyser, H.; Sande, E. Development and evaluation of liquid inoculants. In *Proceedings of the Inoculants and Nitrogen Fixation of Legumes in Vietnam*, Hanoi, Vietnam, 17–18 February 2001; pp. 52–66.
186. Rohr, T. Rheological Study of the Mixture Carboxymethylcellulose/Starch and Its Use As a Vehicle for the Bacterial Inoculation. Ph.D. Thesis, Rural Federal University of Rio de Janeiro, Rio De Janeiro, Brazil, 2007.
187. Manikandan, R.; Saravanakumar, D.; Rajendran, L.; Raguchander, T.; Samiyappan, R. Standardization of liquid formulation of *Pseudomonas fluorescens* Pf1 for its efficacy against Fusarium wilt of tomato. *Biol. Control*. **2010**, *54*, 83–89. [[CrossRef](#)]
188. Ferreira, E.M.; Castro, I.V. e Residues of the cork industry as carriers for the production of legumes inoculants. *Silva. Lusit.* **2005**, *13*, 159–167.
189. Kaljeet, S.; Keyeo, F.; Amir, H. Temperature on survivability of rhizobial inoculant. *Asian J. Plant Sci.* **2011**, *10*, 331–337.
190. Bashan, Y.; de-Bashan, L.E.; Prabhu, S.R.; Hernandez, J.-P. Advances in plant growth-promoting bacterial inoculant technology: Formulations and practical perspectives (1998–2013). *Plant Soil* **2014**, *378*, 1–33. [[CrossRef](#)]
191. Albareda, M.; Rodríguez-Navarro, D.N.; Camacho, M.; Temprano, F.J. Alternatives to peat as a carrier for rhizobia inoculants: Solid and liquid formulations. *Soil Biol. Biochem.* **2008**, *40*, 2771–2779. [[CrossRef](#)]
192. Rivera, D.; Obando, M.; Barbosa, H.; Rojas Tapias, D.; Bonilla Buitrago, R. Evaluation of polymers for the liquid rhizobial formulation and their influence in the *Rhizobium*-Cowpea interaction. *Univ. Sci.* **2014**, *19*, 265–275. [[CrossRef](#)]
193. BrahmaPrakash, G.P.; Sahu, P.K. Biofertilizers for sustainability. *J. Indian Inst. Sci.* **2012**, *92*, 37–62.
194. Vanderghenst, J.; Scher, H.; Guo, H.-Y.; Schultz, D. Water-in-oil emulsions that improve the storage and delivery of the biolarvicide *Lagenidium giganteum*. *BioControl* **2007**, *52*, 207–229. [[CrossRef](#)]
195. Nápoles, M.C.; Carrió, E.G.; Montero, F.; Ferreira, A.; Rossi, A. Role of *Bradyrhizobium japonicum* induced by genistein on soybean stressed by water deficit. *Spanish J. Agric. Res.* **2009**, *7*, 665–671. [[CrossRef](#)]
196. Tank, N.; Saraf, M. Salinity-resistant plant growth promoting rhizobacteria ameliorates sodium chloride stress on tomato plants. *J. Plant Interact.* **2010**, *5*, 51–58. [[CrossRef](#)]
197. Kudoyarova, G.R.; Arkhipova, T.N.; Melent'ev, A.I. Role of bacterial phytohormones in plant growth regulation and their development. In *Bacterial Metabolites in Sustainable Agroecosystem*; Maeshwari, D., Ed.; Springer: Cham, Switzerland, 2015; pp. 69–86.
198. Giller, K.E.; Murwira, M.S.; Dhliwayo, D.K.C.; Mafongoya, P.L.; Mpeperekhi, S. Soybeans and sustainable agriculture in southern Africa. *Int. J. Agric. Sustain.* **2011**, *9*, 50–58. [[CrossRef](#)]
199. Lupwayi, N.Z.; Clayton, G.W.; Rice, W.A. Rhizobial inoculants for legume crops. *J. Crop. Improv.* **2006**, *15*, 289–321. [[CrossRef](#)]
200. Thilakarathna, M.S.; Raizada, M.N. A meta-analysis of the effectiveness of diverse rhizobia inoculants on soybean traits under field conditions. *Soil Biol. Biochem.* **2017**, *105*, 177–196. [[CrossRef](#)]
201. Friesen, M.L. Widespread fitness alignment in the legume—*Rhizobium* symbiosis. *New Phytol.* **2012**, *194*, 1096–1111. [[CrossRef](#)] [[PubMed](#)]
202. Checcucci, A.; Azzarello, E.; Bazzicalupo, M.; Galardini, M.; Lagomarsino, A.; Mancuso, S.; Marti, L.; Marzano, M.C.; Mocali, S.; Squartini, A.; et al. Mixed nodule infection in *Sinorhizobium meliloti*–*Medicago sativa* symbiosis suggest the presence of cheating behavior. *Front. Plant Sci.* **2016**, *7*, 835. [[CrossRef](#)] [[PubMed](#)]
203. West, S.A.; Kiers, E.T.; Simms, E.L.; Denison, R.F. Sanctions and mutualism stability: Why do rhizobia fix nitrogen? *Proc. R. Soc. B Biol. Sci.* **2002**, *269*, 685–694. [[CrossRef](#)] [[PubMed](#)]
204. Westhoek, A.; Field, E.; Rehling, F.; Mulley, G.; Webb, I.; Poole, P.S.; Turnbull, L.A. Policing the legume-Rhizobium symbiosis: A critical test of partner choice. *Sci. Rep.* **2017**, *7*, 1419. [[CrossRef](#)]
205. Kiers, E.T.; Denison, R.F. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annu. Rev. Ecol. Evol. Syst.* **2008**, *39*, 215–236. [[CrossRef](#)]
206. MacLean, A.M.; Finan, T.M.; Sadowsky, M.J. Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiol.* **2007**, *144*, 615–622. [[CrossRef](#)] [[PubMed](#)]

207. Checucci, A.; DiCenzo, G.C.; Ghini, V.; Bazzicalupo, M.; Beker, A.; Decorosi, F.; Dohlemann, J.; Fagorzi, C.; Finan, T.M.; Fondi, M.; et al. Creation and Characterization of a Genomically Hybrid Strain in the Nitrogen-Fixing Symbiotic Bacterium *Sinorhizobium meliloti*. *ACS Synth. Biol.* **2018**, *7*, 296483. [CrossRef]
208. Brewin, N.J.; Wood, E.A.; Young, J.P.W. Contribution of the symbiotic plasmid to the competitiveness of *Rhizobium leguminosarum*. *Microbiology* **1983**, *129*, 2973–2977. [CrossRef]
209. Brewin, N.J.; DeJong, T.M.; Phillips, D.A.; Johnston, A.W.B. Co-transfer of determinants for hydrogenase activity and nodulation ability in *Rhizobium leguminosarum*. *Nature* **1980**, *288*, 77–79. [CrossRef]
210. DeJong, T.M.; Brewin, N.J.; Phillips, D.A. Effects of plasmid content in *Rhizobium leguminosarum* on pea nodule activity and plant growth. *Microbiology* **1981**, *124*, 1–7. [CrossRef]
211. DiCenzo, G.C.; Checucci, A.; Bazzicalupo, M.; Mengoni, A.; Viti, C.; Dziewit, L.; Finan, T.M.; Galardini, M.; Fondi, M. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **2016**, *7*, 12219. [CrossRef]
212. Checucci, A.; Bazzicalupo, M.; Mengoni, A. Exploiting Nitrogen-Fixing Rhizobial Symbionts Genetic Resources for Improving Phytoremediation of Contaminated Soils. In *Enhancing Cleanup of Environmental Pollutants*; Anjum, N., Tuteja, N., Eds.; Springer: Cham, Switzerland, 2017; Volume 1, pp. 275–288.
213. Bradáčová, K.; Florea, A.S.; Bar-Tal, A.; Minz, D.; Yermiyahu, U.; Shawahna, R.; Kraut-Cohen, J.; Zolti, A.; Erel, R.; Dietel, K. Microbial consortia versus single-strain inoculants: An advantage in PGPM-assisted tomato production? *Agronomy* **2019**, *9*, 105. [CrossRef]
214. Bashan, Y. Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnol. Adv.* **1998**, *16*, 729–770. [CrossRef]
215. Nadeem, S.M.; Naveed, M.; Zahir, Z.A.; Asghar, H.N. Plant–microbe interactions for sustainable agriculture: Fundamentals and recent advances. In *Plant Microbe Symbiosis: Fundamentals and Advances*; Arora, N., Ed.; Springer: New Dehli, India, 2013; pp. 51–103.
216. Sekar, J.; Raj, R.; Prabavathy, V.R. Microbial consortial products for sustainable agriculture: Commercialization and regulatory issues in india. In *Agriculturally Importantmicroorganisms*; Sarma, B., Keswani, C., Eds.; Springer: Singapore, 2016; pp. 107–132.
217. Abd-Alla, M.H.; El-Enany, A.-W.E.; Nafady, N.A.; Khalaf, D.M.; Morsy, F.M. Synergistic interaction of *Rhizobium leguminosarum* bv. *viciae* and arbuscular mycorrhizal fungi as a plant growth promoting biofertilizers for faba bean (*Vicia faba* L.) in alkaline soil. *Microbiol. Res.* **2014**, *169*, 49–58. [CrossRef] [PubMed]
218. Tajini, F.; Trabelsi, M.; Drevon, J.-J. Co-inoculation with *Glomus intraradices* and *Rhizobium tropici* CIAT899 increases P use efficiency for N₂ fixation in the common bean (*Phaseolus vulgaris* L.) under P deficiency in hydroaerobic culture. *Symbiosis* **2011**, *53*, 123. [CrossRef]
219. Zhu, R.-F.; Tang, F.; Liu, J.; Liu, F.-Q.; Deng, X.-Y.; Chen, J.-S. Co-inoculation of arbuscular mycorrhizae and nitrogen fixing bacteria enhance alfalfa yield under saline conditions. *Pakistan J. Bot.* **2016**, *48*, 763–769.
220. Meng, L.; Zhang, A.; Wang, F.; Han, X.; Wang, D.; Li, S. Arbuscular mycorrhizal fungi and *Rhizobium* facilitate nitrogen uptake and transfer in soybean/maize intercropping system. *Front. Plant Sci.* **2015**, *6*, 339. [CrossRef] [PubMed]



Review

Harnessing Rhizobia to Improve Heavy-Metal Phytoremediation by Legumes

Camilla Fagorzi ¹, Alice Checcucci ^{1,*}, George C. diCenzo ¹,
Klaudia Debiec-Andrzejewska ², Lukasz Dziewit ³, Francesco Pini ⁴
and Alessio Mengoni ^{1,*}

¹ Department of Biology, University of Florence, Via Madonna del Piano 6, 50019 Sesto Fiorentino, Italy; camilla.fagorzi@unifi.it (C.F.); georgecolin.dicenzo@unifi.it (G.C.D.)

² Laboratory of Environmental Pollution Analysis, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland; k.debiec@biol.uw.edu.pl

³ Department of Bacterial Genetics, Institute of Microbiology, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland; ldziewit@biol.uw.edu.pl

⁴ Department of Agri-food Production and Environmental Science, University of Florence, 50144 Florence, Italy; francesco.pini@unifi.it

* Correspondence: alice.checcucci@unifi.it (A.C.); alessio.mengoni@unifi.it (A.M.);
Tel.: +39-055-457-4738 (A.M.)

Received: 28 September 2018; Accepted: 6 November 2018; Published: 8 November 2018



Abstract: Rhizobia are bacteria that can form symbiotic associations with plants of the Fabaceae family, during which they reduce atmospheric di-nitrogen to ammonia. The symbiosis between rhizobia and leguminous plants is a fundamental contributor to nitrogen cycling in natural and agricultural ecosystems. Rhizobial microsymbionts are a major reason why legumes can colonize marginal lands and nitrogen-deficient soils. Several leguminous species have been found in metal-contaminated areas, and they often harbor metal-tolerant rhizobia. In recent years, there have been numerous efforts and discoveries related to the genetic determinants of metal resistance by rhizobia, and on the effectiveness of such rhizobia to increase the metal tolerance of host plants. Here, we review the main findings on the metal resistance of rhizobia: the physiological role, evolution, and genetic determinants, and the potential to use native and genetically-manipulated rhizobia as inoculants for legumes in phytoremediation practices.

Keywords: soil bioremediation; heavy-metals; serpentine soils; serpentine vegetation; genome manipulation; *cis*-hybrid strains

1. Introduction

Plants are colonized by an extraordinarily high number of (micro)organisms, which may reach numbers much larger than those of plant cells [1]. This is particularly evident in the rhizosphere, the thin layer of soil surrounding and influenced by plant roots, where a staggering diversity of microorganisms is present. The collective communities of plant-associated microorganisms are referred to as the plant microbiota, and include the microbial communities of the rhizosphere, as well as those of the external and internal (the endosphere) plant tissues (for examples see [1–4]). The rhizobiome refers specifically to the microbial community of the rhizosphere, and microbes from this community have been deeply studied for their beneficial effects on plant growth and health. These mainly include mycorrhizal fungi (AMF) and plant-growth promoting rhizobacteria (PGPR), with the latter including the nitrogen fixing legume endosymbiotic bacteria known as rhizobia [5]. Rhizobia are a paraphyletic group of nitrogen fixing bacteria belonging to the Alpha- and Betaproteobacteria classes.

Rhizobia can penetrate plant tissues and establish an intracellular population within specialized tissue (known as a nodule) on the root (or stem in a few cases) of leguminous plants. Once inside the cells, the rhizobia differentiate into forms known as bacteroids, which are able to perform nitrogen fixation (the formation of ammonia from di-nitrogen gas) [6]. This process, termed “symbiotic nitrogen fixation” (SNF), provides the plant with nitrogen to sustain its growth in nitrogen-deficient soils, and has been suggested as one of the factors contributing to the evolutionary success of the Fabaceae plant family [6]. Plant growth and crop yield in agricultural systems emerge as the net results of the interactions between the specific plant cultivar and its associated microbiome [7].

Heavy metals are naturally present in soils; however, their increase over certain thresholds has become a worldwide issue [8]. The major cause of heavy-metal contamination in soil is anthropogenic activities (i.e., atmospheric pollution, industrial and urban waste, mining, and some agricultural practices), while natural contamination is mainly due to weathering of metal-enriched rocks [9]. Plant-associated microbiomes play important roles in phytoremediation, allowing plants to thrive on contaminated soils, alleviating the stress associated with toxic levels of heavy-metals and metalloids (such as As), and increasing phytoextraction and phytostabilization [10–14]. Phytoextraction refers to the plants’ ability to import soil contaminants through their roots, and to accumulate these compounds in the aboveground tissues [15]. Phytostabilization involves the immobilization of pollutants in the soil as a result of either their absorption and accumulation in the roots, their adsorption on the root surface, or their transformation within the rhizosphere into sparingly-soluble compounds [16]. In plants such as legumes, which are generally non-hyperaccumulating species, phytostabilization is likely the more relevant process when considering the remediation of contaminated soils [15–17]. Plant-associated bacteria may promote the chemical transformation, the chelation, or precipitation and sorption of heavy-metals [18] (Figure 1). For instance, some endophytic bacteria may reduce heavy-metal toxicity [19,20]. Improved growth and increased chlorophyll content were detected in several crop plants inoculated with siderophore-producing bacteria [19]. Additionally, enhanced plant biomass production and remediation has been observed in several hyperaccumulating plants following inoculation with rhizosphere or endophytic bacteria with plant growth promoting (PGP) capabilities [21], such as 1-aminocyclopropane-1-carboxylate (ACC) deaminase production (for detailed reviews, please see [11,12]).

The association between leguminous plants and symbiotic rhizobia has stirred the attention of researchers involved in the restoration of heavy-metal-contaminated sites [22]. The possibility to cultivate legumes on marginal and nutrient-poor soils thanks to the intimate association with PGPR, particularly with nitrogen-fixing rhizobia, has been seen as an opportunity to increase phytoremediation efficiencies while simultaneously reducing its costs [23]. Heavy-metals play central roles in symbiotic nitrogen fixation (see [24] for a review of on the role of metals in the symbiosis). Notably, the nitrogenase enzyme is dependent on a cofactor containing molybdenum and iron (FeMo-co), vanadium and iron (VFe-co), or two iron molecules (FeFe-co). There is also evidence for the role of nickel in the symbiosis. For instance, plants inoculated with a deletion mutant of the rhizobium *Sinorhizobium meliloti* lacking the *nreB*-encoded Ni²⁺ efflux system displayed increased growth under controlled conditions [25]. Additionally, a treatment with low doses of Ni²⁺ as the amendment was shown to stimulate nitrogen fixation and plant growth in soybean, and to increase hydrogenase activity in *Rhizobium leguminosarum* bv. *viciae* [26,27]. However, an excess of heavy-metals negatively impacts the symbiosis, reducing the number of symbiotic nodules, the rate of nodulation, and the rate of nitrogen fixation [28,29]. Consequently, in order to promote legume-based phytoremediation through the improvement of the host plant-symbiont partnership, there is a need to discover metal-resistant rhizobia and/or to manipulate existing rhizobial inoculants to increase their level of metal resistance.

In this review, we summarize the main findings on metal resistance in rhizobia: the physiological role, evolution, and genetic determinants of metal resistance, and the perspective to use native and genetically-manipulated rhizobia as inoculants for legumes in phytoremediation practices.

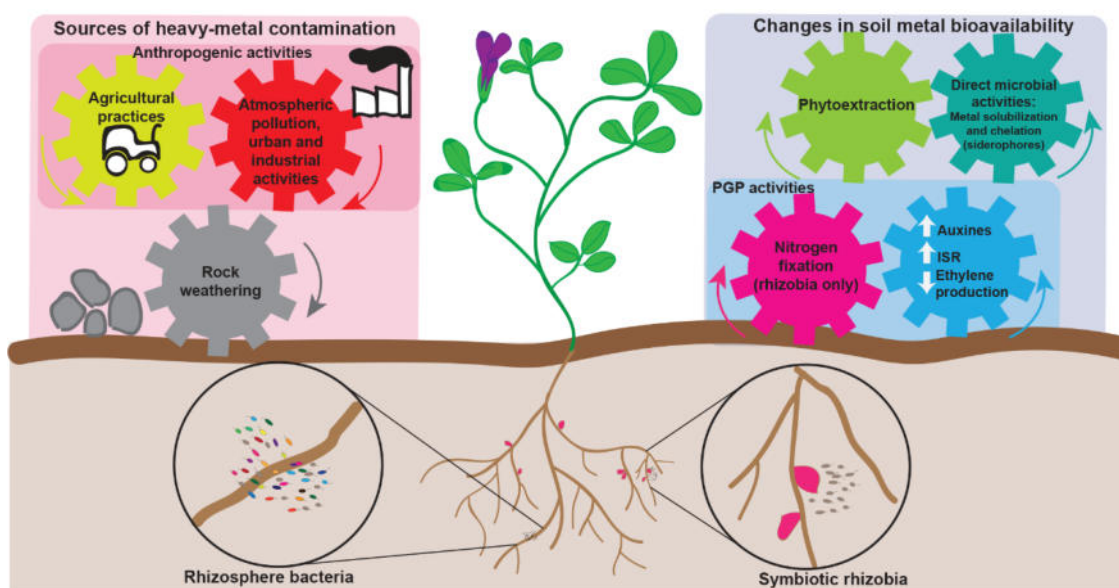


Figure 1. The multiple roles of bacteria in helping plants cope with heavy metals. Plant-associated bacteria may have various roles in both phytostabilization and plant growth. They may influence metal solubility by directly producing molecules for metal chelation (e.g., siderophores), or by influencing plant root growth, resulting in increased production of root exudates. Moreover, both rhizospheric and endophytic bacteria can positively affect plant growth by producing phytohormone molecules (e.g., auxins), alleviating plant stress (e.g., plant ethylene production), or through nitrogen fixation. This latter activity is especially relevant when leguminous plants and their rhizobial microsymbionts are considered. PGP: Plant growth promotion.

2. Legumes in Heavy-Metal Contaminated Areas

The family Leguminosae (Fabaceae) is one of the most diverse among land plants and includes over 700 genera and 20,000 species [30]. Legumes have been proposed as relevant species for phytoremediation, largely due to their ability to colonize marginal lands and nutrient-poor soils [28,31]. In particular, legumes are relevant for phytostabilization, as only a few species have been found to be metal hyperaccumulators (e.g., some species of the genus *Astragalus* isolated in the Western United States are selenium hyperaccumulators) for phytoextraction [23,28,32]. Normally, the symbiosis with rhizobia is inhibited by high levels of heavy-metals in the soil, and genetic engineering techniques have been suggested to improve symbiotic nitrogen fixation under such harsh environmental conditions [33]. However, although such biotechnological proposals are interesting in terms of molecular dissection of the system and theoretical application, currently, there are a number of limitations to the use of genetically-modified microorganisms, including their free release in nature. Analyses on legumes from heavy-metal-contaminated soils have led to the discovery of naturally-resistant rhizobia, which could be used as inoculants in these extreme environments. However, a deeper investigation of leguminous plants growing in metal-enriched sites is required to improve legume-based phytoremediation.

2.1. The Serpentine Vegetation: A Source of Legumes Evolved on Heavy-Metal Rich Soils

Serpentine rocks are an array of ultramafic rock types composed of a hydrous magnesium iron phyllosilicate mineral that originates from metamorphic alterations of peridotite and pyroxene with water. The soils derived from these rocks are characterized by: (i) high levels of nickel, cobalt, and chromium, (ii) low levels of N, P, K, and Ca, and (iii) a high Mg/Ca ratio [34]. This chemical composition strongly limits the growth of most plant species [35], as well as many microorganisms [3]. The presence of serpentine outcrops is scattered across the planet. Along a geological timescale, serpentine outcrops have prompted the evolution of peculiar plant adaptation mechanisms (such as metal hyperaccumulation [36]), which then gave rise to plant differentiation and speciation in a classical

“ecological islands” model [37,38]. Serpentine vegetation in temperate ecosystems includes several leguminous species from various genera, including *Lotus*, *Lupinus*, *Trifolium*, *Vicia*, *Melilotus*, *Medicago*, *Lathyrus*, *Ononis*, *Dorychnium*, *Chamaecytisus*, *Astragalus*, *Anthyllis*, *Cytisus*, and *Acmispon* [39,40]. Serpentine endemic legumes have also been reported, such as *Errazurizia benthamii* [41] in North America, and *Serianthes calycina* [42] in New Caledonia. The microbiomes associated with serpentine plants contain a fraction of microorganisms that appear to have specifically evolved functions to cope with toxic levels of metals present in the soil and in the plant itself [3]. Moreover, some of these microorganisms have been shown to be effective in promoting host plant growth in serpentine soil and, for metal hyperaccumulating plants, to increase metal translocation to the aerial part [43]. Consequently, rhizobia from serpentine endemic legumes (such as Ni-resistant bradyrhizobia from *S. calycina* [42]) may already be adapted to optimizing the fitness of their host in serpentine environments through a long-term natural selection process [44]. Serpentine endemic legumes may therefore represent an ideal source of rhizobia that are naturally highly-competent symbiotic partners in heavy-metal contaminated soils.

2.2. The Search for Heavy-Metal Tolerant Rhizobia and Their Use as Inoculants

Legumes growing in contaminated areas such as mine deposits and serpentine soils have been a source of symbiotic rhizobial strains displaying resistance to heavy-metals, including Zn, Pb, and Cu [45–48]. Table 1 summarizes the main studies on the (positive) effects of rhizobial inoculation on the heavy-metal tolerance of host plants.

Anthyllis vulneraria is one of the most relevant legumes for isolating rhizobia that promote metal-tolerance by the host plant. *A. vulneraria* is a perennial herb from boreo-temperate climate areas in Europe, and it can be found colonizing rocky outcrops and establishing populations on heavy-metal (mainly Zn)-contaminated sites. *Anthyllis* is characterized by determinate nodules, where the meristematic activity disappears shortly after nodule formation, resulting in nodules of spherical shape. *Anthyllis* nodules contain a multilayer cortex: a glycoproteic parenchyma for diffusion, an endodermis, and the outer cortex, which mainly serves as a barrier against pathogens [49]. Nodule bacterial population of leguminous plants grown in Morocco metal-polluted soil displayed a great biodiversity, suggesting that, in these conditions, metal resistant non-rhizobia may efficiently colonize the nodules as endophytes [50]. This highlights the importance of heavy-metal resistance in rhizobia for the establishment of an effective symbiotic interaction in contaminated soils. *A. vulneraria* has been found to be associated with rhizobial symbionts from the genera *Mesorhizobium*, *Rhizobium*, and *Aminobacter*. These include novel rhizobial species, such as *Mesorhizobium metallidurans*, *Rhizobium metallidurans*, and *Aminobacter anthyllidis* [45,47–51]. Interestingly, these novel rhizobial species have so far been identified only in Pb-contaminated environments and not in unpolluted soils [47]. The symbiosis between *A. vulneraria* and its possibly exclusive metal-resistant bacterial species may provide the basis for the establishment of phytoremediation practices. This could involve the use of *A. vulneraria* metal-resistant germplasms, together with its specific natural rhizobial symbionts. Alternatively, the heavy-metal-resistant rhizobia isolated from *A. vulneraria* could be modified, either through laboratory-based experimental evolution studies [52] or direct genetic manipulation, to be capable of entering into an effective symbiosis with other host legumes.

Legumes of the genus *Medicago* have also been deeply investigated for their application in phytoremediation (see Table 1 and references therein). This is mainly because species from this genus are important forage crops for which cultivation techniques and genetics are well established, providing important advantages for future cost-effective applications [53]. Genetically-modified [54,55] and natural [56,57] inocula of *Sinorhizobium* (syn. *Ensifer*) *meliloti* and *Sinorhizobium medicae* [54] have been examined for their abilities to improve plant growth and metal accumulation in the presence of toxic levels of heavy metals such as Cu, Cd, and Zn. However, genetic manipulation is not absolutely required, as interesting results have also been obtained using indigenous *S. meliloti* and *S. medicae* strains directly isolated from contaminated soils [56,57]. For example, inoculation of *Medicago sativa*

plants, grown under field conditions, with wild *S. meliloti* and *S. medicae* strains resulted in active nodulation and the promotion of metal bioaccumulation within the root nodules [56,57]. These results suggest that the exploitation of natural rhizobia could be a valuable tool for promoting land restoration and phytostabilization by legumes.

Legume-based phytoremediation may also be improved through inoculation with a consortium of metal-resistant rhizobia and other PGP bacteria. In metal polluted soil, inoculation of *Lupinus luteus* with *Bradyrhizobium* sp. 750 in consortium with *Pseudomonas* sp. Az13 and *Ochrobactrum cytisi* Azn6.2 increased plant biomass by greater than 100% with respect to uninoculated plants [10]. In contrast, inoculation with only *Bradyrhizobium* sp. 750 increased plant biomass by only 30%. Similarly, co-inoculation of *M. lupina* with *S. meliloti* CCNWSX0020 and *Pseudomonas putida* UW4 resulted in larger plants and greater total Cu accumulation than inoculation with just *S. meliloti* CCNWSX0020 [55]. Inoculation of *Vicia faba*, *Lens culinaris*, and *Sulla coronaria* with consortia of rhizobia and non-rhizobia was also effective at improving plant growth and pod yield when grown in metal-contaminated soil [58]. Moreover, the inoculated *S. coronaria* accumulated significantly more cadmium than uninoculated plants [58]. These results highlight the potential for root-associated microbial communities to influence the success of phytoremediation by rhizobium-inoculated legumes.

It may be concluded that there is great biotechnological potential in increasing the phytoremediation capabilities of legumes by their associated rhizobia. This may be mediated through at least two mechanisms: (i) reducing the toxic effects of the metals, and (ii) promoting the growth of the plant through PGP activities.

Table 1. Studies of phytoremediation mediated by rhizobium-inoculated legumes. NA, not analyzed.

Legume Species	Heavy-Metals in the Soil	Rhizobium Inoculant	Co-Inoculation with Other PGPR?	Evidence for Stimulation of Rhizosphere Microbiota	Type of Study	Effect	Reference
<i>Glycine max</i>	As	<i>Bradyrhizobium</i> sp. Per 3.61	No	NA	Lab scale (pot)	Reduce translocation factor	[59]
<i>Lupinus luteus</i>	Cu, Cd, Pb	<i>Bradyrhizobium</i> sp. 750	Yes	NA	<i>In situ</i>	Increased metal accumulation in root	[10]
<i>Medicago lupulina</i>	Cu	<i>Sinorhizobium meliloti</i> CCNWSX0020	No	NA	<i>In vitro</i> (pot)	Increased plant growth and copper tolerance	[55]
<i>Medicago sativa</i>	Cu	<i>Sinorhizobium meliloti</i> CCNWSX0020	No	NA	<i>In vitro</i>	Increased tolerance of seedlings	[60]
<i>Medicago sativa</i>	Cd	<i>Sinorhizobium meliloti</i> (from contaminated soil [61])	No	NA	Lab scale (pot)	Increased Cd phytoextraction	[56]
<i>Medicago sativa</i>	Zn	<i>Sinorhizobium meliloti</i> (from contaminated soil [61])	No	NA	Lab scale (pot with sterile sand)	Increased Zn accumulation in root	[57]
<i>Medicago truncatula</i>	Cu	<i>Sinorhizobium medicae</i> MA11 (genetically modified with <i>copAB</i> genes)	No	NA	<i>In vitro</i>	Increased metal accumulation in root	[54]
<i>Robinia pseudoacacia</i>	Cd, Zn, Pb	<i>Mesorhizobium loti</i> HZ76	No	Yes	Lab scale (pot)	Increased growth of the plant	[62]
<i>Sulla conoraria</i>	Cu, Zn, Pb	<i>Rhizobium sullae</i>	Yes	NA	<i>In situ</i>	Increased soil Zn stabilization	[58]
<i>Vicia faba</i>	Cu, Zn, Pb	<i>Rhizobium</i> sp. CCNWSX0481	Yes	NA	<i>In situ</i>	Increased soil Cu stabilization	[58]

3. Genetics and Genomics of Heavy-Metal Resistance in Symbiotic Rhizobia

A deep understanding of the genetics and molecular mechanisms of metal resistance remains one of the main goals in environmental biotechnology, with the final aim of promoting the bioremediation (including phytoremediation) of contaminated soils. Table 2 reports the main studies evaluating the genetic determinants of heavy metal resistance in rhizobia. Such studies have most commonly identified the presence of efflux systems that increase metal tolerance by reducing the intracellular concentrations of the metal(s). However, studies employing genome-scale methods, such as transcriptome analyses and transposon mutagenesis, have demonstrated that the cellular response to metal stress involves an intricate genetic network.

Mechanisms mediating resistance to Co and Ni have been identified in many metal resistant rhizobia through the identification of orthologs of metal resistance genes characterized in *Cupriavidus metallidurans* CH34 [63,64]. A gene encoding a DmeF ortholog has been identified in *R. leguminosarum* bv. *viciae* strain UPM791 [65]. DmeF proteins belong to the cation diffusion facilitator (CDF) protein family, which form metal/proton antiport systems to translocate heavy metals across the bacterial membrane [66]. Mutation of the *dmeRF* operon in *R. leguminosarum* resulted in increased sensitivity to Co and Ni, but not to Zn or Cu [65]. The mutant also appeared to be somewhat less effective in symbiosis with pea plants, but not lentil plants, when grown with high concentrations of Co or Ni [65]. Further experiments demonstrated that *dmeR* encodes a Ni- and Co-responsive transcriptional regulator that represses expression of the efflux system in the absence of these metals [65]. Despite being considered a metal-sensitive strain, the *S. meliloti* strain 1021 encodes various metal homeostasis mechanisms, including the DmeRF system, several P-ATPases that are highly common in bacteria, and an ortholog of the *C. metallidurans* NreB protein [25,65,67]. Mutation of *nreB*, encoding a Ni²⁺ efflux protein, resulted in increased sensitivity to Ni, Cu, and low pH, but increased tolerance to urea osmotic stress [25]. The P_{1B-5}-ATPase of *S. meliloti*, termed Nia (nickel iron ATPase), is positively induced by the presence of Ni²⁺ and Fe²⁺, and its expression is higher within nodules relative to free-living cells, which may prevent toxic levels of iron accumulation in the symbiosomes. The wild type protein and recombinants with a deletion of the C-terminal Hr domain have been used to understand the metal specificity of the P_{1B-5}-ATPase family [67].

Genome-wide analyses have been used to investigate the genetics of the resistance mechanisms in *S. meliloti* strain CCNWSX0020, which is resistant to high levels of various heavy-metals (Cu, Zn, Cd and Pb). Gene mutation and transcriptome analyses have suggested the involvement of dozens of genes in the metal-resistance phenotypes of CCNWSX0020, including housekeeping genes [68–70]. Of particular note are the following three operons: the multicopper oxidase (MCO), CopG, and YedYZ operons. The MCO operon is highly expressed following exposure to Cu, and it encodes an outer membrane protein (Omp), the multicopper oxidase CueO, a blue copper azurin-like protein, and a copper chaperone involved in Cu homeostasis [70]. It was proposed that the CueO protein (showing 40% similarity with the CueO protein of *E. coli*) catalyzes Cu(I) oxidation in the periplasmic space, followed by the export of the excessive Cu(II) across the outer membrane [70,71]. The CopG operon consists of four genes: CopG, a CusA-like protein, a FixH-like protein, and a hypothetical protein. Mutation of any of the latter three genes resulted in elevated sensitivity to Zn, Pb, Cd, and Cu, although the mechanism of resistance of this operon remains unknown [70]. The CusA-like protein appears to be a highly-truncated ortholog of the CusA protein of the CusCBA Cu(I) efflux system of *E. coli* [72,73], and may act as a metal binding protein [70]. The FixH-like protein displays similarity to the FixH protein of the FixHGI membrane-bound system, a likely cation transporter that has been shown to be essential for symbiotic nitrogen fixation [74,75]. A FixH-like homolog is also encoded by the pSinB plasmid of *Ensifer* sp. M14 (formerly *Sinorhizobium* sp. M14), where it was also experimentally shown to be involved in metal resistance [76]. Deletion of the *yedYZ* operon resulted in increased sensitivity to Zn, Pb, Cd, and Cu [70]. This was the first report suggesting that YedYZ may be involved in heavy-metal tolerance. In *E. coli*, YedYZ forms a sulfite oxidoreductase [77], and expression of a homologous protein in *S. meliloti* 1021 is induced by taurine and thiosulfate [78]. Thus, the heavy-metal

resistance phenotype may be mediated through disrupting sulfite metabolism, which may influence antioxidant defenses against reactive oxygen species (ROS) generated by heavy metals [70].

Many scientists have used population genetics approaches to identify loci associated with heavy-metal resistance. This was achieved by performing genome-wide association studies on a population's pan-genome, considering allelic variations in the core genome (the set of genes shared by the members of the population), and gene presence/absence in the dispensable genome fraction (the set of genes present in only a fraction of the population). Genomic variants statistically associated with nickel adaptation were identified in a *Mesorhizobium* population using this approach [79]. A population of 47 *Mesorhizobium* strains, isolated from root nodules and soils with different levels of nickel contamination, was studied. Most of the variants associated with metal adaptation were found in the dispensable genome fraction. This work highlights that adaptation to heavy metal stress is likely driven predominately by horizontal gene transfer, and is not due to mutations of pre-existing genes.

Multiple studies have demonstrated that the genetic determinants of metal-resistance in rhizobia are relevant for phytoremediation purposes. Mutation of *ceuO*, *yedYZ*, and the *fixH*-like gene negatively impacted the *M. lupulina* nodulation kinetics of *S. meliloti* CCNWSX0020 in the presence of Cu and/or Zn [70], while deletion of the *cusA*-like gene had a negative effect, even in the absence of heavy metals. It was separately observed that *M. lupulina* plants inoculated with *S. meliloti* CCNWSX0020 strains with independent mutations in five Cu resistance loci were smaller than plants inoculated with the wild type, when grown in the presence of Cu [80]. Notably, *M. lupulina* plants inoculated with any of the *S. meliloti* CCNWSX0020 mutants mentioned above accumulated lower amounts of Cu and/or Ni [78]. Similarly, *Robinia pseudoacacia* plants inoculated with a *Mesorhizobium amorphae* 186 *copA* mutant accumulated 10–15% less Cu than plants inoculated with the wild type [81]; however, no effect on plant growth was observed.

Table 2. Genes for heavy-metal (and metalloids) tolerance in symbiotic rhizobia. A summary of the main genes whose function in tolerance was confirmed experimentally is reported.

Strain	Host Plant	Isolation Site	Method of Identification	Gene(s)	Metal(s) Tolerance	Reference
<i>Bradhyrhizobium</i> spp.	<i>Serianthes calycina</i>	Serpentine (New Caledonia)	PCR amplification, site-directed mutagenesis	<i>cnr/nre</i> systems	Co, Ni	[42]
<i>Mesorhizobium</i> spp.	<i>Acmispon wrangelianus</i>	Serpentine (California)	Association mapping	Various	Ni	[79]
<i>Mesorhizobium metallidurans</i>	<i>Antyllis vulneraria</i>	Zinc mine (France)	Cosmid library	<i>cadA</i> (PIB-2-type ATPase)	Zn, Cd	[82]
<i>Sinorhizobium meliloti</i> 1021	<i>Medicago sativa</i>	Laboratory strain	Site-directed gene deletion	<i>nreB</i> (SMA1641)	Ni	[25]
<i>Sinorhizobium meliloti</i> 1021	<i>Medicago sativa</i>	Laboratory strain	Tn5 insertion, biochemical characterization	SMA1163 (PIB-5-ATPase)	Ni, Fe	[67]
<i>Sinorhizobium meliloti</i> CCNWSX0020	<i>Medicago lupulina</i>	Mine tailings (China)	Site-directed gene deletion and transcriptomics	P1B-type ATPases and others	Cu, Zn	[69,70]
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> UPM1137	<i>Pisum sativum</i>	Serpentine (Italy)	Transposon mutagenesis	14 loci (gene annotation corresponds to Rlv 3841 genome): RL2862, RL2436, RL2322, pRL110066, RL1351, RL4539, pRL90287, RL4188, RL2793, RL2100, RL0615, RL1589, pRL110071, RL1553	Ni, Co	[83]

4. Genomic Manipulation Strategies for Improving Legume Phytoremediation

Various attempts have been made to increase plant growth in the presence of toxic metal concentrations through genetic modification of their rhizobial microsymbionts. One approach is

to introduce new genes conferring heavy-metal resistance into the rhizobium. For example, inoculation of a genetically-modified *M. truncatula* line (which expressed a metallothionein gene from *Arabidopsis thaliana* in its roots) with wild type *S. medicae* resulted in elevated Cu tolerance [84]. Copper tolerance was further increased using a *S. medicae* strain expressing the *P. fluorescence copAB* Cu resistance genes [84]. Inoculation with the latter strain also resulted in elevated Cu accumulation in the plant roots [84]. Similarly, the introduction of an algal As(III) methyltransferase gene (*arsM*) into the chromosome of *R. leguminosarum* bv. *trifolii* produced a strain that was able to methylate and volatilize inorganic arsenic in symbiosis with red clover (with no negative impact on nitrogen fixation) [85]. A second approach is the insertion of genes in rhizobia to modulate phytohormone production, thereby reducing plant stress perception. For example, an ACC deaminase overproducing *S. meliloti* strain increased Cu tolerance and promoted plant growth of the host plant *M. lupulina* [86]. This result was probably due to reduced production of ethylene by the host plant, in turn decreasing stress perception. However, it should be kept in mind that a relatively high number of genes may contribute to the heavy-metal stress response [87–89]. Consequently, a multigenic, genome-wide approach should be considered when attempting to genetically modify competitive rhizobial symbionts to have increased heavy-metal tolerance. One possibility along these lines is the introduction of entire, large resistance plasmids from a non-symbiotic (but highly resistant) strain to a phylogenetically-related, symbiotic metal-sensitive strain. A candidate plasmid for such studies is the pSinA plasmid of the non-symbiotic *Ensifer* sp. M14, which was isolated from an As-contaminated gold mine [76,90,91]. The pSinA plasmid is a self-transmissible replicon with a broad host range. It harbors a genomic island with genes for arsenite oxidation (*aio* genes) and arsenite resistance (*ars* genes), and its transfer to other species results in increased arsenic resistance [90]. Transfer of the pSinA plasmid to closely-related rhizobia, such as *S. meliloti*, may result in the construction of As-tolerant legume symbionts for use in arsenic remediation. Subsequent acquisition of pSinA by other members of the rhizospheric microbiota may further stimulate phytoremediation of arsenic contaminated soils through reducing the arsenic toxicity (oxidizing arsenites to arsenates) and biofortification (increase of the arsenic resistance level) of the autochthonic or augmented microflora.

Similarly, elite and metal-resistant rhizobia may be obtained through combining within one strain genomic elements from the species pangenome. The genomes of most rhizobia are extremely diverse, and many rhizobia have a divided genome structure consisting of at least two large DNA replicons [92]. Although there can be numerous inter-replicon functional, regulatory, and genetic interactions [93–95], in some ways, each replicon in a divided genome could be considered as an independent evolutionary and functional element [94,96–98]. Recently, it was shown that the genome and metabolism of *S. meliloti* is robust to the replacement of the symbiotic megaplasmid with the symbiotic megaplasmid of a different wild-type isolate [99]. Therefore, it may be possible to construct “hybrid” strains (Figure 2) with a collection of replicons derived from various wild-type isolates, potentially allowing for the development of elite strains with improved multifactorial phenotypes (e.g., resistance to heavy-metals, high symbiotic efficiency, and competition toward the indigenous soil microbiota).

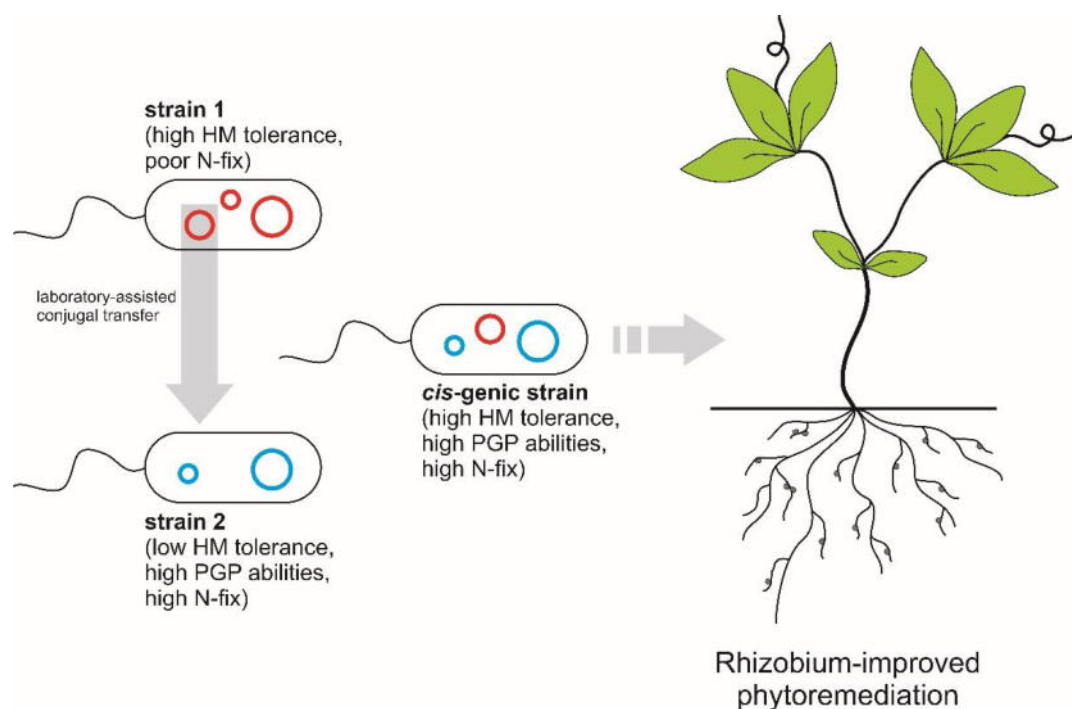


Figure 2. A synthetic biology-based proposal to increase rhizobial-mediated heavy-metal tolerance. Surveys of rhizobial phenotypic and genetic diversity in heavy-metal (HM) rich areas facilitates the discovery of strains (strain 1) with high levels of heavy-metal resistance. However, such strains may not be competitive or good nitrogen-fixers in the crops to be used for phytoremediation. The simultaneous transfer of a large collection of genomic determinants that contribute to HM tolerance, good PGP, and/or nitrogen fixation (N-fix) abilities between two or more strains (strain 2) could create hybrid strains (*cis-genic strain*) with improved features for application in the field for phytoremediation.

5. Conclusions

In recent years, the number of studies related to the potential exploitation of rhizobium–legume symbioses for phytoremediation practices have increased enormously as a result of environmental emergencies. In this brief review, we have presented state-of-the-art studies on heavy-metal tolerant rhizobia, and on their applications in phytoremediation as legume symbionts. A large number of investigations have indicated that rhizobia, and especially heavy-metal resistant rhizobia, can increase legume heavy-metal tolerance and promote improved legume growth in metal-rich soils, thereby resulting in greater removal of heavy-metals from the soil. Heavy-metal resistant rhizobia have been isolated from the nodules of legumes grown in soils that are rich in heavy-metals as a result of geological (e.g., serpentine outcrops) or anthropic causes (e.g., mine deposits). Genetic and genomic studies of heavy-metal resistant rhizobia have shown that although relatively few genes act as the main player in tolerance, a much larger set of genes may be involved in maximizing fitness in heavy metal rich growth conditions. Some of these genes, such as the systems for Ni^{2+} efflux in *S. meliloti*, may also contribute to a linkage between metal homeostasis and nitrogen-fixation efficiency. As such, systems-biology approaches are required to develop an overall picture of heavy-metal resistance and the ways that we can increase and exploit it in biotechnology. It will also be important to keep in mind that the engineering of rhizobia should consider several additional aspects, including the rhizobial genotype, the host plant genotype, and the interactions between the rhizobium with the soil and root microbiota [100].

Going forward, we suggest that large-scale genome-manipulation approaches may be considered in developing rhizobial strains with elite phenotypes (e.g., high heavy-metal resistance, high nitrogen-fixation ability, high competitiveness, etc.) for use in phytoremediation applications. As a

pre-requisite to such studies, it will be necessary to increase efforts at creating culture collections of rhizobial strains from contaminated areas, since the strains isolated from these environments is quite limited in number and in terms of host plant (see also [23]). Such efforts would benefit from exploring areas that have evolved peculiar flora, such as serpentine outcrops, maximizing the chance to find well-adapted strains. Whole genome sequencing, genome-scale mutagenesis (such as Tn-seq or INseq [101]), and metabolic modeling of these strains could then be employed to fully characterize the genomic basis for tolerance against the contaminants.

Author Contributions: Conceptualization, A.M., C.F., A.C. and G.C.D.; Writing-Original Draft Preparation, A.M., C.F., A.C. and G.C.D.; Writing-Review & Editing, A.M., C.F., A.C., G.C.D., L.D. and K.D.A.; Funding Acquisition, A.M.

Funding: This research was partially funded by Ente Cassa di Risparmio di Firenze, grant “2017-0719” and by the University of Florence, project “Dinamiche dell’evoluzione dei genomi batterici: l’evoluzione del genoma multipartito e la suddivisione in moduli funzionali”, call “PROGETTI STRATEGICI DI ATENEO ANNO 2014” to A.M. A.C. was supported by a grant from Fondazione Adriano Buzzati-Traverso. G.C.D. was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a PDF fellowship. K.D. was supported by the European Molecular Biology Organization in the frame of the EMBO Short-Term Fellowship program [grant number 7376] and the National Science Center (Poland) in the frame of the Preludium grant No. 2016/23/N/NZ9/01655.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mendes, R.; Garbeva, P.; Raaijmakers, J.M. The rhizosphere microbiome: Significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* **2013**, *37*, 634–663. [[CrossRef](#)] [[PubMed](#)]
- Bai, Y.; Müller, D.B.; Srinivas, G.; Garrido-Oter, R.; Pothoff, E.; Rott, M.; Dombrowski, N.; Münch, P.C.; Spaepen, S.; Remus-Emsermann, M.; et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **2015**, *528*, 364. [[CrossRef](#)] [[PubMed](#)]
- Mengoni, A.; Schat, H.; Vangronsveld, J. Plants as extreme environments? Ni-resistant bacteria and Ni-hyperaccumulators of serpentine flora. *Plant Soil* **2010**, *331*, 5–16. [[CrossRef](#)]
- Pini, F.; Frascella, A.; Santopolo, L.; Bazzicalupo, M.; Biondi, E.G.; Scotti, C.; Mengoni, A. Exploring the plant-associated bacterial communities in *Medicago sativa* L. *BMC Microbiol.* **2012**, *12*, 78. [[CrossRef](#)] [[PubMed](#)]
- Nadeem, S.M.; Ahmad, M.; Zahir, Z.A.; Javaid, A.; Ashraf, M. The role of mycorrhizae and plant growth promoting rhizobacteria (PGPR) in improving crop productivity under stressful environments. *Biotechnol. Adv.* **2014**, *32*, 429–448. [[CrossRef](#)] [[PubMed](#)]
- Sprent, J.I. *Legume Nodulation: A Global Perspective*; John Wiley & Sons: Hoboken, NJ, USA, 2009; ISBN 1444316397.
- Theis, K.R.; Dheilly, N.M.; Klassen, J.L.; Brucker, R.M.; Baines, J.F.; Bosch, T.C.G.; Cryan, J.F.; Gilbert, S.F.; Goodnight, C.J.; Lloyd, E.A.; et al. Getting the Hologenome Concept Right: An Eco-Evolutionary Framework for Hosts and Their Microbiomes. *Msystems* **2016**, *1*, e00028-16. [[CrossRef](#)] [[PubMed](#)]
- Chibuikwe, G.U.; Obiora, S.C. Heavy metal polluted soils: Effect on plants and bioremediation methods. *Appl. Environ. Soil Sci.* **2014**. [[CrossRef](#)]
- Lebrazi, S.; Fikri-Benbrahim, K. Rhizobium-Legume Symbioses: Heavy metal effects and principal approaches for bioremediation of contaminated soil. In *Legumes for Soil Health and Sustainable Management*; Springer: Heidelberg, Germany, 2018; pp. 205–233. ISBN 978-981-13-0252-7.
- Dary, M.; Chamber-Pérez, M.A.; Palomares, A.J.; Pajuelo, E. “In situ” phytostabilisation of heavy metal polluted soils using *Lupinus luteus* inoculated with metal resistant plant-growth promoting rhizobacteria. *J. Hazard. Mater.* **2010**, *177*, 323–330. [[CrossRef](#)] [[PubMed](#)]
- Kong, Z.; Glick, B.R. The Role of Plant Growth-Promoting Bacteria in Metal Phytoremediation. In *Advanced in Microbial Physiology*, 1st ed.; Elsevier Ltd.: New York, NY, USA, 2017; Volume 71, pp. 97–132. ISBN 0065-2911.
- Sessitsch, A.; Kuffner, M.; Kidd, P.; Vangronsveld, J.; Wenzel, W.W.; Fallmann, K.; Puschenreiter, M. The role of plant-associated bacteria in the mobilization and phytoextraction of trace elements in contaminated soils. *Soil Biol. Biochem.* **2013**, *60*, 182–194. [[CrossRef](#)] [[PubMed](#)]

13. Weyens, N.; Lelie, D. Van Der; Taghavi, S.; Newman, L. Exploiting plant—Microbe partnerships to improve biomass production and remediation. *Trends Biotechnol.* **2009**, *1*, 1–8. [[CrossRef](#)] [[PubMed](#)]
14. Kidd, P.S.; Alvarez-Lopez, V.; Becerra-Castro, C.; Cabello-Conejo, M.; Prieto-Fernandez, A. Potential role of plant-associated bacteria in plant metal uptake and implications in phytotechnologies. In *Advances in Botanical Research*; Academic Press: London UK; New York, NY, USA, 2017; pp. 87–126.
15. Mahar, A.; Wang, P.; Ali, A.; Awasthi, M.K.; Lahori, A.H.; Wang, Q.; Li, R.; Zhang, Z. Challenges and opportunities in the phytoremediation of heavy metals contaminated soils: A review. *Ecotoxicol. Environ. Saf.* **2016**, *126*, 111–121. [[CrossRef](#)] [[PubMed](#)]
16. Bolan, N.S.; Park, J.H.; Robinson, B.; Naidu, R.; Huh, K.Y. Phytostabilization: A green approach to contaminant containment. *Adv. Agron.* **2011**, *112*, 145–204. [[CrossRef](#)]
17. Mahieu, S.; Frérot, H.; Vidal, C.; Galiana, A.; Heulin, K.; Maure, L.; Brunel, B.; Lefebvre, C.; Escarré, J.; Cleyet-Marel, J.-C. *Anthyllis vulneraria*/*Mesorhizobium metallidurans*, an efficient symbiotic nitrogen fixing association able to grow in mine tailings highly contaminated by Zn, Pb and Cd. *Plant Soil* **2011**, *342*, 405–417. [[CrossRef](#)]
18. Gadd, G.M. Accumulation and transformation of metals by microorganisms. In *Biotechnology: Special Processes*; John Wiley & Sons: Hoboken, NJ, USA, 2008; Volume 10, pp. 226–264. ISBN 978-3-52-762093-7.
19. Mastretta, C.; Taghavi, S.; Van Der Lelie, D.; Mengoni, A.; Galardi, F.; Gonnelli, C.; Barac, T.; Boulet, J.; Weyens, N.; Vangronsveld, J. Endophytic bacteria from seeds of *Nicotiana tabacum* can reduce cadmium phytotoxicity. *Int. J. Phytoremediat.* **2009**, *11*, 251–267. [[CrossRef](#)]
20. Etesami, H. Bacterial mediated alleviation of heavy metal stress and decreased accumulation of metals in plant tissues: Mechanisms and future prospects. *Ecotoxicol. Environ. Saf.* **2018**, *147*, 175–191. [[CrossRef](#)] [[PubMed](#)]
21. Novo, L.A.B.; Castro, P.M.L.; Alvarenga, P.; da Silva, E.F. Plant Growth—Promoting Rhizobacteria-Assisted phytoremediation of mine soils. In *Bio-Geotechnologies for Mine Site Rehabilitation*; Elsevier: New York, NY, USA, 2018; pp. 281–295. ISBN 978-0-12-812987-6.
22. Teng, Y.; Wang, X.; Li, L.; Li, Z.; Luo, Y. Rhizobia and their bio-partners as novel drivers for functional remediation in contaminated soils. *Front. Plant Sci.* **2015**, *6*, 32. [[CrossRef](#)] [[PubMed](#)]
23. Checcucci, A.; Bazzicalupo, M.; Mengoni, A. Exploiting nitrogen-fixing rhizobial symbionts genetic resources for improving phytoremediation of contaminated soils. In *Enhancing Cleanup of Environmental Pollutants*; Springer: Heidelberg, Germany, 2017; Volume 1, pp. 275–288. ISBN 978-3-31-955426-6.
24. González-Guerrero, M.; Matthiadis, A.; Saez, Á.; Long, T. Fixating on metals: New insights into the role of metals in nodulation and symbiotic nitrogen fixation. *Front. Plant Sci.* **2014**, *5*, 45. [[CrossRef](#)] [[PubMed](#)]
25. Pini, F.; Spini, G.; Galardini, M.; Bazzicalupo, M.; Benedetti, A.; Chianciani, M.; Florio, A.; Lagomarsino, A.; Migliore, M.; Mocali, S.; et al. Molecular phylogeny of the nickel-resistance gene *nreB* and functional role in the nickel sensitive symbiotic nitrogen fixing bacterium *Sinorhizobium meliloti*. *Plant Soil* **2013**, *377*, 189–201. [[CrossRef](#)]
26. Lavres, J.; Castro Franco, G.; de Sousa Câmara, G.M. Soybean seed treatment with nickel improves biological nitrogen fixation and urease activity. *Front. Environ. Sci.* **2016**, *4*, 37. [[CrossRef](#)]
27. Ureta, A.-C.; Imperial, J.; Ruiz-Argüeso, T.; Palacios, J.M. *Rhizobium leguminosarum* biovar viciae symbiotic hydrogenase activity and processing are limited by the level of nickel in agricultural soils. *Appl. Environ. Microbiol.* **2005**, *71*, 7603–7606. [[CrossRef](#)] [[PubMed](#)]
28. Hao, X.; Taghavi, S.; Xie, P.; Orbach, M.J.; Alwathnani, H.A.; Rensing, C.; Wei, G. Phytoremediation of heavy and transition metals aided by legume-rhizobia symbiosis. *Int. J. Phytoremediat.* **2014**, *16*, 179–202. [[CrossRef](#)] [[PubMed](#)]
29. Ahmad, E.; Zaidi, A.; Khan, M.S.; Oves, M. Heavy metal toxicity to symbiotic nitrogen-fixing microorganism and host legumes. In *Toxicity of Heavy Metals to Legumes and Bioremediation*; Springer: Heidelberg, Germany, 2012; pp. 29–44. ISBN 3709107296.
30. Doyle, J.J.; Luckow, M.A. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol.* **2003**, *131*, 900–910. [[CrossRef](#)] [[PubMed](#)]
31. Bradshaw, A.D.; Chadwick, M.J. *The Restoration of Land: The Ecology and Reclamation of Derelict and Degraded Land*; University of California Press: Berkeley, CA, USA, 1980; ISBN 0520039610.
32. Reeves, R.D.; van der Ent, A.; Baker, A.J.M. Global distribution and ecology of hyperaccumulator plants. In *Agromining: Farming for Metals*; Springer: Heidelberg, Germany, 2018; pp. 75–92, ISBN 978-3-319-61898-2.

33. Pajuelo, E.; Rodríguez-Llorente, I.D.; Lafuente, A.; Caviedes, M.Á. Legume–rhizobium symbioses as a tool for bioremediation of heavy metal polluted soils. In *Biomanagement of Metal-Contaminated Soils*; Springer: Heidelberg, Germany, 2011; pp. 95–123. ISBN 978-94-007-1914-9.
34. Brooks, R.R. *Serpentine and Its Vegetation: A Multidisciplinary Approach*; Croom Helm: London, UK, 1987; ISBN 0709950632.
35. Brady, K.U.; Kruckeberg, A.R.; Bradshaw, H.D., Jr. Evolutionary ecology of plant adaptation to serpentine soils. *Annu. Rev. Ecol. Evol. Syst.* **2005**, *36*, 243–266. [[CrossRef](#)]
36. Words, K. Metal Hyperaccumulation in plants. *Annu. Rev. Plant Biol.* **2010**, *61*, 517–534. [[CrossRef](#)]
37. Harrison, S.; Rajakaruna, N. *Serpentine: The Evolution and Ecology of a Model System*; University of California Press: Berkeley, CA, USA, 2011; ISBN 0520268350.
38. Mengoni, A.; Mocali, S.; Surico, G.; Tegli, S.; Fani, R. Fluctuation of endophytic bacteria and phytoplasmosis in elm trees. *Microbiol. Res.* **2003**, *158*, 363–369. [[CrossRef](#)] [[PubMed](#)]
39. Alexander, E.B.; Coleman, R.G.; Harrison, S.P.; Keeler-Wolfe, T. *Serpentine Geocology of Western North America: Geology, Soils, and Vegetation*; OUP: Oxford, UK, 2007; ISBN 019516508X.
40. Pustahija, F.; Brown, S.C.; Bogunić, F.; Bašić, N.; Muratović, E.; Ollier, S.; Hidalgo, O.; Bourge, M.; Stevanović, V.; Siljak-Yakovlev, S. Small genomes dominate in plants growing on serpentine soils in West Balkans, an exhaustive study of 8 habitats covering 308 taxa. *Plant Soil* **2013**, *373*, 427–453. [[CrossRef](#)]
41. Selvi, F. Diversity, geographic variation and conservation of the serpentine flora of Tuscany (Italy). *Biodivers. Conserv.* **2007**, *16*, 1423–1439. [[CrossRef](#)]
42. Chaintreuil, C.; Rigault, F.; Moulin, L.; Jaffré, T.; Fardoux, J.; Giraud, E.; Dreyfus, B.; Bailly, X. Nickel resistance determinants in *Bradyrhizobium* strains from nodules of the endemic New Caledonia legume *Serianthes calycina*. *Appl. Environ. Microbiol.* **2007**, *73*, 8018–8022. [[CrossRef](#)] [[PubMed](#)]
43. Rajkumar, M.; Narasimha, M.; Prasad, V.; Freitas, H.; Ae, N. Biotechnological applications of serpentine soil bacteria for phytoremediation of trace metals. *Crit. Rev. Biotechnol.* **2009**, *29*, 120–130. [[CrossRef](#)] [[PubMed](#)]
44. Friesen, M.L. Widespread fitness alignment in the legume—Rhizobium symbiosis. *New Phytol.* **2012**, *194*, 1096–1111. [[CrossRef](#)] [[PubMed](#)]
45. Grison, C.M.; Jackson, S.; Merlot, S.; Dobson, A.; Grison, C. *Rhizobium metallidurans* sp. nov., a symbiotic heavy metal resistant bacterium isolated from the anthyllis vulneraria Zn-hyperaccumulator. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 1525–1530. [[CrossRef](#)] [[PubMed](#)]
46. Ye, M.; Liao, B.; Li, J.T.; Mengoni, A.; Hu, M.; Luo, W.C.; Shu, W.S. Contrasting patterns of genetic divergence in two sympatric pseudo-metallophytes: *Rumex acetosa* L. and *Commelina communis* L. *BMC Evol. Biol.* **2012**, *12*, 84. [[CrossRef](#)] [[PubMed](#)]
47. Mohamad, R.; Maynaud, G.; Le Quéré, A.; Vidal, C.; Klonowska, A.; Yashiro, E.; Cleyet-Marel, J.-C.; Brunel, B. Ancient heavy metal contamination in soils as a driver of tolerant *Anthyllis vulneraria* rhizobial communities. *Appl. Environ. Microbiol.* **2016**, *83*, e01735-16. [[CrossRef](#)] [[PubMed](#)]
48. Vidal, C.; Chantreuil, C.; Berge, O.; Mauré, L.; Escarré, J.; Béna, G.; Brunel, B.; Cleyet-Marel, J.C. *Mesorhizobium metallidurans* sp. nov., a metal-resistant symbiont of *Anthyllis vulneraria* growing on metalcolous soil in Languedoc, France. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 850–855. [[CrossRef](#)] [[PubMed](#)]
49. Sujkowska-Rybkowska, M.; Ważny, R. Metal resistant rhizobia and ultrastructure of *Anthyllis vulneraria* nodules from zinc and lead contaminated tailing in Poland. *Int. J. Phytoremediat.* **2018**, *20*, 709–720. [[CrossRef](#)] [[PubMed](#)]
50. El Aafi, N.; Saidi, N.; Maltouf, A.F.; Perez-Palacios, P.; Dary, M.; Brhada, F.; Pajuelo, E. Prospecting metal-tolerant rhizobia for phytoremediation of mining soils from Morocco using *Anthyllis vulneraria* L. *Environ. Sci. Pollut. Res.* **2015**, *22*, 4500–4512. [[CrossRef](#)] [[PubMed](#)]
51. Maynaud, G.; Willems, A.; Soussou, S.; Vidal, C.; Mauré, L.; Moulin, L.; Cleyet-Marel, J.-C.; Brunel, B. Molecular and phenotypic characterization of strains nodulating *Anthyllis vulneraria* in mine tailings, and proposal of *Aminobacter anthyllidis* sp. nov., the first definition of *Aminobacter* as legume-nodulating bacteria. *Syst. Appl. Microbiol.* **2012**, *35*, 65–72. [[CrossRef](#)] [[PubMed](#)]
52. Gilbert, L.B.; Heeb, P.; Gris, C.; Timmers, T.; Batut, J.; Masson-boivin, C. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol.* **2010**, *8*. [[CrossRef](#)]
53. Vamerli, T.; Bandiera, M.; Mosca, G. Field crops for phytoremediation of metal-contaminated land. A review. *Environ. Chem. Lett.* **2010**, *8*, 1–17. [[CrossRef](#)]

54. Delgadillo, J.; Lafuente, A.; Doukkali, B.; Redondo-Gómez, S.; Mateos-Naranjo, E.; Caviedes, M.A.; Pajuelo, E.; Rodríguez-Llorente, I.D. Improving legume nodulation and Cu rhizostabilization using a genetically modified rhizobia. *Environ. Technol.* **2015**, *36*, 1237–1245. [[CrossRef](#)] [[PubMed](#)]
55. Kong, Z.; Glick, B.R.; Duan, J.; Ding, S.; Tian, J.; McConkey, B.J.; Wei, G. Effects of 1-aminocyclopropane-1-carboxylate (ACC) deaminase-overproducing *Sinorhizobium meliloti* on plant growth and copper tolerance of *Medicago lupulina*. *Plant Soil* **2015**, *70*, 5891–5897. [[CrossRef](#)]
56. Ghnaya, T.; Mnassri, M.; Ghabriche, R.; Wali, M.; Poschenrieder, C.; Lutts, S.; Abdelly, C. Nodulation by *Sinorhizobium meliloti* originated from a mining soil alleviates Cd toxicity and increases Cd-phytoextraction in *Medicago sativa* L. *Front. Plant Sci.* **2015**, *6*, 1–10. [[CrossRef](#)] [[PubMed](#)]
57. Zribi, K.; Nouairi, I.; Slama, I.; Talbi-Zribi, O.; Mhadhbi, H. *Medicago sativa*—*Sinorhizobium meliloti* Symbiosis Promotes the Bioaccumulation of Zinc in Nodulated Roots. *Int. J. Phytoremediat.* **2015**, *17*, 49–55. [[CrossRef](#)] [[PubMed](#)]
58. Saadani, O.; Fatnassi, I.C.; Chiboub, M.; Abdelkrim, S.; Barhoumi, F.; Jebara, M.; Jebara, S.H. In situ phytostabilisation capacity of three legumes and their associated Plant Growth Promoting Bacteria (PGPBs) in mine tailings of northern Tunisia. *Ecotoxicol. Environ. Saf.* **2016**, *130*, 263–269. [[CrossRef](#)] [[PubMed](#)]
59. Bianucci, E.; Godoy, A.; Furlan, A.; Peralta, J.M.; Hernández, L.E.; Carpena-Ruiz, R.O.; Castro, S. Arsenic toxicity in soybean alleviated by a symbiotic species of *Bradyrhizobium*. *Symbiosis* **2018**, *74*, 167–176. [[CrossRef](#)]
60. Chen, J.; Liu, Y.Q.; Yan, X.W.; Wei, G.H.; Zhang, J.H.; Fang, L.C. *Rhizobium* inoculation enhances copper tolerance by affecting copper uptake and regulating the ascorbate-glutathione cycle and phytochelatin biosynthesis-related gene expression in *Medicago sativa* seedlings. *Ecotoxicol. Environ. Saf.* **2018**, *162*, 312–323. [[CrossRef](#)] [[PubMed](#)]
61. Zribi, K.; Djéballi, N.; Mrabet, M.; Khayat, N.; Smaoui, A.; Mlayah, A.; Aouani, M.E. Physiological responses to cadmium, copper, lead, and zinc of *Sinorhizobium* sp. strains nodulating *Medicago sativa* grown in Tunisian mining soils. *Ann. Microbiol.* **2012**, *62*, 1181–1188. [[CrossRef](#)]
62. Fan, M.; Xiao, X.; Guo, Y.; Zhang, J.; Wang, E.; Chen, W.; Lin, Y.; Wei, G. Enhanced phytoremediation of *Robinia pseudoacacia* in heavy metal-contaminated soils with rhizobia and the associated bacterial community structure and function. *Chemosphere* **2018**, *197*, 729–740. [[CrossRef](#)] [[PubMed](#)]
63. Van Houdt, R.; Mergeay, M. Genomic context of metal response genes in *Cupriavidus metallidurans* with a focus on strain CH34. In *Metal Response in Cupriavidus Metallidurans*; Springer: Heidelberg, Germany, 2015; pp. 21–44. ISBN 978-3-319-20594-6.
64. Rozycki, T. Von; Nies, A.D.H.; Alcaligenes, W.Á.; Ch, Á.Á.H. *Cupriavidus metallidurans*: Evolution of a metal-resistant bacterium. *Anton. Leeuwenhoek* **2008**, *96*, 115–139. [[CrossRef](#)] [[PubMed](#)]
65. Rubio-Sanz, L.; Prieto, R.I.; Imperial, J.; Palacios, J.M.; Brito, B. Functional and expression analysis of the metal-inducible *dmeRF* system from *Rhizobium leguminosarum* bv. *viciae*. *Appl. Environ. Microbiol.* **2013**, *79*, 6414–6422. [[CrossRef](#)] [[PubMed](#)]
66. Haney, C.J.; Grass, G.; Franke, S.; Rensing, C. New developments in the understanding of the cation diffusion facilitator family. *J. Ind. Microbiol. Biotechnol.* **2005**, *32*, 215–226. [[CrossRef](#)] [[PubMed](#)]
67. Zielazinski, E.L.; González-Guerrero, M.; Subramanian, P.; Stemmler, T.L.; Argüello, J.M.; Rosenzweig, A.C. *Sinorhizobium meliloti* Nia is a P1B-5-ATPase expressed in the nodule during plant symbiosis and is involved in Ni and Fe transport. *Metallomics* **2013**, *5*, 1614–1623. [[CrossRef](#)] [[PubMed](#)]
68. Li, Z.; Lu, M.; Wei, G. An omp gene enhances cell tolerance of Cu(II) in *Sinorhizobium meliloti* CCNWSX0020. *World J. Microbiol. Biotechnol.* **2013**, *29*, 1655–1660. [[CrossRef](#)] [[PubMed](#)]
69. Lu, M.; Li, Z.; Liang, J.; Wei, Y.; Rensing, C.; Wei, G. Zinc resistance mechanisms of P 1B-type ATPases in *Sinorhizobium meliloti* CCNWSX0020. *Sci. Rep.* **2016**, *6*, 1–12. [[CrossRef](#)]
70. Lu, M.; Jiao, S.; Gao, E.; Song, X.; Li, Z.; Hao, X.; Rensing, C.; Wei, G. Transcriptome response to heavy metals in *Sinorhizobium meliloti* CCNWSX0020 reveals new metal resistance determinants that also promote bioremediation by *Medicago lupulina* in metal contaminated soil. *Appl. Environ. Microbiol.* **2017**, *83*. [[CrossRef](#)] [[PubMed](#)]
71. Grass, G.; Rensing, C. CueO is a multi-copper oxidase that confers copper tolerance in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **2001**, *286*, 902–908. [[CrossRef](#)] [[PubMed](#)]
72. Franke, S.; Grass, G.; Rensing, C.; Nies, D.H. Molecular analysis of the copper-transporting efflux system CusCFBA of *Escherichia coli*. *J. Bacteriol.* **2003**, *185*, 3804–3812. [[CrossRef](#)] [[PubMed](#)]

73. Long, F.; Su, C.-C.; Lei, H.-T.; Bolla, J.R.; Do, S.V.; Yu, E.W. Structure and mechanism of the tripartite CusCBA heavy-metal efflux complex. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367*, 1047–1058. [[CrossRef](#)] [[PubMed](#)]
74. Kahn, D.; David, M.; Domergue, O.; Davaeran, M.L.; Ghai, J.; Hirsch, P.R.; Batut, J. *Rhizobium meliloti* fixGHI sequence predicts involvement of a specific cation pump in symbiotic nitrogen fixation. *J. Bacteriol.* **1989**, *171*, 929–939. [[CrossRef](#)] [[PubMed](#)]
75. Batut, J.; Terzaghi, B.; Gherardi, M.; Huguet, M.; Terzaghi, E.; Garnerone, A.M.; Boistard, P.; Huguet, T. Localization of a symbiotic *fix* region on *Rhizobium meliloti* pSym megaplasmid more than 200 kilobases from the *nod-nif* region. *Mol. Gen. Genet. MGG* **1985**, *199*, 232–239. [[CrossRef](#)]
76. Romaniuk, K.; Dziewit, L.; Decewicz, P.; Mielnicki, S.; Radlinska, M.; Drewniak, L. Molecular characterization of the pSinB plasmid of the arsenite oxidizing, metallotolerant *Sinorhizobium* sp. M14—Insight into the heavy metal resistome of sinorhizobial extrachromosomal replicons. *FEMS Microbiol. Ecol.* **2017**, *93*. [[CrossRef](#)] [[PubMed](#)]
77. Brokx, S.J.; Rothery, R.A.; Zhang, G.; Ng, D.P.; Weiner, J.H. Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the role of a conserved active site cysteine in assembly and function. *Biochemistry* **2005**, *44*, 10339–10348. [[CrossRef](#)] [[PubMed](#)]
78. Wilson, J.J.; Kappler, U. Sulfite oxidation in *Sinorhizobium meliloti*. *Biochim. Biophys. Acta (BBA) Bioenerg.* **2009**, *1787*, 1516–1525. [[CrossRef](#)] [[PubMed](#)]
79. Porter, S.S.; Chang, P.L.; Conow, C.A.; Dunham, J.P.; Friesen, M.L. Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J.* **2017**, *11*, 248–262. [[CrossRef](#)] [[PubMed](#)]
80. Li, Z.; Ma, Z.; Hao, X.; Rensing, C.; Wei, G. Genes conferring copper resistance in *Sinorhizobium meliloti* CCNWSX0020 also promote the growth of *Medicago lupulina* in copper-contaminated soil. *Appl. Environ. Microbiol.* **2014**, *80*, 1961–1971. [[CrossRef](#)] [[PubMed](#)]
81. Hao, X.; Xie, P.; Zhu, Y.-G.; Taghavi, S.; Wei, G.; Rensing, C. Copper tolerance mechanisms of *Mesorhizobium amorphae* and its role in aiding phytostabilization by *Robinia pseudoacacia* in copper contaminated soil. *Environ. Sci. Technol.* **2015**, *49*, 2328–2340. [[CrossRef](#)] [[PubMed](#)]
82. Maynaud, G.; Brunel, B.; Yashiro, E.; Mergeay, M.; Cleyet-Marel, J.C.; Le Quéré, A. CadA of *Mesorhizobium metallidurans* isolated from a zinc-rich mining soil is a PIB-2-type ATPase involved in cadmium and zinc resistance. *Res. Microbiol.* **2014**, *165*, 175–189. [[CrossRef](#)] [[PubMed](#)]
83. Rubio-Sanz, L.; Brito, B.; Palacios, J. Analysis of metal tolerance in *Rhizobium leguminosarum* strains isolated from an ultramafic soil. *FEMS Microbiol. Lett.* **2018**, *365*, fny010. [[CrossRef](#)] [[PubMed](#)]
84. Pérez-Palacios, P.; Romero-Aguilar, A.; Delgadillo, J.; Doukkali, B.; Caviedes, M.A.; Rodríguez-Llorente, I.D.; Pajuelo, E. Double genetically modified symbiotic system for improved Cu phytostabilization in legume roots. *Environ. Sci. Pollut. Res.* **2017**, *24*, 14910–14923. [[CrossRef](#)] [[PubMed](#)]
85. Zhang, J.; Xu, Y.; Cao, T.; Chen, J.; Rosen, B.P.; Zhao, F.-J. Arsenic methylation by a genetically engineered *Rhizobium*-legume symbiont. *Plant Soil* **2017**, *416*, 259–269. [[CrossRef](#)] [[PubMed](#)]
86. Kong, Z.; Mohamad, O.A.; Deng, Z.; Liu, X.; Glick, B.R.; Wei, G. Rhizobial symbiosis effect on the growth, metal uptake, and antioxidant responses of *Medicago lupulina* under copper stress. *Environ. Sci. Pollut. Res.* **2015**, *22*, 12479–12489. [[CrossRef](#)] [[PubMed](#)]
87. Rajkumar, M.; Ae, N.; Prasad, M.N.V.; Freitas, H. Potential of siderophore-producing bacteria for improving heavy metal phytoextraction. *Trends Biotechnol.* **2010**, *28*, 142–149. [[CrossRef](#)] [[PubMed](#)]
88. Valls, M.; De Lorenzo, V. Exploiting the genetic and biochemical capacities of bacteria for the remediation of heavy metal pollution. *FEMS Microbiol. Rev.* **2002**, *26*, 327–338. [[CrossRef](#)] [[PubMed](#)]
89. Nies, D.H. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol. Rev.* **2003**, *27*, 313–339. [[CrossRef](#)]
90. Drewniak, L.; Dziewit, L.; Cieczkowska, M.; Gawor, J.; Gromadka, R.; Sklodowska, A. Structural and functional genomics of plasmid pSinA of *Sinorhizobium* sp. M14 encoding genes for the arsenite oxidation and arsenic resistance. *J. Biotechnol.* **2013**, *164*, 479–488. [[CrossRef](#)] [[PubMed](#)]
91. Drewniak, L.; Matlakowska, R.; Sklodowska, A. Arsenite and arsenate metabolism of *Sinorhizobium* sp. M14 living in the extreme environment of the Zloty Stok gold mine. *Geomicrobiol. J.* **2008**, *25*, 363–370. [[CrossRef](#)]
92. DiCenzo, G.C.; Finan, T.M. The divided bacterial genome: Structure, function, and evolution. *Microbiol. Mol. Biol. Rev.* **2017**, *81*, e00019-17. [[CrossRef](#)] [[PubMed](#)]

93. DiCenzo, G.C.; Benedict, A.B.; Fondi, M.; Walker, G.C.; Finan, T.M.; Mengoni, A.; Griffiths, J.S. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genet.* **2018**, *14*, e1007357. [[CrossRef](#)] [[PubMed](#)]
94. DiCenzo, G.C.; Wellappili, D.; Golding, G.B.; Finan, T.M. Inter-replicon gene flow contributes to transcriptional integration in the *Sinorhizobium meliloti* multipartite genome. *G3 Genes Genomes Genet.* **2018**, *8*, 1711–1720. [[CrossRef](#)] [[PubMed](#)]
95. Landeta, C.; Dávalos, A.; Cevallos, M.Á.; Geiger, O.; Brom, S.; Romero, D. Plasmids with a chromosome-like role in Rhizobia. *J. Bacteriol.* **2011**, *193*, 1317–1326. [[CrossRef](#)] [[PubMed](#)]
96. DiCenzo, G.C.; Checcucci, A.; Bazzicalupo, M.; Mengoni, A.; Viti, C.; Dziewit, L.; Finan, T.M.; Galardini, M.; Fondi, M. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **2016**, *7*, 12219. [[CrossRef](#)] [[PubMed](#)]
97. Galardini, M.; Brilli, M.; Spini, G.; Rossi, M.; Roncaglia, B.; Bani, A.; Chiancianesi, M.; Moretto, M.; Engelen, K.; Bacci, G.; et al. Evolution of intra-specific regulatory networks in a multipartite bacterial genome. *PLoS Comput. Biol.* **2015**, *11*, e1004478. [[CrossRef](#)] [[PubMed](#)]
98. Galardini, M.; Pini, F.; Bazzicalupo, M.; Biondi, E.G.; Mengoni, A. Replicon-dependent bacterial genome evolution: The case of *Sinorhizobium meliloti*. *Mol. Biol.* **2013**, *5*, 542–558. [[CrossRef](#)] [[PubMed](#)]
99. Checcucci, A.; diCenzo, G.C.; Ghini, V.; Bazzicalupo, M.; Beker, A.; Decorosi, F.; Dohlemann, J.; Fagorzi, C.; Finan, T.M.; Fondi, M.; et al. Creation and multi-omics characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *bioRxiv* **2018**. [[CrossRef](#)]
100. Checcucci, A.; DiCenzo, G.C.; Bazzicalupo, M.; Mengoni, A. Trade, diplomacy, and warfare: The quest for elite rhizobia inoculant strains. *Front. Microbiol.* **2017**, *8*, 2207. [[CrossRef](#)] [[PubMed](#)]
101. Van Opijnen, T.; Camilli, A. Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* **2013**, *11*, 435–442. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations

Lisa Cangioli¹, Alice Checcucci², Alessio Mengoni¹, and Camilla Fagorzi¹

¹Department of Biology, University of Florence, Via Madonna del Piano, 6, Sesto Fiorentino, 50019, Italy

²Department of Agricultural and Food Science, University of Bologna, Viale Giuseppe Fanin, 40-50, Bologna, 40127, Italy

Address correspondence and requests for materials to Camilla Fagorzi, camilla.fagorzi@unifi.it

Abstract

Mutualistic interactions have great importance in ecology, with genetic information that takes shape through interactions within the symbiotic partners and between the partners and the environment. It is known that variation of the host-associated microbiome contributes to buffer adaptation challenges of the host's physiology when facing varying environmental conditions. In agriculture, pivotal examples are symbiotic nitrogen-fixing rhizobia, known to contribute greatly to host (legume plants) adaptation and host productivity. A holistic view of increasing crop yield and resistance to biotic and abiotic stresses is that of microbiome engineering, the exploitation of a host-associated microbiome through its rationally designed manipulation with synthetic microbial communities. However, several studies highlighted that the expression of the desired phenotype in the host resides in species-specific, even genotype-specific interactions between the symbiotic partners. Consequently, there is a need to dissect such an intimate level of interaction, aiming to identify the main genetic components in both partners playing a role in symbiotic differences/host preferences. In the present paper, while briefly reviewing the knowledge and the challenges in plant-microbe interaction and rhizobial studies, we aim to promote research on genotype x genotype interaction between rhizobia and host plants for a rational design of synthetic symbiotic nitrogen-fixing microbial communities to be used for sustainably improving leguminous plants yield.

Keyword: microbiome engineering, genome x genome, legume-rhizobia mutualism, symbiosis, precision agriculture

Nobody is an island: the relevance of mutualistic interactions with eukaryotic hosts

Mutualistic interactions occur everywhere in the biosphere and are pivotal in evolution, as well as having great importance in ecology (Bronstein, 2015). Mutualism involves co-evolution, innovation, change and the involvement of partners with complementary skills (Lanier et al., 2017). In 2008 the hologenome theory of evolution defined the holobiont as a unit of selection in evolution (Zilber-Rosenberg and Rosenberg, 2008). The eukaryotic host and its associated microbes are considered a superorganism, in which genetic information takes shape through interactions between the partners and between the partners and the environment (Wilson and Sober, 1989; Zilber-Rosenberg and Rosenberg, 2008). For example, plants and animals have extended their metabolic repertoire through the establishment of specific microbial communities associated to roots and gut, respectively. These microbial communities are characterized by the presence of bacteria, archaea, fungi, oomycetes, as well as viruses, and they can be partly considered as the host's extended genome, improving the nutrient acquisition process in the

Citation: Cangioli, L., Checcucci, A., Mengoni, A., and Fagorzi, C. 2021. Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations. *Bio. Comm.* 66(1): 00-00. <https://doi.org/10.21638/spbu03.2021.106>

Authors' information: Lisa Cangioli, PhD Student, orcid.org/0000-0002-7714-1704; Alice Checcucci, PhD, Researcher, orcid.org/0000-0002-0019-0997; Alessio Mengoni, PhD, Associate Professor, Head of Laboratory, orcid.org/0000-0002-1265-8251; Camilla Fagorzi, PhD Student, orcid.org/0000-0002-9996-8868

Manuscript Editor: Anton Nizhnikov, Department of Genetics and Biotechnology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia

Received: October 19, 2020;

Revised: December 21, 2020;

Accepted: January 3, 2021.

Copyright: © 2021 Cangioli et al. This is an open-access article distributed under the terms of the License Agreement with Saint Petersburg State University, which permits to the authors unrestricted distribution, and self-archiving free of charge.

Funding: This work was supported by the grant MICRO4Legumes ("Il microbioma vegetale simbiotico come strumento per il miglioramento delle leguminose foraggere. Acronimo"), D.M.n.89267 (Italian Ministry of Agriculture).

Competing interests: The authors have declared that no competing interests exist.

animal gut and in plant roots (Turner et al., 2013; Hacquard et al., 2015). However, how this can imply coevolution and consideration of the holobiont as a unit of selection is still hotly debated (Moran and Sloan, 2015; Koskella and Bergelson, 2020). Indeed, the host-associated microbiota is not fully stable, and rapid variation in its composition and functioning can take place. Such variations may contribute to buffer adaptation challenges of the host's physiology when facing varying environmental conditions. In other words, the holobiont can allow more time-effective adaptation in rapidly changing environments, reducing the actual selective pressure on the host genome (Zilber-Rosenberg and Rosenberg, 2008; Rosenberg and Zilber-Rosenberg, 2016).

Interactions between the host and microbes are shaped by multiple factors associated with host immunity, phylogeny, environmental features (e.g., pH, presence of biotic or abiotic compounds, temperature), diet and nutrient availability (Brinker et al., 2019).

Classically, plant domestication has involved selection of varieties on the basis of phenotypic features relevant to the farmer and adapted to the agricultural practices, without considering the associated microbiota. Since crops mostly rely on external inputs (e.g., fertilizers) and not on associated microbes, a decrease of microbiota diversity has been observed (Escudero-Martinez and Bulgarelli, 2019; Martínez-Romero et al., 2020), reducing the buffering effect offered by a diverse holobiont. For instance, in legumes it has been demonstrated that in the wild-growing varieties the symbiotic potential is usually higher than in commercial ones (Provorov and Tikhonovich, 2003).

Good diversity at the holobiont level may then be crucial in the search for environmentally sustainable crop production. Here, the authors explored the knowledge and challenges in the promotion of a rational use of symbiotic nitrogen-fixing bacteria for leguminous plant growth.

Biotechnology on the holobiont: cracking the microbiome

The increase of the human world population in the last years (currently 7 billion people, with 9 billion people expected by 2050) has led to the need to increase agricultural production (FAO; 2018). Inorganic nitrogen fertilizers are required to increase field production, but the excessive use of these compounds has deleterious effects on the environment (Yang and Fang, 2015; Zheng et al., 2019).

One of the possible alternative and sustainable approaches to enrich fields with nitrogen is the inoculation of crops with microorganisms able to promote the plant growth and health, the so-called *Plant Growth Promoting Microorganisms* (PGPM) (Lucy et al., 2004;

Schlaeppli and Bulgarelli, 2015). In particular, *Plant Growth Promoting Rhizobacteria* (PGPR) can be applied as biocontrol agents, bio-inoculants and bio-fertilizers (Bloemberg and Lugtenberg, 2001). PGPM and PGPR constitute a relevant fraction of the plant mutualistic microbiome. They thrive on and within plants, using plant-produced organic molecules for their growth and “rewarding” the host with increased nutrient availability for the root apparatus as well as additional phytohormones (Werner et al., 2014).

Modification of the plant-associated microbiome (mainly, but not only, involving PGPR) has an effective potential to improve plant yield and resistance to biotic and abiotic stresses; this potential is stirring the attention of many investigators (Turner et al., 2013). This approach of “*microbiome engineering*” is based on reconstructing synthetic microbial communities after laboratory selection of microbes with the best ability to deliver PGP traits to the plants (Ke et al., 2020). The principal aims of microbiome manipulation are to: 1) reduce the incidence of plant diseases (Malfanova, 2013); 2) increase agricultural production (Bakker et al., 2012); 3) reduce chemical inputs (Adesemoye et al., 2009) and 4) reduce emissions of greenhouse gases (Singh et al., 2010). These goals meet the principles of agronomic sustainability and the world's increasing population (Turner et al., 2013). To date, successful attempts have been reviewed by de Vries et al., 2020; Dubey et al., 2020; Qiu et al., 2019; and Sudheer et al., 2020.

Legume–rhizobia mutualism as a model

Among the “rewards” from microbes, assimilable nitrogen compounds are one of the key components: the supply of nitrogen in the soil represents the classical limiting factors of plant productivity (Fageria and Baliagar, 2005). Bacteria called rhizobia are the typical example of PGPM which provides the host plant assimilable nitrogen. Through a process called biological nitrogen fixation (BNF), involving several genes and signaling molecules, rhizobia fix atmospheric nitrogen into compounds assimilable by plants. The BNF is a paradigmatic example of mutualistic association between plants and bacteria. The symbiotic BNF (SNF, *Symbiotic Nitrogen Fixation*) is a facultative symbiotic association involving rhizobia (bacteria from Alfa- and Betaproteobacteria classes) and some actinobacteria (*Frankia* spp.), which associate with plants from the Fabales (*Leguminosae*), Fagales, Cucurbitales and Rosales orders (Pawlowski and Newton, 2008; Griesmann et al., 2018). During the symbiotic interaction, bacteria induce the formation of a specific structure at the root level, the nodule, then colonize nodule plant cells intracellularly and activate the key component of BNF (i.e., the nitrogenase enzyme), producing ammonia, which is then exchanged for the

photosynthetic products with the host plant (Kereszt et al., 2011).

The symbiosis between legumes and rhizobia is the most diffused and relevant in providing fixed nitrogen to agroecosystems. This interaction is mediated by a plethora of molecular signals that direct bacterial invasion, modulate the host defence response, permit the intracellular colonization, guide the regulation of the cell cycle and regulate the nutrient exchange (Jones et al., 2007; Gibson et al., 2008). One of the most essential signalling compounds produced by rhizobia is the Nod Factor (NF). Rhizobia are able to produce Nod factors characterized by different structures which, consequently, are recognized by specific host plant receptors (Geurts and Bisseling, 2002). Due to this process, evidence has emerged for coevolution between rhizobia and host plant at the population level (Igolkina et al., 2019). However, restricted host specificity is present, and phylogenetical studies confirmed that the limited host range evolved from an ancestral broad range mutualism (Pueppke and Broughton, 1999). Therefore, evolution has shaped the two partners toward species-specific and even strain-specific interactions allowing plant selection for the best beneficial symbionts (Checcucci et al., 2016; Remigi et al., 2016; Westhoek et al., 2017; Sachs et al., 2018).

The two-way exchange of benefits between plant roots and rhizobia offers us the basic knowledge to mold and manipulate this symbiosis interaction, sustaining the ecological and agronomic practices in agricultural systems (Bakker et al., 2012; Checcucci et al., 2017; Soumare et al., 2020).

Forming the symbiotic structure: recognition of the good partners

The multi-step process which allows the symbiotic interaction between nitrogen-fixing rhizobia and leguminous plants is highly regulated, to allow the selection by the plant of the most effective (highly rewarding) partners (Sachs et al., 2018). Many genes and mechanisms involved in the establishment of the symbiosis between rhizobia and host legumes have been identified (Roy et al., 2020).

The early events start with the exchange of signals between the nitrogen-fixing rhizobia and the plant in the rhizosphere. Here, the first specific signals are found in compounds present in root exudates (flavonoids), which bind receptor proteins in the rhizobial cell (NodD). Activated NodD trigger bacterial *nod* genes expression which lead to synthesis of the Nod factor (a lipochitooligosaccharide molecule), which is then secreted by the bacterial cell. The perception of the Nod factor by the plant roots induces developmental changes such as root hair curling, membrane depolarization, intracellular calcium oscillations, and the initiation of cell division in the root cortex, which establishes a meristem and nodule primor-

dium (Liu et al., 2020). Species-specific early recognition is based on the fact that different rhizobial species and strains can have differential activation by flavonoids on NodD proteins and may produce different Nod factor molecules. Different Nod factors are then specifically recognized by plant receptors (LysM), ensuring species-specific recognition on the plant side (Bozsoki et al., 2020).

When the dialogue between plant roots and rhizobia is successful, a small number of bacteria are trapped by the root hair, which begins an inverse tip growth, forming a long and narrow passage called the “infection thread”. The bacteria reach the root cortex as the final destination and reacquire properties of stem cells, which form the lateral organ nodule. The bacteria infect the nodule, enter the cytoplasm of plant cells and differentiate into a distinct cell type called bacteroid, which can fix atmospheric nitrogen into ammonia, establishing an intricate metabolic interchange with the host plant (diCenzo et al., 2020; Roy et al., 2020). Moreover here, nodules where rhizobia fail to produce fixed nitrogen are sanctioned, ensuring that only good mutualist rhizobia are allowed to thrive in root nodules (Kiers and Denison, 2008), though cheating may occur (Checcucci et al., 2016). Though molecular details of sanctions are still elusive, molecular determinants of species-specificity have been discovered. For instance, in the model legume *Medicago truncatula*, the plant response after bacteria invasion of the host cell is modulated by two nodule-specific cysteine-rich (NCR) peptides encoded by the genes *NFS1* and *NFS2* (Wang et al., 2018; Tirichine et al., 2000). More recently, again in *M. truncatula*, MtNPD1, a nodule-specific polycystin lipogenase, has been shown to represent an important determinant of host-strain specificity, with particular regard to plant invasion and nitrogen fixation (Pislariu et al., 2019). Functional symbiotic compatibility between *Mesorhizobium loti* and *Lotus japonicum* is determined by the presence of the nepenthesin-type aspartic peptidase nodule-induced LjAPN1 (Yamaya-Ito et al., 2018).

In *Pisum sativum*, Nod factors perception is mediated by the gene PsSym2, which probably encodes a LysM-RLK. PsSym2 was the first discovered symbiosis-related pea gene and it determines increased selectivity toward rhizobia in the pea cv. Afghanistan (Sulima et al., 2017).

The role of genotype x genotype interactions in shaping the symbiotic partnership

Interactions between different species with different genotypes implies the expression of a phenotype that depends on the interacting organisms (Bose and Schulte, 2014). The effects of (microbial) genotype (G) x (host) genotype (G) interactions can be relevant for both rate and direction of evolutionary selection (Wade, 2007) and have clear importance for rational improvement of plant–rhizobium

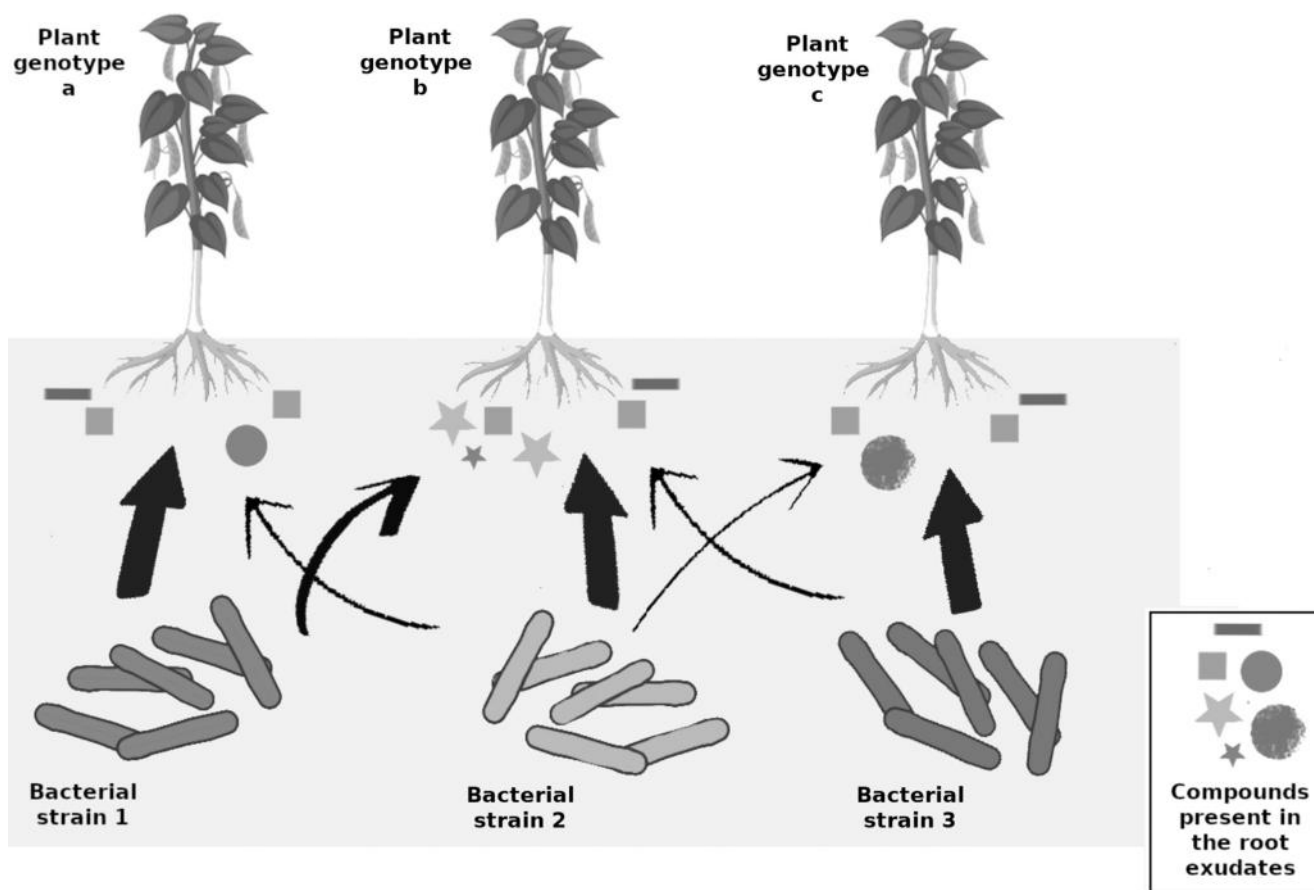


Fig. 1. Simplified model of genotype x genotype rhizobia–legume interaction. Different genotypes of the same species of legume interact with different genotypes of the same rhizobial species. Interactions can be trained by the expression of an intricate network of genes of both symbiotic partners, driving host preference. The rhizobial strain can initiate a mutualistic relationship with multiple legume genotypes, but specific genotype x genotype interaction can result in a more effective symbiotic ability. In the figure, the effectiveness of symbiotic ability is represented by arrows of different thickness (arrows are thicker when symbiosis is more effective) and compounds produced by the different plant genotypes are represented by the symbols near the root system (symbols of the same colour of the bacterial strain are responsible for a more effective symbiotic relationship).

symbiosis. As reviewed above, the development and application of genetic and genomic tools has made it possible to dissect in great detail the molecular determinants involved in plant–microbe interaction (Levy et al., 2018) and rhizobial symbioses (diCenzo et al., 2019).

However, most of the studies involved model strains and plants and their respective mutants, with only a limited evaluation of the genetic bases of the somewhat large phenotypic variation in the symbiotic phenotype due to natural genotype variation. Early studies in the alfalfa rhizobial symbiont *Sinorhizobium meliloti* showed that plant variety exerts a detectable effect on the recruitment of native soil rhizobial population (Carelli et al., 2000; Paffetti et al., s.d.). Clearly, the fitness advantage of the symbiosis (Burghardt, 2020) is influenced by GxG interactions between the two (or more) partners (Heath, 2008). Recently, the use a “select and resequence” approach has disclosed some of the determinants of GxG interaction in the model symbiosis between strains of *Sinorhizobium meliloti* and genotypes of the legume the *Medicago truncatula* (Burghardt et

al., 2018, 2020). With a similar aim, a recent study has been conducted to analyse at the gene expression level the genotype-based variation of the interaction, using *S. meliloti* and the forage legume *M. sativa* (Kang et al., 2020). In this last work, the transcriptional profile of alfalfa inoculated with different *S. meliloti* strains was investigated, finding that the plant differentially expressed genes depending on the *S. meliloti* strain tested. Among the candidate genes influenced by the rhizobial strain genotype, genes encoding nodulins, NCR peptides and proteins belonging to the NBS-LRR family were identified (Kang et al., 2020). On the rhizobial side we recently sequenced the transcriptome of several *S. meliloti* strains incubated in the presence of root exudates of different alfalfa varieties. Results showed an intricate network of genes whose expression is influenced by the genotype of both symbiotic partners, suggesting the presence of a large number of determinants involved in GxG interactions which can have an impact on the outcome of the rhizobium–legume symbiosis (Fig. 1) (Fagorzi et al., 2021). Dissection of such a complex network, iden-

tifying the main components playing a role in symbiotic differences/host preferences, will make it possible to rationally screen and genetically improve the most promising plant and symbiotic bacteria partnerships, as well as to better understand the intricate routes of the evolution of symbioses.

On the plant side, the GxG interactions between selected nitrogen-fixing rhizobia and selected symbiotically active plant genotypes also lend advantages in terms of plant productivity (Provorov & Tikhonovich, 2003).

Holobiont improvement programs toward precision agriculture

Classical studies, such as the ones performed by Bliss (1990) and Barnes et al. (1984), showed the importance of a multidisciplinary approach to improve the symbiotic activity in *Phaseolus vulgaris* and *Medicago sativa*.

The study of plant genotype and breeding programs have been fundamental in understanding plant physiological traits and find directions for productivity improvement. The use of fertilizers can definitely be reduced by exploiting the knowledge of a plant's genome, soil conditions and plant-associated microorganisms. Indeed, genome editing methods allow targeted modification in the plant genotype, making it possible to improve and adapt a crop's genetic traits with the current soil characteristics (Joseph et al., 2020), including rhizosphere resident bacteria and fungi. The advantages of this technological and biological knowledge in the agricultural field were recently discussed in two workshops organized by the USDA National Institute of Food and Agriculture (NIFA)-funded Big Data Driven Agriculture, where the importance of their applicability was assessed on the basis of standardized protocol applications, funding opportunities and involvement of a broader scientific community and farmers (Shakoor et al., 2019). Currently, high-throughput techniques for plants and soil phenotyping, tools for the measurement of crop productivity, genome editing methods, breeding and microbial inoculation technologies form the basis of precision agriculture, which involves both soil/plant science experts and farm management operators.

For SNF, besides the biotic factors shaping legume–rhizobia nitrogen fixation (e.g., selection and interaction with the partner, competition with indigenous microorganisms), abiotic variables can strongly influence the success of the symbiosis. Indeed, nutrient availability, extreme temperature, pH, salinity and drought (Bellabarba et al., 2019) challenge legume crops, stimulating the need for an adjustable agriculture. Precision agriculture can ensure the health and productivity of soil and crops through knowledge and technology advancements, guaranteeing protection and well-being of the environment.

Presently, several diagnostic methods are based on the quantification of nitrogen in legumes, evaluating the effectiveness of rhizobial symbiosis (e.g., measures of nitrogen at the leaf level and measurement of nitrogen abundance in plant tissues through isotope-based methodologies (Thilakarathna and Raizada, 2018). Furthermore, several technologies aimed to optimize biological nitrogen fixation in the field are based on the diagnosis of soil traits (composition, structure, organic matter, landscape position and micronutrients concentration).

The introduction of beneficial microorganisms through crop inoculation is now becoming popular among farmers, both in organic farming and unconventional agricultural practices. Nevertheless, the tuning of inoculants consistent with indigenous PGPM residents in the soil still faces a challenge in precision agriculture (Thilakarathna and Raizada, 2018).

Actually, the success of microbial inoculants' performance can be increased by precision farming methods, such as fertigation, which delivers the bioinoculants in the root area, thereby minimizing loss and interference with other liquid chemical inputs such as fertilizers, herbicides, pesticides and growth hormones, still used periodically (Bharathi et al., 2017). Furthermore, seed coating is widely used in agricultural practice, thanks to its beneficial impact on seed performance and plant productivity. Among coating agents, chemical pesticides, micronutrients, bio-stimulants and marker substances (tracing seeds within the crop supply chain) can be included. However, microbial (including bacteria and fungi) seed coating has been considered a cheap and efficient method for the delivery of inoculants. Recent research in precision agriculture has mainly focused on the optimization of such coating, setting up formulation of multi-effect microbial consortia designed for the already targeted annual crops (such as cereals, legumes, and some vegetables) and/or developed for other agricultural products.

However, it is becoming clear that, apart from soil and climatic features, which determine the choice of a specific plant variety by the farmer, the symbiotic rhizobia (as the other inoculants) genotypes matter. As reported above, GxG effects are present, determining the competitive abilities of the inoculant against the indigenous rhizobial populations. As for genetic screening and breeding programs for crops, there is a need for similar strategies for the symbiotic rhizobia. Several collections of rhizobia are present in the world, such as those at the USDA in the USA, the Centre for Rhizobium Studies (CRS) in Australia or at ARRIAM in the Russian Federation. Genome sequencing programs (e.g., GEBA at DOE-JGI, USA) and analysis of strain genotypes have been conducted. For instance, in *S. meliloti*, more than 280 high quality genome sequences are available on the GenBank database (October 2020). However, informa-

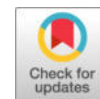
tion on the set of genes linked to the variation in symbiotic partnership performances (such as GxG) are still in their infancy. We need more comprehensive, holobiont-centered and systems biology-oriented studies to provide the basis for a future integration into plant breeding programs, concurrent symbiotic “breeding” programs, resulting in outperformance of the symbiotic partnership.

References


- Adesemoye, A. O., Torbert, H. A., and Kloepper, J. W. 2009. Plant growth-promoting rhizobacteria allow reduced application rates of chemical fertilizers. *Microbial Ecology* 58(4):921–929. <https://doi.org/10.1007/s00248-009-9531-y>
- Bakker, M. G., Manter, D. K., Sheflin, A. M., Weir, T. L., and Vivanco, J. M. 2012. Harnessing the rhizosphere microbiome through plant breeding and agricultural management. *Plant and Soil* 360(1–2):1–13. <https://doi.org/10.1007/s11104-012-1361-x>
- Barnes, D., Heichel, G., Vance, C., and Ellis, W. 1984. A multiple-trait breeding program for improving the symbiosis for N₂ fixation between *Medicago sativa* L. and *Rhizobium meliloti*. *Plant and Soil* 82(3):303–314. <https://doi.org/10.1007/BF02184269>
- Bellarbarba, A., Fagorzi, C., diCenzo, G. C., Pini, F., Viti, C., and Checcucci, A. 2019. Deciphering the symbiotic plant microbiome: translating the most recent discoveries on rhizobia for the improvement of agricultural practices in metal-contaminated and high saline lands. *Agronomy* 9(9):529. <https://doi.org/10.3390/agronomy9090529>
- Bharathi, J., Balachander, D., Kumar, K., and Narayanan, R. 2017. Evaluation of new microbial consortium through biofertilization for precision farming of bhendi (COBH 1). *International Journal of Medical Sciences and Pharmaceutical Research* 1(1):15–24.
- Bliss, F. 1990. Utilization of genetic resources for crop improvement: The common bean. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L., and Weir, B. S. (Eds.), *Plant population genetics, breeding, and genetic resources*, pp. 317–333, Sinauer, Sunderland, MA.
- Bloembergen, G. V. and Lugtenberg, B. J. J. 2001. Molecular basis of plant growth promotion and biocontrol by rhizobacteria. *Current Opinion in Plant Biology* 4(4):343–350. [https://doi.org/10.1016/S1369-5266\(00\)00183-7](https://doi.org/10.1016/S1369-5266(00)00183-7)
- Bose, J. and Schulte, R. D. 2014. Testing GxG interactions between coinfecting microbial parasite genotypes within hosts. *Frontiers in Genetics* 5:124. <https://doi.org/10.3389/fgene.2014.00124>
- Bozsoki, Z., Gysel, K., Hansen, S. B., Lironi, D., Krönauer, C., Feng, F., de Jong, N., Vinther, M., Kamble, M., Thygesen, M. B., Engholm, E., Kofoed, C., Fort, S., Sullivan, J. T., Ronson, C. W., Jensen, K. J., Blaise, M., Oldroyd, G., Stougaard, J., ... Radutoiu, S. 2020. Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity. *Science* 369(6504):663–670. <https://doi.org/10.1126/science.abb3377>
- Brinker, P., Fontaine, M. C., Beukeboom, L. W., and Falcao Salles, J. 2019. Host, symbionts, and the microbiome: The missing tripartite interaction. *Trends in Microbiology* 27(6):480–488. <https://doi.org/10.1016/j.tim.2019.02.002>
- Bronstein, J. L. 2015. *Mutualism*. Oxford University Press, USA. <https://doi.org/10.1093/acprof:oso/9780199675654.001.0001>
- Burghardt, L. T. 2020. Evolving together, evolving apart: Measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. *New Phytologist* 228(1):28–34. <https://doi.org/10.1111/nph.16045>
- Burghardt, L. T., Epstein, B., Guhlin, J., Nelson, M. S., Taylor, M. R., Young, N. D., Sadowsky, M. J., and Tiffin, P. 2018. Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proceedings of the National Academy of Sciences USA* 115(10):2425–2430. <https://doi.org/10.1073/pnas.1714246115>
- Burghardt, L. T., Trujillo, D. I., Epstein, B., Tiffin, P., and Young, N. D. 2020. A select and resequence approach reveals strain-specific effects of *Medicago* nodule-specific PLAT-domain genes. *Plant Physiology* 182(1):463–471. <https://doi.org/10.1104/pp.19.00831>
- Carelli, M., Gnocchi, S., Fancelli, S., Mengoni, A., Paffetti, D., Scotti, C., and Bazzicalupo, M. 2000. Genetic diversity and dynamics of *Sinorhizobium meliloti* populations nodulating different alfalfa cultivars in Italian soils. *Applied and Environmental Microbiology* 66(11):4785–4789. <https://doi.org/10.1128/AEM.66.11.4785-4789.2000>
- Checcucci, A., Azzarello, E., Bazzicalupo, M., Galardini, M., Lagomarsino, A., Mancuso, S., Marti, L., Marzano, M. C., Mocali, S., Squartini, A., Zanardo, M., and Mengoni, A. 2016. Mixed nodule infection in *Sinorhizobium meliloti*-*Medicago sativa* symbiosis suggest the presence of cheating behavior. *Frontiers in Plant Science* 7:835. <https://doi.org/10.3389/fpls.2016.00835>
- Checcucci, A., DiCenzo, G. C., Bazzicalupo, M., and Mengoni, A. 2017. Trade, diplomacy, and warfare: The quest for elite rhizobia inoculant strains. *Frontiers in Microbiology* 8:2207. <https://doi.org/10.3389/fmicb.2017.02207>
- de Vries, F. T., Griffiths, R. I., Knight, C. G., Nicolitch, O., and Williams, A. 2020. Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* 368(6488):270–274. <https://doi.org/10.1126/science.aaz5192>
- diCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A., and Fondi, M. 2020. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nature Communications* 11(1):2574. <https://doi.org/10.1038/s41467-020-16484-2>
- diCenzo, G. C., Zamani, M., Checcucci, A., Fondi, M., Griffiths, J. S., Finan, T. M., and Mengoni, A. 2019. Multidisciplinary approaches for studying rhizobium-legume symbioses. *Canadian Journal of Microbiology* 65(1):1–33. <https://doi.org/10.1139/cjm-2018-0377>
- Dubey, R. K., Tripathi, V., Prabha, R., Chaurasia, R., Singh, D. P., Rao, C. S., El-Keblawy, A., and Abhilash, P. C. 2020. Unravelling the soil microbiome: Perspectives for environmental sustainability. Springer. <https://doi.org/10.1007/978-3-030-15516-2>
- Escudero-Martinez, C. and Bulgarelli, D. 2019. Tracing the evolutionary routes of plant-microbiota interactions. *Current Opinion in Microbiology* 49:34–40. <https://doi.org/10.1016/j.mib.2019.09.013>
- Fageria, N. and Baligar, V. 2005. Enhancing nitrogen use efficiency in crop plants. *Advances in Agronomy* 88:97–185. [https://doi.org/10.1016/S0065-2113\(05\)88004-6](https://doi.org/10.1016/S0065-2113(05)88004-6)
- Fagorzi, C., Bacci, G., Huang, R., Cangioli, L., Checcucci, A., Fini, M., Perrin, E., Natali, C., diCenzo, G. C., and Mengoni, A. 2021. Non-additive transcriptomic signatures of genotype x genotype interactions during the initiation of plant-rhizobium symbiosis. *mSystems* 6(1):e00974-20. <https://doi.org/10.1128/mSystems.00974-20>
- Geurts, R. and Bisseling, T. 2002. *Rhizobium* nod factor perception and signalling. *The Plant Cell* 14(suppl 1):S239. <https://doi.org/10.1105/tpc.002451>
- Gibson, K. E., Kobayashi, H., and Walker, G. C. 2008. Molecular determinants of a symbiotic chronic infection. *Annual Re-*

- view of *Genetics* 42(1):413–441. <https://doi.org/10.1146/annurev.genet.42.110807.091427>
- Griesmann, M., Chang, Y., Liu, X., Song, Y., Haberer, G., Crook, M. B., Billault-Penneteau, B., Lauressergues, D., Keller, J., Imanishi, L., Roswanjaya, Y. P., Kohlen, W., Pujic, P., Battenberg, K., Alloisio, N., Liang, Y., Hilhorst, H., Salgado, M. G., Hocher, V., ... Cheng, S. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361(6398):eaat1743. <https://doi.org/10.1126/science.aat1743>
- Hacquard, S., Garrido-Oter, S., González, A., Spaepen, S., Ackermann, G., Lebeis, S., McHardy, A. C., Dangl, J. L., Knight, R., Ley, R., and Schulze-Lefert, P. 2015. Microbiota and host nutrition across Plant and Animal Kingdoms. *Cell Host and Microbe* 17(5):603–616. <https://doi.org/10.1016/j.chom.2015.04.009>
- Heath, K. D. 2008. The coevolutionary genetics of plant-microbe interactions. *New Phytologist* 180(2):268–270. <https://doi.org/10.1111/j.1469-8137.2008.02633.x>
- Igolkina, A. A., Bazykin, G. A., Chizhevskaya, E. P., Provorov, N. A., and Andronov, E. E. 2019. Matching population diversity of rhizobial *nod* A and legume *NFR5* genes in plant-microbe symbiosis. *Ecology and Evolution* 9(18):10377–10386. <https://doi.org/10.1002/ece3.5556>
- Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., and Walker, G. C. 2007. How rhizobial symbionts invade plants: The *Sinorhizobium-Medicago* model. *Nature Reviews Microbiology* 5(8):619–633. <https://doi.org/10.1038/nrmicro1705>
- Joseph, A., Chandra, J., and Siddharthan, S. 2020. Genome analysis for precision agriculture using artificial intelligence: A survey. In *Data Science and Security*, pp. 221–226. Springer. https://doi.org/10.1007/978-981-15-5309-7_23
- Kang, W., Jiang, Z., Chen, Y., Wu, F., Liu, C., Wang, H., Shi, S., and Zhang, X.-X. 2020. Plant transcriptome analysis reveals specific molecular interactions between alfalfa and its rhizobial symbionts below the species level. *BMC Plant Biology* 20(1):293. <https://doi.org/10.1186/s12870-020-02503-3>
- Ke, J., Wang, B., and Yoshikuni, Y. 2020. Microbiome engineering: synthetic biology of plant-associated microbiomes in sustainable agriculture. *Trends in Biotechnology*. <https://doi.org/10.1016/j.tibtech.2020.07.008>
- Kereszt, A., Mergaert, P., and Kondorosi, E. 2011. Bacteroid development in legume nodules: Evolution of mutual benefit or of sacrificial victims? *Molecular Plant-Microbe Interactions* 24(11):1300–1309. <https://doi.org/10.1094/MPMI-06-11-0152>
- Kiers, E. T. and Denison, R. F. 2008. Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annual Review of Ecology and Systematics* 39(1):215–236. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173423>
- Koskella, B. and Bergelson, J. 2020. The study of host-microbiome (co)evolution across levels of selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1808):20190604. <https://doi.org/10.1098/rstb.2019.0604>
- Lanier, K. A., Petrov, A. S., and Williams, L. D. 2017. The central symbiosis of molecular biology: Molecules in mutualism. *Journal of Molecular Evolution* 85(1–2):8–13. <https://doi.org/10.1007/s00239-017-9804-x>
- Levy, A., Conway, J. M., Dangl, J. L., and Woyke, T. 2018. Elucidating bacterial gene functions in the plant microbiome. *Cell Host and Microbe* 24(4):475–485. <https://doi.org/10.1016/j.chom.2018.09.005>
- Liu, S., Ratet, P., and Magne, K. 2020. Nodule diversity, evolution, organogenesis and identity. *Advances in Botanical Research* 94:119–148. <https://doi.org/10.1016/bs.abr.2019.09.009>
- Lucy, M., Reed, E., and R. Glick, B. 2004. Applications of free living plant growth-promoting rhizobacteria. *Antonie van Leeuwenhoek* 86:1–25. <https://doi.org/10.1023/B:ANTO.0000024903.10757.6e>
- Malfanova, N. 2013. Endophytic bacteria with plant growth promoting and biocontrol abilities. Doctoral Thesis, Leiden University.
- Martínez-Romero, E., Aguirre-Noyola, J. L., Taco-Taype, N., Martínez-Romero, J., and Zuñiga-Dávila, D. 2020. Plant microbiota modified by plant domestication. *Systematic and Applied Microbiology* 43(5):126106. <https://doi.org/10.1016/j.syapm.2020.126106>
- Moran, N. A. and Sloan, D. B. 2015. The hologenome concept: helpful or hollow? *PLoS Biology* 13(12):e1002311. <https://doi.org/10.1371/journal.pbio.1002311>
- Paffetti, D., Daguin, F., Fancelli, S., Gnocchi, S., Lippi, F., Scotti, C., and Bazzicalupo, M. 1998. Influence of plant genotype on the selection of nodulating *Sinorhizobium meliloti* strains by *Medicago sativa*. *Antonie van Leeuwenhoek* 73:3–8. <https://doi.org/10.1023/A:1000591719287>
- Pawlowski, K. and Newton, W. E. 2008. Nitrogen-fixing actinorhizal symbioses. Springer. <https://doi.org/10.1007/978-1-4020-3547-0>
- Pislariu, C. I., Sinharoy, S., Torres-Jerez, I., Nakashima, J., Blancaflor, E. B., and Udvardi, M. K. 2019. The nodule-specific PLAT domain protein NPD1 is required for nitrogen-fixing symbiosis. *Plant Physiology* 180(3):1480–1497. <https://doi.org/10.1104/pp.18.01613>
- Provorov, N. A. and Tikhonovich, I. A. 2003. Genetic resources for improving nitrogen fixation in legume-rhizobia symbiosis. *Genetic Resources and Crop Evolution* 50(1):89–99. <https://doi.org/10.1023/A:1022957429160>
- Pueppke, S. G. and Broughton, W. J. 1999. *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular Plant-Microbe Interactions* 12(4):293–318. <https://doi.org/10.1094/MPMI.1999.12.4.293>
- Qiu, Z., Egidi, E., Liu, H., Kaur, S., and Singh, B. K. 2019. New frontiers in agriculture productivity: Optimised microbial inoculants and in situ microbiome engineering. *Biotechnology Advances* 37(6):107371. <https://doi.org/10.1016/j.biotechadv.2019.03.010>
- Remigi, P., Zhu, J., Young, J. P. W., and Masson-Boivin, C. 2016. Symbiosis within symbiosis: Evolving nitrogen-fixing legume symbionts. *Trends in Microbiology* 24(1):63–75. <https://doi.org/10.1016/j.tim.2015.10.007>
- Rosenberg, E. and Zilber-Rosenberg, I. 2016. Microbes drive evolution of animals and plants: The hologenome concept. *MBio* 7(2):e01395-15. <https://doi.org/10.1128/mBio.01395-15>
- Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., and Udvardi, M. K. 2020. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant Cell* 32(1):15–41. <https://doi.org/10.1105/tpc.19.00279>
- Sachs, J. L., Quides, K. W., and Wendlandt, C. E. 2018. Legumes versus rhizobia: A model for ongoing conflict in symbiosis. *New Phytologist* 219(4):1199–1206. <https://doi.org/10.1111/nph.15222>
- Schlaeppli, K. and Bulgarelli, D. 2015. The plant microbiome at work. *Molecular Plant-Microbe Interactions* 28(3):212–217. <https://doi.org/10.1094/MPMI-10-14-0334-FI>
- Shakoor, N., Northrup, D., Murray, S., and Mockler, T. C. 2019. Big data driven agriculture: Big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity. *The*

- Plant Phenome Journal* 2(1):1–8. <https://doi.org/10.2135/tppj2018.12.0009>
- Singh, B. K., Bardgett, R. D., Smith, P., and Reay, D. S. 2010. Microorganisms and climate change: Terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology* 8(11):779–790. <https://doi.org/10.1038/nrmicro2439>
- Soumare, A., Diedhiou, A. G., Thuita, M., Hafidi, M., Ouhdouch, Y., Gopalakrishnan, S., and Kouisni, L. 2020. Exploiting biological nitrogen fixation: A route towards a sustainable agriculture. *Plants* 9(8):1011. <https://doi.org/10.3390/plants9081011>
- Sudheer, S., Bai, R. G., Usmani, Z., and Sharma, M. 2020. Insights on engineered microbes in sustainable agriculture: Biotechnological developments and future prospects. *Current Genomics* 21(5):321–333. <https://doi.org/10.2174/1389202921999200603165934>
- Sulima, A. S., Zhukov, V. A., Afonin, A. A., Zhernakov, A. I., Tikhonovich, I. A., and Lutova, L. A. 2017. Selection signatures in the first exon of paralogous receptor kinase genes from the Sym2 region of the *Pisum sativum* L. genome. *Frontiers in Plant Science* 8:1957. <https://doi.org/10.3389/fpls.2017.01957>
- Thilakarathna, M. S. and Raizada, M. N. 2018. Challenges in using precision agriculture to optimize symbiotic nitrogen fixation in legumes: Progress, limitations, and future improvements needed in diagnostic testing. *Agronomy* 8(5):78. <https://doi.org/10.3390/agronomy8050078>
- Tirichine, L., de Billy, F., and Hugué, T. 2000. *Mtsym6*, a gene conditioning *Sinorhizobium* strain-specific nitrogen fixation in *Medicago truncatula*. *Plant Physiology* 123(3):845–852. <https://doi.org/10.1104/pp.123.3.845>
- Turner, T. R., James, E. K., and Poole, P. S. 2013. The plant microbiome. *Genome Biology* 14(6):209. <https://doi.org/10.1186/gb-2013-14-6-209>
- Wade, M. J. 2007. The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics* 8(3):185–195. <https://doi.org/10.1038/nrg2031>
- Wang, Q., Liu, J., Li, H., Yang, S., Körmöcz, P., Kereszt, A., and Zhu, H. 2018. Nodule-specific cysteine-rich peptides negatively regulate nitrogen-fixing symbiosis in a strain-specific manner in *Medicago truncatula*. *Molecular Plant-Microbe Interactions* 31(2):240–248. <https://doi.org/10.1094/MPMI-08-17-0207-R>
- Werner, G. D. A., Strassmann, J. E., Ivens, A. B. F., Engelmoer, D. J. P., Verbruggen, E., Queller, D. C., Noe, R., Johnson, N. C., Hammerstein, P., and Kiers, E. T. 2014. Evolution of microbial markets. *Proceedings of the National Academy of Sciences USA* 111(4):1237–1244. <https://doi.org/10.1073/pnas.1315980111>
- Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P. S., and Turnbull, L. A. 2017. Policing the legume-*Rhizobium* symbiosis: A critical test of partner choice. *Scientific Reports* 7(1):1419. <https://doi.org/10.1038/s41598-017-01634-2>
- Wilson, D. S. and Sober, E. 1989. Reviving the superorganism. *Journal of Theoretical Biology* 136(3):337–356. [https://doi.org/10.1016/S0022-5193\(89\)80169-9](https://doi.org/10.1016/S0022-5193(89)80169-9)
- Yamaya-Ito, H., Shimoda, Y., Hakoyama, T., Sato, S., Kaneko, T., Hossain, M. S., Shibata, S., Kawaguchi, M., Hayashi, M., Kouchi, H., and Umehara, Y. 2018. Loss-of-function of ASPARTIC PEPTIDASE NODULE — INDUCED 1 (APN 1) in *Lotus japonicus* restricts efficient nitrogen-fixing symbiosis with specific *Mesorhizobium loti* strains. *The Plant Journal* 93(1):5–16. <https://doi.org/10.1111/tpj.13759>
- Yang, X. and Fang, S. 2015. Practices, perceptions, and implications of fertilizer use in East-Central China. *Ambio* 44(7):647–652. <https://doi.org/10.1007/s13280-015-0639-7>
- Zheng, M., Zhou, Z., Luo, Y., Zhao, P., and Mo, J. 2019. Global pattern and controls of biological nitrogen fixation under nutrient enrichment: A meta-analysis. *Global Change Biology* 25(9):3018–3030. <https://doi.org/10.1111/gcb.14705>
- Zilber-Rosenberg, I. and Rosenberg, E. 2008. Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. *FEMS Microbiology Reviews* 32(5):723–735. <https://doi.org/10.1111/j.1574-6976.2008.00123.x>



Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis

Camilla Fagorzi,^a Giovanni Bacci,^a Rui Huang,^b Lisa Cangioli,^a Alice Checcucci,^{a*} Margherita Fini,^a Elena Perrin,^a Chiara Natali,^a George Colin diCenzo,^b  Alessio Mengoni^a

^aDepartment of Biology, University of Florence, Florence, Italy

^bDepartment of Biology, Queen's University, Kingston, Ontario, Canada

ABSTRACT Rhizobia are ecologically important, facultative plant-symbiotic microbes. In nature, there is a large variability in the association of rhizobial strains and host plants of the same species. Here, we evaluated whether plant and rhizobial genotypes influence the initial transcriptional response of rhizobium following perception of a host plant. RNA sequencing of the model rhizobium *Sinorhizobium meliloti* exposed to root exudates or luteolin (an inducer of *nod* genes, involved in the early steps of symbiotic interaction) was performed on a combination of three *S. meliloti* strains and three alfalfa varieties as host plants. The response to root exudates involved hundreds of changes in the rhizobium transcriptome. Of the differentially expressed genes, 35% were influenced by the strain genotype, 16% were influenced by the plant genotype, and 29% were influenced by strain-by-host plant genotype interactions. We also examined the response of a hybrid *S. meliloti* strain in which the symbiotic megaplasmid (~20% of the genome) was mobilized between two of the above-mentioned strains. Dozens of genes were upregulated in the hybrid strain, indicative of nonadditive variation in the transcriptome. In conclusion, this study demonstrated that transcriptional responses of rhizobia upon perception of legumes are influenced by the genotypes of both symbiotic partners and their interaction, suggesting a wide spectrum of genetic determinants involved in the phenotypic variation of plant-rhizobium symbiosis.

IMPORTANCE A sustainable way for meeting the need of an increased global food demand should be based on a holobiont perspective, viewing crop plants as intimately associated with their microbiome, which helps improve plant nutrition, tolerance to pests, and adverse climate conditions. However, the genetic repertoire needed for efficient association with plants by the microbial symbionts is still poorly understood. The rhizobia are an exemplary model of facultative plant symbiotic microbes. Here, we evaluated whether genotype-by-genotype interactions could be identified in the initial transcriptional response of rhizobium perception of a host plant. We performed an RNA sequencing study to analyze the transcriptomes of different rhizobial strains elicited by root exudates of three alfalfa varieties as a proxy of an early step of the symbiotic interaction. The results indicated strain- and plant variety-dependent variability in the observed transcriptional changes, providing fundamentally novel insights into the genetic basis of rhizobium-plant interactions. Our results provide genetic insights and perspective to aid in the exploitation of natural rhizobium variation for improvement of legume growth in agricultural ecosystems.

KEYWORDS RNA-seq, *Sinorhizobium meliloti*, hybrid strain, plant-microbe interactions

Microbes play a crucial role in the biology and evolution of their eukaryotic hosts (1). Among other activities, microbes contribute to the host's acquisition of


Citation Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali C, diCenzo GC, Mengoni A. 2021. Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-rhizobium symbiosis. *mSystems* 6:e00974-20. <https://doi.org/10.1128/mSystems.00974-20>.

Editor Davide Bulgarelli, University of Dundee

Copyright © 2021 Fagorzi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Giovanni Bacci, giovanni.bacci@unifi.it, or Alessio Mengoni, alessio.mengoni@unifi.it.

* Present address: Alice Checcucci, Department of Agricultural and Food Science, University of Bologna, Bologna, Italy.

 Symbiont tasting: transcriptomic signatures of host plant variety perception by rhizobia

Received 2 October 2020

Accepted 21 December 2020

Published 12 January 2021

nutrients (2), functioning of the host's immune system (3), and protection of the host from predation (4). The rules governing host-microbe interactions remain a topic of intense investigation. In many cases, the eukaryotic host selectively recruits the desired microbial partner: squid light organs are selectively colonized by *Vibrio* symbionts (5), legumes select for effective symbionts by sanctioning noneffective symbionts (6), and the crop microbiome is cultivar dependent (7, 8). The genetic basis determining the quality of a microbial symbiont (i.e., its ability to improve host plant phenotypes such as growth and tolerance) and its ability to effectively colonize its eukaryotic partner is generally not well understood, but evolution experiments and high-throughput genome sequencing projects of host-associated microbes and complete microbiomes are shedding light on this topic (9–14). In the case of plants, such studies have observed an enrichment of certain gene functions in plant-associated microbes, such as genes related to carbohydrate metabolism, secretion systems, phytohormone production, and phosphorus solubilization (11, 12, 15, 16).

The rhizobia are an ecologically important exemplar of facultative host-associated microbes. These soil-dwelling bacteria are able to colonize plants and enter an endosymbiotic association with plants of the family *Fabaceae* (17). This developmentally complex process begins with an exchange of signals between the free-living organisms (18), which leads to the invasion of the plant by the rhizobia (19), and culminates in the formation of a new organ (a nodule) in which the plant cells are intracellularly colonized by N₂-fixing rhizobia (20, 21). Decades of research have identified an intricate network of coordinated gene functions required to establish a successful mutualistic interaction between rhizobia and legumes (21–23). In contrast to the core symbiotic machinery, most of which has been elucidated, much remains unknown about the accessory genes required to optimize the interaction.

In addition to simple gene presence/absence, genotype-by-genotype (GxG) interactions have prominent impacts on symbiotic outcomes (24). The importance of both the plant and bacterial genotypes, and their interaction, in optimizing symbioses between rhizobia and legumes was recognized in early population genetic studies (25–27). More recently, greenhouse studies have directly demonstrated the influence of GxG interactions on the fitness of both the plant and rhizobium partners (28–31). The newly developed select-and-resequence approach is providing a high-throughput approach to uncover the genetic basis underlying GxG interactions for fitness in rhizobium-legume symbioses as well as a way to screen for strain-specific effects of individual genes (32, 33). To date, GxG interaction studies have largely focused on measurements of fitness as a holistic measure of the entire symbiotic process. Nodule formation is a complex developmental process involving several steps, each of which requires a distinct molecular toolkit (34), and in principle, distinct GxG interactions could be acting at each of these developmental stages. Transcriptomic studies have demonstrated that GxG interactions have significant impacts on the gene expression patterns of both partners in mature N₂-fixing nodules (35, 36). However, we are unaware of studies specifically focusing on the role of GxG interactions in early developmental stages, such as during the initial perception of the partners by each other. Such knowledge is critical not only to fully understand the microevolution of host-associated bacteria but also to develop host variety-specific rhizobium bioinoculants that may ensure good nodulation abilities over unwanted (indigenous) rhizobial strains (37, 38).

Here, we evaluated whether GxG interactions could be identified in the initial transcriptional response of rhizobium perception of a host plant. We worked with *Sinorhizobium meliloti*, which is one of the best-studied models for GxG interactions in rhizobia. *S. meliloti* forms N₂-fixing nodules on plants belonging to the tribe *Trigonelleae* (39), which includes alfalfa, a major forage crop grown worldwide for which many varieties have been developed (40). The *S. meliloti* genome comprises three main replicons, a chromosome, a chromid, and a megaplasmid; the latter one harbors most of the essential symbiotic functions, including the genes responsible for the initial molecular dialog with the

host plant (*nod* genes) (41, 42). To address our aim, the gene expression patterns of three strains of *S. meliloti* (each with distinct symbiotic properties) following 4 h of exposure to root exudates derived from three alfalfa cultivated varieties were characterized using RNA sequencing (RNA-seq). The transcriptome following exposure to luteolin (a known inducer of *nod* genes, involved in the early steps of symbiotic interaction [43]) was also analyzed. Additionally, the relevance of the megaplasmid in defining the strain-specific transcriptional responses was analyzed by studying a hybrid *S. meliloti* strain in which the native megaplasmid was replaced with that of another wild-type strain. The results demonstrated that the transcriptional response involved genes on all three replicons and that, even among conserved *S. meliloti* genes, transcriptional patterns were both strain and root exudate specific.

RESULTS

Symbiotic phenotypes differ across rhizobial strain-plant variety combinations.

Symbiotic phenotypes (plant growth and nodule number) and root adhesion of *S. meliloti* strains Rm1021, BL225C, and AK83 were measured during interactions with three varieties of alfalfa (Camporegio, Verbena, and Lodi). The results indicated that these phenotypes are influenced by both the plant and bacterial genotypes (Fig. 1; see also Fig. S1 in the supplemental material). Root adhesion phenotypes (Fig. 1a) were divided by the Scott-Knott test into three main groups reflecting high, medium, and low root colonization. Interestingly, each group was heterogeneous with respect to both plant variety and *S. meliloti* strain, consistent with the specificity of plant variety (i.e., genotype *sensu lato*) and strain individuality (i.e., strain genotype) pairs in root colonization efficiency. For instance, *S. meliloti* BL225C strongly colonized the roots of the Camporegio and Verbena varieties, but it displayed much weaker colonization of the Lodi cultivar. On the other hand, *S. meliloti* AK83 colonized the Lodi and Camporegio varieties better than the Verbena cultivar. Nodules per plant as well as measures of symbiotic efficiency (epicotyl length and shoot dry weight) showed differences among the strain-variety combinations (Fig. 1b to d). However, the extents of the variation were lower than those recorded for plant root adhesion. The highest number of nodules was found on the Lodi variety nodulated by *S. meliloti* AK83, which was previously interpreted as a consequence of its reduced N₂ fixation ability with some alfalfa varieties (44–46). Interestingly, the measures of symbiotic efficiency did not correlate with root adhesion phenotypes (both adhesion versus dry weight and adhesion versus epicotyl length gave nonsignificant Pearson correlation values [$P > 0.18$]). However, we cannot *a priori* exclude that measuring adhesion over the whole root might not reflect adhesion to the root hair extension zone, where rhizobia start the symbiotic interaction. For example, the largest plants were the Lodi variety inoculated with *S. meliloti* BL225C despite the root adhesion of this combination being the lowest. Similarly, the smallest plants were the Verbena variety inoculated with *S. meliloti* Rm1021 despite strong root adhesion in this pairing.

Root exudates differ among alfalfa varieties. Liquid chromatography-mass spectrometry (LC-MS) analysis of the alfalfa root exudates detected a total of 2,688 unique features, including 392 annotated features, across the two platforms: 1,514 hydrophilic features were detected by ultraperformance liquid chromatography (UPLC)-MS in positive mode (PP) (288 annotated), and 1,174 hydrophilic features were detected by UPLC-MS in negative mode (PN) (104 annotated) (see worksheet 1 in Data Set S1 in the supplemental material). In order to clarify if the metabolite compositions of the root exudates from the alfalfa varieties differed, principal-component analysis (PCA) was performed on the two biological replicates of the three cultivars (Fig. 2). The three cultivars clearly grouped separately from each other, suggesting the presence of variety-specific differences in their metabolic compositions. Peaks PP_23583, PP_25608, PP_14051, and PP_23300 were assigned by the PubChem database to liquiritigenin, apigenin, genistein, and apigeninidin, respectively; however, differences in the concentrations of these compounds between the root exudates were not statistically significant (data not shown). Most of the observed differences were related to amino acids, in particular *N*-acetyl-L-leucine, tryptophan, cytosine, 3,5-dihydroxyphenylglycine, and

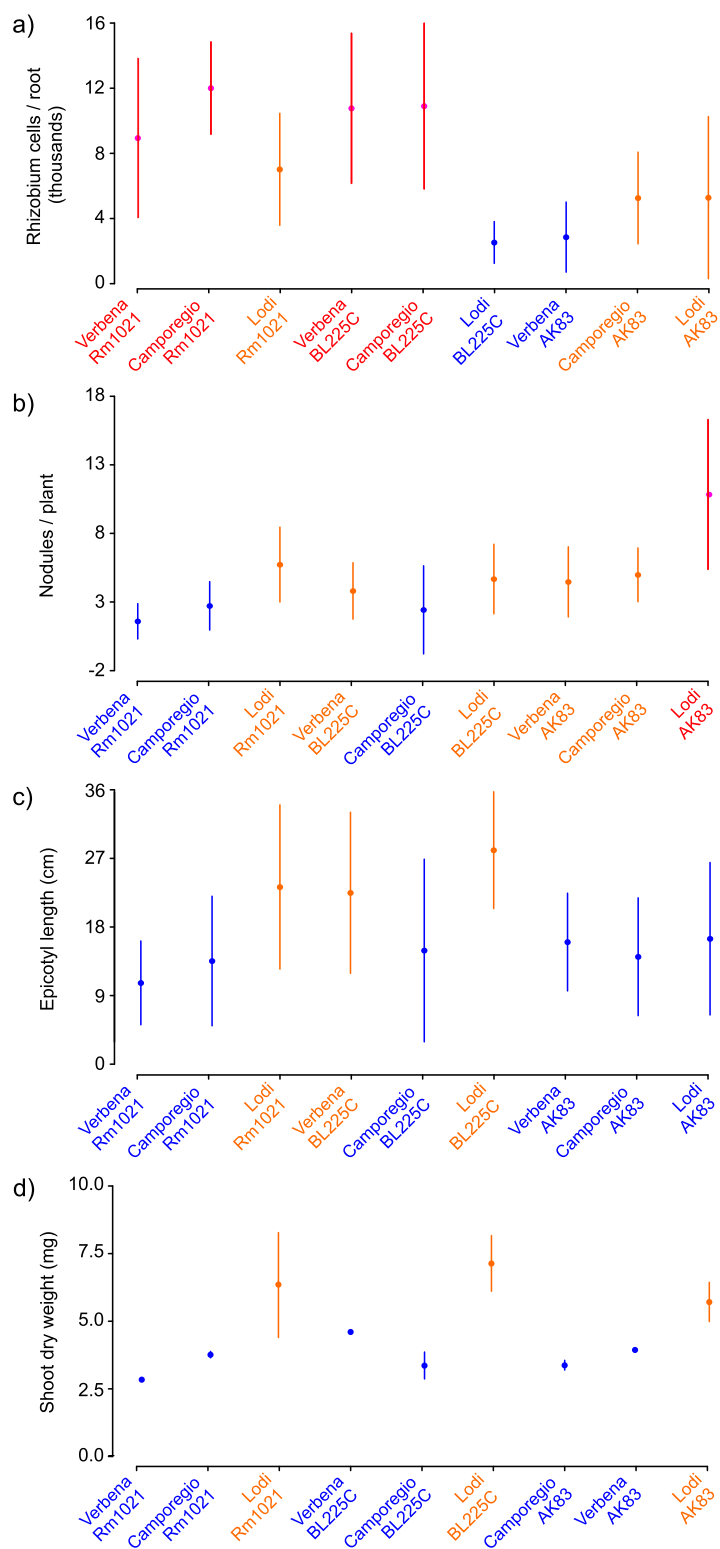


FIG 1 Strain-by-plant variation of symbiosis-associated phenotypes. The number of rhizobium cells retrieved from plant roots (a), number of nodules per plant (b), epicotyl length (c), and plant dry weight (d) are reported. Different colors (pink, orange, and blue) indicate statistically significant groupings ($P < 0.05$) based on a Scott-Knott test. For each condition, the dots indicate the mean values, and the vertical lines indicate the standard deviations.

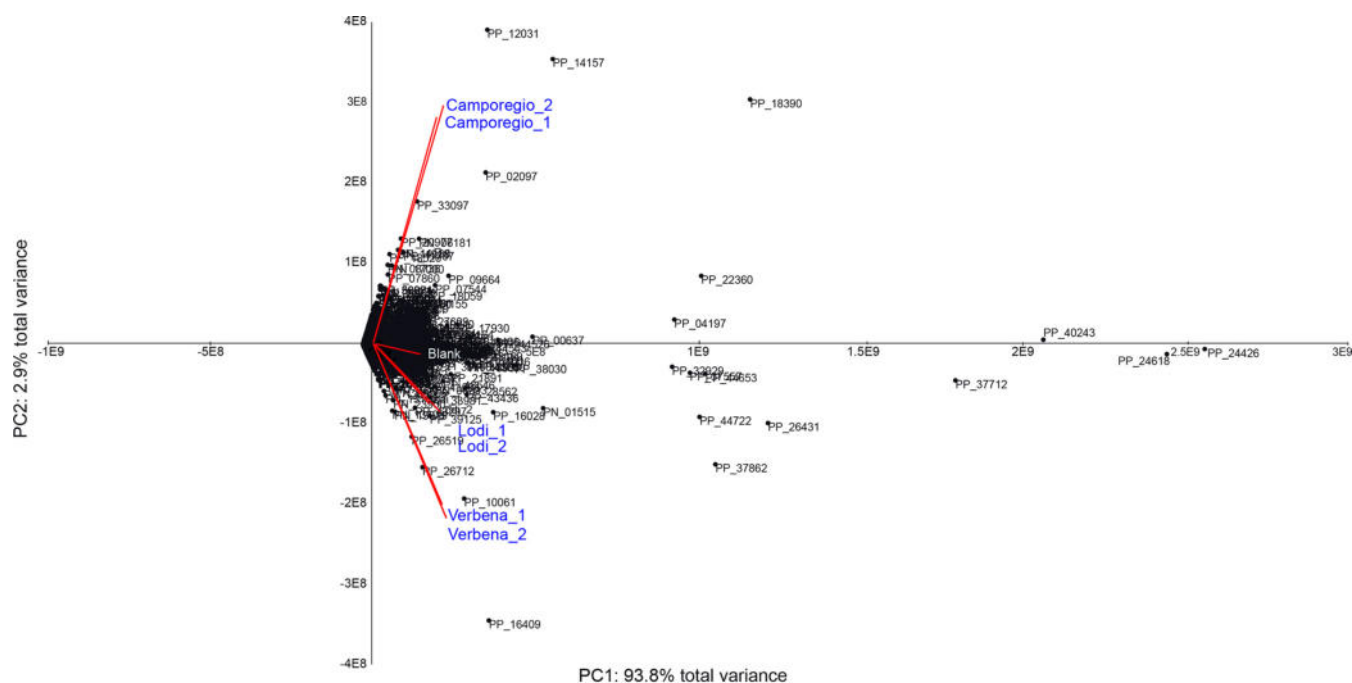


FIG 2 Principal-component analysis biplot from LC-MS analysis of root exudates of the Verbena, Lodi, and Camporegio varieties of alfalfa, including the blank control. Centroids report LC-MS peak IDs (see worksheet 1 in Data Set S1 in the supplemental material), and vectors indicate the loadings of plant varieties.

the dipeptide Val-Ala (Table S1). Multiple flavones and flavonoids, which include known inducers of NodD activation (43) and chemotaxis (47), were potentially identified. These include a peak hypothetically attributed to apigeninidin (PP_23300), which was found in the Verbena and Camporegio root exudates; liquiritigenin (PP_23583), which was found in the Camporegio and Lodi root exudates; as well as apigenin (PP_25608) and genistein (PP_14051), which were found in variable amounts in the root exudates from all three varieties. Elemental analysis (CHNS [carbon, hydrogen, nitrogen, sulfur]) of root exudates was also performed (Table S2), and the results were used to normalize the quantity of root exudates used in the treatment of *S. meliloti* strains based on equalizing the amount of total organic carbon (TOC) added to each culture.

The number of differentially expressed genes changes in strain-condition combinations. The global transcriptional responses of the three *S. meliloti* wild-type strains following a 4-h exposure to luteolin (the model flavone involved in the early steps of symbiotic interaction [43]) or alfalfa root exudates were evaluated using RNA sequencing. In addition, a fourth strain (BM806, referred to as “hybrid” for simplicity) was included (48); the results for this strain are discussed below. A list of differentially expressed genes (DEGs) for each strain and condition (luteolin and the root exudates of the three plant varieties) against the control (blank sample) is reported in Data Set S1, worksheet 2. DEGs were considered to be biologically significant if they had a ≥ 2 -fold change in expression and an adjusted *P* value of < 0.01 . The numbers of DEGs are shown in Table 1 (also see Data Set S1, worksheet 3). Reverse transcriptase quantitative PCR (RT-qPCR) on a panel of seven DEGs validated the reliability of the RNA-seq data (Table S3).

In general, luteolin treatment resulted in the lowest number of DEGs, ranging from 36 to 149 per strain. Concerning the root exudates, the number of DEGs was influenced by both the strain and the alfalfa cultivar. Overall, the Camporegio and Verbena root exudates induced more gene expression changes than the Lodi root exudate. Cluster analyses of all genes that were differentially expressed under at least one condition (fold change of ≥ 2 ; adjusted *P* value of < 0.01) revealed that for each strain, the

TABLE 1 Significant DEGs^a

Strain	No. of significant DEGs (%)			
	Camporegio	Lodi	Verbena	Luteolin
1021	516 (8.79)	32 (0.55)	506 (8.62)	36 (0.61)
AK83	357 (5.84)	66 (1.08)	192 (3.14)	60 (0.98)
BL225C	1,159 (19.33)	76 (1.27)	693 (11.56)	149 (2.49)
Hybrid	503 (8.38)	98 (1.63)	325 (5.41)	52 (0.87)

^aThe number of significant DEGs with respect to the blank control (2-fold change in expression and an adjusted *P* value of ≤ 0.01) and the percentage with respect to the total number of genes are reported.

transcriptional responses to the Verbena and Camporegio root exudates were similar and grouped separately from that of the Lodi cultivar (Fig. 3a; Fig. S2 [see also supplemental File S1 at <https://doi.org/10.5061/dryad.jdfn2z38qj>]). Interestingly, ~80% of the genes upregulated by root exudates were found on the chromosomes of the three strains, whereas ~77% of the downregulated genes were found on the pSymA and pSymB replicons (Data Set S1, worksheet 3). This is consistent with a previous signature-tagged mutagenesis study reporting that 80% of genes required for rhizosphere colonization are chromosomally located in *S. meliloti* Rm1021 (49).

Under all conditions, *S. meliloti* BL225C displayed the largest number of DEGs (with up to 20% of genes differentially expressed) (Fig. 4e and f), while *S. meliloti* AK83 had the fewest (Fig. 4c and d). The majority of DEGs (>75%) had orthologs in all three of the tested strains (Data Set S1, worksheet 4). Interestingly, $\geq 90\%$ of genes upregulated in response to root exudate exposure belonged to the core genome of the three *S. meliloti* strains (Data Set S1, worksheet 2), suggesting that the large majority of genes required for alfalfa rhizosphere colonization are highly conserved. However, expression patterns were not necessarily conserved, and strain-by-strain and condition-dependent variability of the expression pattern on the conserved gene set was observed (Fig. 4; Fig. S3). Indeed, nested likelihood ratio tests (LRTs) indicated that up to 29% of the conserved genes were influenced by strain-condition interactions, consistent with an important role of GxG interactions in the initiation of rhizobium-legume symbioses (Table 2). Moreover, the same analysis emphasized the role of strain genotype in the response to a common condition (35% of associated DEGs).

Stimulons differ in the set of elicited functions. Functional enrichment analyses, based on Kyoto Encyclopedia of Genes and Genomes (KEGG) modules and Clusters of Orthologous Genes (COG) categories, were performed to give a global overview of the functions of the DEGs (Table 3 [see also File S2 at <https://doi.org/10.5061/dryad.jdfn2z38qj>]). Strain- and condition-specific patterns of functional enrichment were observed, consistent with the functional differentiation of the stimulons from each experiment. Nevertheless, a core set of COG categories were commonly over- or under-represented in all three *S. meliloti* strains during exposure to the Camporegio or Verbena root exudates. These included enrichment among the upregulated genes of COG categories J and O related to protein expression and modification, suggesting that the root exudates stimulated major remodeling of the proteome. In addition, for upregulated genes, COG category G (carbohydrate transport and metabolism) was underrepresented, while for the downregulated genes, COG category C (energy production and conversion) was overrepresented. This observation suggests that the root exudates stimulated a global change in the cellular energy production pathways versus growth in our standard minimal medium with succinate as the sole carbon source. A comparison with growth under soil-mimicking conditions would be interesting with respect to interpreting the root exudate-induced changes in an ecological context.

Among the most highly expressed genes in *S. meliloti* Rm1021 during exposure to the Verbena and Camporegio root exudates were *smc03024* and *smc03028*, encoding components of the flagellar apparatus (*flgF* and *flgC*, respectively); the orthologs of these genes were not induced in BL225C or AK83 (Data Set S1, worksheet 2). The

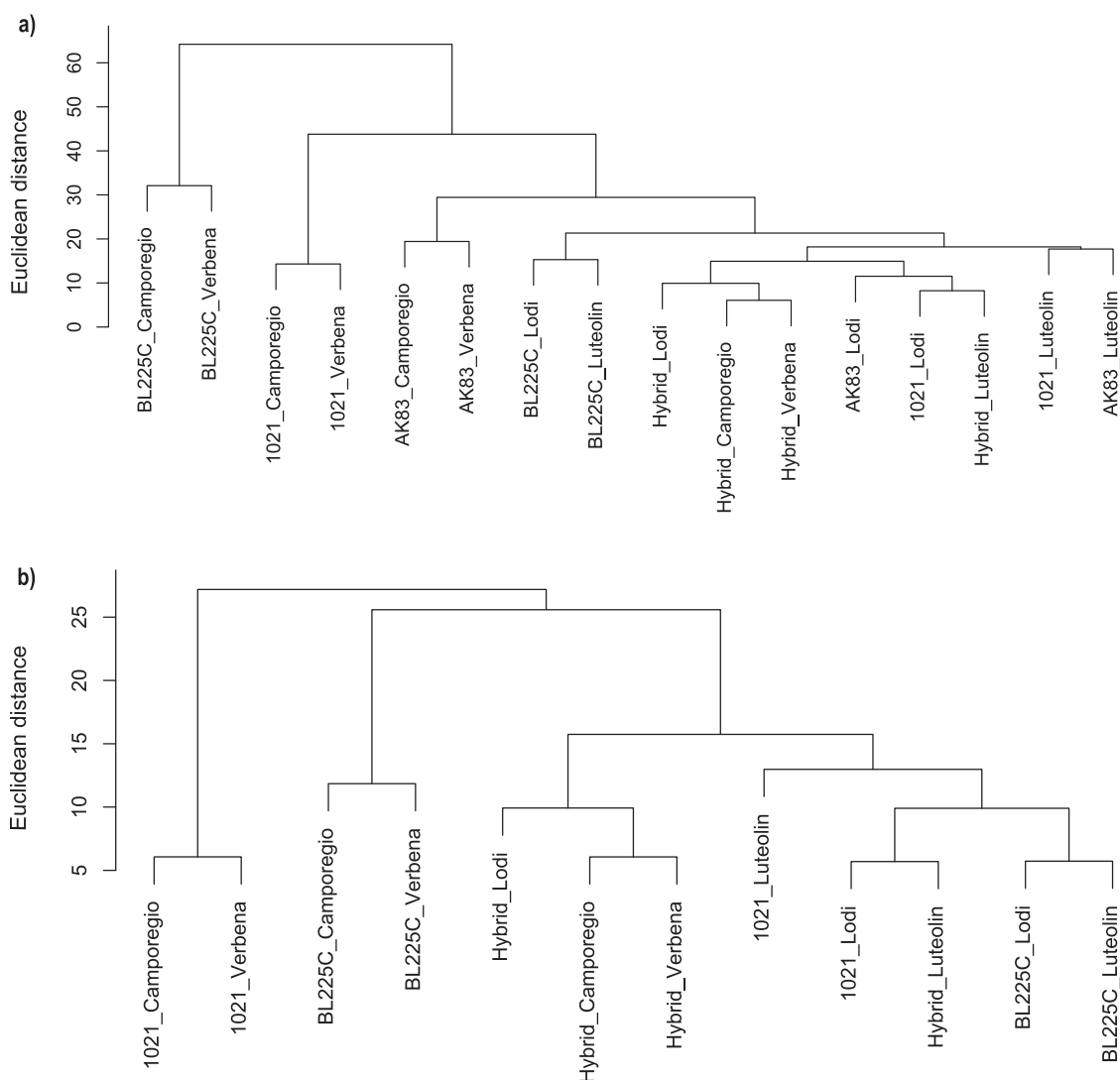


FIG 3 Cluster analyses of the expression profiles of the conserved gene sets of strains grown in the presence of the three root exudates and luteolin. (a) Full data set of DEGs identified in the four strains. (b) DEGs of orthologs of pSymA-like only. Hierarchical clustering was performed in R with the `hclust` function based on Euclidean distance, with the “complete” agglomeration method. Bars represent Euclidean distance.

induction of motility is in contrast to the observation that luteolin alone decreases the motility of the *S. meliloti* Rm1021 strain (50, 51). Presumably, this reflects the presence of additional stimuli in the root exudates. Indeed, amino acids present in root exudates are known to stimulate chemotactic behavior in *S. meliloti* (52), and signature-tagged mutagenesis showed that motility-related genes are relevant during competition for rhizosphere colonization by *S. meliloti* Rm1021 (49).

Differences in the transcriptomes of two *Bradyrhizobium diazoefficiens* strains exposed to root exudates were suggested to be related to differences in their competitive abilities (53). We therefore examined the expression patterns of several genes likely to play a role in competition for rhizosphere colonization and root adhesion. It was previously suggested that the *sin* quorum sensing system is involved in competition in *S. meliloti* (54); in our data, *sinI* (*smc00168*) was repressed in *S. meliloti* Rm1021 in the presence of the Camporegio and Verbena root exudates, but no changes in the expression of the orthologous genes in strain AK83 or BL225C were observed. No evidence was found in any of the strains for changes in the expression of galactoglucan or

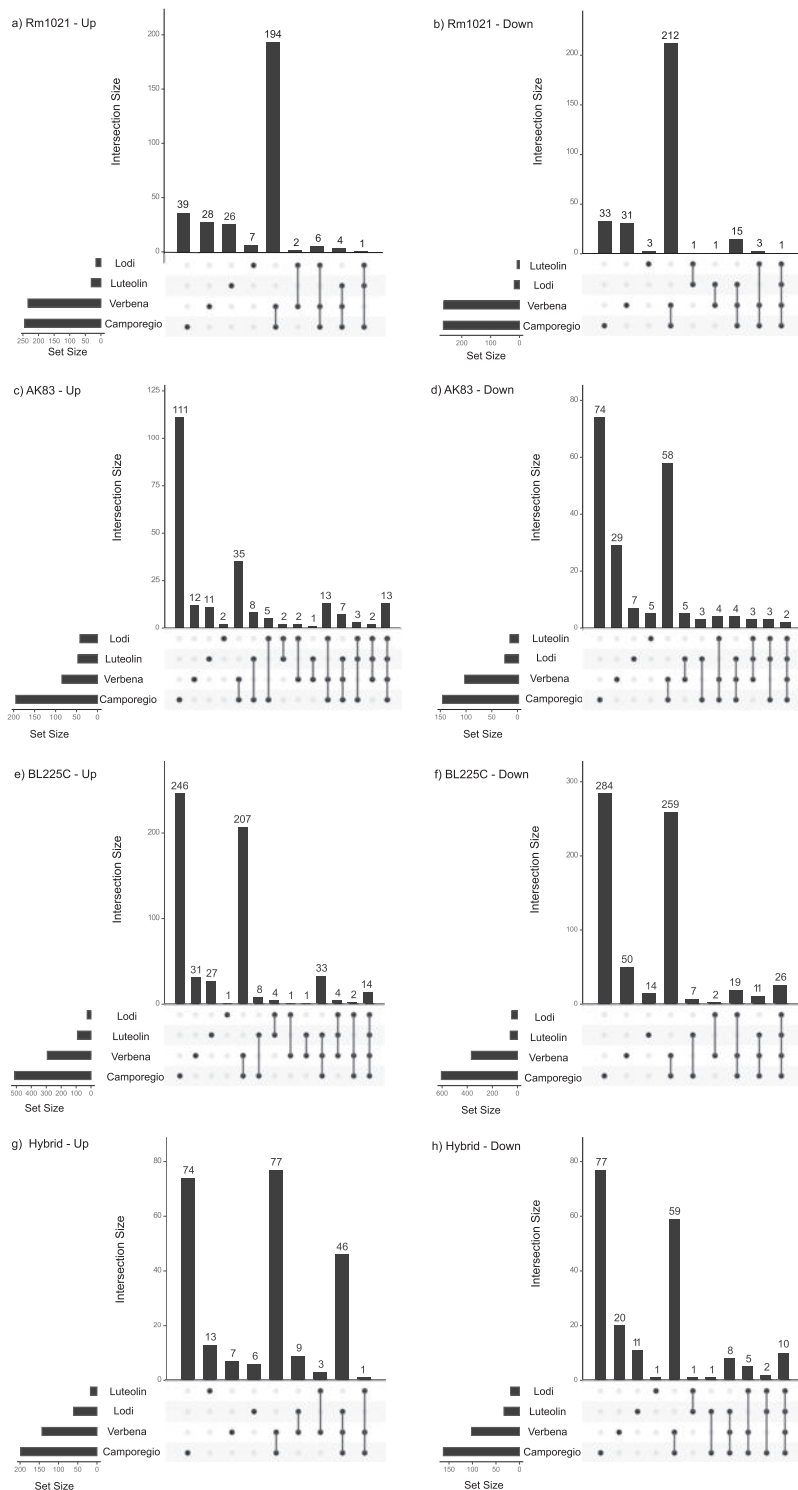


FIG 4 Intersection between up- and downregulated genes under all conditions. The plot reports the numbers of upregulated (a, c, and e) and downregulated (b, d, and f) genes under each condition for each strain. (a and b) Rm1021; (c and d) BL225; (e and f) AK83; (g and h) hybrid strain. Each row of the matrix corresponds to a condition, with the size (number of upregulated/downregulated genes) reported to the left as a bar plot. Each column corresponds to one intersection (similar to a Venn diagram): cells are either empty, indicating that up- or downregulated genes under the specified conditions are not part of the intersection, or filled, indicating that the genes present under the specified conditions are participating at the intersection. Bars on the top show the size of the intersection reported on the bottom.

TABLE 2 Number of expressed genes that showed statistical evidence of each type of expression pattern^a

Parameter	Value for groups of DEGs with significant association							None	Total
	Strain	Condition	Strain and condition	Strain × condition ^b					
Model effect									
Strain	*		*	*	*				
Condition		*	*	*		*			
Strain × condition				*	*	*	*		
No. of DEGs (%)	2,028 (25)	87 (1)	1,201 (15)	1,807 A	436 B	34 C	24 D	2,417 (30)	8,034 (100)

^aDifferential expression related to strain, condition, both strain and condition, or the interaction between strain and condition is reported. Percentages are calculated based on the total number of DEGs. Significance was based on a false discovery rate (FDR)-corrected *P* value of <0.05. * indicates whether the effect of the tested model on the expression of the gene is significant. A gene can be associated with strain (gene differentially expressed only between strains), condition (gene differentially expressed only between different conditions), strain and condition only (gene differentially expressed in relation to strain and condition but not considering the full model strain × condition), or the interaction between strain and condition (strain × condition column). The last situation can be due to a significant association with the three tested models (A), the full model and one of the others (B and C), or the full model only (D).

^bThe total number of DEGs found for strain and condition was 2,301 (29%).

succinoglucan biosynthesis genes such as *wgaA* (*sm_b21319*) and *wgeA* (*sm_b21314*). The Verbena and Camporegio root exudates induced the expression of the rhizobactin transport gene (*sma2337* [*rhtX*]) of Rm1021 and BL225C; this gene is not found in AK83. This may be a consequence of the root exudates chelating the available iron (55), consequently eliciting siderophore production that can inhibit the growth of strains lacking siderophores (56). Plasmid pSINME01 of *S. meliloti* AK83 exhibits similarity with plasmid pHRC017 of *S. meliloti* C017, which confers a competitive advantage for nodule occupancy and host range restrictions (57). Considering that a few of the genes on the plasmids pSINME01 and pSINME02 were differentially expressed upon exposure to root exudates, it is possible that the accessory plasmids of strain AK83 also contribute to competition for rhizosphere colonization (57).

Differences in gene expression patterns across conditions may be related, in part, to differences in the presence of flavonoids. In *S. meliloti*, it is known that root exudates containing flavone molecules activate the transcriptional regulator NodD (43), which triggers the synthesis of Nod factor required for nodule formation. To gain insight into the influence of NodD on the observed stimulons, we compared the *S. meliloti* Rm1021 data to those of the well-known regulons of NodD1 (requiring plant compounds for its activation) and NodD3 (not requiring plant compounds but relying on indirect activation through SyrM and NodD1 [58]) established previously (51, 59). We found that out of the 26 genes of the NodD1 regulon, 7 and 6 were observed in the DEGs in response to the Verbena and Camporegio root exudates, respectively. Camporegio and Verbena root extracts putatively contained apigenin, while the Lodi root extract lacked apigenin, suggesting a role of apigenin in the differential expression pattern observed. For the 226 genes of the NodD3 regulon, 105, 104, and 4 were found in the DEGs in response to the Verbena, Camporegio, and Lodi root exudates, respectively. The presence of a partial overlap of the known *nod* regulons (~20% or fewer of the DEGs under each condition) suggests that most of the observed DEGs belong to *nod*-independent regulons. Moreover, some of these genes showed contrasting patterns of expression, suggesting that the root exudates may also contain antagonistic molecules that repress the *nod* regulon, as previously reported (43, 60).

Mobilization of the symbiotic megaplasmid results in nonadditive changes in stimulons. To evaluate the impact of interreplicon epistatic interactions on the transcriptional response of *S. meliloti* to alfalfa root exudates, we used RNA-seq to characterize the response of a previously constructed *S. meliloti* hybrid strain containing the symbiotic megaplasmid (pSINMEB01) of strain BL225C (48). Cluster analyses clearly demonstrated that the transcriptome (both global and restricted to pSymA-pSINMEB01 orthologs only) of the hybrid strain differed from those of both the BL225C and Rm1021 wild-type strains under all conditions (Fig. 3b; Fig. S4). Of particular interest were the results observed during exposure to the Lodi root exudate. We previously showed that alfalfa cv. Lodi plants inoculated with the hybrid strain were larger than those inoculated

TABLE 3 Selected COG categories over- or underrepresented among the DEGs^a

COG category	Log ₂ fold change											
	Luteolin			Camporegio			Verbena			Lodi		
	Rm1021	BL225C	AK83	Rm1021	BL225C	AK83	Rm1021	BL225C	AK83	Rm1021	BL225C	AK83
Upregulated genes												
G	—	—	—	−1.14	−1.59	−1.67	−1.30	−1.64	—	—	—	—
J	—	—	—	2.81	0.92	2.50	2.80	1.16	1.81	—	—	—
N	—	—	—	4.19	—	—	4.25	—	—	—	—	—
O	—	—	2.38	1.56	1.44	1.50	1.44	1.91	2.21	—	—	2.96
Downregulated genes												
C	—	—	—	1.88	0.88	1.74	1.89	1.11	1.79	2.71	—	—

^aValues represent the log₂ fold changes in the abundances of genes annotated with the given COG category relative to the amount expected by chance. Dashes indicate that the COG category is not statistically different than chance under the given condition (significance threshold of a *P* value of ≤ 0.05).

with either BL225C or Rm2011 (48). Here, we observed that exposure to the Lodi root exudate results in more differentially expressed genes in the hybrid strain (98 genes) than in either Rm1021 or BL225C (32 and 76 genes, respectively) (Table 1). In particular, a cluster of genes was specifically upregulated in the hybrid strain, and the majority of these genes were located on the symbiotic megaplasmid. This peculiar feature of the Lodi-induced transcriptome in the hybrid strain was also highlighted by the cluster analysis of pSymA-pSINMEB01 orthologs; only in the hybrid strain did the Lodi-induced expression profile cluster with those of the Camporegio and Verbena root exudates (Fig. 3b). The presence of these (possibly nonadditive) transcriptional changes may reflect a loss of *cis*-regulation of these megaplasmid genes by chromosomal regulators (61, 62), providing a potential molecular mechanism underlying the improved symbiotic phenotype of the hybrid compared to both wild-type strains.

DISCUSSION

Rhizobium-legume interactions are complex multistep phenomena that begin with an exchange of signals between two partners (18, 63). The rhizobia initially detect the plant through the perception of flavonoids in the root exudate of legumes by NodD proteins, which then triggers the production of lipochitooligosaccharide molecules known as Nod factors. Nod factors are then recognized by specific LysM receptor kinase proteins in plant root cells, triggering the symbiosis signaling pathway and initiating the formation of a nodule. However, root exudates contain a mixture of flavonoids, some of them having different agonistic activities on NodD (43). Root exudates also contain many other molecules that can serve as signals or support rhizobium metabolism, such as amino acids and sugars, that may influence the ability of rhizobia to successfully colonize the rhizosphere and be in a position to enter the symbiosis (64, 65). Consequently, interactions between plant and rhizobium genotypes are expected to influence the success of the initial interaction between the two partners.

Previous works have identified a clear role for GxG interactions in the partnership between *S. meliloti* and *Medicago truncatula* (66), demonstrating that aerial biomass was influenced by the plant and rhizobium genotypes as well as their interaction. Here, we demonstrated that GxG interactions also have a significant impact on the adherence of *S. meliloti* strains to alfalfa roots, as a representative phenotype for an early stage of the interaction between these partners. Rhizosphere colonization appears to have a direct impact on nodule colonization (49, 67); while our data do not address if root adhesion is correlated with competition for nodule occupancy in mixed inocula, they suggest that root adhesion is poorly correlated with overall symbiotic efficiency in single-inoculum studies. Previous studies have demonstrated the influence of GxG interactions on the nodule transcriptome of *Medicago-Sinorhizobium* symbioses (35, 36). Here, we showed that GxG interactions similarly have an important contribution in determining the transcriptional response of *S. meliloti* to the detection of *Medicago sativa* root exudates.

Together, these results demonstrate that GxG interactions have a meaningful impact on the outcome of rhizobium-legume symbioses at multiple stages of development.

The exposure of *B. diazoefficiens* to soybean root exudates resulted in changes in the expression of 450 genes, representing nearly 5.6% of the genome, and the impacts of soybean root exudates differed between the two tested *B. diazoefficiens* strains (53). Similarly, between 0.5% and 20% of *S. meliloti* genes were differentially expressed following exposure to alfalfa root exudates, depending on the host-symbiont combination. The similarities/differences in the responses of the three *S. meliloti* strains to treatments did not appear to depend on the phylogenetic relatedness of the strains (68), although this cannot be definitively concluded without analysis of additional strains. Nevertheless, these results emphasize the importance of transcriptional rewiring during strain diversification in bacteria (62). Similarly, studies with eukaryotic organisms indicate that adaptation has an important role in differentiating the gene expression patterns of organisms (69, 70).

The root exudate stimulons only partially overlapped the stimulons of luteolin, a known inducer of NodD in *S. meliloti* (43), confirming that alfalfa root exudates contain numerous molecular signals aside from flavonoids that may influence the competitiveness of various rhizobium strains. Importantly, the transcriptional patterns induced by alfalfa root exudates differed depending on the cultivar from which they were collected; whether these differences are adaptive requires further investigation. Additionally, although root exudate metabolomic analysis was mainly descriptive, and relatively few peaks could be identified, there was a similar pattern between the differences in the *S. meliloti* gene expression profiles and the overall chemical similarity of the root exudates as measured by LC-MS; the Camporegio and Verbena root exudates induced similar gene expression changes while also being similar along the second principal component of variance (accounting for 30% of the variance) in the PCA of the root exudate composition. In future work, it would be interesting to define which compounds in the root exudates have the greatest impact on the *S. meliloti* transcriptome.

In addition to the impact of GxG interactions on rhizobium-legume symbioses, there is the potential for interreplicon interactions within rhizobium genomes to further influence the symbiosis. Indeed, interreplicon epistatic interactions are abundant in the *S. meliloti* genome (71). To address the contribution of interreplicon interactions to symbiosis, we examined a hybrid strain in which the symbiotic megaplasmid of *S. meliloti* Rm2011 (a strain nearly identical to Rm1021 [72]) was replaced with the symbiotic megaplasmid of *S. meliloti* BL225C. Nonadditive effects on the transcriptional profiles associated with all three replicons were observed in the hybrid strain relative to Rm1021 and BL225C, indicating that megaplasmid mobilization induced a global rewiring of gene expression, likely due to transcriptional cross talk among the replicons (62, 73). Similarly, nonadditive effects on the transcriptome of plant hybrids have been extensively explored (74) and demonstrated as one of the bases for heterosis in crops (75). In previous work looking for regulatory modules where the transcription factor and target genes reside on different replicons in *S. meliloti* Rm1021 (62), we found 17 transcriptional regulators encoded by the chromosome or chromid with predicted target genes on the megaplasmid. Among those transcription factors, systems related to exopolysaccharide production (ExpG), transport (PcaQ), and metabolism (IolR and GlnBK) were present, supporting the hypothesis of a global rewiring of gene expression networks and a wide range of effects of this rewiring. The results with the hybrid led us to hypothesize that the large symbiotic variability observed in natural *S. meliloti* isolates may partly be related to genome-wide transcriptome changes following large-scale horizontal gene transfer followed by natural selection. Moreover, we speculate that while the megaplasmid is the key element for a general response (i.e., cultivar independent) to species-specific host plant associations, the rhizobium chromosome and chromid fine-tune these responses in a genotype-dependent manner. If true, however, this would limit our ability to predict the competitiveness of rhizobium isolates from their simple genome sequence; instead, a more complex understanding of global regulatory network control would be required.

In conclusion, this study demonstrated that the initial perception of legumes by rhizobia leads to hundreds of changes in the rhizobium transcriptome and that these changes are dependent on the plant genotype, the rhizobium genotype, and genotype-by-genotype interactions. These results complement previous studies demonstrating the role of GxG interactions in determining the transcriptome of both the legume and rhizobium partners in mature N_2 -fixing nodules (35, 36). The majority of genes upregulated in response to alfalfa root exudates were conserved in all three strains, supporting the hypothesis that the *S. meliloti* lineage was adapted to rhizosphere colonization before gaining the genes required for symbiotic nitrogen fixation (49). Additionally, the transcriptional response to the perception of alfalfa root exudates involved genes from all three of the *S. meliloti* replicons and seemingly involved nonadditive effects resulting from interreplicon interactions.

MATERIALS AND METHODS

Microbiological methods and plant assays. A list of strains, their host plants of origin, as well as the plant varieties used is reported in Table S4 in the supplemental material. Plant varieties included three contrasting alfalfa genotypes: *Medicago falcata* (Verbena), *M. sativa* (Lodi), and *Medicago* × *varia* (*M. sativa* × *M. falcata*). Strains included *S. meliloti* Rm1021, BL225C, AK83, and a hybrid strain containing the chromosome and pSymb of strain Rm2011 and the symbiotic megaplasmid (pSINMEB01) of strain BL225C. *S. meliloti* Rm2011 is nearly isogenic to Rm1021, both being independent streptomycin-resistant derivatives of the nodule isolate SU47 (76, 77). Details on strains, plant growth, and symbiotic assays are found in Text S1 in the supplemental material. The root adhesion test was performed 5 days following the inoculation of plantlets (Text S1). Differences were evaluated by one-way analysis of variance (ANOVA) Tukey pairwise contrast and using the Scott-Knott procedure as implemented in R (78). All primer pairs used are reported in Table S4.

Root exudate production and metabolomic analyses. Root exudates were produced by growing plants under sterile conditions in water for 14 days, as previously reported (79) and as reported in Text S1. Elemental analysis (CHNS) was performed on crude root exudates (a combined sample for each cultivar) using a carbon, hydrogen, and nitrogen analyzer (CHN-S Flash E1112; Thermo Finnigan, San Jose, CA, USA). Metabolomic analysis was performed by LC-MS, and data from reverse-phase UPLC (RP-UPLC) and UPLC-MS were combined to build the final data matrix. Principal-component analysis (PCA) was performed on the Bray-Curtis dissimilarity obtained from each peak identification (ID) value (Text S1). Statistical differences in single metabolites were assessed by Simper analysis based on the decomposition of the Bray-Curtis dissimilarity obtained from each peak ID value. All statistical analyses were done with the vegan package of R (80). The PubChem database was used for additional peak identification from brute formulas (<https://pubchem.ncbi.nlm.nih.gov/>).

RNA isolation and RNA sequencing. Cultures of *S. meliloti*, grown overnight in M9-succinate medium at 30°C at 130 rpm, were diluted to an optical density at 600 nm (OD_{600}) of 0.05 in 5 ml of M9-succinate medium and incubated until an OD_{600} of 0.4 was reached. Next, either 10 μ M luteolin (Sigma-Aldrich) or one of the alfalfa root exudates (normalized by the total organic carbon as measured by the CHNS analysis) was added to each of the cultures, and the mixture was incubated for an additional 4 h at 30°C with shaking at 130 rpm. Biological replicates were performed for each of the three strains across the five conditions. Total RNA was extracted using RNeasy minikits (Qiagen) from 0.5 ml of the culture and subjected to DNase I treatment. Details on the RNA isolation procedure and quality checks are provided in Text S1. Validation of expression differences was done using reverse transcriptase qPCR as described in Text S1. Protocols for rRNA depletion and library construction are described in Text S1. Libraries were sequenced on an Illumina Novaseq 6000 apparatus with an SP flow cell.

Read mapping, counting, and differential expression analysis. Trimmed and demultiplexed reads were mapped back to transcripts using Salmon (version 1.1.0) (81) (see Text S1 for details). Quantification files produced by Salmon were then imported into R using the tximport package (version 1.10.1) (82). Differential abundance analysis was performed with the DESeq2 version 1.22.2 package (83) on single strains under different conditions.

Statistical analysis of differentially expressed genes. For each *S. meliloti* strain, genes differentially expressed (\log_2 fold change of ≥ 1 ; P value of < 0.01) under at least one condition relative to the control conditions were identified, and all fold change values for these genes were extracted. To compare expression values of genes conserved between Rm1021, AK83, and BL225C, the pangenome of the three strains was calculated using Roary version 3.13.0 (84) with an identity threshold of 90%, and the genes found in all three strains (the core genes) were recorded. Under each condition, core genes differentially expressed in at least one strain relative to the control conditions were identified, and the fold change values for the gene and its orthologs in the other strains were extracted.

All genes of *S. meliloti* strains Rm1021, AK83, and BL225C were functionally annotated using standalone version 2 of eggNOG-mapper (85, 86) with default settings and the following two modifications: the mode was set to diamond, and query cover was set to 20. Methods for the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Genes (COG) category annotations are reported in Text S1.

Nested likelihood ratio tests (LRTs) were used to evaluate the statistical significance of strain, condition, and strain-by-condition interaction effects on gene expression. Transcripts were collapsed into

orthologous groups based on the output of Roary, as described above. Counts produced by Salmon were collapsed following the group ID provided by Roary, producing a single table with ortholog-level quantification of transcripts. The produced table was then used to perform a nested LRT with DESeq2. Strains and conditions were used together with their interaction to build a model for each group. Terms were then removed one by one to test their impact on the likelihood of the full model (as described in the DESeq2 documentation at <http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#likelihood-ratio-test>).

Data availability. Gene expression data are available at GEO under the accession number [GSE151705](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151705). Custom scripts developed for this work can be found in the GitHub repository at <https://github.com/hyhy8181994/Sinorhizobium-RNAseq-2020>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.04 MB.

FIG S1, PDF file, 0.04 MB.

FIG S2, PDF file, 0.2 MB.

FIG S3, PDF file, 0.1 MB.

FIG S4, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.02 MB.

TABLE S2, PDF file, 0.01 MB.

TABLE S3, CSV file, 0.01 MB.

TABLE S4, PDF file, 0.1 MB.

DATA SET S1, XLSX file, 0.8 MB.

ACKNOWLEDGMENTS

We are grateful to Gabriele Brazzini for technical assistance in setting up the root adhesion test and Susanna Pucci, Centro di Microanalisi, University of Florence, Italy, for CHNS analysis.

This work was supported by Fondazione Cassa di Risparmio di Firenze grant number 18204, 2017.0719; by a MICRO4Legumes grant (Italian Ministry of Agriculture); and by a grant for the Dipartimento di Eccellenza 2018-2022 by the Italian Ministry of Education, University, and Research (MIUR). L.C. was supported by a fellowship from MICRO4Legumes (Italian Ministry of Agriculture). Work in the G.C.D. laboratory is supported by funding from Queen's University and the Natural Sciences and Engineering Research Council of Canada.



REFERENCES

- Rosenberg E, Zilber-Rosenberg I. 2016. Microbes drive evolution of animals and plants: the hologenome concept. *mBio* 7:e01395-15. <https://doi.org/10.1128/mBio.01395-15>.
- Mus F, Crook MB, Garcia K, Costas AG, Geddes BA, Kouri ED, Paramasivan P, Ryu M-H, Oldroyd GED, Poole PS, Udvardi MK, Voigt CA, Ané J-M, Peters JW. 2016. Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl Environ Microbiol* 82:3698–3710. <https://doi.org/10.1128/AEM.01055-16>.
- Lee YK, Mazmanian SK. 2010. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 330:1768–1773. <https://doi.org/10.1126/science.1195568>.
- Jones BW, Nishiguchi MK. 2004. Counterillumination in the Hawaiian bobtail squid, *Euprymna scolopes* Berry (Mollusca: Cephalopoda). *Mar Biol* 144:1151–1155. <https://doi.org/10.1007/s00227-003-1285-3>.
- Nyholm SV, McFall-Ngai M. 2004. The winnowing: establishing the squid-vibrio symbiosis. *Nat Rev Microbiol* 2:632–642. <https://doi.org/10.1038/nrmicro957>.
- Kiers ET, Rousseau RA, West SA, Denison RF. 2003. Host sanctions and the legume-rhizobium mutualism. *Nature* 425:78–81. <https://doi.org/10.1038/nature01931>.
- Mwafurirwa L, Baggs EM, Russell J, George T, Morley N, Sim A, de la Fuente Cantó C, Paterson E. 2016. Barley genotype influences stabilization of rhizodeposition-derived C and soil organic matter mineralization. *Soil Biol Biochem* 95:60–69. <https://doi.org/10.1016/j.soilbio.2015.12.011>.
- Escudero-Martinez C, Bulgarelli D. 2019. Tracing the evolutionary routes of plant-microbiota interactions. *Curr Opin Microbiol* 49:34–40. <https://doi.org/10.1016/j.mib.2019.09.013>.
- Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD, Quince C, Huttenhower C, Segata N. 2019. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 176:649–662. e20. <https://doi.org/10.1016/j.cell.2019.01.001>.
- Singh BK, Liu H, Trivedi P. 2020. Eco-holobiont: a new concept to identify drivers of host-associated microorganisms. *Environ Microbiol* 22:564–567. <https://doi.org/10.1111/1462-2920.14900>.
- de Souza RSC, Armanhi JSL, Damasceno NDB, Imperial J, Arruda P. 2019. Genome sequences of a plant beneficial synthetic bacterial community reveal genetic features for successful plant colonization. *Front Microbiol* 10:1779. <https://doi.org/10.3389/fmicb.2019.01779>.
- Levy A, Gonzalez IS, Mittelviehhaus M, Clingenpeel S, Paredes SH, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Alvarez BR, Lundberg DS, Lu T-Y, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. 2018. Genomic features of bacterial adaptation to plants. *Nat Genet* 50:138–150. <https://doi.org/10.1038/s41588-017-0012-9>.
- Harrison E, Brockhurst MA. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol* 20:262–267. <https://doi.org/10.1016/j.tim.2012.04.003>.
- Batstone RT, O'Brien AM, Harrison TL, Frederickson ME. 2020. Experimental evolution makes microbes more cooperative with their local host genotype. *Science* 370:476–478. <https://doi.org/10.1126/science.abb7222>.
- Pini F, Galardini M, Bazzicalupo M, Mengoni A. 2011. Plant-bacteria

- association and symbiosis: are there common genomic traits in Alphaproteobacteria? *Genes (Basel)* 2:1017–1032. <https://doi.org/10.3390/genes2041017>.
16. Guttman D, McHardy AC, Schulze-Lefert P. 2014. Microbial genome-enabled insights into plant-microorganism interactions. *Nat Rev Genet* 15:797–813. <https://doi.org/10.1038/nrg3748>.
 17. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* 16:291–303. <https://doi.org/10.1038/nrmicro.2017.171>.
 18. Oldroyd GED. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol* 11:252–263. <https://doi.org/10.1038/nrmicro2990>.
 19. Gage DJ. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev* 68:280–300. <https://doi.org/10.1128/MMBR.68.2.280-300.2004>.
 20. Udvardi M, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* 64:781–805. <https://doi.org/10.1146/annurev-arplant-050312-120235>.
 21. Kereszt A, Mergaert P, Kondorosí E. 2011. Bacteroid development in legume nodules: evolution of mutual benefit or of sacrificial victims? *Mol Plant Microbe Interact* 24:1300–1309. <https://doi.org/10.1094/MPMI-06-11-0152>.
 22. Benezech C, Doudement M, Gourion B. 2020. Legumes tolerance to rhizobia is not always observed and not always deserved. *Cell Microbiol* 22:e13124. <https://doi.org/10.1111/cmi.13124>.
 23. Gibson KE, Kobayashi H, Walker GC. 2008. Molecular determinants of a symbiotic chronic infection. *Annu Rev Genet* 42:413–441. <https://doi.org/10.1146/annurev.genet.42.110807.091427>.
 24. Burghardt LT. 2020. Evolving together, evolving apart: measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. *New Phytol* 228:28–34. <https://doi.org/10.1111/nph.16045>.
 25. Paffetti D, Scotti C, Gnocchi S, Fancelli S, Bazzicalupo M. 1996. Genetic diversity of an Italian Rhizobium meliloti population from different Medicago sativa varieties. *Appl Environ Microbiol* 62:2279–2285. <https://doi.org/10.1128/AEM.62.7.2279-2285.1996>.
 26. Carelli M, Gnocchi S, Fancelli S, Mengoni A, Paffetti D, Scotti C, Bazzicalupo M. 2000. Genetic diversity and dynamics of Sinorhizobium meliloti populations nodulating different alfalfa cultivars in Italian soils. *Appl Environ Microbiol* 66:4785–4789. <https://doi.org/10.1128/aem.66.11.4785-4789.2000>.
 27. Rangin C, Brunel B, Cleyet-Marel J-C, Perrineau M-M, Béna G. 2008. Effects of Medicago truncatula genetic diversity, rhizobial competition, and strain effectiveness on the diversity of a natural Sinorhizobium species community. *Appl Environ Microbiol* 74:5653–5661. <https://doi.org/10.1128/AEM.01107-08>.
 28. Heath KD, Tiffin P. 2007. Context dependence in the coevolution of plant and rhizobial mutualists. *Proc Biol Sci* 274:1905–1912. <https://doi.org/10.1098/rspb.2007.0495>.
 29. Heath KD, Tiffin P. 2009. Stabilizing mechanisms in a legume-rhizobium mutualism. *Evolution* 63:652–662. <https://doi.org/10.1111/j.1558-5646.2008.00582.x>.
 30. Heath KD. 2010. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution* 64:1446–1458. <https://doi.org/10.1111/j.1558-5646.2009.00913.x>.
 31. Ehinger M, Mohr TJ, Starcevic JB, Sachs JL, Porter SS, Simms EL. 2014. Specialization-generalization trade-off in a Bradyrhizobium symbiosis with wild legume hosts. *BMC Ecol* 14:8. <https://doi.org/10.1186/1472-6785-14-8>.
 32. Burghardt LT, Epstein B, Guhlin J, Nelson MS, Taylor MR, Young ND, Sadowsky MJ, Tiffin P. 2018. Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proc Natl Acad Sci U S A* 115:2425–2430. <https://doi.org/10.1073/pnas.1714246115>.
 33. Burghardt LT, Trujillo DI, Epstein B, Tiffin P, Young ND. 2020. A select and resequence approach reveals strain-specific effects of Medicago nodule-specific PLAT-domain genes. *Plant Physiol* 182:463–471. <https://doi.org/10.1104/pp.19.00831>.
 34. Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, Debelle F, Capela D, De Carvalho-Niebel F, Gouzy J, Bruand C, Gamas P. 2014. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J* 77:817–837. <https://doi.org/10.1111/tpj.12442>.
 35. Heath KD, Burke PV, Stinchcombe JR. 2012. Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol Ecol* 21:4735–4747. <https://doi.org/10.1111/j.1365-294X.2012.05629.x>.
 36. Burghardt LT, Guhlin J, Chun CL, Liu J, Sadowsky MJ, Stupar RM, Young ND, Tiffin P. 2017. Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol Ecol* 26:6122–6135. <https://doi.org/10.1111/mec.14285>.
 37. Triplett EW, Sadowsky MJ. 1992. Genetics of competition for nodulation of legumes. *Annu Rev Microbiol* 46:399–422. <https://doi.org/10.1146/annurev.mi.46.100192.002151>.
 38. Checcucci A, DiCenzo GC, Bazzicalupo M, Mengoni A. 2017. Trade, diplomacy, and warfare: the quest for elite rhizobia inoculant strains. *Front Microbiol* 8:2207. <https://doi.org/10.3389/fmicb.2017.02207>.
 39. Sprent JI, Ardley J, James EK. 2017. Biogeography of nodulated legumes and their nitrogen-fixing symbionts. *New Phytol* 215:40–56. <https://doi.org/10.1111/nph.14474>.
 40. Frame J, Charlton JFL, Laidlaw AS. 1998. Temperate forage legumes. CAB International, Wallingford, United Kingdom.
 41. Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenberg M, Vorholter FJ, et al. 2001. The composite genome of the legume symbiont Sinorhizobium meliloti. *Science* 293:668–672. <https://doi.org/10.1126/science.1060966>.
 42. Harrison PW, Lower RPJ, Kim NKD, Young JPW. 2010. Introducing the bacterial “chromid”: not a chromosome, not a plasmid. *Trends Microbiol* 18:141–148. <https://doi.org/10.1016/j.tim.2009.12.010>.
 43. Peck MC, Fisher RF, Long SR. 2006. Diverse flavonoids stimulate NodD1 binding to nod gene promoters in Sinorhizobium meliloti. *J Bacteriol* 188:5417–5427. <https://doi.org/10.1128/JB.00376-06>.
 44. Galardini M, Mengoni A, Brilli M, Pini F, Fioravanti A, Lucas S, Lapidus A, Cheng J-F, Goodwin L, Pitluck S, Land M, Hauser L, Woyke T, Mikhailova N, Ivanova N, Daligault H, Bruce D, Detter C, Tapia R, Han C, Teshima H, Mocali S, Bazzicalupo M, Biondi EG. 2011. Exploring the symbiotic pangenome of the nitrogen-fixing bacterium Sinorhizobium meliloti. *BMC Genomics* 12:235. <https://doi.org/10.1186/1471-2164-12-235>.
 45. Biondi EG, Tatti E, Comparini D, Giuntini E, Mocali S, Giovannetti L, Bazzicalupo M, Mengoni A, Viti C. 2009. Metabolic capacity of Sinorhizobium (Ensifer) meliloti strains as determined by phenotype microarray analysis. *Appl Environ Microbiol* 75:5396–5404. <https://doi.org/10.1128/AEM.00196-09>.
 46. Checcucci A, Azzarello E, Bazzicalupo M, Galardini M, Lagomarsino A, Mancuso S, Marti L, Marzano MC, Mocali S, Squartini A, Zanardo M, Mengoni A. 2016. Mixed nodule infection in Sinorhizobium meliloti-Medicago sativa symbiosis suggest the presence of cheating behavior. *Front Plant Sci* 7:835. <https://doi.org/10.3389/fpls.2016.00835>.
 47. Caetano-Anollés G, Crist-Estes DK, Bauer WD. 1988. Chemotaxis of Rhizobium meliloti to the plant flavone luteolin requires functional nodulation genes. *J Bacteriol* 170:3164–3169. <https://doi.org/10.1128/jb.170.7.3164-3169.1988>.
 48. Checcucci A, DiCenzo GC, Ghini V, Bazzicalupo M, Becker A, Decorosi F, Döhlemann J, Fagorzi C, Finan TM, Fondi M, Luchinat C, Turano P, Vignolini T, Viti C, Mengoni A. 2018. Creation and characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium Sinorhizobium meliloti. *ACS Synth Biol* 7:2365–2378. <https://doi.org/10.1021/acssynbio.8b00158>.
 49. Salas ME, Lozano MJ, López JL, Draghi WO, Serrania J, Torres Tejerizo GA, Albicoro FJ, Nilsson JF, Pistorio M, Del Papa MF, Parisi G, Becker A, Lagares A. 2017. Specificity traits consistent with legume-rhizobia coevolution displayed by Ensifer meliloti rhizosphere colonization. *Environ Microbiol* 19:3423–3438. <https://doi.org/10.1111/1462-2920.13820>.
 50. Spini F, Decorosi F, Cerboneschi M, Tegli S, Mengoni A, Viti C, Giovannetti L. 2016. Effect of the plant flavonoid luteolin on Ensifer meliloti 3001 phenotypic responses. *Plant Soil* 399:159–178. <https://doi.org/10.1007/s11004-015-2659-2>.
 51. Barnett MJ, Long SR. 2015. The Sinorhizobium meliloti SyrM regulon: effects on global gene expression are mediated by syrA and nodD3. *J Bacteriol* 197:1792–1806. <https://doi.org/10.1128/JB.02626-14>.
 52. Webb BA, Helm RF, Scharf BE. 2016. Contribution of individual chemoreceptors to Sinorhizobium meliloti chemotaxis towards amino acids of

- host and nonhost seed exudates. *Mol Plant Microbe Interact* 29:231–239. <https://doi.org/10.1094/MPMI-12-15-0264-R>.
53. Liu Y, Jiang X, Guan D, Zhou W, Ma M, Zhao B, Cao F, Li L, Li J. 2017. Transcriptional analysis of genes involved in competitive nodulation in *Bradyrhizobium diazoefficiens* at the presence of soybean root exudates. *Sci Rep* 7:10946. <https://doi.org/10.1038/s41598-017-11372-0>.
 54. McIntosh M, Krol E, Becker A. 2008. Competitive and cooperative effects in quorum-sensing-regulated galactoglucan biosynthesis in *Sinorhizobium meliloti*. *J Bacteriol* 190:5308–5317. <https://doi.org/10.1128/JB.00063-08>.
 55. Parker DR, Reichman SM, Crowley DE. 2005. Metal chelation in the rhizosphere, p 57–93. In Zobel RW, Wright SF (ed), *Roots and soil management: interactions between roots and the soil*, vol 48. American Society of Agronomy, Inc, Madison, WI.
 56. diCenzo GC, MacLean AM, Milunovic B, Golding GB, Finan TM. 2014. Examination of prokaryotic multipartite genome evolution through experimental genome reduction. *PLoS Genet* 10:e1004742. <https://doi.org/10.1371/journal.pgen.1004742>.
 57. Crook MB, Lindsay DP, Biggs MB, Bentley JS, Price JC, Clement SC, Clement MJ, Long SR, Griffiths JS. 2012. Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. *Mol Plant Microbe Interact* 25:1026–1033. <https://doi.org/10.1094/MPMI-02-12-0052-R>.
 58. Maillé F, Debelle F, Dénarié J. 1990. Role of the *nodD* and *syrM* genes in the activation of the regulatory gene *nodD3*, and of the common and host-specific *nod* genes of *Rhizobium meliloti*. *Mol Microbiol* 4:1975–1984. <https://doi.org/10.1111/j.1365-2958.1990.tb02047.x>.
 59. Capela D, Carrere S, Batut J. 2005. Transcriptome-based identification of the *Sinorhizobium meliloti* *NodD1* regulon. *Appl Environ Microbiol* 71:4910–4913. <https://doi.org/10.1128/AEM.71.8.4910-4913.2005>.
 60. Zuanazzi JAS, Clergeot PH, Quirion J-C, Husson H-P, Kondorosi A, Ratet P. 1998. Production of *Sinorhizobium meliloti* *nod* gene activator and repressor flavonoids from *Medicago sativa* roots. *Mol Plant Microbe Interact* 11:784–794. <https://doi.org/10.1094/MPMI.1998.11.8.784>.
 61. Ramos JL, Marqués S, Timmis KN. 1997. Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. *Annu Rev Microbiol* 51:341–373. <https://doi.org/10.1146/annurev.micro.51.1.341>.
 62. Galardini M, Brilli M, Spini G, Rossi M, Roncaglia B, Bani A, Chianciani M, Moretto M, Engelen K, Bacci G, Pini F, Biondi EG, Bazzicalupo M, Mengoni A. 2015. Evolution of intra-specific regulatory networks in a multipartite bacterial genome. *PLoS Comput Biol* 11:e1004478. <https://doi.org/10.1371/journal.pcbi.1004478>.
 63. Oldroyd GED, Murray JD, Poole PS, Downie JA. 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* 45:119–144. <https://doi.org/10.1146/annurev-genet-110410-132549>.
 64. Barbour WM, Hattermann DR, Stacey G. 1991. Chemotaxis of *Bradyrhizobium japonicum* to soybean exudates. *Appl Environ Microbiol* 57:2635–2639. <https://doi.org/10.1128/AEM.57.9.2635-2639.1991>.
 65. Webb BA, Compton KK, Martin JS, Taylor D, Sobrado P, Scharf BE. 2017. *Sinorhizobium meliloti* chemotaxis to multiple amino acids is mediated by the chemoreceptor McpU. *Mol Plant Microbe Interact* 30:770–777. <https://doi.org/10.1094/MPMI-04-17-0096-R>.
 66. Mhadhbi H, Jebara M, Limam F, Huguet T, Aouani ME. 2005. Interaction between *Medicago truncatula* lines and *Sinorhizobium meliloti* strains for symbiotic efficiency and nodule antioxidant activities. *Physiol Plant* 124:4–11. <https://doi.org/10.1111/j.1399-3054.2005.00489.x>.
 67. Entcheva P, Phillips DA, Streit WR. 2002. Functional analysis of *Sinorhizobium meliloti* genes involved in biotin synthesis and transport. *Appl Environ Microbiol* 68:2843–2848. <https://doi.org/10.1128/aem.68.6.2843-2848.2002>.
 68. Galardini M, Pini F, Bazzicalupo M, Biondi EG, Mengoni A. 2013. Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. *Genome Biol Evol* 5:542–558. <https://doi.org/10.1093/gbe/evt027>.
 69. López-Maury L, Marguerat S, Bähler J. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* 9:583–593. <https://doi.org/10.1038/nrg2398>.
 70. Whitehead A, Crawford DL. 2006. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol* 15:1197–1211. <https://doi.org/10.1111/j.1365-294X.2006.02868.x>.
 71. diCenzo GC, Benedict AB, Fondi M, Walker GC, Finan TM, Mengoni A, Griffiths JS. 2018. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genet* 14:e1007357. <https://doi.org/10.1371/journal.pgen.1007357>.
 72. Meade HM, Signer ER. 1977. Genetic mapping of *Rhizobium meliloti*. *Proc Natl Acad Sci U S A* 74:2076–2078. <https://doi.org/10.1073/pnas.74.5.2076>.
 73. diCenzo GC, Wellappili D, Golding GB, Finan TM. 2018. Inter-replicon gene flow contributes to transcriptional integration in the *Sinorhizobium meliloti* multipartite genome. *G3 (Bethesda)* 8:1711–1720. <https://doi.org/10.1534/g3.117.300405>.
 74. Bell GDM, Kane NC, Rieseberg LH, Adams KL. 2013. RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol Evol* 5:1309–1323. <https://doi.org/10.1093/gbe/evt072>.
 75. Hochholdinger F, Hoecker N. 2007. Towards the molecular basis of heterosis. *Trends Plant Sci* 12:427–432. <https://doi.org/10.1016/j.tplants.2007.08.005>.
 76. Sallet E, Roux B, Sauviac L, Jardinaud MF, Carrère S, Faraut T, de Carvalho-Niebel F, Gouzy J, Gamas P, Capela D, Bruand C, Schiex T. 2013. Next-generation annotation of prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. *DNA Res* 20:339–353. <https://doi.org/10.1093/dnares/dst014>.
 77. Krol E, Becker A. 2004. Global transcriptional analysis of the phosphate starvation response in *Sinorhizobium meliloti* strains 1021 and 2011. *Mol Genet Genomics* 272:1–17. <https://doi.org/10.1007/s00438-004-1030-8>.
 78. R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
 79. Checcucci A, Azzarello E, Bazzicalupo M, De Carlo A, Emiliani G, Mancuso S, Spini G, Viti C, Mengoni A. 2017. Role and regulation of ACC deaminase gene in *Sinorhizobium meliloti*: is it a symbiotic, rhizospheric or endophytic gene? *Front Genet* 8:6. <https://doi.org/10.3389/fgene.2017.00006>.
 80. Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Wagner H. 2013. *vegan*: community ecology package. R package version 2.0-10.
 81. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14:417–419. <https://doi.org/10.1038/nmeth.4197>.
 82. Sonesson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4:1521. <https://doi.org/10.12688/f1000research.7563.1>.
 83. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
 84. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31:3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>.
 85. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
 86. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* 34:2115–2122. <https://doi.org/10.1093/molbev/msx148>.

Symbiotic and Nonsymbiotic Members of the Genus *Ensifer* (syn. *Sinorhizobium*) Are Separated into Two Clades Based on Comparative Genomics and High-Throughput Phenotyping

Camilla Fagorzi¹, Alexandru Ilie¹, Francesca Decorosi², Lisa Cangiolli¹, Carlo Viti², Alessio Mengoni ^{1,*}, and George C. diCenzo ^{1,3,*}

¹Department of Biology, University of Florence, Sesto Fiorentino, Italy

²Genexpress Laboratory, Department of Agriculture, Food, Environment and Forestry, University of Florence, Sesto Fiorentino, Italy

³Department of Biology, Queen's University, Kingston, Ontario, Canada

*Corresponding authors: E-mails: alessio.mengoni@unifi.it; george.dicenzo@queensu.ca.

Accepted: 12 October 2020

Data deposition: Genome sequences have been deposited at NCBI under the BioProject accession PRJNA622509.

Abstract

Rhizobium–legume symbioses serve as paradigmatic examples for the study of mutualism evolution. The genus *Ensifer* (syn. *Sinorhizobium*) contains diverse plant-associated bacteria, a subset of which can fix nitrogen in symbiosis with legumes. To gain insights into the evolution of symbiotic nitrogen fixation (SNF), and interkingdom mutualisms more generally, we performed extensive phenotypic, genomic, and phylogenetic analyses of the genus *Ensifer*. The data suggest that SNF likely emerged several times within the genus *Ensifer* through independent horizontal gene transfer events. Yet, the majority (105 of 106) of the *Ensifer* strains with the *nodABC* and *nifHDK* nodulation and nitrogen fixation genes were found within a single, monophyletic clade. Comparative genomics highlighted several differences between the “symbiotic” and “nonsymbiotic” clades, including divergences in their pangenome content. Additionally, strains of the symbiotic clade carried 325 fewer genes, on average, and appeared to have fewer rRNA operons than strains of the nonsymbiotic clade. Initial characterization of a subset of ten *Ensifer* strains identified several putative phenotypic differences between the clades. Tested strains of the nonsymbiotic clade could catabolize 25% more carbon sources, on average, than strains of the symbiotic clade, and they were better able to grow in LB medium and tolerate alkaline conditions. On the other hand, the tested strains of the symbiotic clade were better able to tolerate heat stress and acidic conditions. We suggest that these data support the division of the genus *Ensifer* into two main subgroups, as well as the hypothesis that pre-existing genetic features are required to facilitate the evolution of SNF in bacteria.

Key words: mutualism, evolutionary biology, phenomics, comparative genomics, rhizobia, Proteobacteria.

Significance

The bacterial genus *Ensifer* contains ecologically important N₂-fixing symbionts of leguminous plants, as well as nonsymbiotic species. However, the evolutionary dynamics of symbiotic nitrogen fixation within this genus are unclear, and it remains an open question of whether the gain of classical symbiotic N₂-fixation genes is sufficient to allow a bacterium to fix nitrogen. Our results suggest that the symbiotic species of the genus *Ensifer* predominately group separately from the nonsymbiotic species, but that symbiotic abilities were likely acquired multiple times within this group. This study provides new insight into the evolution of symbiotic N₂-fixation in a bacterial genus, while supporting the hypothesis that genetic features aside from the classical symbiotic N₂-fixation genes contribute to the evolution of symbiotic potential.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Symbioses are pervasive phenomena present in all Eukaryotic forms of life (López-García et al. 2017). These includes facultative symbiotic interactions, obligate symbioses, and the evolution of organelles (Douglas 2014), with symbiotic nitrogen fixation (SNF) being a paradigmatic example of the latter (Masson-Boivin and Sachs 2018). SNF (the conversion of N_2 to NH_3) is performed by a polyphyletic group of bacteria from the Alphaproteobacteria and Betaproteobacteria (whose nitrogen-fixing members are collectively called rhizobia) and members of the genus *Frankia* (Masson-Boivin et al. 2009; Wang and Young 2019) while intracellularly housed within specialized organs (nodules) of specific plants in the family *Fabaceae* and the genus *Parasponia*, as well as the actinorhizal plants (Werner et al. 2014; Griesmann et al. 2018; van Velzen et al. 2018). The advantages and evolutionary constraints to SNF have long been investigated in the conceptual framework of mutualistic interactions and the exchange of goods (see, for instance, Heath and Tiffin 2007; Werner et al. 2015; Sørensen et al. 2019), and quantitative estimations with metabolic reconstructions have also been performed (Pfau et al. 2018; diCenzo et al. 2020).

The establishment of a symbiotic nitrogen-fixing interaction requires that the bacterium encode several diverse molecular functions, including those related to signaling and metabolic exchange with the host plant, nitrogenase and nitrogenase-related functions, and escaping or resisting the plant immune system (Oldroyd et al. 2011; Haag et al. 2013; Poole et al. 2018). In general, the primary genes required for SNF (i.e., the *nod*, *nif*, and *fix* genes) are located within mobile genetic elements that include symbiotic islands and symbiotic (mega)-plasmids (Checcucci et al. 2019; Tian and Young 2019; Geddes et al. 2020), facilitating their spread through horizontal gene transfer (HGT) (Sullivan et al. 1995; Barcellos et al. 2007; Pérez Carrascal et al. 2016). Emphasizing the role of HGT in the evolution of rhizobia, rhizobia are dispersed across seven families of the Alphaproteobacteria and one family of the Betaproteobacteria, and most genera with rhizobia also contain nonrhizobia (Garrido-Oter et al. 2018; Wang 2019).

An interesting area of investigation is whether the evolution of mutualistic symbioses, such as SNF, depends on metabolic/genetic requirements ("facilitators," as in Gerhart and Kirschner [2007]) aside from the strict symbiotic genes (Long 2001; Sanjuán 2016; Zhao et al. 2018). In other words, 1) is the acquisition of symbiotic genes present in genomic islands or plasmids sufficient to become a symbiont? or, 2) are metabolic prerequisites or adaptation successive to HGT required? A comparative genomics study of 1,314 Rhizobiales genomes identified no functional difference between rhizobia and nonrhizobia based on Kyoto Encyclopedia of Genes and

Genomes annotations (Garrido-Oter et al. 2018), suggestive of an absence of obvious facilitators. In contrast, experimental studies are generally consistent with an important role of non-symbiotic genes in establishing or optimizing rhizobium–legume symbioses. Several studies have shown that effective symbionts are not produced following the transfer of symbiotic plasmids from rhizobia of the genera *Rhizobium* or *Ensifer* (syn. *Sinorhizobium*) to closely related nonrhizobia from the genera *Agrobacterium* or *Ensifer* (see, for instance, Hooykaas et al. 1982; Finan et al. 1986; Rogel et al. 2001; reviewed in diCenzo et al. [2019]). Similarly, the same symbiotic island is associated with vastly different symbiotic phenotypes depending on the *Mesorhizobium* genotype (Nandasena et al. 2007; Haskett et al. 2016). Further supporting the need for additional adaptations to support SNF, symbiosis plasmid transfer coupled to experimental evolution can lead to the gain of more advanced symbiotic phenotypes (Doin de Moura et al. 2020).

The genus *Ensifer* provides an ideal model to further explore the differentiation, or lack thereof, of symbiotic bacteria from nonsymbionts. This genus comprises rhizobia such as *Ensifer meliloti* and *Ensifer fredii*, as well as nonrhizobia like *Ensifer morelense* and *Ensifer adhaerens*, and many members have been extensively studied producing an abundant set of experimental and genomic data (for a recent review, see diCenzo et al. [2019]). The genus *Ensifer*, as currently defined, resulted from the combination of the genera *Sinorhizobium* and *Ensifer* based on similarities in the 16S rRNA and *recA* sequences of the type strains and the priority of the name *Ensifer* (Willems et al. 2003; Young 2003). Multilocus sequence analysis supported the amalgamation of these genera (Martens et al. 2007), although it was subsequently noted that *E. adhaerens* (the type strain) is an outgroup of this taxon based on whole genome phylogenomics (Ormeño-Orrillo et al. 2015). A more recent taxonomy approach based on genome phylogeny suggests that the genus *Ensifer* should again be split, with the initial type strains of *Ensifer* and *Sinorhizobium* belonging to separate genera (Parks et al. 2018).

In this article, we report an extensive comparative genomic analysis, and initial phenotypic characterization, of legume symbionts and nonsymbionts of the genus *Ensifer*. We identified that SNF likely evolved multiple times through independent HGT events; even so, most symbionts were found in a single clade, consistent with a requirement for pre-existing genetic features to facilitate the evolution of SNF. Moreover, the symbiotic and nonsymbiotic clades differed in their pangenome composition, and tests with a subset of strains suggested they also differed in their substrate utilization and resistance phenotypes as measured by the Phenotype MicroArray platform. We suggest that the data

support the division of the genus *Ensifer* into two subgroups, corresponding to the genera *Ensifer* and *Sinorhizobium* of the Genome Taxonomy Database (Parks et al. 2018).

Materials and Methods

Genome Sequencing, Assembly, and Annotation

Prior to short-read sequencing, all strains were grown to stationary phase at 30 °C in TY medium (5 g l⁻¹ tryptone, 3 g l⁻¹ yeast extract, and 0.4 g l⁻¹ CaCl₂). Total genomic DNA was isolated using a standard cetyltrimethylammonium bromide method (Perrin et al. 2015). Short-read sequencing was performed at IGATech (Udine, Italy) using an Illumina HiSeq2500 instrument with 125-bp paired-end reads. Two independent sequencing runs were performed for *E. morelense* Lc04 and *Ensifer psoraleae* CCBAU 65732, whereas *E. morelense* Lc18 and *Ensifer sesbaniae* CCBAU 65729 were sequenced once. For long-read sequencing, *E. sesbaniae* was grown to midexponential phase at 30 °C in MM9 minimal medium (MOPS buffer [40 mM MOPS, 20 mM KOH], 19.2 mM NH₄Cl, 8.76 mM NaCl, 2 mM KH₂PO₄, 1 mM MgSO₄, 0.25 mM CaCl₂, 1 μg ml⁻¹ biotin, 42 nM CoCl₂, 38 μM FeCl₃, 10 μM thiamine-HCl, and 10 mM sucrose). Total genomic DNA was isolated as described elsewhere (Cowie et al. 2006). Long-read sequencing was performed in-house with a Pacific Biosciences Sequel instrument.

Reads were assembled into scaffolds using SPAdes 3.9.0 (Bankevich et al. 2012; Vasilinetc et al. 2015); in the case of *E. sesbaniae*, long reads were corrected and trimmed using Canu 1.7.1 (Koren et al. 2017) prior to assembly. Scaffolds returned by SPAdes were parsed to remove those with <20× coverage or with a length <200 nucleotides. Using FastANI (Jain et al. 2018), one-way average nucleotide identity (ANI) values of each assembly were calculated against 887 alpha-proteobacterial genomes available through the National Center for Biotechnological Information (NCBI) with an assembly level of complete or chromosome. Based on the FastANI output, each draft genome assembly was further scaffolded using MeDuSa (Bosi et al. 2015) and the reference genomes listed in [supplementary table S1, Supplementary Material](#) online. For most assemblies, scaffolds under 1 kb in length were discarded. The exception was for *S. sesbaniae*, for which case scaffolds <10 kb were discarded. Genome assemblies were annotated using Prokka 1.12-beta (Seemann 2014), annotating coding regions with Prodigal (Hyatt et al. 2010), tRNA with Aragon (Laslett and Canbäck 2004), rRNA with Barrnap (<https://github.com/tseemann/barrnap>; last accessed October 18, 2020), and ncRNA with Infernal (Kolbe and Eddy 2011) and Rfam (Kalvari et al. 2018).

Species Phylogenetic Analyses

All *Ensifer* (and *Sinorhizobium*) genomes were downloaded from the NCBI Genome Database regardless of assembly

level. Strains that either 1) lacked a RefSeq assembly, 2) had genome sizes <1 Gb, or 3) appeared to not belong to the *Ensifer* clade based on preliminary phylogenetic analyses were discarded, leaving a final set of 157 strains ([supplementary data set S1, Supplementary Material](#) online). Eight complete *Rhizobium* genomes ([supplementary data set S2, Supplementary Material](#) online) were downloaded to serve as an outgroup. Genomes were reannotated with prokka to ensure consistent annotation. All genomes were downloaded on November 12, 2018, and associated metadata are available as [supplementary data sets S1 and S2, Supplementary Material](#) online.

To construct an unrooted, core gene phylogeny, the pan-genome of the 157 *Ensifer* strains was calculated using Roary 3.11.3 (Page et al. 2015) with a percent identify threshold of 70%. As part of the running of Roary, the nucleotide sequences of the 1,049 core genes (identified as those found in at least 99% of the genomes; [supplementary data set S3, Supplementary Material](#) online) were individually aligned with PRANK (Löytynoja 2014) and the alignments concatenated. The concatenated alignment was trimmed using TRIMAL 1.2.rev59 (Capella-Gutiérrez et al. 2009) with the automated1 option, and used to construct a maximum likelihood phylogeny (the bootstrap best tree following 100 bootstrap replicates, as determined by the extended majority-rule consensus tree criterion) using RAXML 8.2.9 (Stamatakis 2014) with the GTRCAT model as recommended (<https://cme.h-its.org/exelixis/resource/download/NewManual.pdf>; last accessed October 18, 2020). All phylogenies prepared in this study were visualized with the online iTOL webserver (Letunic and Bork 2016).

To construct a rooted phylogeny, the AMPHORA2 pipeline (Wu and Scott 2012) was used to identify 31 highly conserved bacterial proteins in each *Ensifer* and *Rhizobium* proteome, based on HMMER 3.1b2 (Eddy 2009) and the 31 hidden Markov models (HMMs) that come with AMPHORA2. Custom Perl scripts were then used to remove proteins that were either found in <95% of genomes or were found in multicopy in at least one genome, leaving a set of 30 proteins (Frr, InfC, NusA, Pgc, PyrG, RplA, RplB, RplC, RplD, RplE, RplF, RplK, RplL, RplM, RplN, RplP, RplS, RplT, RpmA, RpoB, RpsB, RpsC, RpsE, RpsI, RpsJ, RpsK, RpsM, RpsS, SmpB, Tsf). Orthologous groups were aligned using MAFFT 7.310 (Katoh and Standley 2013) with the localpair option, following which the alignments were trimmed using TRIMAL 1.2.rev59 with the automated1 option. Alignments were concatenated and used to construct a maximum likelihood phylogeny (the bootstrap best tree following 304 bootstrap replicates, as determined by the extended majority-rule consensus tree criterion) using RAXML with the PROTGAMMAJTTDCMUT model. This model was chosen as a preliminary run using RAXML with the automatic model selection indicated that the best scoring tree was obtained with the selected model.

ANI and AAI Calculations

Pairwise ANI values were calculated for all *Ensifer* strains using FastANI (Jain et al. 2018) with default parameters; a value of 78% was used in cases where no value was returned by FastANI. Pairwise average amino acid identity (AAI) values were calculated with the compareM workflow (<https://github.com/dparks1134/CompareM>; last accessed October 18, 2020). Results were visualized and clustered using the heatmap.2 function of the gplots package in R (Warnes et al. 2016), with average linkage and Pearson correlation distances.

Pangenome Calculation

All proteins of the reannotated *Ensifer* strains were clustered into orthologous groups using CD-HIT 4.6 (Li and Godzik 2006) with a percent identity threshold of 70% and an alignment length of 80% of the longer protein. The output was used to determine core and accessory genomes using a prevalence threshold of 90% as many of the genomes were draft genomes. Gene accumulation curves were produced using the specaccum function of the vegan package of R (Oksanen et al. 2018), with the random method and 500 permutations. Principal component analysis (PCA) was performed with the prcomp function of R, and was visualized with the autoplot function the ggplot2 package (Wickham 2016).

Identification and Phylogenetic Analysis of Common Nod, Nif, and Rep Proteins

The proteomes were collected for the 157 *Ensifer* strains, as well as all strains from the genera *Rhizobium*, *Neorhizobium*, *Agrobacterium*, *Mesorhizobium*, and *Ochrobactrum* with an assembly status of Complete or Chromosome (supplementary data set S4, Supplementary Material online). Additionally, the seed alignments for the HMMs of the nodulation proteins NodA (TIGR04245), NodB (TIGR04243), and NodC (TIGR04242), the nitrogenase proteins NifH (TIGR01287), NifD (TIGR01282), NifK (TIGR01286), and the replicon partitioning proteins RepA (TIGR03453), and RepB (TIGR03454) were downloaded from TIGRFAM (Haft et al. 2012). Seed alignments were converted into HMMs with the HMMBUILD function of HMMER 3.1b2 (Eddy 2009). Each HMM was searched against the complete set of proteins from all 157 reannotated *Sinorhizobium* and *Ensifer* strains using the HMMSEARCH function of HMMER. The amino acid sequences for each hit (regardless of e-value) were collected. Each set of sequences was searched against a HMM database containing all 21,200 HMMs from the Pfam (Finn et al. 2016) and TIGRFAM databases using the HMMSCAN function of HMMER, and the top scoring HMM hit for each query protein was identified. Proteins were annotated as NodA, NodB,

NodC, NifH, NifD, NifK, RepA, or RepB according to supplementary table S2, Supplementary Material online.

The *nodA*, *nodB*, and *nodC* genes are generally found as an operon. Thus, the NodA, NodB, and NodC proteins were putatively associated to operons based on identifying proteins that are encoded by adjacent genes in their respective genomes; orphan proteins not encoded by adjacent genes were discarded as the subsequent phylogenetic analysis was based on concatenated NodA, NodB, and NodC alignments. Each set of orthologs were aligned using MAFFT with the localpair option, and alignments trimmed using TRIMAL and the automated1 algorithm. Alignments were concatenated so as to combine alignments for proteins encoded by adjacent genes, producing a NodABC alignment. The same procedure was followed to produce NifHDK and RepAB alignments. Maximum likelihood phylogenies were built on the basis of each combined alignment using RAxML with the PROTGAMMAJTT (NodABC, NifHDK) or the PROTGAMMALG (RepAB) models. These models were chosen as preliminary runs using RAxML with the automatic model selection indicated that the best scoring trees were obtained with the selected models. The final phylogenies are the bootstrap best trees following 352 bootstrap replicates, as determined by the extended majority-rule consensus tree criterion.

Plant Assays

Phaseolus vulgaris (var. TopCrop, Mangani Sementi, Italy) seeds were surface sterilized in 2.5% HgCl₂ solution for 2 min and washed five times with sterile water. Seeds were germinated in the dark at 23 °C, following which seedlings were placed in sterile polypropylene jars containing vermiculite:perlite (1:1) and nitrogen-free Fåhraeus medium, and grown at 23 °C with a 12-h photoperiod (100 μE m⁻² s⁻¹). One-week-old plantlets were inoculated with 100 μl of the appropriate rhizobium strain (suspended in 0.9% NaCl at an OD₆₀₀ of 1); five plants were inoculated per strain and then grown for 4 weeks at 23 °C with a 12-h photoperiod (100 μE m⁻² s⁻¹). Plant growth assays were repeated three independent times. Nodules were collected and surface sterilized as described elsewhere (Checucci et al. 2016), crushed in sterile 0.9% NaCl solution, and serial dilutions were plated on TY agar plates and incubated at 30 °C for 2 days. PCR amplification of the 16S rRNA gene was performed using crude lysates from single colonies recovered from root nodules, as in Barzanti et al. (2007). Sequencing of the PCR amplified 16S rRNA gene was performed from both the 27f and the 1495r primers using BrilliantDye Terminator Cycle Sequencing chemistry (Nimagen, Nijmegen, The Netherlands) on a 3730xl DNA Analyzer (ThermoFisher Scientific, Waltham, MA).

Phenotype MicroArray

Phenotype MicroArray experiments using Biolog plates PM1 and PM2A (carbon sources), PM9 (osmolytes), and PM10 (pH) were performed as described previously (Biondi et al. 2009). Data were collected over 96 h with an OmniLog instrument. Data analysis was performed with DuctApe (Galardini et al. 2014). Activity index (AV) values were calculated following subtraction of the blank well from the experimental wells. Growth with each compound was evaluated with AV values from 0 (no growth) to 9 (maximal growth), following an elbow test calculation. Phenotype MicroArray experiments were performed once as results for these experiments are highly repeatable (Johnson et al. 2008; Bochner et al. 2010; Dunkley et al. 2019).

Biofilm Assays

Overnight cultures of strains grown in TY and LB (10 g l⁻¹ tryptone, 5 g l⁻¹ yeast extract, 5 g l⁻¹ NaCl) media were diluted to an OD₆₀₀ of 0.02 in fresh media, and six replicates of 100 µl aliquots were transferred to a 96-well microplate. Plates were incubated at 30 °C for 24 h, after which the OD₆₀₀ was measured with a Tecan Infinite 200 PRO (Switzerland). Each well was then stained with 30 µl of a filtered 0.1% (w/v) crystal violet solution for 10 min, and then the medium containing the planktonic cells was gently removed from the wells. Next, the wells were rinsed three times with 200 µl of phosphate-buffered saline (PBS; 0.1 M, pH 7.4) and allowed to dry for 15 min. About 100 µl of 95% (v/v) ethanol was added to each well and then incubated for 15 min at room temperature. The OD₅₄₀ of each well was measured (Rinaudi and González 2009), and biofilm production reported as the ratio of the OD₅₄₀/OD₆₀₀ ratio. Biofilm assays consisted of six replicates, and were performed two independent times.

Growth Curves

Overnight cultures of each strain were grown in the same medium to be used for the growth curve. For minimal media, either 0.2% (w/v) of glucose or succinate was added as the carbon source. Cultures were diluted to an OD₆₀₀ of 0.05 in the same media, and triplicate 150-µl aliquots were added to a 96-well microplate. Microplates were incubated without shaking at 30 °C or 37 °C in a Tecan Infinite 200 PRO, with OD₆₀₀ readings taken every hour for 48 h. Growth rates were evaluated over 2-h windows during the exponential growth phase. All growth curves were performed in triplicate and repeated two independent times.

To evaluate bacterial growth when provided root exudates as a nitrogen source, root exudates were produced from *Medicago sativa* cv. Maraviglia as described elsewhere (Checcucci et al. 2017). Single bacterial colonies from TY plates were resuspended in a 0.9% NaCl solution to an

OD₆₀₀ of 0.5. Then, each well of a 96-well microplate was inoculated with 5 µl of culture, 75 µl of nitrogen-free M9 with 0.2% (w/v) succinate as a carbon source, and 20 µl of root exudate as a nitrogen source as done previously (Checcucci et al. 2017). Triplicates were performed for each strain. Microplates were incubated without shaking at 30 °C in a Tecan Infinite 200 PRO, with OD₆₀₀ readings taken every hour for 48 h. Growth rates were determined as described above.

Results

Genome Sequencing of Four *Rhizobiaceae* Strains

Draft genomes of *E. morelense* Lc04, *E. morelense* Lc18, *E. sesbaniae* CCBAU 65729, and *E. psoraleae* CCBAU 65732 (Wang et al. 1999, 2013) were generated to increase the species diversity available for our analyses. Summary statistics of the assemblies are provided in [supplementary table S3, Supplementary Material](#) online. The genome sequences confirmed the presence of nodulation and nitrogen-fixing genes in *E. morelense* Lc18, *E. sesbaniae* CCBAU 65729, and *E. psoraleae* CCBAU 65732, whereas these genes appeared absent in the *E. morelense* Lc04 assembly. Strains Lc04, CCBAU 65729, and CCBAU 65732 were confirmed to belong to the genus *Ensifer*, as one-way ANI comparisons revealed that the most similar alpha-proteobacterial genomes were from the genus *Ensifer*. However, the genome of strain Lc18 was most similar to genomes from the genera *Rhizobium* and *Agrobacterium*, consistent with an earlier 16S rRNA gene restriction fragment length polymorphism analysis (34). Thus, we propose renaming *E. morelense* Lc18 to *Rhizobium* sp. Lc18. As this strain does not belong to the genus *Ensifer*, it was excluded from further analyses.

Symbiotic and Nonsymbiotic *Ensifer* Strains Segregate Phylogenetically

An unrooted, core gene phylogeny of 157 *Ensifer* strains was prepared to evaluate the phylogenetic relationships between the symbiotic and nonsymbiotic strains (fig. 1). A rooted phylogeny based on a multilocus sequence analysis was also prepared ([supplementary fig. S1, Supplementary Material](#) online). Each of the 157 strains was annotated as symbiotic or nonsymbiotic based on the presence of the common *nodABC* nodulation genes and the *nifHDK* nitrogenase genes. Consistent with previous work (Garrido-Oter et al. 2018), both phylogenies revealed a clear division of the symbiotic and nonsymbiotic strains into two well-defined clades. Nevertheless, a few exceptions were noted. *Ensifer sesbaniae* was found within the nonsymbiotic clade; however, *E. sesbaniae* was reported to be a symbiont of legumes such as *P. vulgaris* (Wang et al. 2013), and the ability of *E. sesbaniae* to

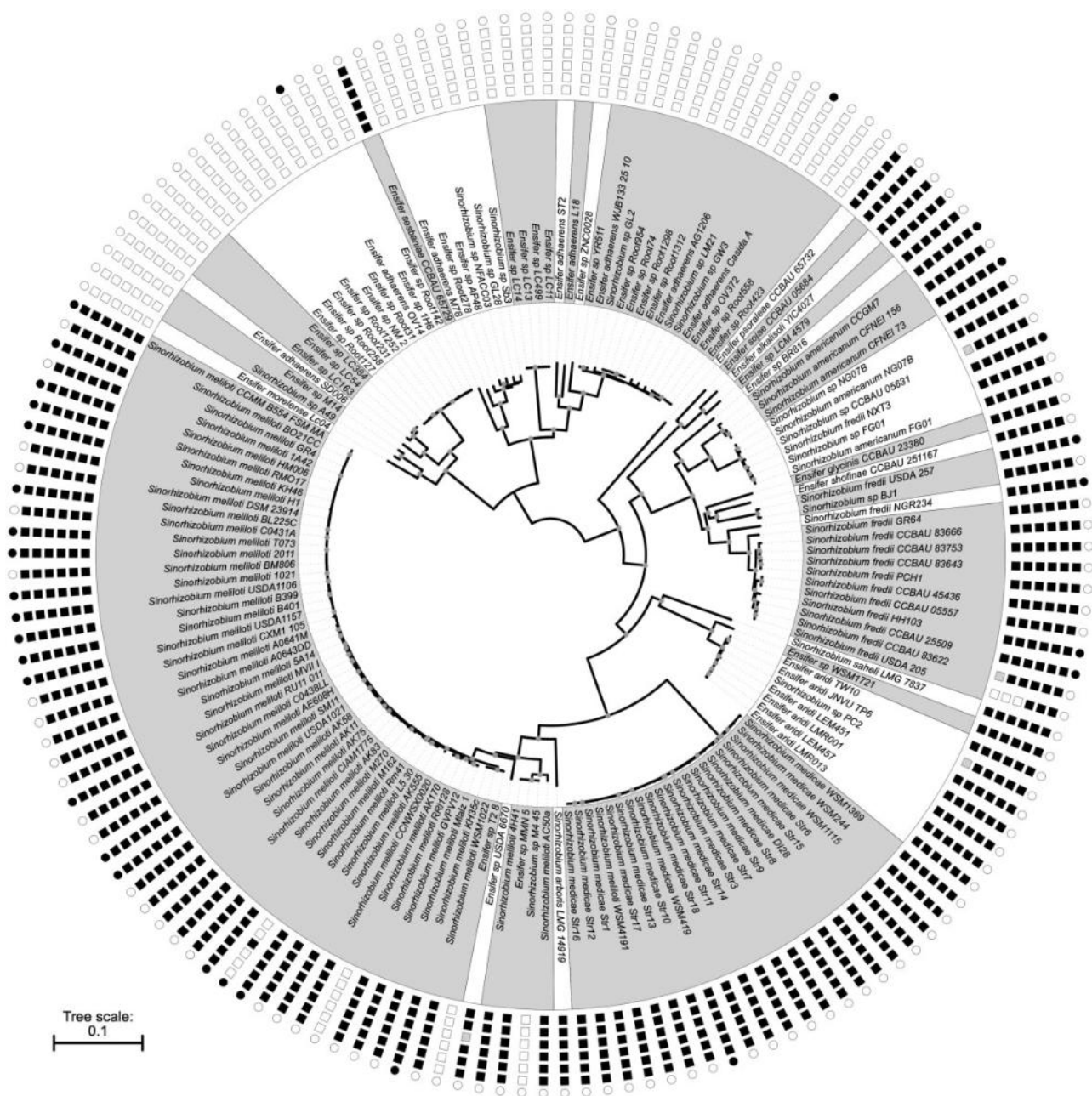


Fig. 1.—Unrooted phylogeny of the genus *Ensifer*. A maximum likelihood phylogeny of 157 strains was prepared from a concatenated alignment of 1,049 core genes. Nodes with a bootstrap value of 100 are indicated with the gray dots. The scale represents the mean number of nucleotide substitutions per site. The white and gray shading is used to group strains into genospecies on the basis of ANI and AAI results (supplementary figs. S3 and S4, Supplementary Material online), using ANI and AAI genospecies threshold of 95% and 96%, respectively. From outside to inside, rings represent the genome assembly level (black, finished; white, draft), and the presence (black) or absence (white) of *NodA*, *NodB*, *NodC*, *NifH*, *NifD*, and *NifK*. Gray boxes indicate the presence of a truncated version of the corresponding gene (as a result of incomplete genome assembly) detected through inspection of the RefSeq annotations. Strains are named as recorded in NCBI at the time of collection.

nodulate *P. vulgaris* was confirmed in this study (supplementary fig. S2, Supplementary Material online). Similarly, at least one of the six symbiotic proteins were not detected in five strains of the symbiotic group,

although we cannot rule out that these are false negatives due to incomplete genome assemblies or genome assembly errors. ANI (genospecies threshold: 95%) and AAI (genospecies threshold: 96%) calculations suggested

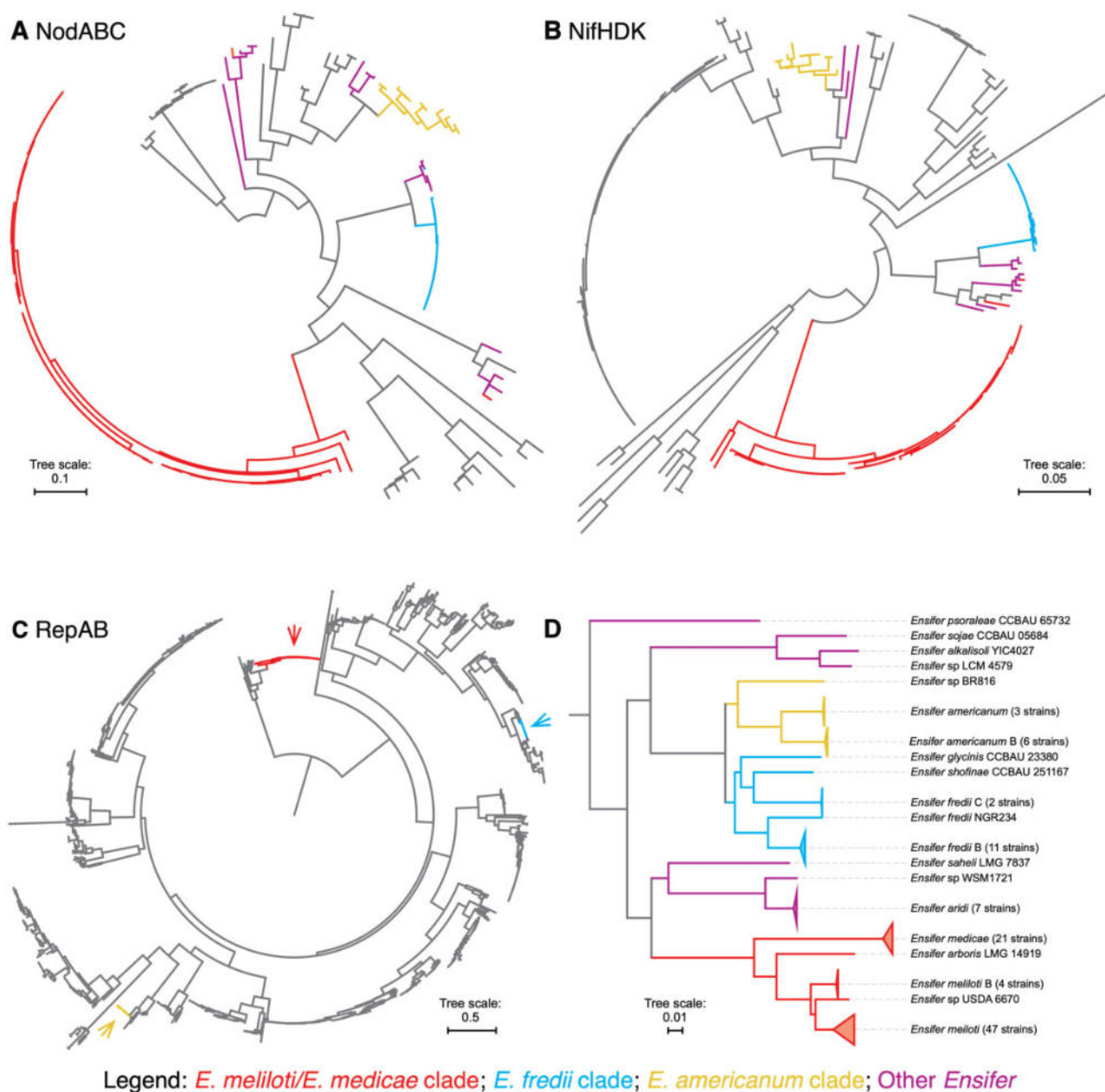


FIG. 2.—Evolution of SNF within the genus *Ensifer*. Maximum likelihood phylogenies of concatenated alignments of (A) NodABC nodulation proteins, (B) NifHDK nitrogenase proteins, and (C) RepAB replicon partitioning proteins of the order Rhizobiales. Branches corresponding to proteins from the genus *Ensifer* are indicated with color. (D) A subtree of the core gene species phylogeny of figure 1. Colors denote taxa whose symbiotic proteins are predicted to have been vertically acquired from a common ancestor. The scale bars represent the mean number of amino acid (A–C) or nucleotide (D) substitutions per site.

the presence of 12 and 20 genospecies within the nonsymbiotic and symbiotic groups, respectively (fig. 1 and supplementary figs. S3 and S4, Supplementary Material online), confirming that the nonsymbiotic clade is not an artifact of low species diversity. Thus, we conclude that the genus *Ensifer* consists of two well-defined clusters, each consisting predominately of either symbiotic or nonsymbiotic strains.

SNF Likely Arose Multiple Times within the Genus *Ensifer*

A possible explanation for the phylogenetic segregation of SNF within the genus *Ensifer* was that the symbiotic genes were gained once through a single HGT event. To test this hypothesis, the phylogenetic relationships of the NodABC and NifHDK proteins of the order Rhizobiales were examined (fig. 2A and B). SNF genes are situated on megaplasmids in

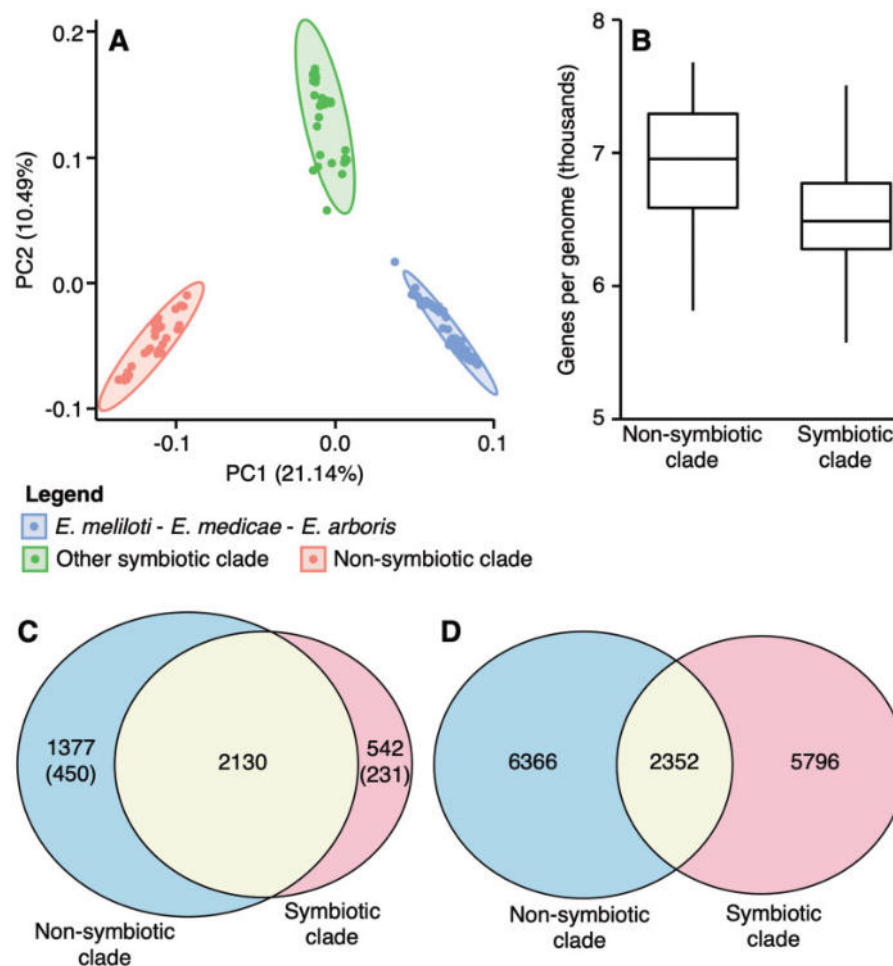


Fig. 3.—Global genome properties of the genus *Ensifer*. (A) A PCA plot based on the presence and absence of all orthologous protein groups in each of the 157 *Ensifer* strains. (B) Box-and-whisker plots displaying the number of genes per genome in the symbiotic and nonsymbiotic *Ensifer* clades. (C) A Venn diagram displaying the overlap in the core genomes of the symbiotic and nonsymbiotic *Ensifer* clades. (D) A Venn diagram displaying the overlap in the accessory genomes of the symbiotic and nonsymbiotic *Ensifer* clades.

the genus *Ensifer*; thus, a phylogeny of RepAB partitioning proteins of the order Rhizobiales was prepared as a proxy of the evolutionary relationships among the symbiotic megaplasms (fig. 2C). We predicted that the NodABC, NifHDK, and RepAB proteins of the genus *Ensifer* would form a single, monophyletic clade in each phylogeny, if the above hypothesis were true. This was not observed. Instead, all three phylogenies were inconsistent with a single origin of SNF within the genus *Ensifer* as the *Ensifer* strains were predominantly split into three clades: 1) *E. melliloti* and *Ensifer medicae*, 2) *E. fredii* and related strains, and 3) *Ensifer americanum* and related strains. As the same clades are observed in the species tree (fig. 2D), this observation suggests SNF was independently acquired through HGT in each clade. The relationships between the SNF genes of the remaining *Ensifer* species (e.g., *Ensifer aridi* and *E. psoralae*) was not clear; however, the most parsimonious solution was that there were three

additional acquisitions of SNF (fig. 2D). Overall, the phylogenetic analyses of the NodABC, NifHDK, and RepAB proteins support the hypothesis that there were multiple, independent acquisition of symbiosis genes (hence SNF) by lineages within the genus *Ensifer*; however, the gain (and/or maintenance) of symbiosis genes preferentially occurred within one monophyletic group of species.

The Genomic Features of the Symbiotic and Nonsymbiotic Clades Differ

The pangenome of the 157 *Ensifer* strains was calculated to evaluate if there were global genomic differences between the symbiotic and nonsymbiotic clades. Both clades had open pangenomes (supplementary fig. S5, Supplementary Material online). A PCA based on gene presence/absence revealed a clear separation of the two clades (fig. 3A), suggesting a

Table 1

Ensifer Strains Phenotypically Characterized in This Study

Strain	Original Source	SNF ^a	Ensifer Clade ^b	Reference
<i>Ensifer adhaerens</i> Casida A	Isolated from a Pennsylvania soil sample	No	Nonsymbiotic	Casida (1982)
<i>Ensifer adhaerens</i> OV14	Isolated from the rhizosphere of <i>Brassica napus</i>	No	Nonsymbiotic	Wendt et al. (2012)
<i>Ensifer</i> sp. M14	Isolated from arsenic-rich sediments of a gold mine	No	Nonsymbiotic	Drewniak et al. (2008)
<i>Ensifer morelense</i> Lc04	Isolated from root nodules of <i>Leucaena leucocephala</i>	No	Nonsymbiotic	Wang et al. (1999)
<i>Ensifer sesbaniae</i> CCBAU 65729	Isolated from root nodules of <i>Sesbania cannabina</i>	Yes	Nonsymbiotic	Wang et al. (2013)
<i>Ensifer fredii</i> NGR234	Isolated from root nodules of <i>Lablab purpureus</i>	Yes	Symbiotic	Trinick (1980)
<i>Ensifer sojae</i> CCBAU 05684	Isolated from root nodules of <i>Glycine max</i> grown in saline-alkaline soils	Yes	Symbiotic	Li et al. (2011)
<i>Ensifer americanum</i> CFNEI 156	Isolated from root nodules of <i>Acacia acatensis</i>	Yes	Symbiotic	Toledo et al. (2003)
<i>Ensifer psoraleae</i> CCBAU 65732	Isolated from root nodules of <i>Psoralea corylifolia</i>	Yes	Symbiotic	Wang et al. (2013)
<i>Ensifer medicae</i> WSM419	Isolated from root nodules of <i>Medicago murex</i>	Yes	Symbiotic	Howieson and Ewing (1986)

^aThis column indicates if the strain can (Yes) or cannot (No) form nitrogen-fixing nodules on legumes.

^bThis column indicates if the strain belongs to the symbiotic or nonsymbiotic clade of the genus *Ensifer* as defined in this study.

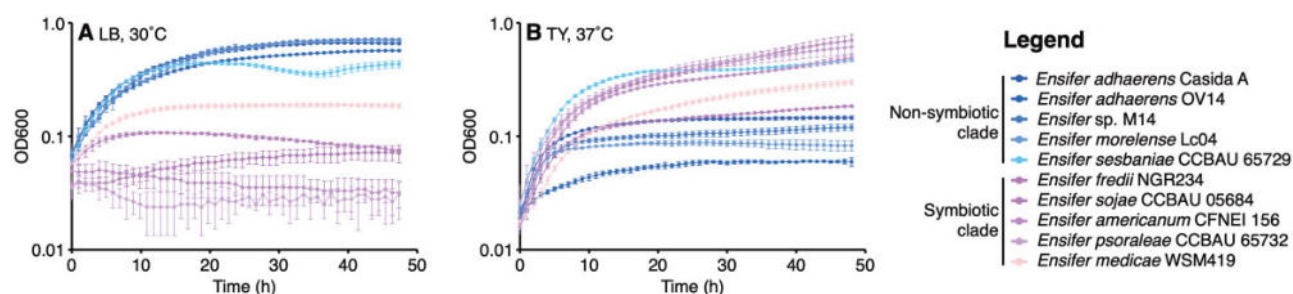


Fig. 4.—Growth properties of phylogenetically diverse *Ensifer* strains. *Ensifer* strains were grown in microplates without shaking. Data points represent the average of triplicate samples, whereas the error bars indicate the SD. Shades of pink are used for strains of the symbiotic clade, whereas shades of blue are used for strains of the nonsymbiotic clade. (A) Growth in LB medium at 30°C. (B) Growth in TY medium during heat stress (37°C).

divergence of the pangenomes of these clades. The symbiotic clade was subdivided into two groups along the second component of the PCA (fig. 3A), which may suggest further levels of genomic separation. About 2,130 genes were found in the core genomes of both clades, whereas 20% (542 genes) and 40% (1,377 genes) of the core genomes of the symbiotic and nonsymbiotic clades, respectively, were absent from the core genome of the other clade; of these, about a third were completely absent from the other clade’s pangenome (fig. 3C). Of the 14,514 accessory genes (defined as genes found in at least 10% of at least one clade, excluding the 2,130 *Ensifer* core genes), only 2,352 (16%) were found in the pangenomes of both the symbiotic and the nonsymbiotic clades. Moreover, a statistically significant difference (Wilcoxon rank-sum test, $P < 0.0001$) in the genome sizes of the two clades was observed (fig. 3B); strains of the nonsymbiotic clade carried 325 more genes, on average, than strains of the symbiotic clade (median difference of 470). Finally, based on the limited number of strains with finished genomes, strains of the symbiotic clade appear to generally have three copies of the rRNA operon whereas strains of the nonsymbiotic clade appear to have a norm of five copies of their rRNA operon. Together, these multiple lines of data are

consistent with there being a broad genomic divergence of the symbiotic and nonsymbiotic clades of the genus *Ensifer*.

Phenotypic Features of the Symbiotic and Nonsymbiotic Clades Differ

A subset of ten strains (table 1), five each from the symbiotic and nonsymbiotic clades, were subjected to a panel of assays to investigate how phenotypes vary across the genus *Ensifer*. These ten strains were chosen so as to provide broad phylogenetic coverage of the genus, while excluding strains for which extensive phenotypic characterizations have been previously published. No statistically significant differences were observed in the ability of members of the two clades to form biofilms (supplementary fig. S6 and table S4, Supplementary Material online). However, the tested strains clearly differed in their ability to grow in LB media; whereas the tested strains of the nonsymbiotic clade displayed robust growth in LB, the tested strains of the symbiotic clade largely failed to grow (fig. 4A). Tested strains of the nonsymbiotic clade also displayed a slightly faster specific growth rate, on average, than the tested strains of the symbiotic clade in TY media (nonsymbiotic clade: $0.54 \pm 0.03 \text{ h}^{-1}$; symbiotic clade:

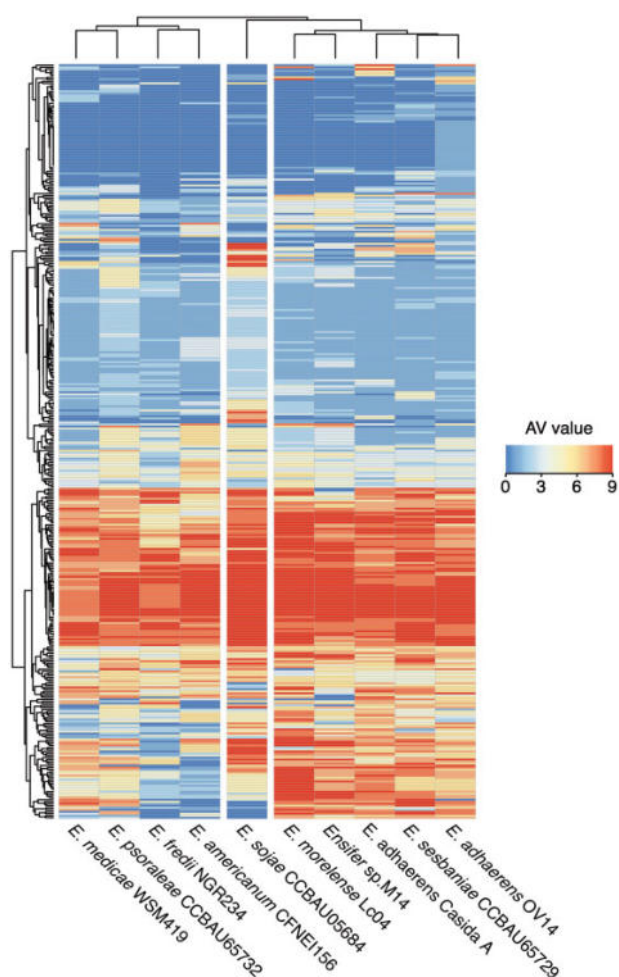


FIG. 5.—Phenotypic properties of phylogenetically diverse *Ensifer* strains. Ten *Ensifer* strains were screened for their ability to catabolize 190 carbon sources, and to grow in 96 osmolyte and 96 pH conditions using Biolog Phenotype MicroArray plates PM1, PM2, PM9, and PM10. Growth in each well was summarized on a scale of 0 (dark blue) through 9 (dark red), with higher numbers representing more robust growth. A larger version of this figure, in which each condition is labeled along the Y axis, is provided as [supplementary figure S8, Supplementary Material online](#).

$0.44 \pm 0.08 \text{ h}^{-1}$; $P=0.03$ from an ANOVA followed by Tukey's test; [supplementary fig. S7A and table S5, Supplementary Material online](#)). On the other hand, tested strains of the symbiotic clade were, on average, better able to withstand heat stress (37°C) in TY media ([fig. 4B](#)). No statistically significant difference in the average ability of the tested strains of the two clades to grow in minimal media with succinate or glucose as a carbon source, or with *M. sativa* (a symbiotic partner of *E. meliloti* and *E. medicae*) root exudates as a nitrogen source, was detected ([supplementary fig. S7 and table S5, Supplementary Material online](#)).

The phenotypic properties of the genus *Ensifer* were further examined through evaluating the ability of the same ten

strains to catabolize 190 carbon sources, and to grow in 96 osmolyte and 96 pH conditions, through the use of Biolog Phenotype MicroArrays. Clustering the strains based on growth properties largely separated the tested strains of the symbiotic clade and nonsymbiotic clade into distinct groups ([fig. 5 and supplementary fig. S8, Supplementary Material online](#)). The exception was *Ensifer sojae*, which formed its own intermediate group in the phenotype data. To aid in identifying which conditions best separate the tested strains of the symbiotic clade (including *E. sojae*) from the tested strains of the nonsymbiotic clade, a linear discriminant analysis (LDA) was run over the AV values summarizing growth in each condition ([supplementary data set S5, Supplementary Material online](#)). In general, tested strains from the nonsymbiotic clade better tolerated high pH (pH 9.0–9.5) than did the tested strains from the symbiotic cluster. In contrast, tested strains of the symbiotic clade had better tolerance to low pH conditions (pH 3.5–4.5). In addition, the tested strains from the two clades clearly differed in their overall metabolic abilities with tested strains of the nonsymbiotic clade generally having a broader metabolic capacity than those of the symbiotic clade ([supplementary table S6, Supplementary Material online](#)). Whereas tested strains of the symbiotic clade displayed robust growth on 65 carbon sources on average, the tested strains of the nonsymbiotic clade grew on an average of 81 carbon sources ($P < 0.05$, Student's *t*-test). Overall, these initial experiments provide support for the hypothesis that a variety of phenotypes, not just the ability to nodulate legumes, differ between the symbiotic and nonsymbiotic clades of the genus *Ensifer*.

Discussion

We investigated the evolution of SNF within the genus *Ensifer*, which includes a mix of nitrogen-fixing bacteria and bacteria unable to fix nitrogen, as a model for the evolution of interkingdom mutualisms. Our results indicate that, despite SNF having likely evolved multiple times within the genus *Ensifer*, the symbiotic and nonsymbiotic strains are largely separated into two phylogenetic clades reminiscent of the general division of pathogenic and environmental strains between the genera *Burkholderia* and *Paraburkholderia* (Sawana et al. 2014). Although it is possible that this result will fail to remain true as more genome sequences are published, we believe the result to be robust as the current set of genome sequences are of strains isolated from diverse legumes and other diverse environments including the rhizosphere, pristine caves, and an abandoned mine. In addition to the prevalence of SNF, the two clades differ with respect to their genomic composition (pangenome content and genome size) and phenotypic properties (metabolic capacity and stress tolerance) based on initial tests of a subset of strains. There have been several revisions to the taxonomy of the genus *Ensifer*. Recently, a

genome-based approach was proposed to standardize bacterial taxonomy (Parks et al. 2018) that splits the genus *Ensifer* into two genera: *Sinorhizobium* and *Ensifer*. The symbiotic and nonsymbiotic clades identified here correspond with the genera *Sinorhizobium* and *Ensifer*, respectively, supporting the proposal to divide the genus *Ensifer* into two genera.

Our analyses revealed a complex evolutionary history of SNF within the genus *Ensifer*. In addition to SNF emerging a predicted six times or more, we detected possible losses of SNF and allele switches. Between one and six of the NodABC and NifHDK proteins were not detected in five of the strains in the symbiotic clade (fig. 1). Although this may be indicative of multiple losses of symbiosis, we cannot rule out that these are false negatives due to incomplete genome assemblies or genome assembly errors; five of the six genomes were draft genomes, and the one strain with a complete genome (*E. meliloti* M162) can nodulate ten of 27 tested *Medicago truncatula* genotypes suggesting it does contain *nod* and *nif* genes (Sugawara et al. 2013). Based on the RepAB phylogeny (fig. 2), the symbiotic megaplasmid of *Ensifer arboris* likely shares common ancestry with the symbiotic megaplasmids of the sister species *E. meliloti* and *E. medicae*. Yet, the nodulation and nitrogen fixation genes appeared distinct. Thus, we hypothesize that there was a recent replacement of the symbiotic genes in *E. arboris*, or alternatively, in *E. meliloti* and *E. medicae*. This hypothesis is supported by the observation that the host ranges of these rhizobia differ; unlike *E. meliloti* and *E. medicae*, *E. arboris* does not nodulate plants of the genus *Medicago* (Zhang et al. 1991).

The reason for the phylogenetic bias in the evolution of SNF within the genus *Ensifer* remains unclear, especially considering that strains from both clades are plant-associated (Bai et al. 2015). One hypothesis is that each clade occupies distinct niches within the soil and plant-associated environments. Indeed, analysis of a subset of strains suggested that species of the nonsymbiotic clade have a broader metabolic capacity (supplementary table S6, Supplementary Material online), as well as a larger average genome size (fig. 3B), which is consistent with these species being more capable of adapting to fluctuating nutritional environments. This is further supported by the apparently higher number of rRNA operons in strains of the nonsymbiotic clade, which is generally thought to allow bacteria to more quickly respond to changing nutrient conditions (Stevenson and Schmidt 2004; Roller et al. 2016). Moreover, the nonsymbiotic clade could be differentiated from the symbiotic clade based on its pangenome content (fig. 3A), which leads us to hypothesize that strains of these clades acquire genes from distinct gene pools, further supporting the hypothesis that they belong to distinct gene-cohesive groups and ecological niches. This hypothesis may then be interpreted in the framework of the stable ecotype model (sensu Cohan 2006), where the symbiotic and nonsymbiotic clades represent two, ecologically distinct and

monophyletic groups and where periodic selection events (e.g., fitness for SNF) are recurrent.

An alternate, but not mutually exclusive, hypothesis is that the symbiotic *Ensifer* clade contains “facilitator” genes required to support SNF, similar to the theory that ancestral legumes contained a genetic “predisposition” necessary for the eventual evolution of rhizobium symbioses (Soltis et al. 1995; Doyle 2011; Werner et al. 2014). Conversely, evolution of SNF may also require the absence of “inhibitor” genes, such as the absence of virulence factors (Marchetti et al. 2010). As we did not evaluate cause-and-effect relationships, our data set does not definitely address these hypotheses. However, we observed numerous genotypic and likely phenotypic differences between the symbiotic and nonsymbiotic clade, providing some support for these hypotheses. For example, the tested strains of the symbiotic clade appeared to have higher tolerance to low pH (supplementary fig. S8 and data set S5, Supplementary Material online), which is notable as the curled root hair is an acidic environment (Hawkins et al. 2017). At the genomic level, 231 of the core genes of the symbiotic clade were absent from the pangenome of the nonsymbiotic clade and thus are good candidates as possible facilitators and follow-up studies. However, facilitators and inhibitors could also take the form of polymorphisms within highly conserved genes, as shown for *bacA* and the *Sinorhizobium*–*Medicago* symbiosis (diCenzo et al. 2017).

In summary, we show that the legume symbionts and nonsymbionts of the genus *Ensifer* are largely segregated into two phylogenetically distinct clades that differ in their genomic and phenotypic properties. We suggest that these observations, which follow the guidelines recently reported for rhizobia and agrobacteria (de Lajudie et al. 2019), support the division of the genus into two genera: *Ensifer* for the nonsymbiotic clade and *Sinorhizobium* for the symbiotic clade. However, formal descriptions and publication of the genera in the International Journal of Systematic and Evolutionary Microbiology (IJSSEM) are still required. We also provide evidence that SNF genes were likely acquired several independent times within this genus, but predominately within one monophyletic clade. These observations suggest that the presence or absence of other genomic features (“facilitators” or “inhibitors”) aside from the core symbiotic genes could be required for the establishment of an effective symbiosis. This suggestion is supported by the ability to differentiate the strains of the two clades based on their pangenome content and, at least for the tested subset, their phenotypic properties. However, as cause-and-effect relationships were not examined, follow-up study is required to more directly test this facilitators hypothesis.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We are grateful to E. Martínez-Romero for providing strains *Ensifer morelense* Lc04 and *E. morelense* Lc18, to E. Mullins (Teagasc, MTA2018233) for *Ensifer adhaerens* OV14, to C.-F. Tian for *Sinorhizobium fredii* NGR 234, and to L. Dziejwit for *Ensifer* sp. M14. A.M. was supported by the Fondazione Cassa di Risparmio di Firenze (Grant No. 18204, 2017.0719), by the “MICRO4Legumes” grant (Italian Ministry of Agriculture), and by the grant “Dipartimento di Eccellenza 2018–2022” from the Italian Ministry of Education, University and Research (MIUR). L.C. was supported by the MICRO4Legumes grant (Italian Ministry of Agriculture). G.C.D. was supported by a postdoctoral fellowship from the Natural Science and Engineering Research Council of Canada (NSERC), funding from Queen’s University, and a NSERC Discovery Grant.

Data Availability

Scripts to repeat the computational analyses reported in this study are available at https://github.com/diCenzo-GC/Ensifer_phylogenomics (last accessed October 18, 2020).

Literature Cited

- Bai Y, et al. 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* 528(7582):364–369.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477.
- Barcellos FG, Menna P, Batista JS, da S, Hungria M. 2007. Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium (Ensifer) fredii* and *Bradyrhizobium elkanii* in a Brazilian savannah soil. *Appl Environ Microbiol*. 73(8):2635–2643.
- Barzanti R, et al. 2007. Isolation and characterization of endophytic bacteria from the nickel hyperaccumulator plant *Alyssum bertolonii*. *Microb Ecol*. 53(2):306–316.
- Biondi EG, et al. 2009. Metabolic capacity of *Sinorhizobium (Ensifer) meliloti* strains as determined by phenotype MicroArray analysis. *Appl Environ Microbiol*. 75(16):5396–5404.
- Bochner B, Gomez V, Ziman M, Yang S, Brown SD. 2010. Phenotype microarray profiling of *Zymomonas mobilis* ZM4. *Appl Biochem Biotechnol*. 161(1–8):116–123.
- Bosi E, et al. 2015. MeDuSa: a multi-draft based scaffold. *Bioinformatics* 31(15):2443–2451.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
- Casida LE. 1982. *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Evol Microbiol*. 32:339–345.
- Checucci A, et al. 2016. Mixed nodule infection in *Sinorhizobium meliloti*–*Medicago sativa* symbiosis suggest the presence of cheating behavior. *Front Plant Sci*. 7:835.
- Checucci A, et al. 2017. Role and regulation of ACC deaminase gene in *Sinorhizobium meliloti*: is it a symbiotic, rhizospheric or endophytic gene? *Front Genet*. 8:6.
- Checucci A, diCenzo GC, Perrin E, Bazzicalupo M, Mengoni A. 2019. Genomic diversity and evolution of rhizobia. In: Das S, Dash HR, editors. *Microbial diversity in the genomic era*. London, United Kingdom: Academic Press. p. 37–46.
- Cohan FM. 2006. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos Trans R Soc B*. 361(1475):1985–1996.
- Cowie A, et al. 2006. An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. *Appl Environ Microbiol*. 72(11):7156–7167.
- de Lajudie PM, et al. 2019. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *Int J Syst Evol Microbiol*. 69(7):1852–1863.
- diCenzo GC, Tesi M, Pfau T, Mengoni A, Fondi M. 2020. Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat Commun* 11:2574.
- diCenzo GC, et al. 2019. Multi-disciplinary approaches for studying rhizobium–legume symbioses. *Can J Microbiol*. 65(1):1–33.
- diCenzo GC, Zamani M, Ludwig HN, Finan TM. 2017. Heterologous complementation reveals a specialized activity for BacA in the *Medicago*–*Sinorhizobium meliloti* symbiosis. *Mol Plant Microbe Interact*. 30(4):312–324.
- Doin de Moura GG, Remigi P, Masson-Boivin C, Capela D. 2020. Experimental evolution of legume symbionts: what have we learnt? *Genes* 11(3):339.
- Douglas AE. 2014. Symbiosis as a general principle in eukaryotic evolution. *Cold Spring Harb Perspect Biol*. 6(2):a016113.
- Doyle JJ. 2011. Phylogenetic perspectives on the origins of nodulation. *Mol Plant Microbe Interact*. 24(11):1289–1295.
- Drewniak L, Matlakowska R, Sklodowska A. 2008. Arsenite and arsenate metabolism of *Sinorhizobium* sp. M14 living in the extreme environment of the Zloty Stok gold mine. *Geomicrobiol J*. 25(7–8):363–370.
- Dunkley EJ, Chalmers JD, Cho S, Finn TJ, Patrick WM. 2019. Assessment of phenotype microarray plates for rapid and high-throughput analysis of collateral sensitivity networks. *PLoS One* 14(12):e0219879.
- Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform*. 23:205–211.
- Finan TM, Kunkel B, De Vos GF, Signer ER. 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J Bacteriol*. 167(1):66–72.
- Finn RD, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44(D1):D279–D285.
- Galardini M, et al. 2014. DuctApe: a suite for the analysis and correlation of genomic and OmniLog™ phenotype microarray data. *Genomics* 103(1):1–10.
- Garrido-Oter R, et al. 2018. Modular traits of the Rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host Microbe*. 24(1):155–167.
- Geddes BA, Kearsley J, Morton R, diCenzo GC, Finan TM. 2020. The genomes of rhizobia. In: Frenzo P, Frugier F, Masson-Boivin C, editors. *Regulation of nitrogen-fixing symbioses in legumes*. Vol. 94. London, United Kingdom: Academic Press. p. 213–249.
- Gerhart J, Kirschner M. 2007. The theory of facilitated variation. *Proc Natl Acad Sci U S A*. 104(Suppl 1):8582–8589.
- Griesmann M, et al. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*. 361(6398):eaat1743.
- Haag AF, et al. 2013. Molecular insights into bacteroid development during *Rhizobium*–legume symbiosis. *FEMS Microbiol Rev*. 37(3):364–383.
- Haft DH, et al. 2012. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res*. 41(D1):D387–D395.

- Haskett TL, et al. 2016. Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc Natl Acad Sci U S A*. 113(43):12268–12273.
- Hawkins JP, Geddes BA, Oresnik IJ. 2017. Succinoglycan production contributes to acidic pH tolerance in *Sinorhizobium meliloti* Rm1021. *Mol Plant Microbe Interact*. 30(12):1009–1019.
- Heath KD, Tiffin P. 2007. Context dependence in the coevolution of plant and rhizobial mutualists. *Proc R Soc B*. 274(1620):1905–1912.
- Hooykaas PJ, Snijderwint FG, Schilperoort RA. 1982. Identification of the Sym plasmid of *Rhizobium leguminosarum* strain 1001 and its transfer to and expression in other rhizobia and *Agrobacterium tumefaciens*. *Plasmid* 8(1):73–82.
- Howieson JG, Ewing MA. 1986. Acid tolerance in the *Rhizobium meliloti*–*Medicago symbiosis*. *Aust J Agric Res*. 37(1):55–64.
- Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11(1):119.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 9(1):7200.
- Johnson DA, et al. 2008. High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. *PLoS Genet*. 4(10):e1000211.
- Kalvari I, et al. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*. 46(D1):D335–D342.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kolbe DL, Eddy SR. 2011. Fast filtering for RNA homology search. *Bioinformatics* 27(22):3102–3109.
- Koren S, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 27(5):722–736.
- Laslett D, Canbäck B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res*. 32(1):11–16.
- Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 44(W1):W242–W245.
- Li QQ, et al. 2011. *Ensifer sojae* sp. nov., isolated from root nodules of *Glycine max* grown in saline-alkaline soils. *Int J Syst Evol Microbiol*. 61(Pt 8):1981–1988.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
- Long SR. 2001. Genes and signals in the rhizobium–legume symbiosis. *Plant Physiol*. 125(1):69–72.
- López-García P, Eme L, Moreira D. 2017. Symbiosis in eukaryotic evolution. *J Theor Biol*. 434:20–33.
- Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell D, editor. *Methods in molecular biology (methods and protocols)*. Vol. 1079. Totowa (NJ): Humana Press. p. 155–170.
- Marchetti M, et al. 2010. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol*. 8(1):e1000280.
- Martens M, et al. 2007. Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol*. 57(3):489–503.
- Masson-Boivin C, Giraud E, Perret X, Batut J. 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol*. 17(10):458–466.
- Masson-Boivin C, Sachs JL. 2018. Symbiotic nitrogen fixation by rhizobia—the roots of a success story. *Curr Opin Plant Biol*. 44:7–15.
- Nandasena KG, O'Hara GW, Tiwari RP, Sezmiş E, Howieson JG. 2007. *In situ* lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ Microbiol*. 9(10):2496–2511.
- Oksanen J, et al. 2018. vegan: community ecology package. R package version 2.5-3. Available from: <https://CRAN.R-project.org/package=vegan>. Accessed October 18, 2020.
- Oldroyd GED, Murray JD, Poole PS, Downie JA. 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet*. 45(1):119–144.
- Ormeño-Orrillo E, et al. 2015. Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics. *Syst Appl Microbiol*. 38(4):287–291.
- Page AJ, et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22):3691–3693.
- Parks DH, et al. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 36(10):996–1004.
- Pérez Carrascal OM, et al. 2016. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. *Environ Microbiol*. 18(8):2660–2676.
- Perrin E, et al. 2015. Genomes analysis and bacteria identification: the use of overlapping genes as molecular markers. *J Microbiol Methods*. 117:108–112.
- Pfau T, et al. 2018. The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci Rep*. 8(1):12504.
- Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol*. 16(5):291–303.
- Rinaudi LV, González JE. 2009. The low-molecular-weight fraction of exopolysaccharide II from *Sinorhizobium meliloti* is a crucial determinant of biofilm formation. *J Bacteriol*. 191(23):7216–7224.
- Rogel MA, Hernández-Lucas I, Kuykendall LD, Balkwill DL, Martínez-Romero E. 2001. Nitrogen-fixing nodules with *Ensifer adhaerens* harboring *Rhizobium tropici* symbiotic plasmids. *Appl Environ Microbiol*. 67(7):3264–3268.
- Roller BRK, Stoddard SF, Schmidt TM. 2016. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat Microbiol*. 1(11):16160.
- Sanjuán J. 2016. Towards the minimal nitrogen-fixing symbiotic genome. *Environ Microbiol*. 18(8):2292–2294.
- Sawana A, Adeolu M, Gupta RS. 2014. Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet*. 5:429.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
- Soltis DE, et al. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proc Natl Acad Sci U S A*. 92(7):2647–2651.
- Sørensen M, et al. 2019. The role of exploitation in the establishment of mutualistic microbial symbioses. *FEMS Microbiol Lett*. 366(12):fnz148.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stevenson BS, Schmidt TM. 2004. Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl Environ Microbiol*. 70(11):6670–6677.
- Sugawara M, et al. 2013. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol*. 14(2):R17.
- Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. 1995. Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci U S A*. 92(19):8985–8989.
- Tian CF, Young JPW. 2019. Genomics and evolution of rhizobia. In: Wang ET, Tian CF, Chen WF, Young JPW, Chen WX,

- editors. Ecology and evolution of rhizobia: principles and applications. Singapore: Springer. p. 103–119.
- Toledo I, Lloret L, Martínez-Romero E. 2003. *Sinorhizobium americanus* sp. nov., a new *Sinorhizobium* species nodulating native *Acacia* spp. in Mexico. *Syst Appl Microbiol*. 26(1):54–64.
- Trinick MJ. 1980. Relationships amongst the fast-growing rhizobia of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* spp., *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with other rhizobial groups. *J Appl Bacteriol*. 49(1):39–53.
- van Velzen R, et al. 2018. Comparative genomics of the nonlegume *Parasponia* reveals insights into evolution of nitrogen-fixing rhizobium symbioses. *Proc Natl Acad Sci U S A*. 115:E4700–E4709.
- Vasilinetc I, Prijbelski AD, Gurevich A, Korobeynikov A, Pevzner PA. 2015. Assembling short reads from jumping libraries with large insert sizes. *Bioinformatics* 31(20):3262–3268.
- Wang ET. 2019. Current systematics of rhizobia. In: Wang ET, Tian CF, Chen WF, Young JPW, Chen WX, editors. Ecology and evolution of rhizobia: principles and applications. Singapore: Springer. p. 41–102.
- Wang ET, Romero JM, Romero EM. 1999. Genetic diversity of rhizobia from *Leucaena leucocephala* nodules in Mexican soils. *Mol Ecol*. 8(5):711–724.
- Wang ET, Young JPW. 2019. History of rhizobial taxonomy. In: Wang ET, Tian CF, Chen WF, Young JPW, Chen WX, editors. Ecology and evolution of rhizobia: principles and applications. Singapore: Springer. p. 23–39.
- Wang YC, et al. 2013. Proposal of *Ensifer psoraleae* sp. nov., *Ensifer sesbaniae* sp. nov., *Ensifer morelense* comb. nov. and *Ensifer americanum* comb. nov. *Syst Appl Microbiol*. 36(7):467–473.
- Warnes GR, et al. 2016. gplots: various R programming tools for plotting data. R package version 3.0.1. Available from: <https://CRAN.R-project.org/package=gplots>. Accessed October 18, 2020.
- Wendt T, Doohan F, Mullins E. 2012. Production of *Phytophthora infestans*-resistant potato (*Solanum tuberosum*) utilising *Ensifer adhaerens* OV14. *Transgenic Res*. 21(3):567–578.
- Werner GDA, Cornwell WK, Cornelissen JHC, Kiers ET. 2015. Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. *Proc Natl Acad Sci U S A*. 112(33):10262–10269.
- Werner GDA, Cornwell WK, Sprent J, Kattge J, Kiers ET. 2014. A single evolutionary innovation drives the deep evolution of symbiotic N₂-fixation in angiosperms. *Nat Commun*. 5(1):4087.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.
- Willems A, et al. 2003. Description of new *Ensifer* strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an opinion. *Int J Syst Evol Microbiol*. 53(4):1207–1217.
- Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28(7):1033–1034.
- Young JM. 2003. The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination '*Sinorhizobium adhaerens*' (Casida 1982) Willems et al. 2003 legitimate? Request for an opinion. *Int J Syst Evol Microbiol*. 53:2107–2110.
- Zhang X, Harper R, Karisto M, Lindström K. 1991. Diversity of *Rhizobium* bacteria isolated from the root nodules of leguminous trees. *Int J Syst Evol Microbiol*. 41(1):104–113.
- Zhao R, et al. 2018. Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. *ISME J*. 12(1):101–111.

Associate editor: Esperanza Martínez-Romero

Article

Genomic and Biotechnological Characterization of the Heavy-Metal Resistant, Arsenic-Oxidizing Bacterium *Ensifer* sp. M14

George C diCenzo ^{1,*}, Klaudia Debiec ^{2,*}, Jan Krzysztoforski ³, Witold Uhrynowski ², Alessio Mengoni ¹, Camilla Fagorzi ¹, Adrian Gorecki ⁴, Lukasz Dziewit ⁴, Tomasz Bajda ⁵, Grzegorz Rzepa ⁵ and Lukasz Drewniak ²

¹ Laboratory of Microbial Genetics, Department of Biology, University of Florence, via Madonna del Piano 6, 50019 Sesto Fiorentino, Italy; alessio.mengoni@unifi.it (A.M.); camilla.fagorzi@unifi.it (C.F.)

² Laboratory of Environmental Pollution Analysis, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland; w.uhrynowski@biol.uw.edu.pl (W.U.); ldrewniak@biol.uw.edu.pl (L.Dr.)

³ Faculty of Chemical and Process Engineering, Warsaw University of Technology, Warynskiego 1, 00-645 Warsaw, Poland; jan.krzysztoforski@pw.edu.pl

⁴ Department of Bacterial Genetics, Institute of Microbiology, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland; agorecki@biol.uw.edu.pl (A.G.); ldziewit@biol.uw.edu.pl (L.Dz.)

⁵ Department of Mineralogy, Petrography and Geochemistry, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Mickiewicza 30, 30-059 Krakow, Poland; bajda@agh.edu.pl (T.B.); gprzepa@cyf-kr.edu.pl (G.R.)

* Correspondence: georgecolin.dicenzo@unifi.it (G.C.d.); k.debiec@biol.uw.edu.pl (K.D.)

† These authors contributed equally to this paper.

Received: 14 June 2018; Accepted: 25 July 2018; Published: 27 July 2018



Abstract: *Ensifer* (*Sinorhizobium*) sp. M14 is an efficient arsenic-oxidizing bacterium (AOB) that displays high resistance to numerous metals and various stressors. Here, we report the draft genome sequence and genome-guided characterization of *Ensifer* sp. M14, and we describe a pilot-scale installation applying the M14 strain for remediation of arsenic-contaminated waters. The M14 genome contains 6874 protein coding sequences, including hundreds not found in related strains. Nearly all unique genes that are associated with metal resistance and arsenic oxidation are localized within the pSinA and pSinB megaplasmids. Comparative genomics revealed that multiple copies of high-affinity phosphate transport systems are common in AOBs, possibly as an As-resistance mechanism. Genome and antibiotic sensitivity analyses further suggested that the use of *Ensifer* sp. M14 in biotechnology does not pose serious biosafety risks. Therefore, a novel two-stage installation for remediation of arsenic-contaminated waters was developed. It consists of a microbiological module, where M14 oxidizes As(III) to As(V) ion, followed by an adsorption module for As(V) removal using granulated bog iron ores. During a 40-day pilot-scale test in an abandoned gold mine in Zloty Stok (Poland), water leaving the microbiological module generally contained trace amounts of As(III), and dramatic decreases in total arsenic concentrations were observed after passage through the adsorption module. These results demonstrate the usefulness of *Ensifer* sp. M14 in arsenic removal performed in environmental settings.

Keywords: *Ensifer* (*Sinorhizobium*) sp. M14; arsenic-oxidizing bacteria; heavy metal resistance; draft genome sequence; comparative genomic analysis; biosafety; biotechnology for arsenic removal; adsorption; water treatment; in situ (bio)remediation

1. Introduction

The development and implementation of bioremediation technologies based on bioaugmentation requires the selection of appropriate microbial strains. A basic requirement of strains used as

bioaugmentation agents is their ability to survive in the environment into which they are introduced. Thus, such strains are usually characterized by high tolerance to heavy metals [1,2], resistance and ability to use organic (sometimes toxic) compounds [3,4], resistance to antibiotics [5], and an ability to thrive in the presence of local bacteriophages and microorganisms. Another important feature of strains used in bioaugmentation is their ability to perform effective transformation of the particular compound under changing environmental conditions (e.g., temperature, humidity, and pH). This is always the critical limitation, as many strains effective under laboratory conditions are, in fact, ineffective in field applications. Microorganisms suitable in bioremediation should maintain their activity in various seasons and under variable substrate inflow. A very important factor influencing the decision to apply a given microorganism in practice is also its interaction with the environment [6]. Strains that contribute to the uncontrolled release of contaminants, dissemination of antibiotic resistance genes, or disrupt the functioning of the ecosystem (e.g., by eliminating key microorganisms) should not be applied in open (uncontrolled) usage.

In this study, we provide a detailed characterization of *Sinorhizobium* sp. M14 (renamed here to *Ensifer* sp. M14 due to its phylogenetic positioning within the *Ensifer* clade), which is a strain with high potential to be used in bioremediation technologies for the removal of arsenic from contaminated waters and wastewaters. *Ensifer* sp. M14 is a psychrotolerant strain that was isolated from the microbial mats present in the arsenic-rich bottom sediments of an abandoned gold mine in Zloty Stok (Poland) [7]. The arsenic concentration in the mine waters reaches $\sim 6 \text{ mg L}^{-1}$, while in the microbial mats the level of accumulated arsenic is close to 20 g L^{-1} [8]. Previous physiological studies showed that *Ensifer* sp. M14 tolerates extremely high concentrations of arsenate [As(V)—up to 250 mM] and arsenite [As(III)—up to 20 mM], and is able to oxidize As(III) both chemolithoautotrophically [using arsenite or arsenopyrite (FeAsS) as a source of energy] and heterotrophically [7]. Batch experiments performed under various conditions of pH, temperature, and arsenic concentration confirmed the high adaptive potential of *Ensifer* sp. M14 [9]. The strain was capable of intensive growth and efficient biooxidation in a wide range of conditions, including low temperature [As(III) oxidation rate = $0.533 \text{ mg L}^{-1} \text{ h}^{-1}$ at $10 \text{ }^\circ\text{C}$]. Continuous flow experiments under environment-like conditions (2 L flow bioreactor) showed that *Ensifer* sp. M14 efficiently transforms As(III) into As(V) [24 h of residence time was sufficient to oxidize 5 mg L^{-1} of As(III)], but its activity depended mainly on the retention time in the bioreactor, which may be accelerated by stimulation with yeast extract as a source of nutrients [9].

Analysis of the extrachromosomal replicons of *Ensifer* sp. M14 revealed that its arsenic metabolism properties are linked with the presence of the mega-sized plasmid pSinA (109 kbp) [10]. The loss of the pSinA plasmid from *Ensifer* sp. M14 cells (using a target-oriented replicon curing technique [11]) eliminated the ability to oxidize As(III), and caused deficiencies in resistance to arsenic and heavy metals (Cd, Co, Zn, and Hg). In turn, the introduction of this plasmid into other representatives of the *Alphaproteobacteria* showed that cells with pSinA acquired the ability to oxidize arsenite and exhibited higher tolerance to arsenite than their parental, pSinA-less, wild-type strains. Horizontal transfer of arsenic metabolism genes by *Ensifer* sp. M14 was also confirmed in microcosm experiments [10]. The plasmid pSinA was successfully transferred via conjugation into indigenous bacteria of *Alpha*- and *Gammaproteobacteria* classes from the microbial community of As-contaminated soils. Transconjugants carrying plasmid pSinA expressed arsenite oxidase and stably maintained pSinA in their cells after approximately 60 generations of growth under nonselective conditions [10].

The second mega-sized replicon of *Ensifer* sp. M14—plasmid pSinB (300 kbp)—also plays an important role in the adaptation of the host to the mine environment. Structural and functional analysis of this plasmid showed that it carries gene clusters involved in heavy metals resistance. Among these are genes encoding efflux pumps, permeases, transporters, and copper oxidases, which are responsible for resistance to arsenic, cobalt, zinc, cadmium, iron, mercury, nickel, copper, and silver [12].

In this paper, we obtained a draft genomic sequence of *Ensifer* sp. M14 and performed complex genome-guided characterization of this bacterium. Special considerations were given to (i) determination of the metabolism of phosphate, sulfur, iron, and one-carbon substrates,

and (ii) investigation of the biosafety of *Ensifer* sp. M14 in the context of its release to the environment (e.g., determination of the presence of virulence and antibiotic resistance genes). These analyses revealed hints about the potential application of this strain in biotechnological applications; for example, the ability of it to survive environmental stresses, and whether it is likely to pose a safety risk. As the genomic analyses were consistent with *Ensifer* sp. M14 having potential application in biotechnology, we performed a large-scale simulation of the usage of M14 in the biological and chemical removal of arsenic from contaminated waters. The results support that the developed low-cost approach is an efficient method for the removal of arsenic from contaminated water.

2. Materials and Methods

2.1. Genome Sequencing, Assembly, and Annotation

Ensifer sp. M14 (available on request from the authors) was grown at 30 °C to stationary phase in TY medium (5 g L⁻¹ tryptone, 3 g L⁻¹ yeast extract, and 0.4 g L⁻¹ calcium chloride). Genomic DNA was isolated from the culture using a cetyltrimethylammonium bromide (CTAB) method [13] modified for bacterial DNA isolation as described by the Joint Genome Institute [14]. Sequencing was performed at IGATech (Udine, Italy) using an Illumina HiSeq2500 instrument with 125-bp paired-end reads. Two independent sequencing runs were performed. Reads were assembled into scaffolds using SPAdes v3.9.0 [15,16]. The scaffolds returned by SPAdes were parsed to remove those with less than 10× coverage or with a length below 200 nucleotides. Using FastANI [17], one-way average nucleotide identity (ANI) of the *Ensifer* sp. M14 assembly was calculated against the 887 alpha-proteobacterial genomes available through the National Center for Biotechnological Information (NCBI) with an assembly level of ‘complete’ or ‘chromosome’. The 10 genomes most closely related to *Ensifer* sp. M14 were identified on the basis of the ANI results. These 10 genomes, together with the complete pSinA and pSinB plasmid sequences [10,12], were used as reference genomes for further scaffolding of the assembly using MeDuSa [18]. The *Ensifer* sp. M14 assembly was then annotated using prokka version 1.12-beta [19], annotating coding regions with Prodigal [20], tRNA with Aragon [21], rRNA with Barrnap (github.com/tseemann/barrnap), and ncRNA with Infernal [22] and Rfam [23]. The predicted coding sequences were associated with Cluster of Orthologous Genes (COG) categories, Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms, and eggNOG annotations using eggNOG-mapper version 0.99.2-3-g41823b2 [24]. The assembly was deposited to NCBI with the GenBank accession QJNR000000000 (the version described in this paper is version QJNR01000000) and the BioSample accession SAMN09254189.

2.2. Phylogenetic Analysis

Initially, all 133 *Sinorhizobium/Ensifer* genomes available through NCBI, regardless of assembly level, were downloaded. FastANI [17] was used to calculate one-way ANI values between *Ensifer* sp. M14 and each of the 133 downloaded genomes. Only the strains meeting at least one of the following two requirements were kept for further analyses: (i) had a genome assembly level of ‘complete’ or ‘chromosome’, or (ii) had an ANI value of at least 85% compared to *Ensifer* sp. M14. This resulted in a final set of 46 strains, when including *Ensifer* sp. M14.

The pangenome of the 46 strains was calculated using Roary version 3.11.3 [25], as described below, following re-annotation with prokka version 1.12-beta [19]. Included in the Roary output was a concatenated nucleotide alignment of the 1652 core genes, each individually aligned with PRANK [26]. The core gene alignment was used to build a maximum likelihood phylogeny with RAxML version 8.2.9 [27] using the following command:

```
raxmlHPC-HYBRID-SSE3-T 5-s input.fasta-N autoMRE-n output-f a-p 12345-x 12345-m GTRCAT.
```

The final tree is the bootstrap best tree following 50 bootstrap replicates, and was visualized using the online iTOL (Interactive Tree of Life) webserver [28].

Strains were grouped into putative species on the basis of ANI and average amino acid identity (AAI) values, using thresholds of 96% for both measures. Groupings for ANI were the same at thresholds of 96% and 94%. Pairwise ANI values were calculated between each strain using FastANI [17], and the values in both directions were averaged. The CompareM workflow (github.com/dparks1134/CompareM) was used for calculating the AAI values. In the CompareM workflow, orthologous proteins were first identified using DIAMOND with the sensitive setting [29], and thresholds of 40% identity over 70% the length of the protein and a maximum e-value of $1e^{-12}$ were applied, as these are the thresholds used in the myTaxa program [30].

2.3. *Sinorhizobium/Ensifer* Pangenome Calculation

All 46 strains included in the phylogenetic analyses were reannotated using prokka version 1.12-beta [19], to ensure consistent annotation. The pangenome of the 46 reannotated strains was then determined with Roary version 3.11.3 [25], using an amino acid identity threshold of 80% and the following command:

```
roary-p 20-f Output-e-I 80-g 100,000 Input/*.gff.
```

For comparison of the gene content of *Ensifer* sp. M14, *Ensifer* sp. A49, *Ensifer adhaerens* OV14, and *Ensifer adhaerens* Casida A, the data was extracted from the full 46-strain pangenome. The complete gene presence/absence output from Roary is provided as Data Set S1. Several short proteins of *Ensifer* sp. M14 were not present in the output of the Roary analysis; these proteins were not considered when identifying unique genes.

2.4. Comparative Genomics of Arsenic Oxidizing Bacteria

The genomes of *Agrobacterium tumefaciens* 5A [31], *Agrobacterium tumefaciens* Ach5 [32], *Ensifer adhaerens* OV14 [33], *Neorhizobium galegae* HAMB1 540 [34], and *Rhizobium* sp. NT-26 [35] were downloaded from NCBI GenBank and reannotated using prokka, as described above for *Ensifer* sp. M14. The GenBank files of the re-annotated genomes, and the *Ensifer* sp. M14 genome, were uploaded to the KBase webserver [36], and OrthoMCL [37] was run on the KBase server using an e-value threshold of $1e^{-12}$. Identification of phosphate transport and arsenic resistance genes in other bacterial genomes (*Achromobacter arsenitoxydans* SY8 [38], *Herminiimonas arsenicoxydans* ULPAs1 [39], and *Pseudomonas stutzeri* TS44 [40]) was accomplished by manually searching the GenBank file of the RefSeq annotated genomes [41].

2.5. Identification of Prophage Loci

PhiSpy version 3.2 [42], implemented in Python, was used to predict phage genes. The *Ensifer* sp. M14 GenBank file produced with prokka was converted to SEED format using the `genbank_to_seed.py` script. The converted file was then used as input for the `PhiSpy.py` script, using the generic test set for training.

2.6. Identification of Putative Antibiotic Resistance Genes

To identify putative antibiotic resistance genes, the Resistance Gene Identifier (RGI) in the Comprehensive Antibiotic Resistance Database (CARD) software was used [43]. Hits showing at least 50% identity with the reference protein were considered significant. Each hit was verified manually using BLASTp analysis.

2.7. Analysis of the Antimicrobial Susceptibility Patterns

To determine the antimicrobial susceptibility patterns of *Ensifer* sp. M14, minimum inhibitory concentrations (MICs) of 11 antimicrobial agents were assessed using Etest™ (Liofilchem, Roseto degli Abruzzi, Italy). The analysis was conducted according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) recommendations [44]. The following antibiotics (selected based on the bioinformatic analyses that identified putative antibiotic resistance genes) were used:

(i) aminoglycosides–gentamicin (GN; concentration of antibiotic: 0.064–1024 $\mu\text{g mL}^{-1}$ Roseto degli Abruzzi¹); (ii) β -lactams (penicillin derivatives)–ampicillin (AMP; 0.016–256 $\mu\text{g mL}^{-1}$); (iii) β -lactams (cephalosporins)–cefixime (CFM; 0.016–256 $\mu\text{g mL}^{-1}$); (iv) β -lactams (cephalosporins)–cefotaxime (CTX; 0.016–256 $\mu\text{g mL}^{-1}$); (v) β -lactams (cephalosporins)–ceftriaxone (CRO; 0.016–256 $\mu\text{g mL}^{-1}$); (vi) fluoroquinolones–ciprofloxacin (CIP; 0.002–32 $\mu\text{g mL}^{-1}$); (vii) fluoroquinolones–moxifloxacin (MXF; 0.002–32 $\mu\text{g mL}^{-1}$); (viii) phenicols–chloramphenicol (C; 0.016–256 $\mu\text{g mL}^{-1}$); (ix) rifamicyns–rifampicin (RD; 0.016–256 $\mu\text{g mL}^{-1}$); (x) sulfonamides–trimethoprim (TM; 0.002–32 $\mu\text{g mL}^{-1}$); and (xi) tetracyclines–tetracycline (TE; 0.016–256 $\mu\text{g mL}^{-1}$). The susceptibility testing was performed at 30 °C for 20 h. After incubation, plates were photographed and MICs were defined. Antimicrobial susceptibility data were interpreted according to the EUCAST breakpoint table version 8.0 [45].

2.8. Search for Symbiotic Proteins

A custom pipeline based on the use of hidden Markov models (HMM) was used to search the proteomes of all 46 *Sinorhizobium/Ensifer* strains for the presence of the nodulation proteins NodA, NodB, and NodC, as well as for the nitrogenase proteins NifH, NifD, and NifK. This pipeline is dependent on HMMER version 3.1b2 [46], and the complete Pfam-A version 31.0 (16,712 HMMs) and TIGERFAM version 15.0 (4488 HMMs) databases [47,48]. After downloading the HMM databases, hmmconvert was used to ensure consistent formatting. The two databases were combined into a single HMM database, and then converted into a searchable database with hmmpress. Additionally, the HMM seed alignments for NodA (TIGR04245), NodB (TIGR04243), NodC (TIGR04242), NifH (TIGR01287), NifD (TIGR01282), and NifK (TIGR01286) were downloaded from the TIGRFAM database [47].

For each HMM seed alignment, a HMM was built using hmmbuild, and the output was then searched against the complete set of *Sinorhizobium/Ensifer* proteins using hmmsearch. The output was parsed, and the amino acid sequences for each of the hits (regardless of e-value) were collected. Each set of sequences were then searched against the combined HMM database using hmmscan, and the output parsed to identify the top scoring HMM hit for each query protein. Proteins were annotated as follows: NodA if the top hit was TIGR04245 (TIGRFAM) or NodA (Pfam); NodB if the top hit was TIGR04243 (TIGRFAM); NodC if the top hit was TIGR04242 (TIGRFAM); NifH if the top hit was TIGR01287 (TIGRFAM) or Fer4_NifH (Pfam); NifD if the top hit was TIGR01282 (TIGRFAM), TIGR01860 (TIGRFAM), or TIGR01861 (TIGRFAM); NifK if the top hit was TIGR02932 (TIGRFAM), TIGR02931 (TIGRFAM), or TIGR01286 (TIGRFAM).

2.9. Cluster of Orthologous Genes Functional Annotation

Proteomes were annotated with COG functional categories using eggNOG-mapper version 0.99.2-3-g41823b2 [24]. The output of eggNOG-mapper was parsed with a custom Perl script to count the percentage of proteins annotated with each functional category. Fisher exact tests, performed using MATLAB R2016b (www.mathworks.com), were performed to identify statistically significant differences ($p < 0.05$) between *Ensifer* sp. M14 and the other strains.

2.10. In Silico Metabolic Reconstruction and Constraint-Based Modelling

Metabolic reconstruction steps and constraint-based metabolic modeling were performed in MATLAB 2017a (Mathworks, Natick, MA, USA), using the Gurobi 7.0.2 solver (gurobi.com), SBMLToolbox 4.1.0 [49], libSBML 5.13.0 [50], and scripts from the COBRA Toolbox [51] and the Tn-Core Toolbox [52]. The ability of the *Ensifer* sp. M14 model to grow when individually provided with 163 carbon sources was tested using flux balance analysis (FBA) as implemented in the 'optimizeCbModel' function of the COBRA Toolbox.

An initial draft metabolic reconstruction was prepared using the online KBase webserver [36]. The *Ensifer* sp. M14 genome was uploaded and re-annotated with RAST functions using the 'annotate microbial genome' function. The re-annotated genome was used to build a draft model with the 'build

metabolic model' function, performing gap-filling on a glucose minimal medium, and with automatic biomass template selection. This reconstruction was downloaded in SBML format, and then imported into MATLAB as a COBRA formatted metabolic model for further manipulation. After removing duplicate genes from the gene list and updating the gene-reaction rules appropriately, the model was expanded based on the reaction content of the curated iGD1575 and iGD726 metabolic reconstructions of the closely related species *Sinorhizobium meliloti* [53,54]. First, a BLAST bidirectional best hit approach was used to identify putative orthologs (at least 70% identity over at least 70% the protein length) between *S. meliloti* Rm1021 and *Ensifer* sp. M14. All *S. meliloti* genes without a putative ortholog in *Ensifer* sp. M14 were deleted from the iGD1575 and iGD726 models, and the constrained reactions removed. Next, the reactions of iGD726 and the draft *Ensifer* sp. M14 model were compared based on their equations, and all reactions unique to iGD726 were identified and transferred to the *Ensifer* sp. M14 model. Exceptions were iGD726 reactions that differed from a reaction in the *Ensifer* sp. M14 model only in the presence/absence of a proton or in metabolite stoichiometry. This process was then repeated, transferring the unique reactions of iGD1575 to the partially expanded model. When transferring reactions, associated genes were also transferred and changed to the name of the *Ensifer* sp. M14 orthologs. Following the expansions, all reactions producing dead-end metabolites were iteratively removed from the model. The final model contained 1491 genes, 1561 reactions, and 1105 metabolites, and is available in Data Set S2.

2.11. Prediction of Secondary Metabolism

Loci encoding secondary metabolic pathways were predicted in the *Ensifer* sp. M14 genome using the antiSMASH webserver [55]. The *Ensifer* sp. M14 GenBank file was uploaded to the bacterial version of antiSMASH, and the analysis was run with all options selected with default parameters.

2.12. Construction of a Pilot-Scale Installation for Arsenic Bioremediation

A pilot-scale installation for the removal of arsenic from contaminated waters was developed. The installation was operated using water from a dewatering system of a former gold mine located in the Zloty Stok area (SW Poland), which is highly polluted with arsenic. The total arsenic concentration, arsenic speciation, as well as detailed chemical and physical characteristics of the water are presented elsewhere [56]. The installation consisted of two modules: the microbiological module and the adsorption module (Figure 1).

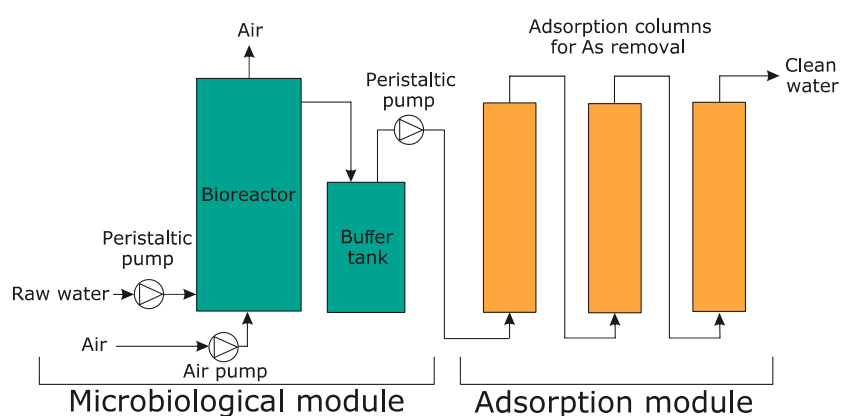


Figure 1. The pilot-scale installation used for remediation of arsenic contaminated water. The image is a schematic representation of the pilot-scale installation developed as part of this work. Both the microbiological and adsorption modules are shown.

The microbiological module was based on the activity of *Ensifer* sp. M14, which was used as an arsenite biooxidizer. This module included a 200 L bioreactor with an electric heater. The contaminated water flowing out from the gold mine was fed into the bioreactor through a pressure

reducer and a peristaltic pump at a volume flow rate of 8.33 L h^{-1} , corresponding to a residence time of 24 h in the bioreactor. Outflow of the water occurred as overflow in the upper part of the bioreactor. To increase the effectiveness of the arsenite biooxidation, the bioreactor was equipped with an additional aeration system that consisted of an air pump producing compressed air. The additional aeration system was included in our previous study and showed that the arsenite oxidation efficiency of *Ensifer* sp. M14 is higher in the presence of additional aeration during continuous culturing [9]. Moreover, yeast extract was added to the bioreactor as a source of vitamins (growth supplements). Fifty grams of powdered yeast extract (Sigma-Aldrich, St. Louis, MO, USA) was added to the bioreactor twice a week. This was done as we previously observed that the presence of yeast extract led to an increase in the growth and efficiency of arsenite biooxidation of *Ensifer* sp. M14 during continuous culturing [9]. This relationship was also confirmed in other papers [10,57]. The supply of air also contributed to the mixing of the bioreactor content. The bioreactor was equipped with a multifunctional electrode dedicated to controlling the chemical and physical parameters of water, specifically, to monitor pH, redox potential, and temperature (Hydrolab HL4, OTT Hydromet, Kempten, Germany). The water leaving the bioreactor was fed into a 60 L buffer tank, which functioned as the connecting element between the bioreactor and the adsorption module. The inclusion of the buffer tank helped maintain a constant water level in the adsorption columns and ensured a constant flow of water from the bioreactor to the adsorption columns.

The adsorption module consisted of three columns (17 L volume each) filled with granulated bog iron ores (about 15 kg per column) and connected in series (Figure 1). The detailed chemical and physical parameters, chemical composition, and stability of the adsorbent were presented previously [56]. Contaminated water from the buffer tank (after passing through the microbiological module) was fed into the first column using a second peristaltic pump at a volume flow rate of 8.33 L h^{-1} , which corresponded to approximately one hour of residence time per column.

The installation was also equipped with a process control system (operated at the location of the pilot plant or remotely via a Global System for Mobile Communications (GSM)) that monitored and controlled key process parameters including the volume flow rate of the water, the water temperature at the inlet, in the bioreactor, and at the outlet of the pilot plant, as well as the ambient temperature.

2.13. Installation Start-Up

Scale-up of the installation (from laboratory scale to pilot scale) required the development of procedures for successful start-up based on the results of our previous study [9]. The first step of the start-up of the microbiological module was inoculation of the bioreactor with an appropriate amount of *Ensifer* sp. M14. The bioreactor filled with arsenic contaminated water was inoculated with 200 mL of a highly concentrated overnight culture of *Ensifer* sp. M14 suspended in 0.85% NaCl solution. The initial OD_{600} in the bioreactor was 0.01. In earlier experiments, it was determined that a starting cell density of 10^8 CFU mL^{-1} (which corresponds to an OD_{600} of 0.1) is required for the installation to work properly [9]. To increase the density of the *Ensifer* sp. M14, the water in the bioreactor was supplemented with powdered yeast extract to a final concentration of 0.04%. Additionally, aeration was applied. Finally, the temperature of the water was increased (from 10 to 22 °C) with the use of an electric heater placed in the bioreactor. Application of all these treatments led to an OD_{600} value of 0.1 within 24 h.

Start-up procedures related to the adsorption module mainly concerned the preparation of the adsorbent for its usage. After filling the columns with granulated bog iron ores, it was necessary to condition the adsorbent (rinsing the adsorbent with the tap water without arsenic) to remove all the loosely bound fractions.

2.14. Biological and Chemical Analyses

Arsenic speciation was investigated with the use of ion chromatography on an IonPac AS18 (2 mm, Dionex, Lübeck, Germany) column on an ICS Dionex 3000 (Lübeck, Germany) instrument

equipped with an ASRS[®] 2 mm suppressor, which was coupled to a ZQ 2000 mass spectrometer via an electrospray source (Waters, Milford, MA, USA) according to the method described by Debiec et al. [9]. In the adsorption module, the total arsenic concentration was investigated. Total arsenic concentration was measured using a Graphite Furnace Atomic Absorption Spectrometry (GFAAS; AA Solaar M6 Spectrometer, TJA Solutions, Waltham, MA, USA). Arsenic standard solutions (Merck, Darmstadt, Germany) were prepared in 3% HNO₃. The pH and redox potential were measured only in the microbiological module. Samples of raw water, water from the bioreactor, as well as water at the inflow and outflow of each adsorption column were collected once a day during the first 8 days, and then three times a week up to day 40. Samples taken from the bioreactor were stored at −20 °C, while samples collected from the adsorption module were stored at 4 °C. This experiment was repeated twice.

3. Results and Discussion

3.1. Sequencing of the *Ensifer* sp. M14 Genome

The draft genome sequence of *Ensifer* sp. M14 was obtained as described in the Materials and Methods, and the general genomic features are described in Table 1.

Table 1. Features of the *Ensifer* sp. M14 genome assembly.

Length	7,345,249 bp
G + C content	61.47%
CDS	6874
rRNA	3
tRNA	53
Miscellaneous RNA	33
Scaffolds	45
Scaffold N50 (L50)	4400,487 (1)
CDS with COG terms ^{*,†}	64.00%
CDS with GO terms [*]	28.70%
CDS with KEGG pathway terms [*]	35.50%
CDS with eggNOG annotations ^{*,‡}	80.50%
CDS with no similarity [*]	9.40%

^{*} As determined using eggNOG-mapper [24]. Those genes not returned in the eggNOG-mapper output were said to have no similarity; [†] Excluding those annotated with COG category S (unknown function); [‡] Excluding those annotated as protein/domain of unknown/uncharacterized function. CDS (Coding Sequences); COG (Cluster of Orthologous Genes); KEGG (Kyoto Encyclopedia of Genes and Genomes); GO (Gene Ontology).

The assembly consists of 7,345,249 bp spread over 45 scaffolds at an average coverage of 118×. Of the 45 scaffolds, 12 are over 40 kbp in size and account for 98.7% of the assembly. Based on similarity searches of the scaffolds, previous plasmid profiling of *Ensifer* sp. M14 [10,12], and the finished genomes of related strains [33,58], we predict that the *Ensifer* sp. M14 genome consists of one chromosome (at least 4.4 Mbp in size), two additional large replicons (chromids and/or large megaplasmids, at least 1.6 Mbp and 0.6 Mbp in size), and the two previously reported smaller megaplasmids (pSinA and pSinB, 109 kbp and 300 kbp, respectively, based on previous papers [10,12]). A total of 6874 coding sequences were predicted, which is more than the 6218 predicted in *S. meliloti* Rm1021 and the 6641 of *E. adhaerens* Casida A, but less than the 7033 predicted in *E. adhaerens* OV14 [33,58,59]. Six putative prophages were identified on Scaffold 4 (the chromosome) using PhiSpy [42]; these ranged in size from 21 to 65 genes, and accounted for a total of 292 genes (Data Set S3). However, no CRISPR loci were detected during annotation with prokka [19]; a questionable, short CRISPR with one spacer was detected with CRISPRfinder [59], but its location within a predicted coding region suggests it is unlikely to be a true CRISPR locus. No evidence for the presence of the common nodulation genes *nodABC* or the nitrogenase genes *nifHDK* was found using a hidden Markov model based approach. The *Ensifer* sp. M14 assembly has been deposited in GenBank under the accession QJNR000000000, as part of the BioSample SAMN09254189.

3.2. Taxonomic Analysis of *Ensifer* sp. M14

Phylogenetic analyses were performed to identify the relationships between *Ensifer* sp. M14 and previously sequenced *Sinorhizobium*/*Ensifer* strains. Forty-five *Sinorhizobium*/*Ensifer* genomes were downloaded from the NCBI database (see Materials and Methods for criteria for strain inclusion), and a maximum likelihood phylogeny of these strains plus *Ensifer* sp. M14 was built based on 1652 core genes (Figure 2).

The 46 strains were grouped into putative species on the basis of whole genome ANI and AAI values (Figures S1 and S2). The results revealed that *Ensifer* sp. M14 is closely related to *Ensifer* (*Sinorhizobium*) sp. A49 (98.5% ANI and 98.9% AAI), and that these strains likely belong to a new species. *Ensifer* sp. A49 was previously isolated from soil of the Fureneset Rural Development Centre of Fjaler, Norway [60]. However, the pSinA and pSinB plasmids, carrying genes involved in arsenic oxidation and heavy metal resistance [10,12], appear to be specific to *Ensifer* sp. M14 and may therefore have been gained during growth in the Zloty Stok gold mine [7]. The most closely related named species is *Ensifer adhaerens*, which includes bacterial predators capable of feeding on organisms such as *Micrococcus luteus* [58,61].

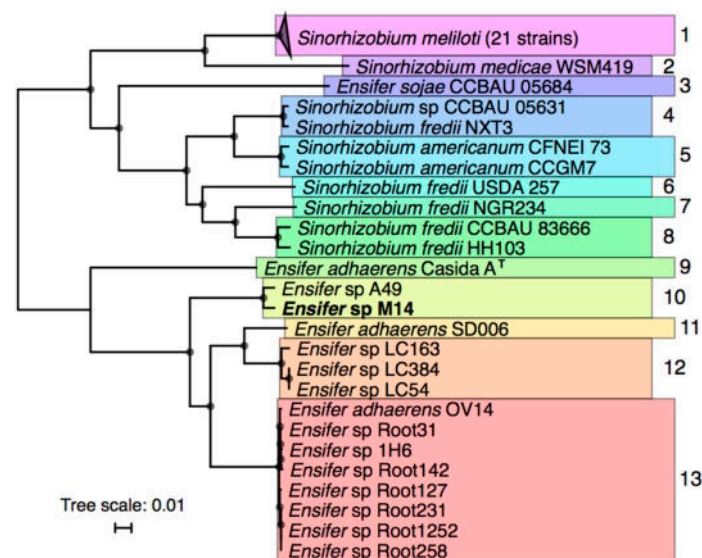


Figure 2. Phylogeny of a selected 46 *Sinorhizobium*/*Ensifer* strains with a publicly available whole genome sequence. An unrooted RAxML maximum likelihood phylogeny of 46 *Sinorhizobium*/*Ensifer* strains was prepared on the basis of the concatenated nucleotide alignments of 1652 core genes. The presented tree is the bootstrap best tree following 50 bootstrap replicates, and the scale represents the mean number of nucleotide substitutions per site. Nodes with 100% bootstrap support are indicated by the black circles. The colors and numbers to the right of the tree are used to indicate strains that group into putative species on the basis of average nucleotide identity (>96% ANI; same results were obtained with >94% ANI) and average amino acid identity (>96% AAI), as described in the Materials and Methods. Type strains are indicated by the ‘T’. The accessions for all strains included in this figure are provided in Table S1.

3.3. Identification of Unique Features of the *Ensifer* sp. M14 Genome

A global, functional analysis of the *Ensifer* sp. M14 proteome was performed using COG categories, and the proteome was compared with closely related species to identify general functional biases. This analysis was performed with the goal of identifying recently acquired genomic islands that may contribute to the adaptation of *Ensifer* sp. M14 to the gold mine environment. When compared with *Ensifer* sp. A49, *E. adhaerens* OV14, and *E. adhaerens* Casida A, no statistically significant biases (pairwise Fisher’s exact tests, $p > 0.05$ in all cases) in COG category abundances were detected in

the *Ensifer* sp. M14 proteome (Figure 3A). However, there was a slight, but statistically insignificant (pairwise Fisher's exact tests, $p > 0.05$), enrichment in inorganic ion transport and metabolism (COG P) in the proteomes of *Ensifer* sp. M14 and *Ensifer* sp. A49 compared to the other two strains (Figure 3A). These results suggest no gross functional changes in the *Ensifer* sp. M14 genome occurred during adaptation to growth in the Zloty Stok gold mine, at least at the general level of COG categories.

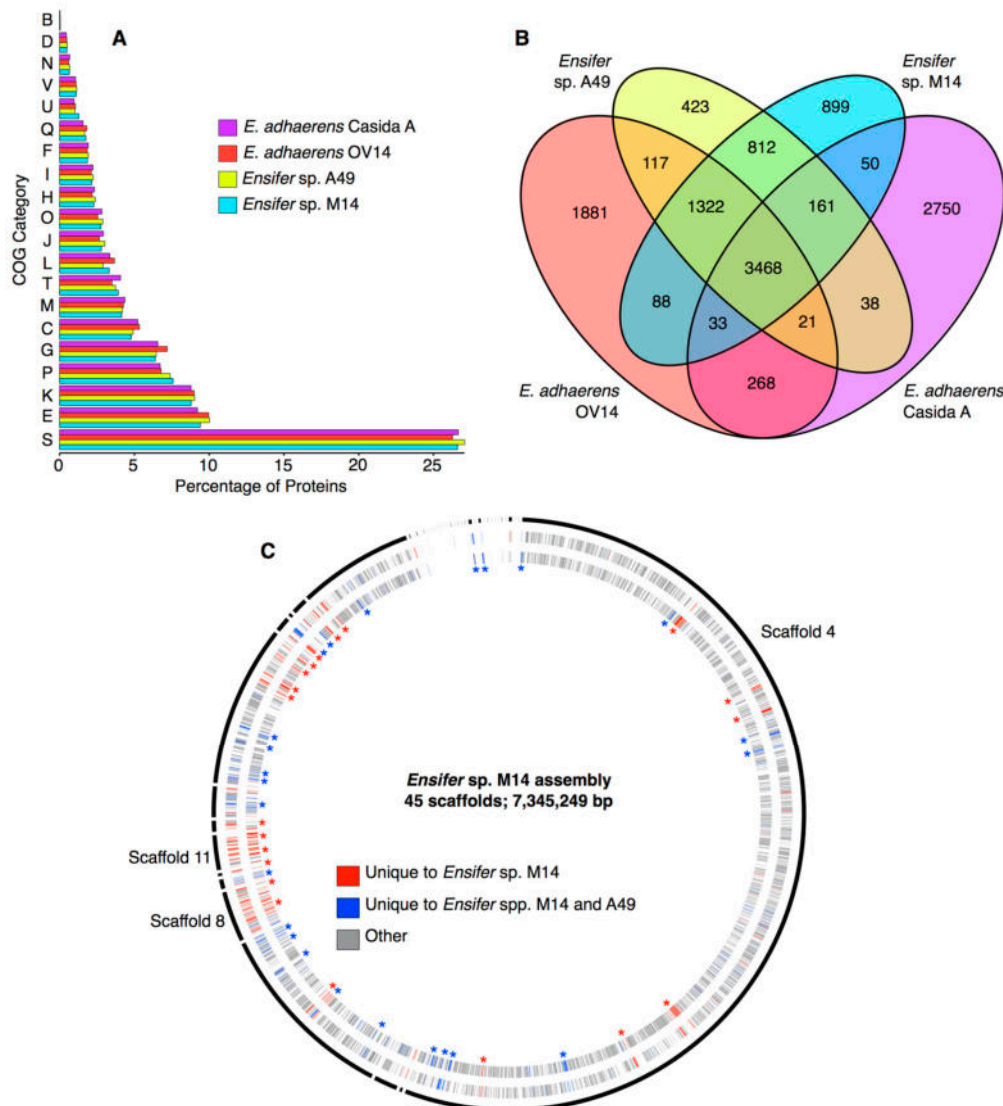


Figure 3. General features of the genome of *Ensifer* sp. M14 and related strains. **(A)** The percentage of proteins encoded by each strain annotated with each COG (Cluster of Orthologous Genes) functional category. COG categories not represented in the proteome are excluded from the graph. COG category definitions are provided in Table S4. **(B)** A Venn diagram indicating the number of genes shared among these four strains, as extracted from the pangenome of the 46 strains shown in Figure 2. **(C)** A circular plot, prepared with Circos version 0.67-7 [62], showing the scaffolds of the *Ensifer* sp. M14 assembly (outer black curved lines) including the plasmids, and the predicted coding sequences on the positive strand (outer ring) and negative strand (inner ring). Scaffolds are drawn proportional to their size, and they are presented in the order they are numbered. Scaffold 4 (chromosome), 8 (pSinB), and 11 (pSinA) are labelled. The coding regions are colored according to their conservation level, with red indicating genes unique to *Ensifer* sp. M14, and yellow indicating species common and unique to *Ensifer* spp. M14 and A49. Some multi-gene loci unique to M14 (red asterisks) or unique to M14 and A49 (blue asterisks) are indicated.

Despite the similarity in COG abundances, the *Ensifer* sp. M14 genome contains a large number of unique genes. There are 899 genes found in *Ensifer* sp. M14 but not in *Ensifer* sp. A49, *E. adhaerens* OV14, or *E. adhaerens* Casida A, while an additional 812 are found in M14 and A49, but not OV14 or Casida A (Figure 3B). Of the 899 genes specific to *Ensifer* sp. M14, 656 (9.4% of the genome) were not detected in any of the other 45 *Sinorhizobium/Ensifer* strains included in the phylogenetic analysis (Data Sets S1 and S4). Five hundred and ninety of the 656 unique proteins had a blast hit (e-value $\leq 1e^{-10}$) when queried against the NCBI non-redundant protein database, consistent with the corresponding genes being real genes that were likely acquired from other organisms through horizontal gene transfer (HGT). Mapping the location of the 656 unique genes across the assembly revealed the presence of several putative genomic islands (GIs) likely acquired through recent HGT since the divergence of *Ensifer* sp. M14 from *Ensifer* sp. A49 (Figure 3C, Data Set S4). Scaffolds 11 and 8, which correspond to the pSinA and pSinB plasmids, respectively, were not surprisingly enriched in unique genes, and together account for 217 (33%) of the unique genes. As described in detail elsewhere, these plasmids carry numerous functions associated with arsenic oxidation [10] and heavy metal resistance [10,12]. Of the 439 unique genes spread among the other scaffolds, 309 (70.4%) were annotated as hypothetical genes. Little else of interest was detected among the unique genes (Data Set S4); however, scaffold 36 was predicted to encode a zinc transporting ATPase, and a few genes related to stress resistance or drug resistance were found (discussed later). Overall, these results suggest that essentially all of the recently acquired traits associated with heavy metal resistance, arsenic oxidation, and adaptation to the stressful conditions of the Zloty Stock gold mine are associated with the pSinA and pSinB plasmids.

3.4. Metabolism of *Ensifer* sp. M14

Detailed phenotypic characterization of *Ensifer* sp. M14 was previously reported [7]. To further evaluate (in silico) the metabolic and transport potential of *Ensifer* sp. M14, a draft metabolic reconstruction was prepared encompassing 1491 genes and 1289 gene-associated reactions (Data Set S2). As expected based on the metabolism of related organisms [63], glycolysis in *Ensifer* sp. M14 is predicted to proceed through the Entner–Duodoroff pathway (Figures S3–S5). Growth simulations using Flux Balance Analysis suggested that *Ensifer* sp. M14 has a broad metabolic capacity, with a predicted ability to catabolize 72 carbon sources, including a variety of sugars, sugar alcohols, and organic acids (Table S2). This is consistent with previous work, which found that *Ensifer* sp. M14 could grow on 12 of 16 tested carbon substrates, including glucose, xylose, and lactate [7]. The following paragraphs provide a description of several metabolic capabilities that may be relevant to survival in the stressful environment of the Zloty Stok gold mine, and/or to resistance to elevated arsenic concentrations.

3.4.1. Phosphate Transport

The metabolic reconstruction indicated that *Ensifer* sp. M14 encodes two copies of the PstSCAB-PhoU high-affinity phosphate transporter (*BLJAPNOD_00112* through *BLJAPNOD_00116*; and *BLJAPNOD_05453* through *BLJAPNOD_05457*). Further examination of the *Ensifer* sp. M14 genome additionally revealed two copies of the PhnCDE(T) high-affinity phosphate and phosphonate transport system (*BLJAPNOD_04783* through *BLJAPNOD_04786*; and *BLJAPNOD_05447* through *BLJAPNOD_05450*). Notably, one copy of PstSCAB-PhoU and one copy of PhnCDE(T) were adjacent to the arsenic oxidation gene cluster within pSinA. This led us to explore the presence of phosphate transport systems in other arsenic-oxidizing bacteria (AOB). Using OrthoMCL [37], orthologous proteins were identified among six strains from the family *Rhizobiaceae* (Table S3): these included three AOB (*Ensifer* sp. M14, *A. tumefaciens* 5A, and *Rhizobium* sp. NT-26), as well as three related strains that are not AOB (*E. adhaerens* OV14, *N. galegae* HAMBI 540, and *A. tumefaciens* Ach5). Thirteen proteins were found to be common and specific to the three AOB, which not surprisingly included the arsenic oxidation gene cluster [10]. Notably, included within these 13 proteins were subunits of the PstSCAB-PhoU and PhnCDE(T) transporters. While all six strains encoded orthologous versions of PstSCAB-PhoU and PhnCDE(T), all three AOB encoded additional copies adjacent to their arsenic

oxidation loci. Examining the genomes of three additional diverse AOB (*H. arsenicoxydans* ULPAs1, *A. arsenitoxydans* SY8, and *P. stutzeri* TS44) revealed that the first two also contained a second copy of the PstSCAB transporter in close proximity to arsenite related genes.

Based on the above results, we predict that phosphate transport genes are commonly associated with arsenite resistance loci [64]. Arsenates and phosphate are chemical analogs, with the toxicity of arsenic being a result of arsenic replacing phosphate in key biological molecules [65]. Similarly, arsenic competes with phosphate for transport through phosphate transport systems, including the PstSCAB and PhnCDE(T) systems [66–68], potentially resulting in phosphate starvation. However, the phosphate periplasmic binding proteins of at least some PstSCAB-PhoU systems, such as from the arsenic-resistant strain *Halomonas* strain GFAJ-1, display a strong preference for binding phosphate over arsenic [68]. Thus, the presence of additional high-affinity phosphate systems in AOB may be a mechanism to increase the rate (and selectivity) of phosphate import, thereby reducing the toxic effects of elevated environmental arsenic concentrations.

3.4.2. Sulfur Metabolism

We evaluated sulfur metabolism by *Ensifer* sp. M14, as sulfur compounds, such as sulfide, can be abundant in gold mines, and the arsenic oxidase enzyme contains an iron-sulfur subunit [64]. *Ensifer* sp. M14 appears to have a variety of mechanisms for sulfate assimilation. Based on the metabolic reconstruction, the genome is predicted to encode multiple sulfate and thiosulfate transporters. It is further predicted to encode several putative thiosulfate sulfurtransferases and a hydrogen sulfide oxidoreductase (*BLJAPNOD_03089*); in contrast, a sulfite oxidoreductase was not identified. Genes *BLJAPNOD_05764* through *BLJAPNOD_05768* may encode for the transport and metabolism of taurine, while *BLJAPNOD_05769* may encode the TauR taurine transcriptional regulator. *Ensifer* sp. M14 is also predicted to encode an alkanesulfonate monooxygenase (*BLJAPNOD_06609*). At least one copy of each of the subunits of the SsuABC alkanesulfonate ABC-type transporter are also predicted to be encoded in the genome; however, no locus appeared to contain all three.

3.4.3. One-Carbon Metabolism

Ensifer sp. M14 is capable of growing with carbon dioxide or bicarbonate as the sole source of carbon [7], although the underlying metabolic pathway for this capability has not been examined. The metabolic reconstruction identified a putative formamide amidohydrolase (*BLJAPNOD_04973*) and putative formate dehydrogenases (*BLJAPNOD_00952* and *BLJAPNOD_03433*), suggestive of the utilization of these one-carbon compounds. No clear evidence for genes associated with methanol or methylamine metabolism were found. However, the mechanism underlying one-carbon metabolism remains unclear. Unlike *S. meliloti* [69], *Ensifer* sp. M14 does not appear to encode the Calvin–Benson–Bassham cycle, nor were we able to identify any of the complete carbon-fixation pathways [70]. However, multiple enzymes potentially involved in the incorporation of bicarbonate were identified. These include putative acetyl-CoA carboxylases (*BLJAPNOD_03269*, *BLJAPNOD_04937*, *BLJAPNOD_04938*), a putative 3-oxopropanoate oxidoreductase (*BLJAPNOD_03990*), putative propanoyl-CoA carboxylases (*BLJAPNOD_06206*, *BLJAPNOD_06208*), a putative pyruvate carboxylase (*BLJAPNOD_00700*), and a phosphoenolpyruvate carboxylase (*BLJAPNOD_01050*).

3.4.4. Iron Transport and Metabolism

Due to the involvement of iron in arsenic oxidation, the transport and metabolism of this metal was examined. *Ensifer* sp. M14 is predicted to encode several transporters of iron or iron containing compounds. The genes *BLJAPNOD_01755* and *BLJAPNOD_01831* are predicted to encode a ferrous iron (Fe²⁺) permease (EfeU) and a ferrous iron efflux pump (FieF), respectively. Genes *BLJAPNOD_05889* through *BLJAPNOD_05891* may encode a FecBDE ferric dicitrate transporter, while *BLJAPNOD_05888* may encode the FecA ferric dicitrate outer membrane receptor protein. The genes *BLJAPNOD_00861*

through *BLJAPNOD_00863* may encode a second ferric dicitrate transporter. Additionally, the genes *BLJAPNOD_05777*, *BLJAPNOD_05780*, and *BLJAPNOD_05781* may form an ABC-type transport system for iron or an iron complexes. Moreover, three putative FhuA ferrichrome (iron containing siderophore) transporting outer membrane proteins (*BLJAPNOD_04144*, *BLJAPNOD_04445*, *BLJAPNOD_05778*), and a FcuA ferrichrome receptor (*BLJAPNOD_05962*) are predicted to be encoded in the genome. A putative FepCDG ferric enterobactin transporter (*BLJAPNOD_04147*, *BLJAPNOD_04148*, *BLJAPNOD_04149*) and a PfeA enterobactin receptor (*BLJAPNOD_05560*) are also annotated. Aside from transport, *Ensifer* sp. M14 is predicted to encode a ferric reductase (*BLJAPNOD_01976-fhuF*), a ferrous oxidoreductase (*BLJAPNOD_01631*), and a ferric-chelate reductase (*BLJAPNOD_02273*). Additionally, the five gene operon (*BLJAPNOD_05798-BLJAPNOD_05802*) was predicted (using antiSMASH [55]) to encode a siderophore (aerobacin-like) biosynthetic pathway. Finally, the ferric uptake regulator (Fur) is predicted to be encoded by *BLJAPNOD_00930*.

3.4.5. Halotolerance

The *Ensifer* sp. M14 genome was searched for genes relevant to halotolerance as *Ensifer* sp. M14 has been shown to grow in highly saline environments with up to 20 mg L⁻¹ NaCl [10]. Examination of the *Ensifer* sp. M14 genome with antiSMASH [55] identified a 13 gene locus (*BLJAPNOD_06859* to *BLJAPNOD_06872*) in which 12 of the genes showed similarity to 12 of the 15 genes of a known salectan biosynthetic cluster. Salectan is a water-soluble β -glucan also produced by the salt tolerant strain *Agrobacterium* sp. ZX09 [71]. Thus, this locus in *Ensifer* sp. M14 may encode for the biosynthesis of salectan, or another carbohydrate, that contributes to halotolerance. Additionally, *Ensifer* sp. M14 is predicted to be capable of synthesizing the compatible solute betaine from choline using the BetA (*BLJAPNOD_01468*, *BLJAPNOD_03726*, *BLJAPNOD_06536*) and BetB (*BLJAPNOD_00678*, *BLJAPNOD_03725*, *BLJAPNOD_05671*) pathway, as well as from choline-O-sulfate with BetC (*BLJAPNOD_02271*, *BLJAPNOD_03724*). The genome is further predicted to encode numerous proteins related to glycine betaine and proline betaine transport. Finally, as previously reported [10], pSinA encodes a putative NhaA pH-dependent sodium/proton antiporter (*BLJAPNOD_05431*), which may contribute to adaptation to high salinity [72].

3.4.6. Heavy Metal Resistance

Ensifer sp. M14 displays high resistance to numerous heavy metals [7]. Previous work identified eight modules related to heavy metal resistance on the pSinB replicon of *Ensifer* sp. M14 [12]. These modules were involved in resistance to arsenic, cadmium, cobalt, copper, iron, mercury, nickel, silver, and zinc [12]. Additionally, pSinA contains a locus involved in resistance to cadmium, zinc, cobalt, and mercury [10]. Our analyses reported above suggested that the majority, if not all, genes relevant to adaptation to the heavy metal-rich environment in the Zloty Stok gold mine are located on the pSinA and pSinB plasmids [10,12].

3.5. Biosafety Considerations of *Ensifer* sp. M14

The *Sinorhizobium/Ensifer* group of bacteria contain numerous plant symbionts and other biotechnologically relevant strains, but it lacks known pathogens. Considering this, and the observation that none of the genomic islands detected in *Ensifer* sp. M14 appear to be pathogenicity islands, it is unlikely that *Ensifer* sp. M14 is pathogenic. Therefore, the environmental release of *Ensifer* sp. M14 is not expected to pose a biosafety risk from that perspective. Additionally, analysis of the secondary metabolism of *Ensifer* sp. M14 with antiSMASH [55] did not identify antibiotic synthesis loci. However, *Ensifer* sp. M14 may carry several antimicrobial resistance (AMR) genes. The analysis applying the RGI analyzer revealed the presence of 12 putative antibiotic resistance genes/gene clusters (Table 2). It is worth mentioning that the best hits were found for four *acrAB(-TolC)* modules encoding resistance-nodulation-cell division (RND) type multidrug efflux systems, while the remaining eight genes were much more divergent compared with the reference proteins (they were detected only when

applying the LOOSE algorithm of the RGI analyzer). This may suggest that these hits are accidental, and that the identified genes are not truly AMR genes, or that these are novel, emergent threats and more distant homologs of known reference genes.

Table 2. Putative antimicrobial resistance genes found in the *Ensifer* sp. M14 genome.

Scaffold	Gene ID	CARD Database Hit	Predicted Resistance to	Tested Antibiotics
Scaffold_4	BLJAPNOD_00187- BLJAPNOD_00188	acrAB	Fluoroquinolone Tetracyclines	CIP (S); MXF (S) TE (S/R)
Scaffold_4	BLJAPNOD_00458	cmlA/floR	Chloramphenicol	C (R)
Scaffold_4	BLJAPNOD_00485- BLJAPNOD_00487	acrAB-TolC	Tetracyclines Cephalosporins Penams Phenicols Rifamycins Fluoroquinolones	TE (S/R) CFM (S); CRO (S); CTX (S) AMP (R) C (R) RD (R) CIP (S); MXF (S)
Scaffold_4	BLJAPNOD_00960	aph(3')-IIa	Aminoglycosides	CN (S)
Scaffold_4	BLJAPNOD_01284	adeF	Fluoroquinolones Tetracyclines	CIP (S); MXF (S) TE (S/R)
Scaffold_4	BLJAPNOD_02256	bla _{OXA}	Cephalosporins Penams	CFM (S); CRO (S); CTX (S) AMP (R)
Scaffold_4	BLJAPNOD_02798	aph(6)-Ic	Aminoglycosides	CN (S)
Scaffold_7	BLJAPNOD_04982	aph(3')-Ib	Aminoglycosides	CN (S)
Scaffold_8	BLJAPNOD_05149- BLJAPNOD_05151	acrAB-TolC	Tetracyclines Cephalosporins Penams Phenicols Rifamycins Fluoroquinolones	TE (S/R) CFM (S); CRO (S); CTX (S) AMP (R) C (R) RD (R) CIP (S); MXF (S)
Scaffold_14	BLJAPNOD_05841- BLJAPNOD_05842	acrAB	Fluoroquinolone Tetracyclines	CIP (S); MXF (S) TE (S/R)
Scaffold_17	BLJAPNOD_06442	dfrA12	Trimethoprim	TM (S)
Scaffold_18	BLJAPNOD_06615	aph(6)-Ic	Aminoglycosides	CN (S)

The most significant hits, defined with the usage of the STRICT algorithm of the RGI analyzer, are bolded. Abbreviations: AMP—ampicillin; C—chloramphenicol; CN—gentamicin; CFM—cefixime; CTX—cefotaxime, CRO—ceftriaxone; CIP—ciprofloxacin; TE—tetracycline; TM—trimethoprim; MXF—moxifloxacin; RIF—rifampicin; R—resistant; S—susceptibility; S/R—inability of interpretation of the result (threshold value).

Previous analyses revealed that the closely related organism *E. adhaerens* OV14 displays resistance to numerous antibiotics, including, among others, ampicillin, spectinomycin, kanamycin, and carbenicillin [73]. Therefore, to check whether the predicted antibiotic resistance genes truly associated with antibiotic resistance in *Ensifer* sp. M14, the MICs of 11 antibiotics were determined using Etests. Results from the Etests showed that *Ensifer* sp. M14 is resistant to ampicillin (MIC: 12.0 mg L⁻¹), chloramphenicol (MIC: 8.0 mg L⁻¹), and rifampicin (MIC 4.0 mg L⁻¹), while it is susceptible to cefixime, cefotaxime, ceftriaxone, ciprofloxacin, gentamicin, moxifloxacin, and trimethoprim. In the case of tetracycline, the MIC values fluctuated around the threshold for classification as resistant (1–4 mg L⁻¹); hence, precise interpretation of this result is not possible. Resistance to antibiotics belonging to the penams, phenicols, and rifamycins families may be explained by the presence of efflux pumps belonging to the RND family. These multidrug resistance systems are highly prevalent in Gram-negative bacteria, and play an important role in resistance to various types of stress factors, including antibiotics [74]. It is also worth mentioning, that environmental isolates of *Alphaproteobacteria* usually possess several copies of genetic modules encoding RND type multidrug efflux systems, which may be linked with their adaptation to the heterogeneity of the soil habitat [75,76]. Therefore, we think that the environmental release of *Ensifer* sp. M14 is unlikely to pose a biosafety risk.

3.6. Development of a Pilot-Scale Installation for Arsenic Bioremediation

The genomic analyses suggested that *Ensifer* sp. M14 contains several genetic features that may allow it to be successfully used in environmental bioremediation applications. In addition, previous experimental studies demonstrated that this strain can efficiently transform As(III) into As(V) (24 h of residence time was sufficient to oxidize 5 mg L⁻¹ of As(III) in the laboratory) [9]. We therefore attempted to prepare an installation for environmental bioremediation of arsenic contaminated water involving *Ensifer* sp. M14. The purification of arsenic contaminated waters constitutes a serious environmental challenge, as most of the available chemical and physical methods are dedicated to the selective removal of As(V), and are inefficient with regard to As(III). Thus, the aim of the microbiological module of the installation was to harness the arsenite oxidation capabilities of *Ensifer* sp. M14 to ensure efficient oxidation of As(III) to facilitate its subsequent removal. We reasoned that combining a biological approach with an appropriate physicochemical process (i.e., adsorption) could overcome the constraints and reservations of the conventional methods dedicated to the removal of arsenic from contaminated waters [77,78].

3.7. The Activity and Characterization of the Microbiological Module of the Pilot-Scale Installation

In our preliminary study [9], we observed that efficient functioning of the laboratory-scale installation required a high density of *Ensifer* sp. M14 (OD₆₀₀ between 0.1 and 0.2). This is in part because the quantity of *Ensifer* sp. M14 usually decreases quite intensively during the first hours/days of continuous culturing in the bioreactor [9]. Although appropriate growth conditions and length of residence time during continuous cultures were previously determined [9], the move from the laboratory-scale to pilot-scale installation meant it was necessary to re-evaluate them. In particular, replacement of the synthetic medium by natural arsenic contaminated water, as well as increasing the scale of application, may result in a deceleration of bacterial growth and a decrease in the efficiency of the biooxidation processes [79].

3.7.1. Microbial Growth and Efficiency of Arsenic Biooxidation in the Bioreactor

Using the start-up procedures described in the Materials and Methods, the initial quantity of bacteria in the bioreactor after yeast extract augmentation was about 10⁸ CFU mL⁻¹ (Figure 4). The value was almost nine orders of magnitude higher compared to raw arsenic-contaminated water, where the CFU mL⁻¹ (when plated on Luria-Bertani agar medium) was about 10⁰.

As expected based on our preliminary study [9], the density of bacteria decreased systematically during the first few days of operation, reaching a density on the magnitude of 10³ CFU mL⁻¹ on day seven (Figure 4A). After this point, the density of bacteria largely stabilized, with the exception of a few days (days 17–20), when an ~100-fold drop in bacterial density was observed (Figure 4A). A bacterial concentration of 10³ CFU mL⁻¹ in the bioreactor generally appeared sufficient for efficient biooxidation of the arsenite in the contaminated water, as there was generally little to no arsenite detected in the water following passage through the bioreactor (Figure 4D). The exceptions were five of the nine measurements taken between days 15 and 31, inclusive, when arsenite accounted for up to 62.86% of the total arsenic concentration; this corresponded with the drop in the density of bacteria within the bioreactor (Figure 4A). Thus, the low arsenite concentration throughout the majority of the experiment suggests that the microbiological module efficiently converted the As(III) to As(V).

Recently, Tardy et al. [52] showed that efficient arsenite biooxidation in environmental samples of water at 20 °C occurred after eight days of culture (batch experiment), and the quantity of bacteria at the end of their experiment was about 10⁵ CFU mL⁻¹. On the other hand, Kamde et al. [80] reported that arsenic removal was most intensive when the quantity of bacteria was about 28 CFU mL⁻¹ (batch cultures with the use on synthetic medium). The higher quantity of bacteria in the abovementioned papers in comparison with our study is presumably related to differences in culture conditions (various media and/or culturing methods).

Our data (Figure 4A,D) is also consistent with a relationship between the quantity of *Ensifer* sp. M14 and the efficiency of arsenic biooxidation, as were our preliminary experiments in batch cultures (data not shown). Indeed, many studies have observed a positive correlation between the density of bacteria and the rate of metal metabolism or biotransformation for arsenic compounds and other elements [80–82].

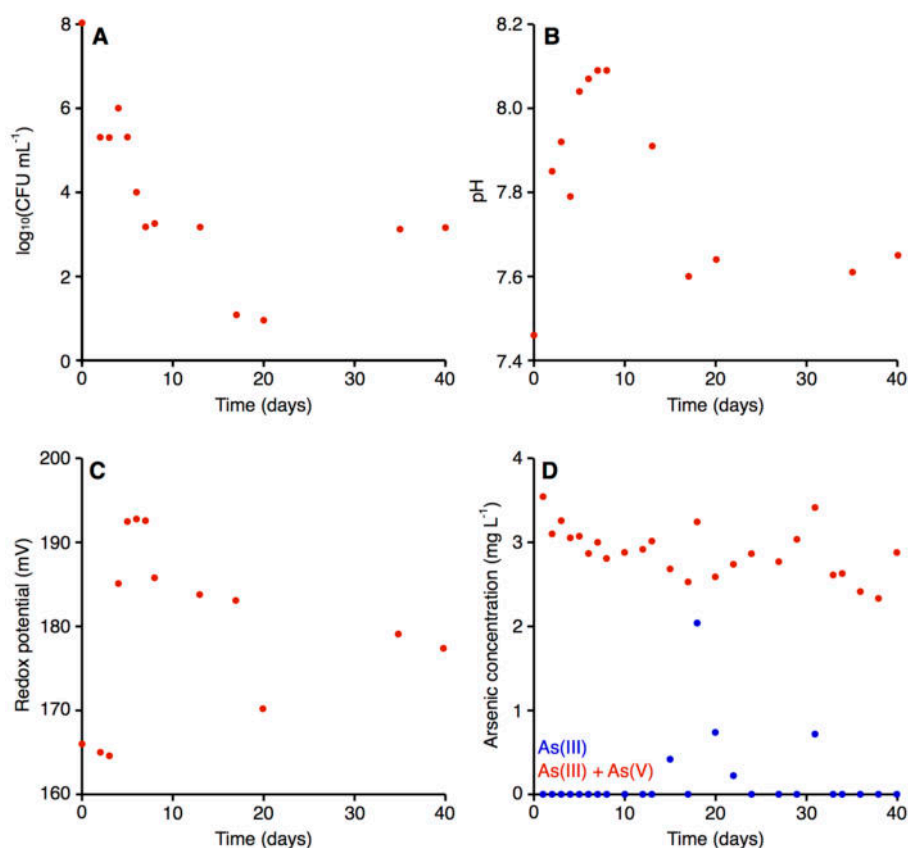


Figure 4. Parameters of the water in the bioreactor of the microbiological module. The graphs display (A) the quantity of bacteria, (B) the pH of the water, (C) the redox potential of the water, and (D) the concentration of As(III) (blue) and total arsenic (red) in the water.

3.7.2. Physical and Chemical Characterization of the Bioreactor

Previous studies have observed that there is a relationship between pH and redox potential with the arsenite/arsenate ratio; arsenites are the predominant form in reducing conditions and lower pH values, as the concentration of the arsenate form increases, both pH and redox potential also increase [83,84]. We therefore evaluated the pH and the redox potential in the treated water. For both parameters, the biological treatment had a small but noticeable effect. In the raw water, the pH and the redox potential were 7.48 and 170.90 mV, respectively [56]. In the case of the pH, the value in the bioreactor systematically increased up to the eighth day, with the treated water reaching a pH of 8.09 (Figure 4B). The pH returned to 7.60 by day 17, following which the pH stabilized in the range of 7.60 to 7.65 until the end of the experiment (day 40). In general, the redox potential remained relatively stable (Figure 4C). For the first three days, a value around 155.00 mV was observed, following which the redox potential increased and stabilized (with a slight, gradual decrease) within a range from 177.00 mV and 193.00 mV, with the exception of day 20. Water for human consumption is expected to have a pH in the range of 6.50–9.00 [85] and a redox potential between 100 and 300.00 mV [86]. Thus, both the pH and the redox potential of the water treated with our installation fell within the acceptable range for drinking water.

3.8. Effectiveness of the Adsorption Module of the Pilot-Scale Installation

Granulated bog iron ores are characterized by high arsenic adsorption capacity (up to 5.72 mg kg^{-1} , depending on the adsorbate concentration), short residence time (20 min) [56], they display high chemical stability, and they are resistant to bioweathering processes [87]. These properties allow this material to function as an effective adsorbent for removal of arsenics from contaminated water in both passive and active remediation systems, as demonstrated in our earlier work [56]. Here, we have coupled the use of granulated bog iron ores as an input to the adsorption module as well as the microbiological module described above, as a way to ensure the efficient conversion of As(III) to As(V) by *Ensifer* sp. M14, followed by the removal of As(V) by the bog iron ores. The pre-conversion of As(III) to As(V) is important as bog iron ores saturated with As(V) display higher chemical stability than bog iron ores saturated with As(III) [87].

Treatment of the arsenic contaminated water with the pilot-scale installation led to a dramatic decrease in arsenic concentrations, going from $2400 \mu\text{g L}^{-1}$ in the raw water to less than $10 \mu\text{g L}^{-1}$ (Figure 5). Analysis of the breakthrough curves for each of the adsorption columns indicated that the adsorbent in none of the columns reached equilibrium saturation during the 40-day experiment (Figure 5). Equilibrium saturation is herein defined as the maximum adsorption capacity (full saturation) of the adsorbent at a given concentration of the adsorbate; i.e., when the arsenic concentration in the input and output water is equal. Upon reaching equilibrium saturation, the adsorbent would be completely consumed and unable to further remove arsenic from the water, and it would therefore require regeneration or replacement. As the total arsenic concentration in water after each column was lower than the water entering the column, none of the columns reached equilibrium saturation. Thus, under the tested environmental conditions, the pilot-scale installation is expected to have been able to effectively continue the bioremediation process for much longer than the 40 days of the experiment (during which, 8 m^3 of water flowed through the system).

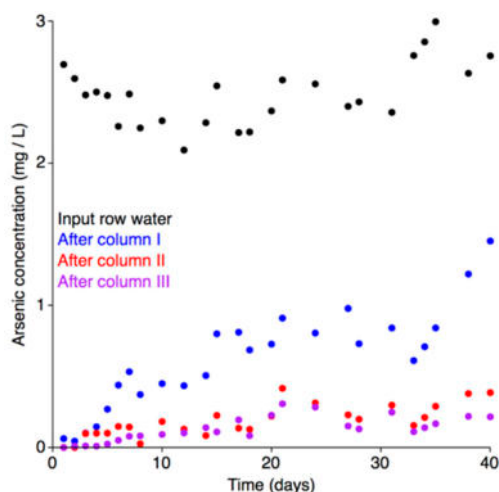


Figure 5. Arsenic adsorption breakthrough curves. The arsenic adsorption breakthrough curves for each column of the adsorption module are shown. Total arsenic concentrations in the raw water (black), and after column I (blue), column II (red), and column III (purple) are shown.

In Poland, the Regulation of the Polish Ministry of the Environment [88] currently sets the upper limit for arsenic contamination in water for use in technological purposes at $100 \mu\text{g L}^{-1}$. In the experiment reported here, the arsenic concentration in the treated water remained below $100 \mu\text{g L}^{-1}$ for the first ten days of the experiment (Figure 5), and never exceeded $220 \mu\text{g L}^{-1}$ during the 40 day test. Thus, at least the first 2.0 m^3 of water treated by pilot-scale installation was below the Polish limit for use in technological purposes. However, if pooling the treated water (and thus averaging the

arsenic concentration), it is likely that the cumulative concentration of arsenic in the 4.0 m³ of water treated over the first 20 days remained below the limit.

The local adsorption capacity of the adsorbent varied between the columns and depended on the arsenic concentration of the inflowing water. The adsorbent from the first column was characterized by the highest adsorption capacity, which was 0.500 mg kg⁻¹. Adsorbent placed in the second and third columns had lower adsorption capacities of 0.031 and 0.021 mg kg⁻¹, respectively. Likely, these differences are due to the later columns adsorbing less arsenic and being farther from reaching equilibrium saturation. The adsorption capacities recorded in the current study were significantly lower than those described in our previous work, presumably due to the adsorbent not reaching equilibrium saturation [56].

4. Conclusions

Here, we reported the draft genome sequence of *Ensifer* sp. M14 in order to gain insights into the genomic adaptation of this organism to the stressful environment of the abandoned Zloty Stok gold mine from which it was isolated. In addition, we were interested in the genetic basis of the strains arsenic oxidation and resistance capabilities, resistance to arsenic and other heavy metals, and the biosafety of the strain for use in biotechnological applications. The results revealed hundreds of genes present in *Ensifer* sp. M14 that are not found in related species, and these genes are often colocalized in genomic islands. However, the majority of these genes encoded hypothetical proteins of unknown function. Based on the genome sequence, it appears that the majority of the genes have been acquired to deal with the hostile environment of the Zloty Stok gold mine, i.e., conferring resistance to heavy metals, and enabling arsenic oxidation, are located on the self-transmissible pSinA and pSinB megaplasmids. Additionally, analysis of the *Ensifer* sp. M14 genome suggested that this strain should be safe for use in biotechnology and bioremediation. However, it was noted that several putative antibiotic resistance genes are present in the genome, as is also true for the related strain *Ensifer adhaerens* OV14 that is used in biotechnological applications [89]. This property of *Ensifer* sp. M14 should be kept in mind during its application in order to limit the spread of antimicrobial resistance. The results of these genomic analyses provide hints into the genetic potential of *Ensifer* sp. M14. They will help focus future experimental research aimed at further characterizing the biology of this organism, and may contribute to the development of procedures for large-scale cultivation of this strain.

This study also reports the construction and validation of a pilot-scale installation designed for the remediation of arsenic contaminated waters. This novel installation couples a microbiological module, based on the arsenic oxidation abilities of *Ensifer* sp. M14, with an adsorption module, based on the use of granulated bog iron ores. The underlying principle is to use *Ensifer* sp. M14 to efficiently oxidize the As(III) ions to As(V), followed by the removal of the As(V) through adsorption by the bog iron ores. Characterization of the arsenic contaminated water following passage through the microbiological module generally revealed little to no As(III), consistent with the *Ensifer* sp. M14 generally ensuring effective conversion of As(III) to As(V). Additionally, a dramatic decrease (from 10-fold to greater than 250-fold) in the arsenic concentration of the water was observed following passage of this water through the adsorption module. These results therefore confirm the effectiveness of the tested installation for the remediation of arsenic contaminated waters, which pose risks to both the environmental and human health. Future work will be aimed at further developing and optimizing this system, which could involve, for example, the addition of beads to the reactor containing *Ensifer* sp. M14 biofilms.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/9/8/379/s1>. Table S1: Accession numbers for all genomes used in this work, Table S2: In silico test of the metabolic capacity of *Ensifer* sp. M14, Table S3: Orthologous groupings of six *Rhizobiaceae* strains, including three AOB and three strains that are not AOB, Table S4: COG category descriptions, Figure S1: Average nucleotide identity matrix. A matrix of the two-way ANI values for 46 *Sinorhizobium/Ensifer* strains is shown. Clustering was performed along both axes using hierarchical clustering with Pearson distance and average linkage, Figure S2: Average amino acid identity matrix. A matrix of the two-way AAI values for 46 *Sinorhizobium/Ensifer* strains is shown.

Clustering was performed along both axes using hierarchical clustering with Pearson distance and average linkage, Figure S3: Entner–Duodoroff pathway and the pentose phosphate pathway. A modified version of the KEGG pathway map ko00030 [90] displaying the Entner–Duodoroff pathway and the pentose phosphate pathway is shown. Reactions encoded by the *Ensifer* sp. M14 genome are colored green; those in white are missing. The figure was prepared using the KAAS webserver [91] using BLAST search with the bi-directional best hit assignment method, and with the default organism list for ‘prokaryotes’ plus *Sinorhizobium meliloti* Rm1021, Figure S4: Gluconeogenesis. A modified version of the KEGG pathway map ko00010 [90] displaying the pathway for gluconeogenesis is shown. Reactions encoded by the *Ensifer* sp. M14 genome are colored green; those in white are missing. The figure was prepared using the KAAS webserver [91] using BLAST search with the bi-directional best hit assignment method, and with the default organism list for ‘prokaryotes’ plus *Sinorhizobium meliloti* Rm1021, Figure S5: Tricarboxylic acid cycle. A modified version of the KEGG pathway map ko00020 [90] displaying the tricarboxylic acid (TCA) cycle is shown. Reactions encoded by the *Ensifer* sp. M14 genome are colored green; those in white are missing. The figure was prepared using the KAAS webserver [91] using BLAST search with the bi-directional best hit assignment method, and with the default organism list for ‘prokaryotes’ plus *Sinorhizobium meliloti* Rm1021, Data Set S1: *Sinorhizobium/Ensifer* gene presence and absence. This file contains the gene presence/absence output data from Roary for the pangenome analysis of 46 *Sinorhizobium/Ensifer* strains. Details on the information provided in the file is available at: <https://sanger-pathogens.github.io/Roary/>, Data Set S2: Metabolic reconstruction of *Ensifer* sp. M14. This archive contains the expanded, draft metabolic reconstruction of *Ensifer* sp. M14. The reconstruction is provided in COBRA format as a MATLAB file, as well as in a table within an Excel workbook. A readme file is included to explain the contents, Data Set S3: PhiSpy phage prediction. This file contains the PhiSpy phage prediction output for all *Ensifer* sp. M14 genes, as well as separate sheets for each of the putative prophage loci. Details on the information provided in the file is available at: <https://github.com/linsalrob/PhiSpy>, Data Set S4: Functional annotation of the *Ensifer* sp. M14 genome. This file contains the genome annotation and the eggNOG-mapper output for three sets of genes: (i) all genes in the *Ensifer* sp. M14 genome; (ii) all genes unique to the *Ensifer* sp. M14 genome; and (iii) all genes unique and common to the *Ensifer* sp. M14 and A49 genomes. Details on the eggNOG-mapper output provided in the file is available at: <https://github.com/jhcepas/eggnog-mapper/wiki/Results-Interpretation>.

Author Contributions: Conceptualization, G.C.d., K.D., A.M., L.D., and L.Dr.; Methodology, G.C.d., K.D., and J.K.; Software, G.C.d., A.M., C.F., and L.Dz.; Validation, G.C.d., K.D., A.M., L.Dz., and L.Dr.; Formal Analysis, G.C.d. and K.D.; Investigation, K.D., W.U., A.G., G.R., and T.B.; Resources, K.D., A.M., and L.Dr.; Data Curation, G.C.d., A.M., C.F., and L.Dz.; Writing-Original Draft Preparation, G.C.d., K.D., and L.Dr.; Writing-Review & Editing, W.U., A.M., L.Dz., T.B., and G.R.; Visualization, G.C.d., K.D., and J.K.; Supervision, A.M. and L.Dr.; Project Administration, A.M. and L.Dr.; Funding Acquisition, K.D., A.M., and L.Dr.

Funding: This research was funded by National Centre for Research and Development (Poland) grant number LIDER/043/403/L-4/12/NCBR/2013, National Science Center (Poland) grant no. 2016/23/N/NZ9/01655 and intramural funding from the University of Florence, call “PROGETTI STRATEGICI DI ATENE ANNO 2014”.

Acknowledgments: G.C.d. was supported by a Post-Doctoral Fellowship from the Natural Sciences and Engineering Council of Canada. K.D. was supported by the European Molecular Biology Organization in the frame of the EMBO Short-Term Fellowship program [grant number 7376].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alisi, C.; Musella, R.; Tasso, F.; Ubaldi, C.; Manzo, S.; Cremisini, C.; Sprocati, A.R. Bioremediation of diesel oil in a co-contaminated soil by bioaugmentation with a microbial formula tailored with native strains selected for heavy metals resistance. *Sci. Total Environ.* **2009**, *407*, 3024–3032. [CrossRef] [PubMed]
2. Kuppusamy, S.; Thavamani, P.; Megharaj, M.; Lee, Y.B.; Naidu, R. Polyaromatic hydrocarbon (PAH) degradation potential of a new acid tolerant, diazotrophic P-solubilizing and heavy metal resistant bacterium *Cupriavidus* sp. MTS-7 isolated from long-term mixed contaminated soil. *Chemosphere* **2016**, *162*, 31–39. [CrossRef] [PubMed]
3. Yu, D.; Yang, J.; Teng, F.; Feng, L.; Fang, X.; Ren, H. Bioaugmentation treatment of mature landfill leachate by new isolated ammonia nitrogen and humic acid resistant microorganism. *J. Microbiol. Biotechnol.* **2014**, *24*, 987–997. [CrossRef] [PubMed]
4. Pepper, I.L.; Gentry, T.J.; Newby, D.T.; Roane, T.M.; Josephson, K.L. The role of cell bioaugmentation and gene bioaugmentation in the remediation of co-contaminated soils. *Environ. Health Perspect.* **2002**, *110* (Suppl. 6), 943–946. [CrossRef] [PubMed]
5. Feng, Z.; Li, X.; Lu, C.; Shen, Z.; Xu, F.; Chen, Y. Characterization of *Pseudomonas mendocina* LR capable of removing nitrogen from various nitrogen-contaminated water samples when cultivated with *Cyperus alternifolius* L. *J. Biosci. Bioeng.* **2012**, *114*, 182–187. [CrossRef] [PubMed]

6. Lee, P.K.H.; Warnecke, F.; Brodie, E.L.; Macbeth, T.W.; Conrad, M.E.; Andersen, G.L.; Alvarez-Cohen, L. Phylogenetic microarray analysis of a microbial community performing reductive dechlorination at a TCE-contaminated site. *Environ. Sci. Technol.* **2012**, *46*, 1044–1054. [[CrossRef](#)] [[PubMed](#)]
7. Drewniak, L.; Matlakowska, R.; Sklodowska, A. Arsenite and arsenate metabolism of *Sinorhizobium* sp. M14 living in the extreme environment of the Zloty Stok gold mine. *Geomicrobiol. J.* **2008**, *25*, 363–370. [[CrossRef](#)]
8. Drewniak, L.; Krawczyk, P.S.; Mielnicki, S.; Adamska, D.; Sobczak, A.; Lipinski, L.; Burec-Drewniak, W.; Sklodowska, A. Physiological and metagenomic analyses of microbial mats involved in self-purification of mine waters contaminated with heavy metals. *Front. Microbiol.* **2016**, *7*, 1252. [[CrossRef](#)] [[PubMed](#)]
9. Debiec, K.; Krzysztoforski, J.; Uhrynowski, W.; Sklodowska, A.; Drewniak, L. Kinetics of arsenite oxidation by *Sinorhizobium* sp. M14 under changing environmental conditions. *Int. Biodeterior. Biodegrad.* **2017**, *119*, 476–485. [[CrossRef](#)]
10. Drewniak, L.; Dziewit, L.; Cieczkowska, M.; Gawor, J.; Gromadka, R.; Sklodowska, A. Structural and functional genomics of plasmid pSinA of *Sinorhizobium* sp. M14 encoding genes for the arsenite oxidation and arsenic resistance. *J. Biotechnol.* **2013**, *164*, 479–488. [[CrossRef](#)] [[PubMed](#)]
11. Dziewit, L.; Bartosik, D. Comparative analyses of extrachromosomal bacterial replicons, identification of chromids, and experimental evaluation of their indispensability. *Methods Mol. Biol.* **2015**, *1231*, 15–29. [[CrossRef](#)] [[PubMed](#)]
12. Romaniuk, K.; Dziewit, L.; Decewicz, P.; Mielnicki, S.; Radlinska, M.; Drewniak, L. Molecular characterization of the pSinB plasmid of the arsenite oxidizing, metallotolerant *Sinorhizobium* sp. M14—Insight into the heavy metal resistome of sinorhizobial extrachromosomal replicons. *FEMS Microbiol. Ecol.* **2017**, *93*, fiw215. [[CrossRef](#)] [[PubMed](#)]
13. Doyle, J.; Doyle, J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
14. Joint Genome Institute JGI Bacterial DNA Isolation CTAB Protocol. Available online: <https://jgi.doe.gov/user-program-info/pmo-overview/protocols-sample-preparation-information/jgi-bacterial-dna-isolation-ctab-protocol-2012/> (accessed on 20 July 2018).
15. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)] [[PubMed](#)]
16. Vasilinet, I.; Pribelski, A.D.; Gurevich, A.; Korobeynikov, A.; Pevzner, P.A. Assembling short reads from jumping libraries with large insert sizes. *Bioinformatics* **2015**, *31*, 3262–3268. [[CrossRef](#)] [[PubMed](#)]
17. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High-throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *bioRxiv* **2017**, 225342. [[CrossRef](#)]
18. Bosi, E.; Donati, B.; Galardini, M.; Brunetti, S.; Sagot, M.-F.; Lió, P.; Crescenzi, P.; Fani, R.; Fondi, M. MeDuSa: A multi-draft based scaffold. *Bioinformatics* **2015**, *31*, 2443–2451. [[CrossRef](#)] [[PubMed](#)]
19. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)] [[PubMed](#)]
20. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)] [[PubMed](#)]
21. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [[CrossRef](#)] [[PubMed](#)]
22. Kolbe, D.L.; Eddy, S.R. Fast filtering for RNA homology search. *Bioinformatics* **2011**, *27*, 3102–3109. [[CrossRef](#)] [[PubMed](#)]
23. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.R.; Bateman, A.; Finn, R.D.; Petrov, A.I. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **2018**, *46*, D335–D342. [[CrossRef](#)] [[PubMed](#)]
24. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)] [[PubMed](#)]
25. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)] [[PubMed](#)]

26. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **2014**, *1079*, 155–170. [[CrossRef](#)] [[PubMed](#)]
27. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
28. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [[CrossRef](#)] [[PubMed](#)]
29. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [[CrossRef](#)] [[PubMed](#)]
30. Luo, C.; Rodriguez-R, L.M.; Konstantinidis, K.T. MyTaxa: An advanced taxonomic classifier for genomic and metagenomic sequences. *Nucleic Acids Res.* **2014**, *42*, e73. [[CrossRef](#)] [[PubMed](#)]
31. Hao, X.; Lin, Y.; Johnstone, L.; Liu, G.; Wang, G.; Wei, G.; McDermott, T.; Rensing, C. Genome sequence of the arsenite-oxidizing strain *Agrobacterium tumefaciens* 5A. *J. Bacteriol.* **2012**, *194*, 903. [[CrossRef](#)] [[PubMed](#)]
32. Henkel, C.V.; den Dulk-Ras, A.; Zhang, X.; Hooykaas, P.J.J. Genome sequence of the octopine-type *Agrobacterium tumefaciens* strain Ach5. *Genome Announc.* **2014**, *2*, e00225-14. [[CrossRef](#)] [[PubMed](#)]
33. Rudder, S.; Doohan, F.; Creevey, C.J.; Wendt, T.; Mullins, E. Genome sequence of *Ensifer adhaerens* OV14 provides insights into its ability as a novel vector for the genetic transformation of plant genomes. *BMC Genom.* **2014**, *15*, 268. [[CrossRef](#)] [[PubMed](#)]
34. Österman, J.; Marsh, J.; Laine, P.K.; Zeng, Z.; Alatalo, E.; Sullivan, J.T.; Young, J.P.W.; Thomas-Oates, J.; Paulin, L.; Lindström, K. Genome sequencing of two *Neorhizobium galegae* strains reveals a *noeT* gene responsible for the unusual acetylation of the nodulation factors. *BMC Genom.* **2014**, *15*, 500. [[CrossRef](#)] [[PubMed](#)]
35. Andres, J.; Arsène-Ploetze, F.; Barbe, V.; Brochier-Armanet, C.; Cleiss-Arnold, J.; Coppée, J.-Y.; Dillies, M.-A.; Geist, L.; Joublin, A.; Koechler, S.; et al. Life in an arsenic-containing gold mine: Genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium* sp. NT-26. *Genome Biol. Evol.* **2013**, *5*, 934–953. [[CrossRef](#)] [[PubMed](#)]
36. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **2018**, *36*, 566–569. [[CrossRef](#)] [[PubMed](#)]
37. Li, L.; Stoekert, C.J.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)] [[PubMed](#)]
38. Li, X.; Hu, Y.; Gong, J.; Lin, Y.; Johnstone, L.; Rensing, C.; Wang, G. Genome sequence of the highly efficient arsenite-oxidizing bacterium *Achromobacter arsenitoxydans* SY8. *J. Bacteriol.* **2012**, *194*, 1243–1244. [[CrossRef](#)] [[PubMed](#)]
39. Muller, D.; Médigue, C.; Koechler, S.; Barbe, V.; Barakat, M.; Talla, E.; Bonnefoy, V.; Krin, E.; Arsène-Ploetze, F.; Carapito, C.; et al. A tale of two oxidation states: Bacterial colonization of arsenic-rich environments. *PLoS Genet.* **2007**, *3*, e53. [[CrossRef](#)] [[PubMed](#)]
40. Li, X.; Gong, J.; Hu, Y.; Cai, L.; Johnstone, L.; Grass, G.; Rensing, C.; Wang, G. Genome sequence of the moderately halotolerant, arsenite-oxidizing bacterium *Pseudomonas stutzeri* TS44. *J. Bacteriol.* **2012**, *194*, 4473–4474. [[CrossRef](#)] [[PubMed](#)]
41. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)] [[PubMed](#)]
42. Akhter, S.; Aziz, R.K.; Edwards, R.A. PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **2012**, *40*, e126. [[CrossRef](#)] [[PubMed](#)]
43. McArthur, A.G.; Wagleichner, N.; Nizam, F.; Yan, A.; Azad, M.A.; Baylay, A.J.; Bhullar, K.; Canova, M.J.; De Pascale, G.; Ejim, L.; et al. The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **2013**, *57*, 3348–3357. [[CrossRef](#)] [[PubMed](#)]
44. Kahlmeter, G.; Brown, D.F.J.; Goldstein, F.W.; MacGowan, A.P.; Mouton, J.W.; Odenholt, I.; Rodloff, A.; Soussy, C.-J.; Steinbakk, M.; Soriano, F.; et al. European Committee on Antimicrobial Susceptibility Testing (EUCAST) Technical notes on antimicrobial susceptibility testing. *Clin. Microbiol. Infect.* **2006**, *12*, 501–503. [[CrossRef](#)] [[PubMed](#)]

45. EUCAST European Committee on Antimicrobial Susceptibility Testing. Available online: <http://www.eucast.org> (accessed on 20 July 2018).
46. Eddy, S.R. A new generation of homology search tools based on probabilistic inference. *Genome Inform.* **2009**, *23*, 205–211. [[CrossRef](#)] [[PubMed](#)]
47. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [[CrossRef](#)] [[PubMed](#)]
48. Haft, D.H.; Selengut, J.D.; Richter, R.A.; Harkins, D.; Basu, M.K.; Beck, E. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* **2012**, *41*, D387–D395. [[CrossRef](#)] [[PubMed](#)]
49. Keating, S.M.; Bornstein, B.J.; Finney, A.; Hucka, M. SBMLToolbox: An SBML toolbox for MATLAB users. *Bioinformatics* **2006**, *22*, 1275–1277. [[CrossRef](#)] [[PubMed](#)]
50. Bornstein, B.J.; Keating, S.M.; Jouraku, A.; Hucka, M. LibSBML: An API library for SBML. *Bioinformatics* **2008**, *24*, 880–881. [[CrossRef](#)] [[PubMed](#)]
51. Schellenberger, J.; Que, R.; Fleming, R.M.T.; Thiele, I.; Orth, J.D.; Feist, A.M.; Zielinski, D.C.; Bordbar, A.; Lewis, N.E.; Rahmanian, S.; et al. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6*, 1290–1307. [[CrossRef](#)] [[PubMed](#)]
52. DiCenzo, G.; Mengoni, A.; Fondi, M. Tn-Core: Context-specific reconstruction of core metabolic models using Tn-seq data. *bioRxiv* **2017**, 221325. [[CrossRef](#)]
53. DiCenzo, G.C.; Checucci, A.; Bazzicalupo, M.; Mengoni, A.; Viti, C.; Dziewit, L.; Finan, T.M.; Galardini, M.; Fondi, M. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **2016**, *7*, 12219. [[CrossRef](#)] [[PubMed](#)]
54. DiCenzo, G.C.; Benedict, A.B.; Fondi, M.; Walker, G.C.; Finan, T.M.; Mengoni, A.; Griffiths, J.S. Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium *Sinorhizobium meliloti*. *PLoS Genet.* **2018**, *14*, e1007357. [[CrossRef](#)] [[PubMed](#)]
55. Blin, K.; Wolf, T.; Chevrette, M.G.; Lu, X.; Schwalen, C.J.; Kautsar, S.A.; Suarez Duran, H.G.; de los Santos, E.L.C.; Kim, H.U.; Nave, M.; et al. antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017**, *45*, W36–W41. [[CrossRef](#)] [[PubMed](#)]
56. Debiec, K.; Rzepa, G.; Bajda, T.; Uhrynowski, W.; Sklodowska, A.; Krzysztoforski, J.; Drewniak, L. Granulated bog iron ores as sorbents in passive (bio)remediation systems for arsenic removal. *Front. Chem.* **2018**, *6*, 54. [[CrossRef](#)] [[PubMed](#)]
57. Tardy, V.; Casiot, C.; Fernandez-Rojo, L.; Resongles, E.; Desoeuvre, A.; Joulian, C.; Battaglia-Brunet, F.; Héry, M. Temperature and nutrients as drivers of microbially mediated arsenic oxidation and removal from acid mine drainage. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 2413–2424. [[CrossRef](#)] [[PubMed](#)]
58. Williams, L.E.; Baltrus, D.A.; O'Donnell, S.D.; Skelly, T.J.; Martin, M.O. Complete genome sequence of the predatory bacterium *Ensifer adhaerens* Casida A. *Genome Announc.* **2017**, *5*, e01344-17. [[CrossRef](#)] [[PubMed](#)]
59. Grissa, I.; Vergnaud, G.; Pourcel, C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **2007**, *35*, W52–W57. [[CrossRef](#)] [[PubMed](#)]
60. Lycus, P.; Lovise Bøthun, K.; Bergaust, L.; Peele Shapleigh, J.; Reier Bakken, L.; Frostegård, Å. Phenotypic and genotypic richness of denitrifiers revealed by a novel isolation strategy. *ISME J.* **2017**, *11*, 2219–2232. [[CrossRef](#)] [[PubMed](#)]
61. Casida, L.E. *Ensifer adhaerens* gen. nov., sp. nov.: A bacterial predator of bacteria in soil. *Int. J. Syst. Bacteriol.* **1982**, *32*, 339–345. [[CrossRef](#)]
62. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)] [[PubMed](#)]
63. Geddes, B.A.; Oresnik, I.J. Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*. *Can. J. Microbiol.* **2014**, *60*, 491–507. [[CrossRef](#)] [[PubMed](#)]
64. Silver, S.; Phung, L.T. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl. Environ. Microbiol.* **2005**, *71*, 599–608. [[CrossRef](#)] [[PubMed](#)]
65. Hughes, M.F. Arsenic toxicity and potential mechanisms of action. *Toxicol. Lett.* **2002**, *133*, 1–16. [[CrossRef](#)]
66. Voegelé, R.T.; Bardin, S.; Finan, T.M. Characterization of the *Rhizobium* (*Sinorhizobium*) *meliloti* high- and low-affinity phosphate uptake systems. *J. Bacteriol.* **1997**, *179*, 7226–7232. [[CrossRef](#)] [[PubMed](#)]
67. Yuan, Z.-C.; Zaheer, R.; Finan, T.M. Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. *J. Bacteriol.* **2006**, *188*, 1089–1102. [[CrossRef](#)] [[PubMed](#)]

68. Elias, M.; Wellner, A.; Goldin-Azulay, K.; Chabriere, E.; Vorholt, J.A.; Erb, T.J.; Tawfik, D.S. The molecular basis of phosphate discrimination in arsenate-rich environments. *Nature* **2012**, *491*, 134–137. [[CrossRef](#)] [[PubMed](#)]
69. Pickering, B.S.; Oresnik, I.J. Formate-dependent autotrophic growth in *Sinorhizobium meliloti*. *J. Bacteriol.* **2008**, *190*, 6409–6418. [[CrossRef](#)] [[PubMed](#)]
70. Fuchs, G. Alternative pathways of carbon dioxide fixation: Insights into the early evolution of life? *Annu. Rev. Microbiol.* **2011**, *65*, 631–658. [[CrossRef](#)] [[PubMed](#)]
71. Xiu, A.; Kong, Y.; Zhou, M.; Zhu, B.; Wang, S.; Zhang, J. The chemical and digestive properties of a soluble glucan from *Agrobacterium* sp. ZX09. *Carbohydr. Polym.* **2010**, *82*, 623–628. [[CrossRef](#)]
72. Padan, E.; Tzuber, T.; Herz, K.; Kozachkov, L.; Rimon, A.; Galili, L. NhaA of *Escherichia coli*, as a model of a pH-regulated Na⁺/H⁺ antiporter. *Biochim. Biophys. Acta Bioenerg.* **2004**, *1658*, 2–13. [[CrossRef](#)] [[PubMed](#)]
73. Rathore, D.S.; Lopez-Vernaza, M.A.; Doohan, F.; Connell, D.O.; Lloyd, A.; Mullins, E. Profiling antibiotic resistance and electro-transformation potential of *Ensifer adhaerens* OV14; a non-*Agrobacterium* species capable of efficient rates of plant transformation. *FEMS Microbiol. Lett.* **2015**, *362*, fnv126. [[CrossRef](#)] [[PubMed](#)]
74. Nikaido, H. Multidrug resistance in bacteria. *Annu. Rev. Biochem.* **2009**, *78*, 119–146. [[CrossRef](#)] [[PubMed](#)]
75. Xiong, X.H.; Han, S.; Wang, J.H.; Jiang, Z.H.; Chen, W.; Jia, N.; Wei, H.L.; Cheng, H.; Yang, Y.X.; Zhu, B.; et al. Complete genome sequence of the bacterium *Ketogulonicigenium vulgare* Y25. *J. Bacteriol.* **2011**, *193*, 315–316. [[CrossRef](#)] [[PubMed](#)]
76. Martin, F.A.; Posadas, D.M.; Carrica, M.C.; Cravero, S.L.; O’Callaghan, D.; Zorreguieta, A. Interplay between two RND systems mediating antimicrobial resistance in *Brucella suis*. *J. Bacteriol.* **2009**, *191*, 2530–2540. [[CrossRef](#)] [[PubMed](#)]
77. Barakat, M.A. New trends in removing heavy metals from industrial wastewater. *Arab. J. Chem.* **2011**, *4*, 361–377. [[CrossRef](#)]
78. Lièvremon, D.; N’negue, M.A.; Behra, P.; Lett, M.C. Biological oxidation of arsenite: Batch reactor experiments in presence of kutnahorite and chabazite. *Chemosphere* **2003**, *51*, 419–428. [[CrossRef](#)]
79. Hong, J.; Silva, R.A.; Park, J.; Lee, E.; Park, J.; Kim, H. Adaptation of a mixed culture of acidophiles for a tank biooxidation of refractory gold concentrates containing a high concentration of arsenic. *J. Biosci. Bioeng.* **2016**, *121*, 536–542. [[CrossRef](#)] [[PubMed](#)]
80. Kamde, K.; Pandey, R.A.; Thul, S.T.; Dahake, R.; Shinde, V.M.; Bansawal, A. Microbially assisted arsenic removal using *Acidithiobacillus ferrooxidans* mediated by iron oxidation. *Environ. Technol. Innov.* **2018**, *10*, 78–90. [[CrossRef](#)]
81. Wang, G.; Xie, S.; Liu, X.; Wu, Y.; Liu, Y.; Zeng, T. Bio-oxidation of a high-sulfur and high-arsenic refractory gold concentrate using a two-stage process. *Miner. Eng.* **2018**, *120*, 94–101. [[CrossRef](#)]
82. Wang, H.; Zhang, X.; Zhu, M.; Tan, W. Effects of dissolved oxygen and carbon dioxide under oxygen-rich conditions on the biooxidation process of refractory gold concentrate and the microbial community. *Miner. Eng.* **2015**, *80*, 37–44. [[CrossRef](#)]
83. Wang, S.; Zhao, X. On the potential of biological treatment for arsenic contaminated soils and groundwater. *J. Environ. Manag.* **2009**, *90*, 2367–2376. [[CrossRef](#)] [[PubMed](#)]
84. Katsoyiannis, I.A.; Zouboulis, A.I. Application of biological processes for the removal of arsenic from groundwaters. *Water Res.* **2004**, *38*, 17–26. [[CrossRef](#)] [[PubMed](#)]
85. Polish Minister of Health. *Regulation of the Polish Minister of Health on the Quality of Water Intended for Human Consumption*; No. 1989; Polish Minister of Health: Warsaw, Poland, 2015.
86. Goncharuk, V.V.; Bagrii, V.A.; Mel’nik, L.A.; Chebotareva, R.D.; Bashtan, S.Y. The use of redox potential in water treatment processes. *J. Water Chem. Technol.* **2010**, *32*, 1–9. [[CrossRef](#)]
87. Debiec, K.; Rzepa, G.; Bajda, T.; Zych, L.; Krzysztoforski, J.; Sklodowska, A.; Drowniak, L. The influence of thermal treatment on bioweathering and arsenic sorption capacity of a natural iron (oxyhydr)oxide-based adsorbent. *Chemosphere* **2017**, *188*, 99–109. [[CrossRef](#)] [[PubMed](#)]
88. Polish Ministry of the Environment. *Regulation on the Conditions to Be Met during Introducing Sewage into Waters or into the Ground, and in on Substances Particularly Harmful to the Aquatic Environment*; No. 06.137.984; Polish Ministry of the Environment: Warsaw, Poland, 2006.
89. Zuniga-Soto, E.; Mullins, E.; Dedicova, B. *Ensifer*-mediated transformation: An efficient non-*Agrobacterium* protocol for the genetic modification of rice. *Springerplus* **2015**, *4*, 600. [[CrossRef](#)] [[PubMed](#)]

90. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)] [[PubMed](#)]
91. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Title: Taxonomy of *Rhizobiaceae* revisited: proposal of a new framework for genus delimitation

Authors: Nemanja Kuzmanović¹, Camilla Fagorzi², Alessio Mengoni², Florent Lassalle^{3,*}, George C diCenzo^{4,*}

Affiliations: ¹ Julius Kühn Institute, Federal Research Centre for Cultivated Plants (JKI), Institute for Plant Protection in Horticulture and Forests, Braunschweig, Germany

² Department of Biology, University of Florence, Florence, Italy

³ Parasites and Microbes, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

⁴ Department of Biology, Queen's University, Kingston, Ontario, Canada

* **Corresponding authors:** Florent Lassalle (fl4@sanger.ac.uk) and George diCenzo (george.dicenzo@queensu.ca)

Keywords: *Rhizobiaceae*, *Xaviernesmea*, *Ensifer*, *Sinorhizobium*, *Rhizobium*, genus boundaries

ABSTRACT

The alphaproteobacterial family *Rhizobiaceae* is highly diverse, with 168 species with validly published names classified into 17 genera with validly published names. Most named genera in this family are delineated based on genomic relatedness and phylogenetic relationships, but some historically named genera show inconsistent distribution and phylogenetic breadth. Most problematic is *Rhizobium*, which is notorious for being highly paraphyletic, as most newly described species in the family being assigned to this genus without consideration for their proximity to existing genera, or the need to create novel genera. In addition, many *Rhizobiaceae* genera lack synapomorphic traits that would give them biological and ecological significance. We propose a common framework for genus delimitation within the family *Rhizobiaceae*. We propose that genera in this family should be defined as monophyletic groups in a core-genome gene phylogeny, that are separated from related species using a pairwise core-proteome average amino acid identity (cpAAI) threshold of approximately 86%. We further propose that the presence of additional genomic or phenotypic evidence can justify the division of species into separate genera even if they all share greater than 86% cpAAI. Applying this framework, we propose to reclassify *Rhizobium rhizosphaerae* and *Rhizobium oryzae* into the new genus *Xaviernesmea* gen. nov. Data is also provided to support the recently proposed genus “*Peteryoungia*”, and the reclassifications of *Rhizobium yantingense* as *Endobacterium yantingense* comb. nov., *Rhizobium petrolearium* as *Neorhizobium petrolearium* comb. nov., *Rhizobium arenae* as *Pararhizobium arenae* comb. nov., *Rhizobium tarimense* as *Pseudorhizobium tarimense* comb. nov., and *Rhizobium azooxidifex* as *Mycoplana azooxidifex* comb. nov. Lastly, we present arguments that the unification of the genera *Ensifer* and *Sinorhizobium* in Opinion 84 of the Judicial Commission is no longer justified by current genomic and phenotypic data. We thus argue that the genus *Sinorhizobium* is not illegitimate and now encompasses 17 species.

INTRODUCTION

The family *Rhizobiaceae* of the order *Alphaproteobacteria* was proposed in 1938 and has since undergone numerous, and at times contentious, taxonomic revisions [1, 2]. Currently, this family comprises the genera *Agrobacterium*, *Allorhizobium*, *Ciceribacter*, *Endobacterium*, *Ensifer* (syn. *Sinorhizobium*), *Gellertiella*, *Georhizobium*, *Hoeflea*, *Lentilitoribacter*, *Liberibacter*, *Martelella*, *Mycoplana*, “*Neopararhizobium*”, *Neorhizobium*, *Pararhizobium*, “*Peteryoungia*”, *Pseudorhizobium*, *Rhizobium*, and *Shinella* (syn. *Crabtreeella*) (<https://lpsn.dsmz.de/>; [3]). The *Rhizobiaceae* family contains phenotypically diverse organisms, including N₂-fixing legume symbionts (known as rhizobia), plant pathogens, bacterial predators, and other soil bacteria. The agricultural and ecological significance of the family *Rhizobiaceae* has prompted the isolation and whole genome sequencing of hundreds of strains at a rate outpacing taxonomic refinement of the family. As a result, some species and genera within the family are well known to be paraphyletic [4], while others that are monophyletic likely represent multiple species/genera [5]. In addition, most currently named genera have been delineated based on genomic relatedness – as per current taxonomic guidelines [6] – but lack synapomorphic traits that would give them biological and ecological significance [7].

To aid in the taxonomic classification of this family, here we propose a general framework for defining genera in the family *Rhizobiaceae*. This framework is based on a set of baseline genomic relatedness measures meant to serve as minimal thresholds for genus demarcation, while allowing for more closely related species to be divided into separate genera when supported by supplemental genomic and/or biological data. By applying this framework, we propose the formation of a new genus – *Xaviernesmea* – on the basis of the genomic relatedness measures, and provide support for the recently proposed genus *Peteryoungia*. In addition, despite genomic relatedness values above the proposed baseline thresholds, we argue that current phylogenetic, genomic (e.g., pentanucleotide frequency), and biological (e.g., division by budding) data indicate that the genera *Ensifer* Casida *et al.* 1982 [8] and *Sinorhizobium* Chen 1988 [9] are not synonymous, meaning that the unification of the genera *Ensifer* and *Sinorhizobium* in Opinion 84 of the Judicial Commission is no longer justified.

METHODS

Dataset

The analysis was performed on a dataset of 94 genomes of *Rhizobiaceae* strains, among which the majority were type strains of the corresponding species (**Dataset S1**). As an outgroup, we included the genomes of three *Mesorhizobium* spp. strains, belonging to the related family *Phyllobacteriaceae*. Moreover, for calculation of some overall genome relatedness indexes to support additional taxonomic revisions, genomes of two *Pararhizobium* and two *Pseudorhizobium* strains were included (**Dataset S1**). To verify the authenticity of genomes used for taxonomic reclassifications proposed in this paper, we compared the reference 16S rRNA gene sequences (as well as housekeeping gene sequences in ambiguous cases) associated with the original species publication with the sequences retrieved from genome sequences (**Dataset S1**). Whole genome

sequences (WGSs) generated to support new species description in original publications were considered as authentic (**Dataset S1**).

Core-genome gene phylogeny

The core genome phylogeny was obtained using the GET_HOMOLOGUES software package version 10032020 [10] and the GET_PHYLOMARKERS software package Version 2.2.8_18Nov2018 [11], as described previously [12]. As a result, a set of 170 non-recombining single-copy core marker genes was selected (**Dataset S2**), and a concatenation of their codon-based alignments was used as input for IQ-Tree ModelFinder, with which a search for the best sequence evolution model was conducted. The model ‘GTR+F+ASC+R8’ was selected based on a Bayesian information criterion. The maximum likelihood (ML) core genome phylogeny was inferred under this model using IQ-TREE [13], with branch supports assessed with approximate Bayes test (-abayes) and ultrafast bootstrap with 1000 replicates (-bb 1000).

Overall Genome Relatedness Indexes calculations

Whole-proteome amino-acid identity (wpAAI; usually simply known as AAI) was computed using the CompareM software (github.com/dparks1134/CompareM) using the aai_wf command with default parameters, i.e., ortholog identification with DIAMOND [14], e-value < 1e-3, percent identity > 30%, and alignment length > 70% the length of the protein.

Core-proteome amino-acid identity (cpAAI) was computed as the proportion of substitutions in pairwise comparisons of sequences from the 170 non-recombining, single-copy core marker genes identified using GET_PHYLOMARKERS [11], using a custom R script that notably relied on the dist.aa() function from the ‘ape’ package [15].

Percentage of conserved proteins (POCP) was determined using publicly available code (github.com/hoelzer/pocp) and the ortholog identification thresholds defined by Qin *et al.* [16], namely, e-value < 1e-5, percent identity > 40%, and alignment length > 50% the length of the protein. This pipeline involved the reannotation of genomes with Prodigal version 2.6.3 [17] and ortholog identification using the BLAST+ package, version 2.10.1 [18].

The average nucleotide identity (ANI) comparisons were conducted using PyANI version 0.2.9, with scripts employing the BLAST+ (ANiB) algorithm to align the input sequences (github.com/widdowquinn/pyani). The *in silico* DNA-DNA hybridization (isDDH) computations were performed with the Genome-to-Genome Distance Calculator (GGDC 2.1; ggdc.dsmz.de/distcalc2.php) using the recommended BLAST+ alignment and formula 2 (identities/HSP length) [19].

16S rRNA gene phylogeny

The RNA fasta files for the 157 *Sinorhizobium* or *Ensifer* strains analyzed in our recent study [20] were downloaded from the National Center for Biotechnology Information database, and all 16S rRNA gene sequences ≥ 1000 nt were extracted. The 16S rRNA gene sequences were aligned using MAFFT version 7.3.10 with the localpair option [21], and trimmed using trimAl version

1.4.rev22 with the automated1 option [22]. A ML phylogeny was prepared using raxmlHPC-HYBRID-AVX2 version 8.2.12 with the GTRCAT model [23]. The final phylogeny is the bootstrap best tree following 756 bootstrap replicates, as determined by the extended majority-rule consensus tree criterion.

RESULTS AND DISCUSSION

Overall genomic relatedness indexes measurements in the family *Rhizobiaceae*

To develop a framework for genus demarcation within the family *Rhizobiaceae*, we examined a selection of 94 genomes of *Rhizobiaceae* isolates, most of which are species type strains. *Liberibacter*, an obligate intra-cellular pathogen with a highly reduced genome, was excluded from our selection of organisms to avoid biasing the analysis by overly reducing the conserved gene set. We reasoned that good practices for genome sequence-based genus delineation should consider both phylogenetic relatedness of species based on a concatenated alignment of core-genome genes (**Figures 1, S1**), and one or more overall genomic relatedness index (OGRI) measurement. We initially considered three OGRIs, calculated as described in the Methods: (i) whole-proteome average amino acid identity (wpAAI); (ii) core-proteome average amino acid identity (cpAAI) based on the proportion of substitutions between the concatenated translated sequences of the core marker gene set used for the core-genome phylogeny; and (iii) the percentage of conserved proteins (POCP) as defined by Qin *et al.* [16]. Average nucleotide identity (ANIb; **Dataset S3**) and *in silico* DNA-DNA hybridization (isDDH; **Dataset S4**) were performed for some strains to verify they represented distinct species; however, these OGRIs were not considered when defining genera as it has been argued that these measures are more appropriate for intra-genus analyses rather than inter-genus comparisons [16].

We reasoned that an OGRI threshold for delimiting genera should correspond to a drop in the OGRI frequency distribution. We therefore plotted histograms of all pairwise comparisons to identify potential genera boundaries (**Figures 2, S2, S3**). It was previously suggested that a 50% POCP threshold is a good measure of genus boundaries in other families [16]. However, we found that 3,885 out of the 4,371 (89%) pairwise comparisons in our *Rhizobiaceae* dataset gave a POCP value $\geq 50\%$, with no clear breaks in the frequency distribution (**Figure S2**). We therefore concluded that POCP is not a useful OGRI measurement for defining genera in the family *Rhizobiaceae*.

cpAAI data was recently used to delineate genera among other bacterial families [24] and stands as a promising metric for genus demarcation in the family *Rhizobiaceae* (**Figure 2**). We observed a break in the frequency distribution at $\sim 93\%$ to $\sim 94\%$, but it was too stringent to use for genus demarcation as it would result in the majority of the 94 strains being classified into their own genera. Likewise, the break at $\sim 73\%$ to $\sim 74\%$ was too lenient for genus delimitation as all strains would be grouped as a single genus, except for those belonging to the genera *Martellella*, *Hoeflea*, “*Neopararhizobium*”, and *Georhizobium*. Instead, the drop in the frequency distribution at 86% to 86.7% (inclusive), within which only five of the 4,371 pairwise comparisons fell, appeared to be a reasonable threshold to aid with defining genera in the *Rhizobiaceae* family

(**Figure 2**). Using a cpAAI threshold of $\sim 86\%$, combined with the phylogeny of **Figure 1**, we were able to recover the genera *Agrobacterium*, *Ciceribacter*, *Endobacterium*, *Ensifer* (as previously defined), *Gellertiella*, *Georhizobium*, *Mycoplana*, “*Neopararhizobium*”, “*Peteryoungia*”, *Pseudorhizobium*, *Pararhizobium*, and *Shinella* (**Figures 3, S4**). Such a threshold would, however, split the genera *Allorhizobium*, *Hoeflea*, *Martellella*, *Neorhizobium*, and *Rhizobium sensu stricto* into two or more genera.

There was less support for the presence of a genus-level drop in the wpAAI frequency distribution (**Figure S3**). However, the wpAAI frequency distribution density increased sharply below 76.5%. Additionally, although noisier, a wpAAI threshold of 76.5% returned similar genus demarcations as did a cpAAI threshold of $\sim 86\%$, with the exceptions that the genus *Martellella* was recovered as a single genus, *Hoeflea* was split into fewer genera, and the separation between the genus *Ciceribacter* and its sister taxon was less clear (**Figures S5, S6**).

Proposal for a framework for genus delineation in the family *Rhizobiaceae*

Based on the results summarized above, we propose that genera within the family *Rhizobiaceae* be defined as monophyletic groups (as determined by a phylogenetic reconstruction using a core-genome analysis approach; **Figure 1**) separated from related species using a pairwise cpAAI threshold of approximately 86% calculated as described in the methods. We strongly recommend the use of cpAAI over wpAAI due to 1) its natural agreement – by construction – with the core-genome gene phylogeny, 2) clearer gaps in its distribution of values among *Rhizobiaceae*, and 3) the fact that it would not be sensitive to the wide genome size variation within the *Rhizobiaceae*, notably due to the variation in presence of large mobile genetic elements, including symbiotic megaplasms. We do not, however, propose that cpAAI serve as the sole information source for genus demarcation as nearly all biological rules have exceptions. We therefore propose that genus demarcation using a cpAAI threshold higher than 86% can be justified by the presence of alternate genomic or phenotypic evidence (as proposed below for splitting of the genus *Ensifer*), while a lower cpAAI threshold may be appropriate when considering historical classifications of genera within the family.

Taxonomic implications of the proposed framework

Following the criteria for genus demarcation outline above would notably lead to the formation of several new genera for species currently assigned to the genus *Rhizobium*, which is notoriously paraphyletic. They also imply that a few genera (*Neorhizobium*, *Allorhizobium*, *Martellella*, *Hoeflea*, and *Rhizobium sensu stricto*) may be candidates for division. We also note that there is a clear break in the distribution of cpAAI values at $\sim 73\%$ to $\sim 74\%$ that may represent an appropriate threshold for delimiting the family *Rhizobiaceae*. If adopted, this threshold would result in the genera *Martellella* and *Hoeflea* being transferred to their own families, while the genera *Georhizobium* and “*Neopararhizobium*” would form another family. However, a proposal for family-level demarcations in the order *Rhizobiales* is outside the scope of this work.

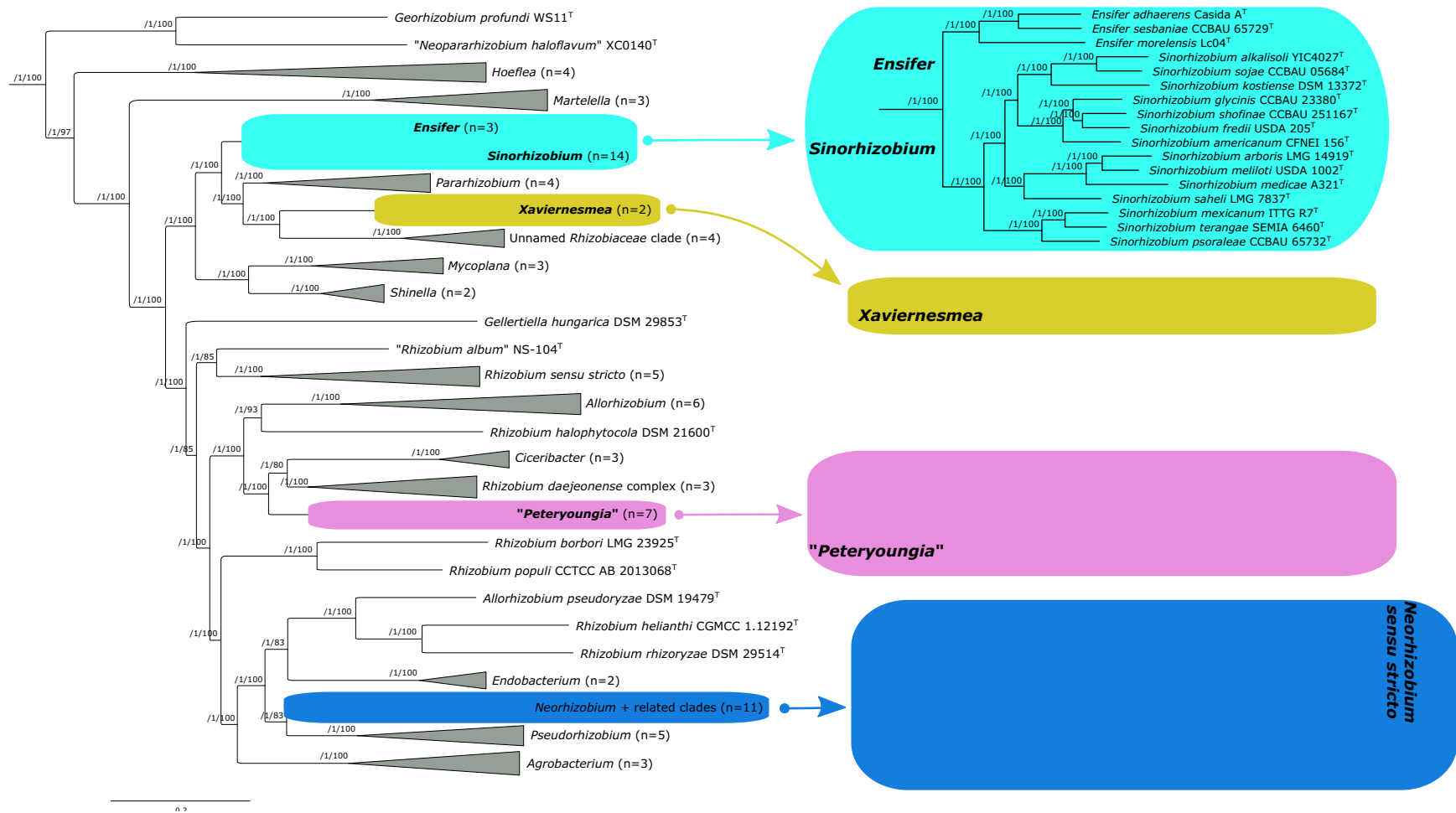


Figure 1. Maximum likelihood core-genome phylogeny of the family Rhizobiaceae. A maximum likelihood phylogeny 94 *Rhizobiaceae* strains is shown. The number of strains included in each collapsed clade is indicated. Clades of focus in the current study are expanded along the righthand side of the figure. The phylogeny is built from the concatenated alignments of 170 nonrecombinant loci using IQ-TREE [13]. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value). The tree was rooted using three *Mesorhizobium* spp. sequences as the outgroup. The scale bar represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R8 model. An expanded phylogeny is provided as Figure S1.

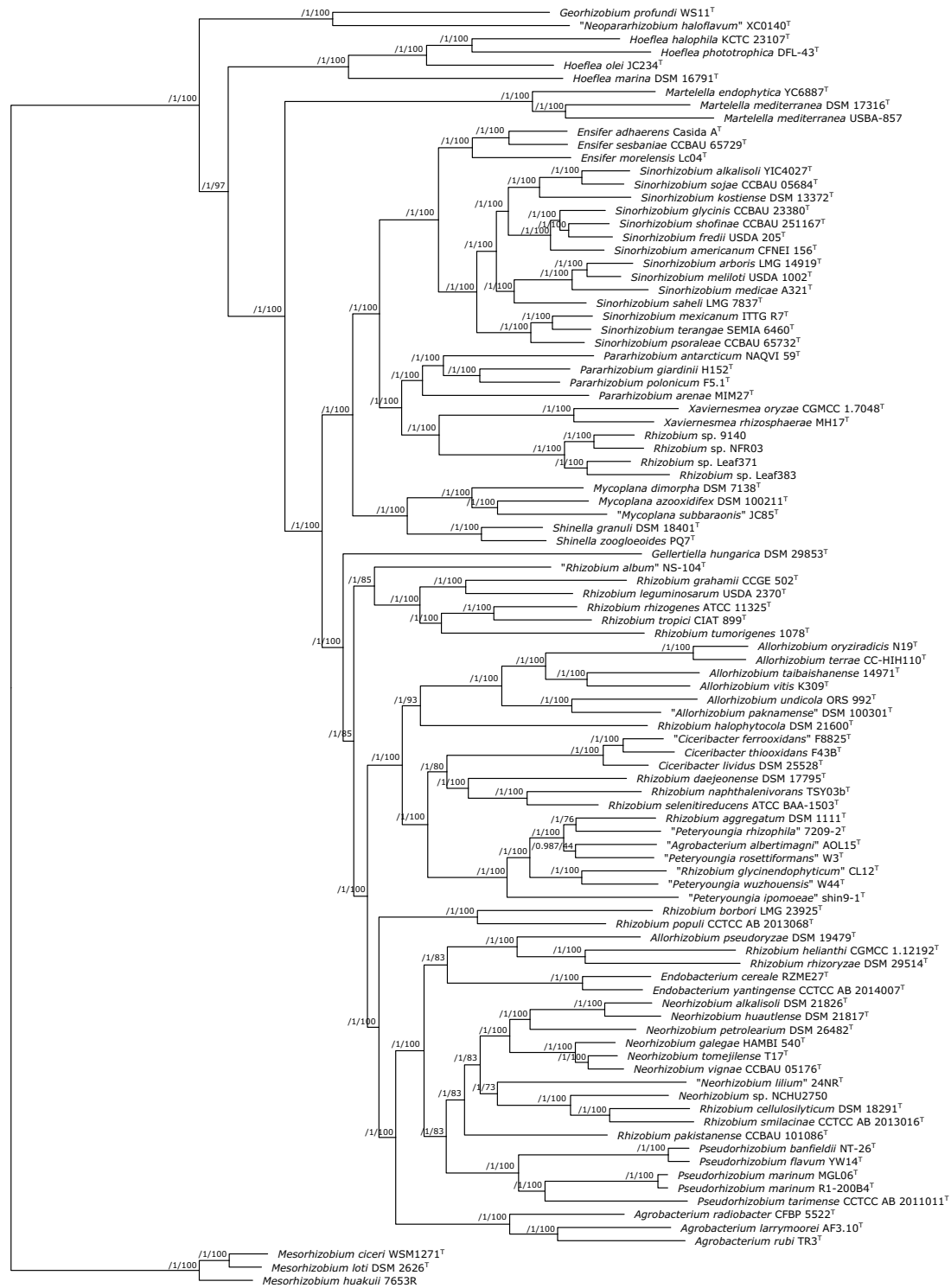


Figure S1. Maximum likelihood core-genome phylogeny of the family Rhizobiaceae. A maximum likelihood phylogeny 94 *Rhizobiaceae* strains, and three *Mesorhizobium* spp., is shown. The phylogeny is built from the concatenated alignments of 170 nonrecombinant loci using IQ-TREE. The numbers on the nodes indicate the approximate Bayesian posterior probabilities support values (first value) and ultra-fast bootstrap values (second value). The scale bar represents the number of expected substitutions per site under the best-fitting GTR+F+ASC+R8 model.

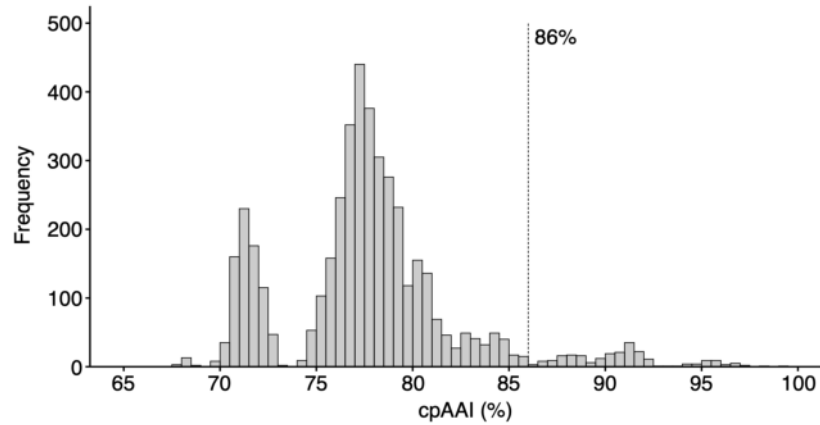


Figure 2. Distribution of core-proteome AAI (cpAAI) comparisons of the family *Rhizobiaceae*. Pairwise AAI values were calculated based on 170 nonrecombinant loci from the core-genome of 94 members of the family *Rhizobiaceae*. Results are summarized as a histogram with a bin width of 0.5%.

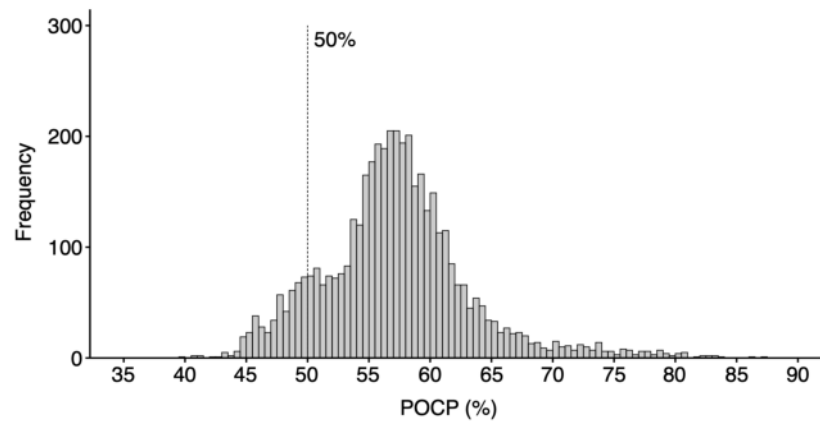


Figure S2. Distribution of percentage of conserved proteins (POCP) comparisons of the family *Rhizobiaceae*. Pairwise POCP values were calculated between 94 members of the family *Rhizobiaceae* as described in the Methods, and the results are summarized as a histogram with a bin width of 0.5%.

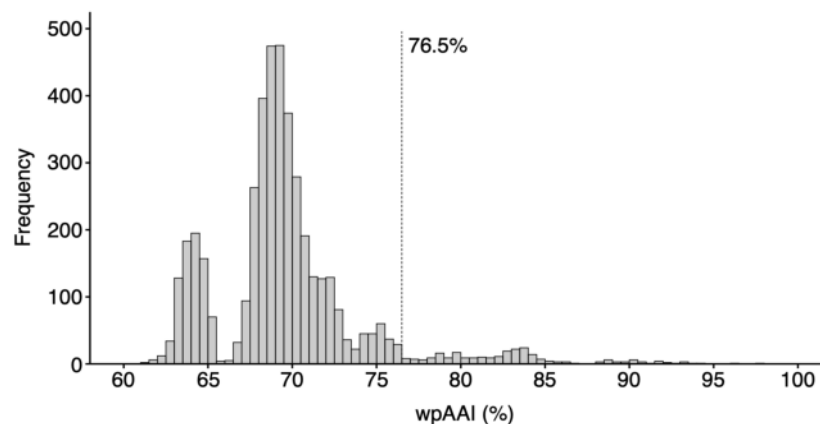


Figure S3. Distribution of whole-proteome AAI (wpAAI) comparisons of the family *Rhizobiaceae*. Pairwise wpAAI values were calculated between 94 members of the family *Rhizobiaceae* using the CompareM workflow, and the results are summarized as a histogram with a bin width of 0.5%.

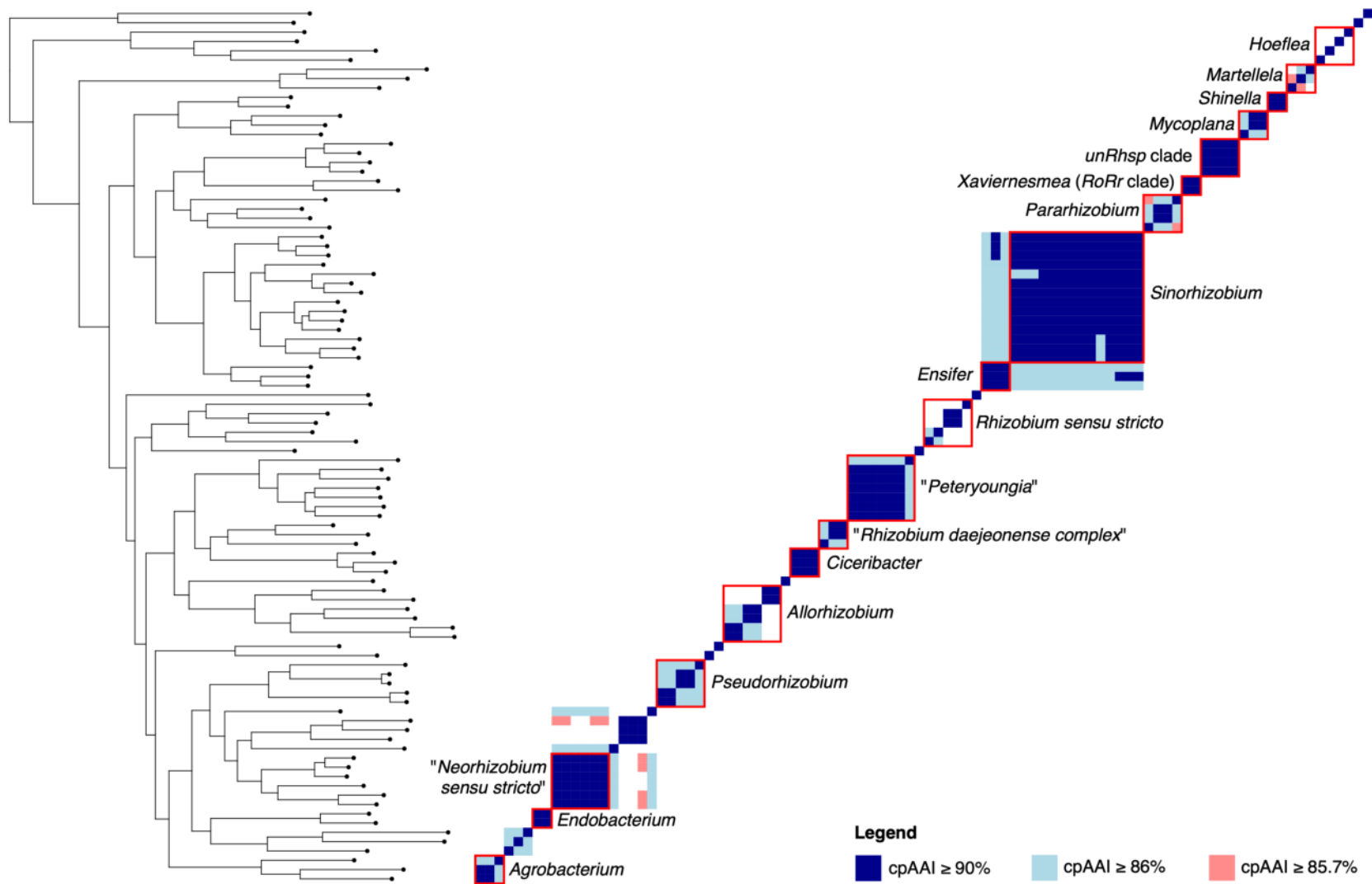


Figure 3. Core-proteome AAI (cpAAI) matrix of the family *Rhizobiaceae*. A matrix showing the pairwise cpAAI values for each pair of 94 members of the family *Rhizobiaceae*. Values were clustered using the core-genome gene phylogeny of Figures 1 and S1. Several named genera are indicated with red boxes, as indicated. A version of this matrix with a colour scheme representing the full range of cpAAI values is provided as Figure S4.

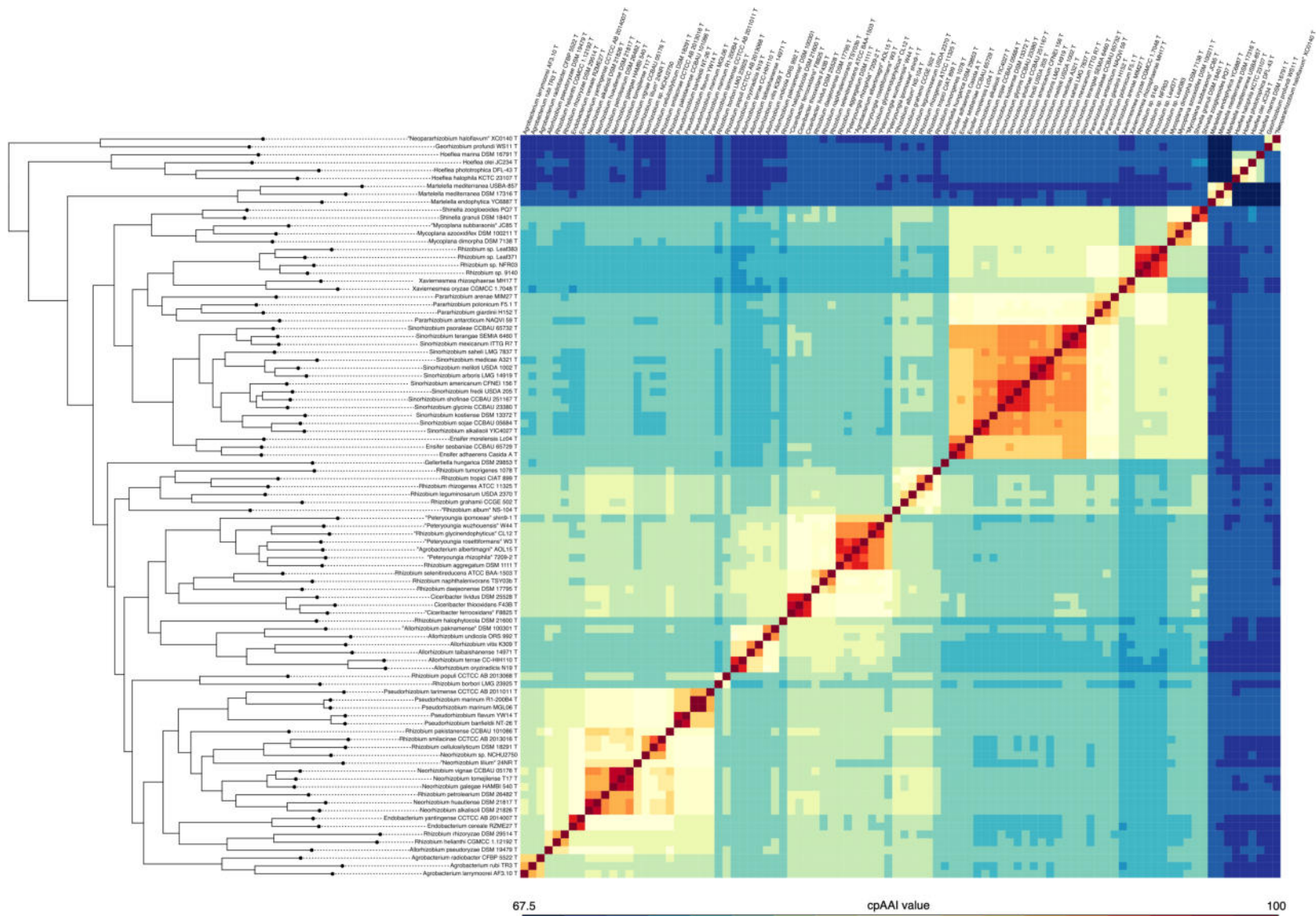


Figure S4. Core-proteome AAI (cpAAI) matrix of the family *Rhizobiaceae*. A matrix showing the pairwise cpAAI values for each pair of 94 members of the family *Rhizobiaceae*. Values were clustered using the core-genome gene phylogeny of Figures 1 and S1.

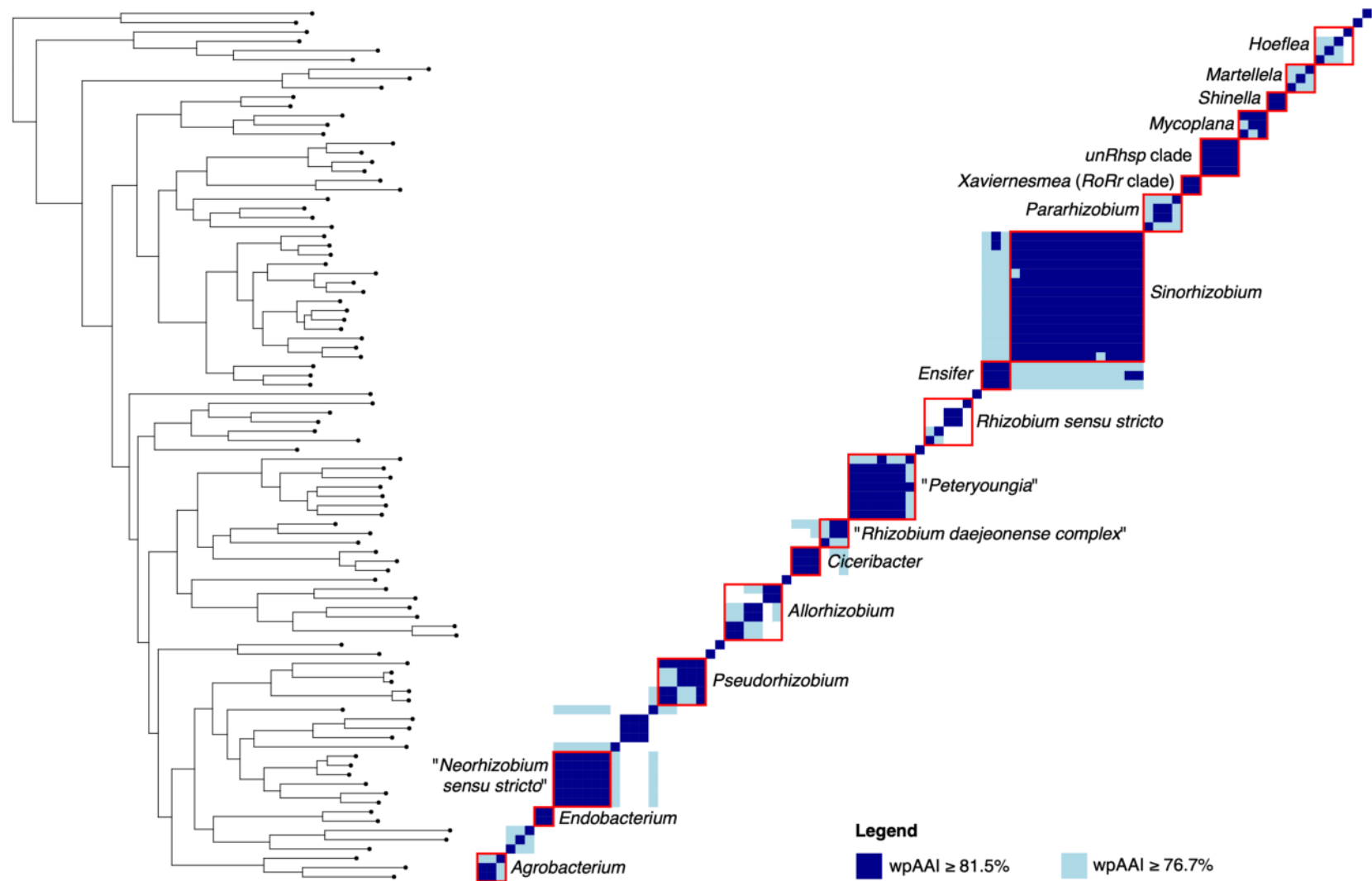


Figure S5. Whole-proteome AAI (wpAAI) matrix of the family *Rhizobiaceae*. A matrix showing the pairwise wpAAI values for each pair of 94 members of the family *Rhizobiaceae*. Values were clustered using the core-genome gene phylogeny of Figures 1 and S1. Several named genera are indicated with red boxes, as indicated. A version of this matrix with a colour scheme representing the full range of cpAAI values is provided as Figure S6.

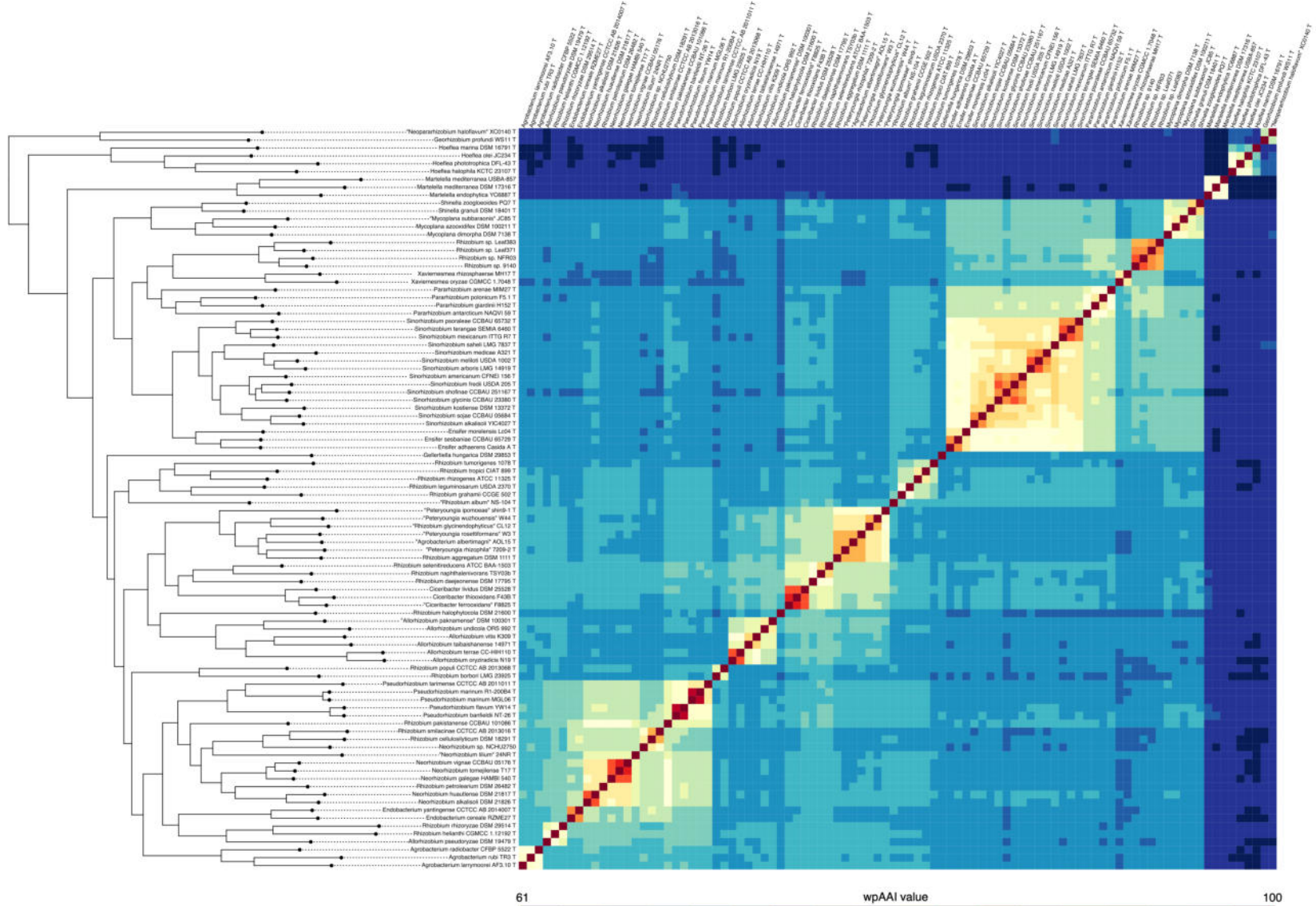


Figure S6. Whole-proteome AAI (wpAAI) matrix of the family *Rhizobiaceae*. A matrix showing the pairwise wpAAI values for each pair of 94 members of the family *Rhizobiaceae*. Values were clustered using the core-genome gene phylogeny of Figures 1 and S1.

Proposal of a new genus encompassing the species *R. oryzae* and “*R. rhizosphaerae*”

In a recent study presenting a phylogeny of 571 *Rhizobiaceae* and *Aurantimonadaceae* strains (ML tree based on 155 concatenated core proteins) [25], the type strains of the species *R. oryzae* (*Allorhizobium oryzae*) [26] and “*R. rhizosphaerae*” formed a well-delineated clade (with 100% bootstrap support) that was clearly separated from the closest validly published genus type, i.e. *Pararhizobium giardinii* strain H152^T. This pattern was also evident from a ML phylogeny of 797 *Rhizobiaceae* produced in another study based on the concatenation of 120 near-universal bacterial core genes [5]. The analyses presented in the current study further support the separation of the *R. oryzae*/“*R. rhizosphaerae*” clade (*RoRr* clade; two species type strains) not only from the *Pararhizobium* clade (four species type strains), but also from a sister clade consisting solely of rhizobial strains from unnamed species (*unRhsp* clade; including *Rhizobium* sp. strains Leaf383, Leaf371, 9140 and NFR03) (**Figure 1**); all three clades in the phylogenetic tree are supported by 100% bootstrap value. All within-clade pairwise cpAAI values were above 85.7%, 91%, and 94% for the *Pararhizobium* clade, *RoRr* clade, and the unnamed *Rhizobium* clade, respectively (**Figure 3**). In contrast, all pairwise cpAAI values between the *RoRr* clade and the *Pararhizobium* or *unRhsp* clades were less than 81%, while pairwise cpAAI values between the *Pararhizobium* and *unRhsp* clades were below 83.5% (**Figure 3**). All three clades thus represent separate genera according to the criteria proposed above, and this remains true when the analysis is repeated with an expanded set of strains (**Figure S7**). We therefore propose to define a new genus encompassing the *RoRr* clade, for which we propose the name *Xaviernesmea* (see below for formal description). As no strains belonging to the *unRhsp* clade have been deposited in any international culture collection, we leave the task of describing new species and genera within this clade to others who have access to these strains.

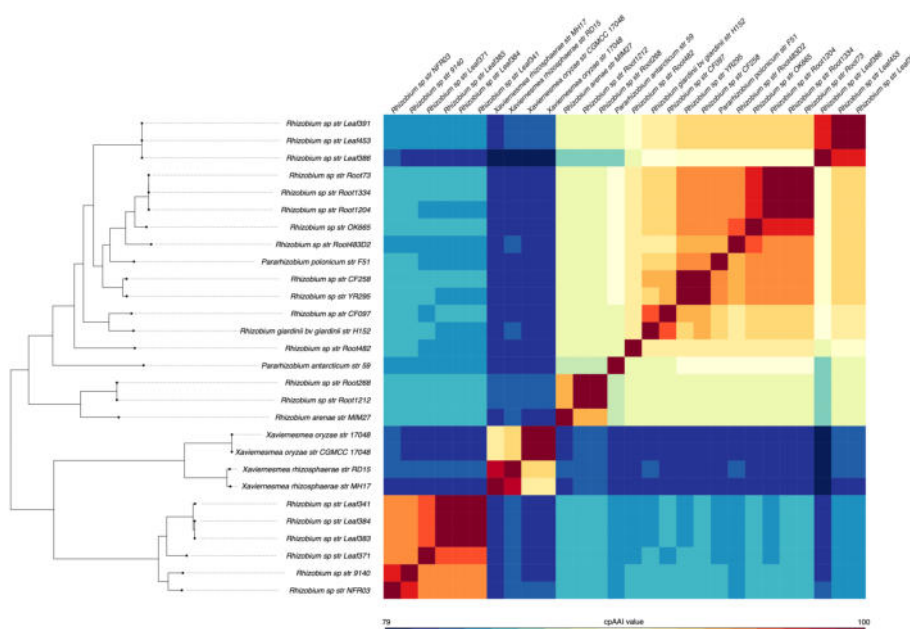


Figure S7. Whole-proteome AAI (wpAAI) matrix of the genus *Xaviernesmea* and related taxa. A matrix showing the pairwise wpAAI values for four members of the genus *Xaviernesmea* together with 24 related organisms.

Taxonomy of the “*R. aggregatum* complex”

The “*Rhizobium aggregatum* complex” was initially identified as a sister taxon of the genus *Agrobacterium* [26], with subsequent work demonstrating that it is instead located on a clade neighbouring the genus *Allorhizobium* [12]. Moreover, the latter study suggested that “*R. aggregatum* complex” includes members of the genus *Ciceribacter* and that it may represent a novel genus on the basis of phylogenetic and multiple OGRI data, although the authors advised that further investigation was required [12]. It was recently suggested that the “*R. aggregatum* complex” be split into two genera [27]. It was proposed that *R. daejeonense*, *R. naphthalenivorans*, and *R. selenitireducens* be transferred to the genus *Ciceribacter*, while *R. ipomoeae*, *Rhizobium rhizophilum*, *R. rosettiformans*, *R. wuzhouense*, and “*Peteryoungia desertarenae*” be transferred to the novel genus “*Peteryoungia*” [27].

The analyses presented in the current study included 13 strains belonging to the “*R. aggregatum* complex” (**Figure 1**). The genus demarcation framework proposed here supports the previous studies indicating that the “*R. aggregatum* complex” is separate from the genus *Allorhizobium*. A group of seven species that included all “*Peteryoungia*” species present in our analysis (*Rhizobium aggregatum* DSM 1111^T, “*Agrobacterium albertimagni*” AOL15^T, “*Rhizobium glycinendophyticum*” CL12^T, “*Peteryoungia ipomoeae*” shin9-1^T, “*Peteryoungia rhizophila*” 7209-2^T, “*Peteryoungia rosettiformans*” W3^T, and “*Peteryoungia wuzhouensis*” W44^T) formed a monophyletic group with 100% bootstrap support (**Figure 1**). All pairwise cpAAI values within this group were > 88%, while all pairwise cpAAI values against the other six “*R. aggregatum* complex” species were < 84.7% (**Figure 3**). These results support the proposal for the genus “*Peteryoungia*” [27], which should also primarily include *R. aggregatum*, as well as “*A. albertimagni*”, and “*R. glycinendophyticum*” once their names are validly published.

The remaining six “*R. aggregatum* complex” species formed a monophyletic group that could be further sub-divided into two clades. One clade corresponded to the three *Ciceribacter* species, while the other clade contained *R. daejeonense* DSM 17795^T, *R. naphthalenivorans* TSY03b^T, and *R. selenitireducens* ATCC BAA-1503^T (**Figure 1**). All within-group cpAAI values were > 86.5% while all between-group cpAAI values were ≤ 85.4%, providing support for these two clades representing separate genera. However, the bootstrap support for the split of these two clades in the phylogeny is only 80%, and the topology of the tree in this region (**Figure 1**) differs from the tree reported by Rahi *et al.*, wherein *R. daejeonense*, *R. naphthalenivorans*, and *R. selenitireducens* were not monophyletic (see Figure 2 of [27]). Overall, the data presented here are not in agreement with the proposal to transfer *R. daejeonense*, *R. naphthalenivorans*, and *R. selenitireducens* to the genus *Ciceribacter*. Instead, we propose this clade be referred to as the “*R. daejeonense* complex” pending further study – enabled by the availability of additional genomes of strains belonging to these clades – to resolve whether these species belong to a novel genus or whether they should be transferred to the genus *Ciceribacter*.

Proposal for the emendation of the genus *Sinorhizobium* as a distinct genus from *Ensifer*

Taxonomy of the *Ensifer* / *Sinorhizobium* genus has been the subject of discussion since the early

2000s. The genus *Ensifer* was proposed in 1982 to describe *Ensifer adhaerens*, a bacterial predator [8]. Subsequently, the genus *Sinorhizobium* was proposed in 1988 when *Rhizobium fredii* was reclassified as *Sinorhizobium fredii* [9], which was followed by the emendation of this genus by de Lajudie *et al.*, 1994 [28]. In 2002, as the 16S rRNA gene sequence of *E. adhaerens* became available, the Subcommittee on the Taxonomy of *Agrobacterium* and *Rhizobium* (hereafter “the subcommittee”) of the International Committee on Systematics of Prokaryotes (ICSP) noted that this taxon is a part of *Sinorhizobium* [29]. Although the subcommittee pointed out that the name *Ensifer* has priority, conservation of the name *Sinorhizobium* was endorsed in contravention of the rules of the International Code of Nomenclature of Prokaryotes (ICNP). Neighbour-joining trees constructed from 16S rRNA gene sequences or partial *recA* gene sequences, together with phenotypic data, provided further data interpreted as supporting the synonymy and unification of the genera *Sinorhizobium* and *Ensifer*, leading Willems *et al.* (2003) to propose the new combination “*Sinorhizobium adhaerens*” [30]. Accordingly, in their Request for an Opinion to the Judicial Commission, Willems *et al.* officially proposed to conserve the name *Sinorhizobium* [30]. As the primary argument for conservation of the name *Sinorhizobium*, the authors indicated that the name *Ensifer* would cause misunderstanding and confusion in the scientific community. A few months later, in his Request for an Opinion to the Judicial Commission, J. M. Young argued that *Ensifer*, not *Sinorhizobium*, was the valid name for the unified genus, as *Ensifer* had priority [31]. At the same time, J. M. Young emended the description of the genus *Ensifer*, and transferred previously described *Sinorhizobium* species to this genus [31]. The Judicial Commission of the ICSP (Judicial Opinion 84) later confirmed that *Ensifer* had priority over *Sinorhizobium*, pointed out that the name “*Sinorhizobium adhaerens*” is not validly published, and supported the transfer of members of the genus *Sinorhizobium* to *Ensifer* [32]. In this Opinion, it was claimed that the transfer of the members of the genus *Sinorhizobium* to the genus *Ensifer* would not cause confusion. The subcommittee, however, disagreed with this justification [33]. J. M. Young criticized these actions of the subcommittee [34], which was also later acknowledged by Tindall [35]. As predicted by Willems *et al.* [30], adoption of the genus name *Ensifer* continues to be met with resistance from many rhizobiologists [36].

Earlier phylogenetic studies noted that *E. adhaerens* was an outgroup of the genus *Ensifer* [37], providing some support that *E. adhaerens* represented a distinct genus; however, it was suggested that further evidence would be required prior to redefining genera within this clade [37]. Significant phylogenomic and phenotypic data now exists providing strong evidence that the genera *Ensifer* and *Sinorhizobium* as defined Casida 1982 [8] and Chen *et al.* 1988 [9], respectively, refer to closely related, yet separate, taxa. At least seven studies, including the Genome Taxonomy Database, have presented phylogenetic trees containing two well-defined clades within the genus *Ensifer* [20, 25, 38–42]. These phylogenies were built on the basis of gene (up to 1,652 genes) or protein (up to 155 proteins) sequences using ML or Bayesian inference analysis approaches, indicating that the observed clades are robust to the choice of phylogenetic approach. Notably, our recent study presents a ML phylogeny where the genus *Ensifer* is split into two clades of 12 and 20 genospecies with 100% bootstrap support for the split, which we then

defined as the “nonsymbiotic” and “symbiotic” clades, respectively [20]. The split is also observed in a ML phylogeny of the 16S rRNA genes of the same strains, with 62% bootstrap support (**Figure S8**). We similarly see a split of the genus *Ensifer* into two clades of three species type strains (including *E. adhaerens* Casida A^T, the type strain of the type species of the genus *Ensifer* Casida 1982) and 12 species type strains (including *E. fredii* USDA 205^T, the type strain of the type species of the genus *Sinorhizobium* Chen *et al.* 1988) in our core-genome gene phylogeny, with 100% bootstrap support (**Figure 1**), representing the nonsymbiotic and symbiotic clades, respectively. However, all between-clade cpAAI values were above the suggested 86% threshold as a baseline criterion for genus delimitation. Despite this, and following our proposed framework, we argue that there is sufficient other genomic and phenotypic data supporting the division of this genus (cf. Figures 3, 4, and 5 of [20]). We describe the distinctive traits and respective synapomorphies of these clades in **Table 1** and below.

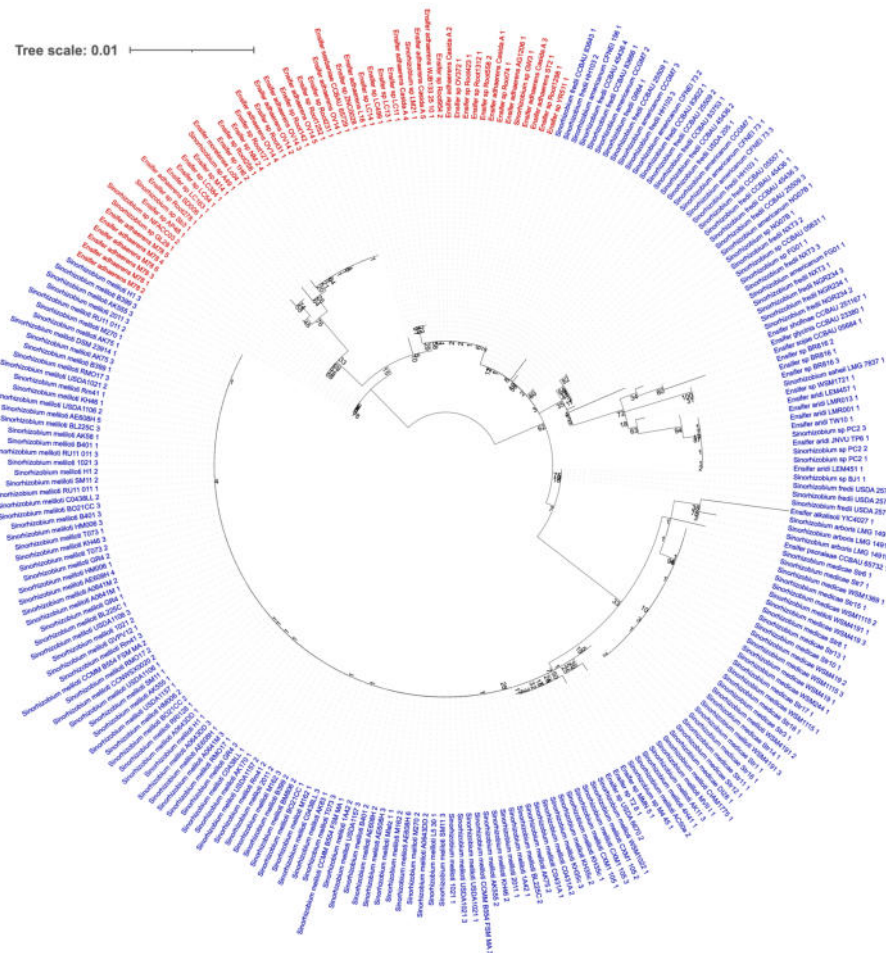


Figure S8. 16S rRNA gene phylogeny of the genera *Ensifer* and *Sinorhizobium*. A maximum likelihood phylogeny of 258 16S rRNA genes from 157 strains of the genera *Ensifer* (red) and *Sinorhizobium* (blue) is shown. Strains are named as recorded in NCBI at the time of collection. The phylogeny was created using RAxML as described in the Materials and Methods. Values represent bootstrap support from 756 bootstrap replicates. The scale represents the mean number of nucleotide substitutions per site. One 16S rRNA gene from each of *E. sp.* GL28 and *S. meliloti* BM806 are not shown due to extremely long branch lengths.

Table 1. Characteristics differentiating the previously-defined nonsymbiotic and symbiotic clades of the genus *Ensifer*, corresponding to the emended genera *Ensifer* and *Sinorhizobium*, respectively.

Characteristic	Nonsymbiotic clade (emended genus <i>Ensifer</i>)	Symbiotic clade (emended genus <i>Sinorhizobium</i>)	Reference
GANTC sites per kb	0.9 to 1.3 (mean: 1.06)	1.5 to 1.8 (mean: 1.70)	[43]
Number of CDS	5816 to 7682 (mean: 6876)	5516 to 8629 (mean: 6550)	[20]
Ribosomal RNA operons	5	3	[20]
Carries <i>nod</i> and <i>nif</i> genes	No*	Yes	[20]
Bacterial predation ability	Yes	No	[8, 44, 45]
Division by budding	Yes	No	[46]
Growth in unmodified LB medium	Yes	Poor	[20]
Starch hydrolysis	Yes	No	[47]
Growth at 37°C	No (generally)	Yes (generally)	[20]
Fatty acids	More 16:0 3OH	More 18:1 ω 9c	[47]
Carbon sources used (Biolog PM1/PM2)	69 to 87 (mean: 81)	50 to 81 (mean: 65)	[20]
pH tolerance (Biolog PM9)	Better low pH tolerance	Better high pH tolerance	[20]

* Except for the species *E. sesbaniae*, whose nine strains are legume symbionts.

The genome-wide frequency of the pentanucleotide GANTC is higher in all genomes of the symbiotic clade compared to all genomes of the nonsymbiotic clade, with a statistically significant average difference of 60% (1.70 vs 1.06 GANTC sites per kb, p -value < 1×10^{-10} using a two-sample t -test) [43]. As the GANTC motif is methylated by the highly conserved cell cycle-regulated methyltransferase CcrM [48, 49], this difference may reflect an important difference in the cell cycle biology of these two clades [43]. Indeed, species of the nonsymbiotic clade (*E. adhaerens* and *E. morelensis*) are capable of division by budding unlike species of the symbiotic clade [46]. It has also been shown that the ability to hydrolyze starch [47] and robustly grow in LB broth lacking Ca^{2+} and Mg^{2+} ion supplementation [20] is specific to the nonsymbiotic clade. Stress tolerance of the two clades also differs (based on an analysis of 10 representative strains), with strains of the nonsymbiotic clade generally being more tolerant to alkaline conditions while strains of the symbiotic clade were generally more acid-tolerant and heat-tolerant [20]. Although many catabolic abilities could be found in at least a subset of each clade, which is unsurprising given both clades have open pangenomes, species of the nonsymbiotic clade are capable of catabolizing an average of 81 (out of 190 tested) carbon sources compared to an average of 65 for the symbiotic clade [20]. These differences in general phenotypic traits, together with the additional genomic and phenotypic differences outlined in **Table 1**, indicate marked differences in the biology of strains from these two clades. Indeed, at least two genospecies of the nonsymbiotic clade have been described as bacterial predators [8, 44, 45], a lifestyle that has not been attributed to any members of the symbiotic clade. Moreover, these two clades display significant differences in relation to their interactions with plant species, specifically, a biased distribution of the *nod* and *nif* genes required for establishment of nitrogen-fixing symbiosis with legumes [20, 41]. A recent study showed that whereas the core *nodABC* and *nifHDK* genes were present in strains from all 20 genospecies of the symbiotic clade, they were observed in just one of the 12 genospecies belonging to the nonsymbiotic clade (*E. sesbaniae*, with all 9 reported strains, isolated from three different geographic origins, being symbiotic) [20, 47]. Symbiotic traits are linked to the presence of an accessory megaplasmid in the genome, and thus should not be considered relevant in delineating

taxa [6]. However, this almost unique ability of genomes from the symbiotic clade to host symbiotic megaplasmids with respect to their relatives from the nonsymbiotic clade likely reflects differences in their genetic background. These discrepancies in symbiotic potential could thus be interpreted as a further marker of differentiated biology between these two clades.

Taken together, these genomic and phenotypic data suggest that the organisms in these two clades significantly differ in their biology and ecology, reminiscent of the stable ecotype model for bacterial species [50]. Notably, the type species of the genus *Ensifer* (*E. adhaerens*) is found within the nonsymbiotic clade, while the original type species of the genus *Sinorhizobium* (*S. fredii*) is found within the symbiotic clade. Given we established the taxonomic position of these type species-containing clades to be well separated, we argue that the proposal of Willems *et al.* (2003) [30] to unify the genera *Ensifer* Casida 1982 and *Sinorhizobium* Chen *et al.* 1988, and the Judicial Opinion 84 enacting the transfer of the members of the genus *Sinorhizobium* to the genus *Ensifer* [32], are no longer supported. Instead, we propose that *Ensifer* Casida 1982 and *Sinorhizobium* Chen *et al.* 1988 refer to closely related sister genera, of which *Ensifer* and *Sinorhizobium*, respectively, are the legitimate names in accordance with Rules 51a and 23a of the ICNP [51]. We note that the subcommittee has previously indicated support for this proposal [52], while also stating they are not in favour of creating subgenera for these taxa [36]. Formal genus and species emendations and circumscriptions are provided below.

Taxonomy of the genus *Neorhizobium*

More study is required to resolve the taxonomic relationships between the “*Neorhizobium sensu stricto*” clade (**Figure 1**) – which includes *N. vignae*, *N. alkalisolii*, *N. hautlense*, *N. galegae* and *N. tomejilense*, as well as *Rhizobium petrolearium* – and related taxa. The core-genome gene phylogeny (**Figure 1**) and the cpAAI data (**Figure 3**) suggest that “*Neorhizobium lilium*” represents a new genus, as does the clade formed by *Neorhizobium* sp. NCHU2750, *Rhizobium smilacinae*, and *Rhizobium cellulosilyticum*. However, because bootstrap values provided only moderately good support for the topology of the tree in the extended *Neorhizobium* clade and the clades were not well-resolved by the cpAAI data, we defer the proposal of new genera until publication of further genomic evidence.

Additional taxonomic implications of the proposed framework for genus delimitation

R. petrolearium DSM 26482^T formed a clade with “*Neorhizobium sensu stricto*” (**Figure 1**). Pairwise cpAAI values were all > 90% when *R. petrolearium* was compared against “*Neorhizobium sensu stricto*” species type strains (**Figure 3**). We therefore propose that *R. petrolearium* be transferred to the genus *Neorhizobium* (see below for formal description).

Rhizobium tarimense CCTC AB 2011022^T formed a clade with the genus *Pseudorhizobium* (**Figure 1**). Pairwise cpAAI values were all > 88% when *R. tarimense* was compared against the four *Pseudorhizobium* species type strains, but < 85% when compared against all other species (**Figure 3**). We therefore propose that *R. tarimense* be transferred to the genus *Pseudorhizobium* (see below for formal description).

Rhizobium arenae MIM27^T formed a clade with the genus *Pararhizobium*. Pairwise cpAAI values were 87.4%, 87.4%, and 85.7% when *R. arenae* was compared against the three *Pararhizobium* species type strains, but < 84% when compared against all other species (**Figure 3**). We therefore propose that *R. arenae* be transferred to the genus *Pararhizobium* (see below for formal description).

Rhizobium azooxidifex DSM 100211^T formed a clade with “*Mycoplana subbaraonis*” JC85^T and *Mycoplana dimorpha* DSM 7138^T (**Figure 1**). Pairwise cpAAI values between these three species type strains were all > 89%, while cpAAI values against strains outside of this clade were all < 85% (**Figure 3**). We therefore propose that *R. azooxidifex* be transferred to the genus *Mycoplana* (see below for formal description).

Rhizobium yangtingense CCTCC AB 2014007^T formed a clade with *Endobacterium cereale* (**Figure 1**). The pairwise cpAAI value between these two species was > 95%, while cpAAI values against strains outside of this clade were all < 84% (**Figure 3**). We therefore propose that *R. yangtingense* be transferred to the genus *Endobacterium* (see below for formal description).

To confirm the distinct taxonomic positions of the above-mentioned species and support their transfer to the respective genera, we compared them to other genus members using ANIb and isDDH indices. These indices are regarded as standard measures of relatedness between prokaryotic species that were widely used for species delimitation [53, 54]. In all cases, the obtained values were clearly below the thresholds for species delimitation (95–96% for ANI or 70% for DDH) (**Datasets S3 and S4**), confirming the authenticity of these species.

In addition, the following species are candidates as type species for new genera: “*R. album*”, *R. populii*, *R. borbori*, and *R. halophytocola*. The reclassification of *R. album* into a new genus was also suggested by Young *et al.* 2021 [5]. Moreover, *R. helianthi* CGMC 1.12192^T, *R. rhizoryzae* DSM 29514^R, and *Allorhizobium pseudoryzae* DSM 19479^T formed a monophyletic group as a sister taxon to the genus *Endobacterium* (**Figure 1**). This clade of three species type strains is another candidate for reclassification as a new genus as pairwise cpAAI values within the clade were all > 86.5% while all cpAAI values with species outside of the clade were all < 83% (**Figure 3**).

Description of *Xaviernesmea* gen. nov.

Xaviernesmea (gza.vje.nem'e.a.; N.L. fem. n., in honour of Dr. Xavier Nesme, taxonomist of agrobacteria and rhizobia who pioneered the use of reverse ecology approaches to infer the ecology of *Agrobacterium* genomic species from comparative genomic analyses [7]).

Cells are Gram-negative, rod-shaped, and aerobic. Oxidase and catalase positive. Can utilise adonitol, D-raffinose and succinic acid and pH range for growth is 5.0–11.0 [55]. The G+C content of the genomic DNA is in the range 62.8–64.7 %. The genus *Xaviernesmea* has been separated from other *Rhizobiaceae* genera based on a core-genome phylogeny and whole- and core-proteome relatedness indexes (wpAAI and cpAAI).

The type species is *Xaviernesmea oryzae*.

Description of *Xaviernesmea oryzae* comb. nov.

Xaviernesmea oryzae (o.ry'zae. L. gen. fem. n. *oryzae*, of rice, referring to the host of isolation of the type strain).

Basonym: *Rhizobium oryzae* Peng *et al.* 2008 [56].

Homotypic synonym: *Allorhizobium oryzae* (Peng *et al.*, 2008) Mousavi *et al.* 2015.

The description is as provided by Peng *et al.* 2008 [56] and Mousavi *et al.* 2015 [26]. *X. oryzae* can be differentiated from other species of the genus *Xaviernesmea* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.8 %. Its approximate genome size is 5.39 Mbp.

The type strain is Alt 505^T (= LMG 24253^T = CGMCC 1.7048^T), isolated from *Oryza alta* growing in the Wild Rice Core Collection Nursery of South China Agricultural University. The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001938935.1.

Description of *Xaviernesmea rhizosphaerae* sp. nov.

Xaviernesmea rhizosphaerae (rhi.zo.sphae'rae. N.L. gen. fem. n. *rhizosphaerae*, of the rhizosphere, referring to host plant compartment of isolation of the type strain).

The description is as provided by Zhao *et al.* 2017 [55]. *X. rhizosphaerae* can be differentiated from other species of the genus *Xaviernesmea* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 64.7 %. Its approximate genome size is 5.18 Mbp.

The type strain is MH17^T (= ACCC 19963^T = KCTC 52414^T), which was isolated from root of rice collected from Beijing, China. The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001938945.1. We note that the name “*Rhizobium rhizosphaerae*” that was proposed in the original publication [55] has yet to be validly published.

Emended description of the genus *Ensifer* Casida 1982

The description is as given by Casida 1982 [8] with the following emendations. The optimal growth temperature is 27-28°C. Some species can grow at 37°C. Capable of growth in unmodified Lysogeny Broth (LB). Capable of hydrolyzing starch. Resistant to multiple antibiotics including ampicillin and erythromycin. The genomic G+C content is ~ 61-63 %. The genomic GANTC pentanucleotide frequency is ~ 0.9-1.3 sites per kb. Most strains carry 5 rRNA operons. The genus can be differentiated from other genera based on core-genome gene phylogenies.

The type species is *Ensifer adhaerens*.

The emended genus contains the species *E. adhaerens*, *E. morelensis*, and *E. sesbaniae*.

The species *E. alkalisoli*, *E. americanum*, *E. arboris*, *E. fredii*, *E. garamanticum*, *E. glycinis*, *E. kostiensis*, *E. kummerowiae*, *E. medicae*, *E. meliloti*, *E. mexicanum*, *E. numidicum*, *E. psoraleae*, *E. sahelii*, *E. sofinae*, *E. sojae*, and *E. teranga* are transferred to the genus *Sinorhizobium*.

Emended description of the genus *Sinorhizobium* Chen et al. 1988 emend. de Lajudie et al. 1994

The description is as given by de Lajudie *et al.* 1994 [28] with the following emendations, drawing also from Young 2003 [31]. The optimal growth temperature is 25-33°C, but some strains can grow at 12°C and others can grow at 44°C. Optimum pH is 6-7, but some strains can grow at pH 5.0 and others at pH 10.0. Starch is not utilized. Ammonium salts, nitrate, nitrite, and many amino acids can serve as nitrogen sources for most strains. Most strains produce cytochrome oxidase and catalase. The genomic G+C content is ~ 61-64 %. The genomic GANTC pentanucleotide frequency is ~ 1.5-1.8 sites per kb. Most strains carry 3 rRNA operons. The genus can be differentiated from other genera based on core-genome gene phylogenies.

The type species is *Sinorhizobium fredii*.

The emended genus contains the species *S. alkalisoli*, *S. americanum*, *S. arboris*, *S. fredii*, *S. garamanticum*, *S. glycinis*, *S. kostiense*, *S. kummerowiae*, *S. medicae*, *S. meliloti*, *S. mexicanum*, *S. numidicum*, *S. psoraleae*, *S. saheli*, *S. sofinae*, *S. sojiae*, and *S. terangae*.

Description of *Sinorhizobium alkalisoli* comb. nov.

Sinorhizobium alkalisoli (al.ka.li.so'li. N.L. neut. n. *alkali*, alkali (from Arabic al-qaliy); L. neut. n. *solum*, soil; N.L. gen. neut. n. *alkalisoli*, of alkaline soil, referring to the saline-alkali soil where the bacterium was isolated).

Basonym: *Ensifer alkalisoli* Li *et al.* 2016.

The description is as provided by Li *et al.* 2016 [57]. *S. alkalisoli* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.2 %. Its approximate genome size is 6.13 Mbp.

The type strain is YIC4027^T (=HAMBI 3655^T =LMG 29286^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_008932245.1.

Emended description of *Sinorhizobium americanum* Toledo et al. 2004

Sinorhizobium americanum (a.me.ri.ca'num. N.L. neut. adj. *americanum*, American, referring to the isolation of the type strain from the Colorado Plateau).

Homotypic synonym: *Ensifer americanus* (Toledo *et al.* 2004) Wang *et al.* 2015 emend. Hördt *et al.* 2020.

The description is as provided by Hördt *et al.* 2020 [4]. *S. americanum* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.3 %. Its approximate genome size is 6.75 Mbp.

The type strain is CFNEI 156^T (=ATCC BAA-532^T =CIP 108390^T =DSM 15007^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001889105.1.

Emended description of *Sinorhizobium arboris* Nick et al. 1999

Sinorhizobium arboris (ar'bo.ris. L. fem. n. *arbor*, tree; L. gen. fem. n. *arboris*, of a tree).

Homotypic synonym: *Ensifer arboris* (Nick *et al.* 1999) Young 2003 emend. Hördt *et al.* 2020.

The description is as provided by Hördt et al. 2020 [4]. *S. arboris* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.0 %. Its approximate genome size is 6.85 Mbp.

The type strain is HAMBI 1552^T (=ATCC BAA-226^T =DSM 13375^T =LMG 14919^T =NBRC 100383^T =TTR 38^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_000427465.1.

Emended description of *Sinorhizobium fredii* (Scholla and Elkan 1984) Chen et al. 1988

Sinorhizobium fredii (fred'i.i. N.L. gen. neut. n. *fredii*, of Fred, named after of E.B. Fred).

Homotypic synonym: *Ensifer fredii* (Scholla and Elkan 1984) Young 2003 emend. Hördt et al. 2020.

Heterotypic synonym: *Ensifer xinjiangensis* (Chen et al. 1988) Young 2003 [39]

The description is as provided by Hördt et al. 2020 [4]. *S. fredii* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.2 %. Its approximate genome size is 7.15 Mbp.

The type strain is USDA 205^T (=ATCC 35423^T =CCUG 27877^T =DSM 5851^T =HAMBI 2075^T =ICMP 11139^T =IFO 14780^T =JCM 20967^T =LMG 6217^T =NBRC 14780^T =NRRL B-14241^T =NRRL B-14594^T =PRC 205^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_009601405.1.

Description of *Sinorhizobium garamanticum* comb. nov.

Sinorhizobium garamanticum (ga.ra.man'ti.cum. N.L. neut. adj. *garamanticum*, pertaining to Garamante, Garamantian, the country of Garamantes, from which the strains were isolated).

Basonym: *Ensifer garamanticus* Merabet et al. 2010.

The description is as provided by Merabet et al. 2010 [58]. *S. garamanticum* can be differentiated from other species of the genus *Sinorhizobium* by phylogenetic analysis based on several housekeeping (*recA*, *glnA*, *gltA*, *thrC* and *atpD*) genes and 16S rRNA gene sequencing. The genomic G+C content of the type strain is approximately 62.4 % (HPLC).

The type strain is ORS 1400^T (=CIP 109916^T =LMG 24692^T).

Description of *Sinorhizobium glycinis* comb. nov.

Sinorhizobium glycinis (gly.ci'nis. N.L. gen. n. *glycinis*, of the botanical genus *Glycine*, the soybean, named for its nodulation characteristics and symbiotic genes).

Basonym: *Ensifer glycinis* Yan et al. 2016.

The description is as provided by Yan et al. 2016 [59]. *S. glycinis* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.4 %. Its approximate genome size is 6.04 Mbp.

The type strain is CCBAU 23380^T (=HAMBI 3645^T =LMG 29231^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001651865.1.

Emended description of *Sinorhizobium kostiense* Nick *et al.* 1999

Sinorhizobium kostiense (kos.ti.en'se. N.L. masc./fem. adj. *kostiense*, pertaining to Kosti, the region in Sudan where most of these organisms have been isolated).

Homotypic synonym: *Ensifer kostiensis* (Nick *et al.* 1999) Young 2003.

The description is as provided by Nick *et al.* 1999 [60]. *S. kostiense* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 61.7 %. Its approximate genome size is 6.33 Mbp.

The type strain is DSM 13372^T (=ATCC BAA-227^T =HAMBI 1489^T =LMG 15613^T =LMG 19227^T =NBRC 100382^T =TTR 15^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001651865.1.

Emended description of *Sinorhizobium kummerowiae* Wei *et al.* 2002

Sinorhizobium kummerowiae (kum.me.ro'wi.ae. N.L. gen. fem. n. *kummerowiae*, of *Kummerowia*, a genus of leguminous plants, referring to the host from which the bacterium was isolated).

Homotypic synonym: *Ensifer kummerowiae* (Wei *et al.* 2002) Young 2003.

The description is as provided by Wei *et al.* 2002 [61]. *S. kummerowiae* can be differentiated from other species of the genus *Sinorhizobium* by phylogenetic analysis based on 16S rRNA gene sequences. The genomic G+C content of the type strain is approximately 61.6 % (*Tm*).

The type strain is CCBAU 71714^T (=CGMCC 1.3046^T =CIP 108026^T =NBRC 100799^T).

Emended description of *Sinorhizobium medicae* Rome *et al.* 1996

Sinorhizobium medicae (me'di.cae. L. gen. fem. n. *medicae*, of/from lucern (plant belonging to the genus *Medicago*)).

Homotypic synonym: *Ensifer medicae* (Rome *et al.* 1996) Young 2003.

The description is as provided by Rome *et al.* 1996 [62]. *S. medicae* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 61.2 %. Its approximate genome size is 6.53 Mbp.

The type strain is A 321^T (=11-3 21a^T =HAMBI 2306^T =ICMP 13798^T =LMG 19920^T =NBRC 100384^T =USDA 1037^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_007827695.

Emended description of *Sinorhizobium meliloti* (Dangeard 1926) de Lajudie *et al.* 1994

Sinorhizobium meliloti (me.li.lo'ti. N.L. masc. n. *Melilotus*, generic name of sweet clover; N.L. gen. masc. n. *meliloti*, of *Melilotus*).

Homotypic synonym: *Ensifer meliloti* (Dangeard 1926) Young 2003.

The description is as provided by de Lajudie *et al.* 1994 [28]. *S. meliloti* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 62.0 %. Its approximate genome size is 7.34 Mbp.

The type strain is USDA 1002^T (=ATCC 9930^T =CCUG 27879^T =CFBP 5561^T =DSM 30135^T =HAMBI 2148^T =ICMP 12623^T =IFO 14782^T =JCM 20682^T =LMG 6133^T =NBRC 14782^T).

=NCAIM B.01520^T =NRRL L-45^T =NZIP 4027^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_007827695.1.

Description of *Sinorhizobium mexicanum* comb. nov.

Sinorhizobium mexicanum (me.xi.ca'num. N.L. neut. adj. *mexicanum*, of or belonging to Mexico, where the strains were isolated).

Basonym: *Ensifer mexicanus* Lloret *et al.* 2011.

The description is as provided by Lloret *et al.* 2011 [63]. *S. mexicanum* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 61.4 %. Its approximate genome size is 7.14 Mbp.

The type strain is ITTG R7^T (=ATCC BAA-1312^T =CFN ER1001^T =CIP 109033^T =DSM 18446^T =HAMB1 2910^T). The NCBI Assembly accession number for the genome sequence is GCF_013488225.1.

Description of *Sinorhizobium numidicum* comb. nov.

Sinorhizobium numidicum (nu.mi'di.cum. N.L. neut. adj. *numidicum*, pertaining to the country of Numidia, Numidian, the Roman denomination of the region in North Africa from which the majority of the organisms were isolated).

Basonym: *Ensifer numidicus* Merabet *et al.* 2010.

The description is as provided by Merabet *et al.* 2010 [58]. *S. numidicus* can be differentiated from other species of the genus *Sinorhizobium* by phylogenetic analysis based on several housekeeping (*recA*, *glnA*, *gltA*, *thrC* and *atpD*) genes and 16S rRNA gene sequencing. The genomic G+C content of the type strain is approximately 62.8 % (HPLC).

The type strain is ORS 1407^T (=CIP 109916^T =LMG 24692^T).

Description of *Sinorhizobium psoraleae* comb. nov.

Sinorhizobium psoraleae (pso.ra.le'a.e N.L. gen. fem. n. *psoraleae*, of *Psoralea*, referring to the main host of the species).

Basonym: *Ensifer psoraleae* Wang *et al.* 2013.

The description is as provided by Wang *et al.* 2013 [47]. *S. psoraleae* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 61.3 %. Its approximate genome size is 7.43 Mbp.

The type strain is CCBAU 65732^T (=HAMB1 3286^T =LMG 26835^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_013283645.1.

Emended description of *Sinorhizobium saheli* de Lajudie *et al.* 1994

Sinorhizobium saheli (sa'heli.i. N.L. gen. neut. n. *saheli*, of the Sahel, the region in Africa from which they were isolated).

Homotypic synonym: *Ensifer saheli* (de Lajudie *et al.* 1994) Young 2003 emend. Hördt *et al.* 2020.

The description is as provided by Hördt et al. 2020 [4]. *S. saheli* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 63.6 %. Its approximate genome size is 5.99 Mbp.

The type strain is LMG 7837^T (=ATCC 51690^T =DSM 11273^T =HAMBI 215^T =ICMP 13648^T =NBRC 100386^T =ORS 609^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001651875.1.

Description of *Sinorhizobium shofinae* comb. nov.

Sinorhizobium shofinae (sho.fi'nae. N.L. fem. gen. n. *shofinae* from Shofine, a company name, referring to the fact that the type strain of this species was isolated from root nodule of soybean grown in the farm of the company, Shandong Shofine Seed Technology Company Ltd., located in Jiaxiang County, Shandong Province of China).

Basonym: *Ensifer shofinae* Chen *et al.* 2017 emend. Hördt et al. 2020.

The description is as provided by Hördt et al. 2020 [4]. *S. shofinae* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 59.9 %. Its approximate genome size is 6.21 Mbp.

The type strain is CCBAU 251167^T (=HAMBI 3507^T =LMG 29645^T =ACCC 19939^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001704765.1.

Description of *Sinorhizobium sojae* comb. nov.

Sinorhizobium sojae (so'ja.e. N.L. gen. n. *sojae*, of soja, of soybean, referring to the source of the first isolates).

Basonym: *Ensifer sojae* Li *et al.* 2011 emend. Hördt et al. 2020.

The description is as provided by Hördt et al. 2020 [4]. *S. sojae* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 60.9 %. Its approximate genome size is 6.09 Mbp.

The type strain is CCBAU 5684^T (=DSM 26426^T =HAMBI 3098^T =LMG 25493^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_002288525.1.

Emended description of *Sinorhizobium terangae* de Lajudie et al. 1994

Sinorhizobium terangae (te'ran.gae. N.L. n. *terangae*, hospitality (from West African Wolof n. *terenga*, hospitality); N.L. gen. n. *terangae*, of hospitality, intended to mean that this species is isolated from different host plants).

Homotypic synonym: *Ensifer terangae* (de Lajudie et al. 1994) Young 2003.

The description is as provided by de Lajudie *et al.* 1994 [28]. *S. terangae* can be differentiated from other species of the genus *Sinorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 61.4 %. Its approximate genome size is 6.79 Mbp.

The type strain is SEMIA 6460^T (=ATCC 51692^T =DSM 11282^T =HAMBI 220^T =ICMP 13649^T =LMG 7834^T =NBRC 100385^T =ORS 1009^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_014197705.1.

Description of *Endobacterium yantingense* comb. nov.

Endobacterium yantingense (yan. ting. en'se. N.L. neut. adj. yantingense referring to Yanting district, Sichuan Province, PR China, where the organism was isolated).

Basonym: *Rhizobium yantingense* Chen *et al.* 2015.

The description is as provided by Chen *et al.* 2015 [64]. *E. yantingense* can be differentiated from another species of the genus *Endobacterium* (*Endobacterium cereale* corrig. Menéndez *et al.* 2021) based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 59.5 %. Its approximate genome size is 5.82 Mbp.

The type strain is H66^T (=CCTCC AB 2014007^T =LMG 28229^T), which was isolated from the surfaces of weathered rock (purple siltstone) in Yanting (Sichuan, PR China). The JGI IMG accession number for the genome sequence is Ga0196656.

Description of *Neorhizobium petrolearium* comb. nov.

Neorhizobium petrolearium (pe.tro.le.a'ri.um. L. fem. n. *petra*, rock; L. neut. *olearium* related to oil, of or belonging to oil; N.L. neut. adj. *petrolearium* related to mineral oil).

Basonym: *Rhizobium petrolearium* Zhang *et al.* 2012.

The description is as provided by Zhang *et al.* 2012 [65]. *N. petrolearium* can be differentiated from another species of the genus *Neorhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 60.5 %. Its approximate genome size is 6.97 Mbp.

The type strain, SL-1^T (=ACCC 11238^T =KCTC 23288^T), was isolated from petroleum-contaminated sludge samples in Shengli oilfield, Shandong Province, China. The JGI IMG accession number for the genome sequence is Ga0196653.

Description of *Pararhizobium arenae* comb. nov.

Pararhizobium arenae (a.re'nae. L. fem. gen. n. *arenae* of sand, the isolation source of the type strain).

Basonym: *Rhizobium arenae* Zhang *et al.* 2017.

The description is as provided by Zhang *et al.* 2017 [66]. *P. arenae* can be differentiated from another species of the genus *Pararhizobium* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 59.8 %. Its approximate genome size is 4.94 Mbp.

The type strain is MIM27^T (=KCTC 52299^T =MCCC 1K03215^T), isolated from sand of the Mu Us Desert, PR China. The NCBI RefSeq Assembly accession number for the genome sequence is GCF_001931685.1.

Description of *Pseudorhizobium tarimense* comb. nov.

Pseudorhizobium tarimense (ta.rim.en'se. N.L. neut. adj. *tarimense*, pertaining to Tarim basin in Xinjiang Uyghur autonomous region of China, where the type strain was isolated).

Basonym: *Rhizobium tarimense* Turdahon *et al.* 2013.

The description is as provided by Turdahon *et al.* 2013 [67]. *P. tarimense* can be differentiated from other species of the genus *Pseudorhizobium* based on OGRI calculations (ANI and isDDH).

The genomic G+C content of the type strain is 61.2 %. Its approximate genome size is 4.83 Mbp.

The type strain is CCTCC AB 2011011^T (=NRRL B-59556^T =PL-41^T). The JGI IMG accession number for the genome sequence is Ga0196649.

Description of *Mycoplana azooxidifex* comb. nov.

Mycoplana azooxidifex (a.zo.o.xi'di.fex. N.L. azooxidum, dinitrogenmonoxide; L. masc. suff. -fex, the maker; N.L. masc. n. azooxidifex, the dinitrogenmonoxide maker (nominative in apposition)).

Basonym: *Rhizobium azooxidifex* Turdahon *et al.* 2013.

The description is as provided by Turdahon *et al.* 2013 [67]. *M. azooxidifex* can be differentiated from other species of the genus *Mycoplana* based on OGRI calculations (ANI and isDDH). The genomic G+C content of the type strain is 64.3 %. Its approximate genome size is 5.89 Mbp.

The type strain is DSM 100211^T (=Po 20/26^T =LMG 28788^T). The NCBI RefSeq Assembly accession number for the genome sequence is GCF_014196765.1.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Prof. George M. Garrity (Michigan State University, East Lansing, MI, USA) for valuable help interpreting the ICNP, and Prof. Peter Young for helpful feedback on the manuscript. This research was enabled, in part, through computational resources provided by Compute Ontario (computeontario.ca), Compute Canada (computeCanada.ca), and BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B). The work of NK was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 429677233. Research in the G.C.D laboratory is supported by a NSERC Discovery Grant, and funding from Queen's University. A.M. and C.F. are supported by MICRO4Legumes grant (Italian Ministry of Agriculture) and ALL-IN project (H2020 ERA-NETs SUSFOOD2 and CORE Organic Cofund, under the Joint SUSFOOD2/CORE Organic Call 2019). This research was funded in part by the Wellcome Trust Grant [206194]. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

DATASETS

Dataset S1. A list of strains included in our analyses, as well as the corresponding genome assembly accessions.

Dataset S2. A representative sequence for each of the 170 nonrecombinant genes used to construct the maximum likelihood phylogeny.

Dataset S3. Average nucleotide identity (ANI_b) values for select taxa from the family *Rhizobiaceae*.

Dataset S4. *In silico* DNA-DNA hybridization (isDDH) values for select taxa from the family *Rhizobiaceae*.

REFERENCES

1. **Conn HJ.** Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938;36:320–321.
2. **Alves LMC, Souza JAM de, Varani A de M, Lemos EG de M.** The Family *Rhizobiaceae*. *The Prokaryotes* 2014;419–437.
3. **Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M 2020.** List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol*;70:5607–5612.
4. **Hördt A, López MG, Meier-Kolthoff JP, Schleuning M, Weinhold L-M, et al.** Analysis of 1,000+ type-strain genomes substantially improves taxonomic classification of *Alphaproteobacteria*. *Front Microbiol* 2020;11:468.
5. **Young JPW, Moeskjær S, Afonin A, Rahi P, Maluk M, et al.** Defining the *Rhizobium leguminosarum* species complex. *Genes* 2021;12:111.
6. **de Lajudie PM, Andrews M, Ardley J, Eardly B, Jumas-Bilak E, et al.** Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *Int J Syst Evol Microbiol* 2019;69:1852–1863.
7. **Lassalle F, Muller D, Nesme X.** Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. *Res Microbiol* 2015;166:729–741.
8. **Casida LE.** *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Evol Microbiol* 1982;32:339–345.
9. **Chen WX, Yan GH, Li JL.** Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. *Int J Syst Evol Microbiol* 1988;38:392–397.
10. **Contreras-Moreira B, Vinuesa P.** GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 2013;79:7696–7701.
11. **Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B.** GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pangenome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Front Microbiol* 2018;9:771.
12. **Kuzmanović N, Biondi E, Overmann J, Puławska J, Verborg S, et al.** Revisiting the taxonomy of *Allorhizobium vitis* (i.e. *Agrobacterium vitis*) using genomics - emended description of *All. vitis sensu stricto* and description of *Allorhizobium ampelinum* sp. nov. *bioRxiv*. Epub ahead of print 21 December 2020. DOI: 10.1101/2020.12.19.423612.
13. **Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
14. **Buchfink B, Xie C, Huson DH.** Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;12:59–60.

15. **Paradis E, Schliep K.** ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 2019;35:526–528.
16. **Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, et al.** A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.
17. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, et al.** Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 2010;11:119.
18. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al.** BLAST+: architecture and applications. *BMC Bioinform* 2009;10:421.
19. **Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M.** Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform* 2013;14:60.
20. **Fagorzi C, Ilie A, Decorosi F, Cangiali L, Viti C, et al.** Symbiotic and nonsymbiotic members of the genus *Ensifer* (syn. *Sinorhizobium*) are separated into two clades based on comparative genomics and high-throughput phenotyping. *Genome Biol Evol* 2020;12:2521–2534.
21. **Katoh K, Standley DM.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
22. **Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T.** trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.
23. **Stamatakis A.** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
24. **Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, et al.** A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 2020;70:2782–2858.
25. **Lassalle F, Dastgheib SMM, Zhao F-J, Zhang J, Verbarq S, et al.** Phylogenomics reveals the basis of adaptation of *Pseudorhizobium* species to extreme environments and supports a taxonomic revision of the genus. *Syst Appl Microbiol* 2021;44:126165.
26. **Mousavi SA, Willems A, Nesme X, de Lajudie P, Lindström K.** Revised phylogeny of *Rhizobiaceae*: Proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations. *Syst Appl Microbiol* 2015;38:84–90.
27. **Rahi P, Khairnar M, Hagir A, Narayan A, Jain KR, et al.** *Peteryoungia* gen. nov. with four new species combinations and description of *Peteryoungia desertarenae* sp. nov., and taxonomic revision of the genus *Ciceribacter* based on phylogenomics of *Rhizobiaceae*. *Arch Microbiol*. Epub ahead of print 9 May 2021. DOI: 10.1007/s00203-021-02349-9.
28. **de Lajudie, Philippe P, Willems A, Pot B, Dewettinck D, Maestrojuan G, et al.** Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *Int J Syst Evol Microbiol* 1994;44:715–733.

29. **Lindstrom K, Martínez-Romero ME.** International Committee on Systematics of Prokaryotes: Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*. Minutes of the meeting, 4 July 2001, Hamilton, Canada. *Int J Syst Evol Microbiol* 2002;52:2337.
30. **Willems A, Fernández-López M, Muñoz-Adelantado E, Goris J, De Vos P, et al.** Description of new *Ensifer* strains from nodules and proposal to transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an opinion. *Int J Syst Evol Microbiol* 2003;43:1207–1217.
31. **Young JM.** The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination ‘*Sinorhizobium adhaerens*’ (Casida 1982) Willems et al. 2003 legitimate? Request for an opinion. *Int J Syst Evol Microbiol* 2003;53:2107–2110.
32. **Judicial Commission of the International Committee on Systematics of Prokaryotes.** The genus name *Sinorhizobium* Chen et al. 1988 is a later synonym of *Ensifer* Casida 1982 and is not conserved over the latter genus name, and the species name ‘*Sinorhizobium adhaerens*’ is not validly published. Opinion 84. *Int J Syst Evol Microbiol* 2008;58:1973–1973.
33. **Lindstrom K, Young JPW.** International Committee on Systematics of Prokaryotes; Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*. Minutes of the meetings, 31 August 2008, Gent, Belgium. *Int J Syst Evol Microbiol* 2009;59:921–922.
34. **Young JM.** *Sinorhizobium* versus *Ensifer*: may a taxonomy subcommittee of the ICSP contradict the Judicial Commission? *Int J Syst Evol Microbiol* 2010;60:1711–1713.
35. **Tindall BJ.** The correct name of the taxon that contains the type strain of *Rhodococcus equi*. *Int J Syst Evol Microbiol* 2014;64:302–308.
36. **de Lajudie P, Young JPW.** International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Rhizobia and Agrobacteria Minutes of the meeting by video conference, 11 July 2018. *Int J Syst Evol Microbiol* 2019;69:1835–1840.
37. **Martens M, Delaere M, Coopman R, De Vos P, Gillis M, et al.** Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* 2007;57:489–503.
38. **Kumar HKS, Gan HM, Tan MH, Eng WWH, Barton HA, et al.** Genomic characterization of eight *Ensifer* strains isolated from pristine caves and a whole genome phylogeny of *Ensifer* (*Sinorhizobium*). *J Genomics* 2017;5:12–15.
39. **Martens M, Dawyndt P, Coopman R, Gillis M, De Vos P, et al.** Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int J Syst Evol Microbiol* 2008;58:200–214.
40. **diCenzo GC, Debiec K, Krzysztoforski J, Uhrynowski W, Mengoni A, et al.** Genomic and biotechnological characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14. *Genes* 2018;9:379.
41. **Garrido-Oter R, Nakano RT, Dombrowski N, Ma K-W, AgBiome Team, et al.** Modular traits of the Rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. *Cell Host Microbe* 2018;24:155-167.e5.

42. **Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, et al.** A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.
43. **diCenzo GC, Cangioli L, Nicoud Q, Cheng JHT, Blow MJ, et al.** DNA methylation patterns in bacteria of the genus *Ensifer* during free-living growth and during nitrogen-fixing symbiosis with *Medicago* spp. *bioRxiv*. Epub ahead of print 8 March 2021. DOI: 10.1101/2021.03.08.434416.
44. **Wilks M.** *Predation Mediated Carbon Turnover in Nutrient-Limited Cave Environments*. MSc Thesis; University of Akron; 2013.
45. **Martin MO.** Predatory prokaryotes: an emerging research opportunity. *J Mol Microbiol Biotechnol* 2002;4:467–477.
46. **Wang ET, Tan ZY, Willems A, Fernández-López M, Reinhold-Hurek B, et al.** *Sinorhizobium morelense* sp. nov., a *Leucaena leucocephala*-associated bacterium that is highly resistant to multiple antibiotics. *Int J Syst Evol Microbiol* 2002;52:1687–1693.
47. **Wang YC, Wang F, Hou BC, Wang ET, Chen WF, et al.** Proposal of *Ensifer psoraleae* sp. nov., *Ensifer sesbaniae* sp. nov., *Ensifer morelense* comb. nov. and *Ensifer americanum* comb. nov. *Syst Appl Microbiol* 2013;36:467–473.
48. **Zweiger G, Marczynski G, Shapiro L.** A *Caulobacter* DNA methyltransferase that functions only in the predivisive cell. *J Mol Biol* 1994;235:472–485.
49. **Wright R, Stephens C, Shapiro L.** The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J Bacteriol* 1997;179:5869–5877.
50. **Cohan FM.** Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Phil Trans R Soc B* 2006;361:1985–1996.
51. **Parker CT, Tindall BJ, Garrity GM.** International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2015;65:4284–4287.
52. **de Lajudie P, Mousavi SA, Young JPW.** International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Rhizobia and Agrobacteria. Minutes of the closed meeting by videoconference, 6 July 2020. *Int J Syst Evol Microbiol* 2021;71:004784.
53. **Richter M, Rosselló-Móra R.** Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009;106:19126–19131.
54. **Meier-Kolthoff JP, Klenk H-P, Göker M.** Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* 2014;64:352–356.
55. **Zhao J-J, Zhang J, Zhang R-J, Zhang C-W, Yin H-Q, et al.** *Rhizobium rhizosphaerae* sp. nov., a novel species isolated from rice rhizosphere. *Antonie van Leeuwenhoek* 2017;110:651–656.
56. **Peng G, Yuan Q, Li H, Zhang W, Tan Z.** *Rhizobium oryzae* sp. nov., isolated from the wild rice *Oryza alta*. *Int J Syst Evol Microbiol* 2008;58:2158–2163.

57. **Li Y, Yan J, Yu B, Wang ET, Li X, et al.** *Ensifer alkalisoli* sp. nov. isolated from root nodules of *Sesbania cannabina* grown in saline–alkaline soils. *Int J Syst Evol Microbiol* 2016;66:5294–5300.
58. **Merabet C, Martens M, Mahdhi M, Zakhia F, Sy A, et al.** Multilocus sequence analysis of root nodule isolates from *Lotus arabicus* (Senegal), *Lotus creticus*, *Argyrolobium uniflorum* and *Medicago sativa* (Tunisia) and description of *Ensifer numidicus* sp. nov. and *Ensifer garamanticus* sp. nov. *Int J Syst Evol Microbiol* 2010;60:664–674.
59. **Yan H, Yan J, Sui XH, Wang ET, Chen WX, et al.** *Ensifer glycinis* sp. nov., a rhizobial species associated with species of the genus *Glycine*. *Int J Syst Evol Microbiol* 2016;66:2910–2916.
60. **Nick G, de Lajudie P, Eardly BD, Suomalainen S, Paulin L, et al.** *Sinorhizobium arboris* sp. nov. and *Sinorhizobium kostiense* sp. nov., isolated from leguminous trees in Sudan and Kenya. *Int J Syst Evol Microbiol* 1999;49:1359–1368.
61. **Wei GH, Wang ET, Tan ZY, Zhu ME, Chen WX.** *Rhizobium indigoferae* sp. nov. and *Sinorhizobium kummerowiae* sp. nov., respectively isolated from *Indigofera* spp. and *Kummerowia stipulacea*. *Int J Syst Evol Microbiol* 2002;52:2231–2239.
62. **Rome S, Fernandez MP, Brunel B, Normand P, Cleyet-Marel JC.** *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* 1996;46:972–980.
63. **Lloret L, Ormeño-Orrillo E, Rincón R, Martínez-Romero J, Rogel-Hernández MA, et al.** *Ensifer mexicanus* sp. nov. a new species nodulating *Acacia angustissima* (Mill.) Kuntze in Mexico. *Syst Appl Microbiol* 2007;30:280–290.
64. **Chen W, Sheng X-F, He L-Y, Huang Z.** *Rhizobium yantingense* sp. nov., a mineral-weathering bacterium. *Int J Syst Evol Microbiol* 2015;65:412–417.
65. **Zhang X, Li B, Wang H, Sui X, Ma X, et al.** *Rhizobium petrolearium* sp. nov., isolated from oil-contaminated soil. *Int J Syst Evol Microbiol* 2012;62:1871–1876.
66. **Zhang S, Yang S, Chen W, Chen Y, Zhang M, et al.** *Rhizobium arenae* sp. nov., isolated from the sand of Desert Mu Us, China. *Int J Syst Evol Microbiol* 2017;67:2098–2103.
67. **Turdahon M, Osman G, Hamdun M, Yusuf K, Abdurehim Z, et al.** *Rhizobium tarimense* sp. nov., isolated from soil in the ancient Khyik River. *Int J Syst Evol Microbiol* 2013;63:2424–2429.

Creation and Characterization of a Genomically Hybrid Strain in the Nitrogen-Fixing Symbiotic Bacterium *Sinorhizobium meliloti*

Alice Checcucci,[†] George C. diCenzo,[†] Veronica Ghini,[‡] Marco Bazzicalupo,[†] Anke Becker,[§] Francesca Decorosi,^{||} Johannes Döhlemann,[§] Camilla Fagorzi,[†] Turlough M. Finan,[⊥] Marco Fondi,^{*,†} Claudio Luchinat,^{‡,#} Paola Turano,^{‡,#} Tiziano Vignolini,[∇] Carlo Viti,^{||} and Alessio Mengoni^{*,†}

[†]Department of Biology, University of Florence, 50019 Sesto Fiorentino, Italy

[‡]CERM & CIRMMMP, University of Florence, 50019 Sesto Fiorentino, Italy

[§]LOEWE – Center for Synthetic Microbiology, 35043 Marburg, Germany

^{||}Department of Agri-food Production and Environmental Science, University of Florence, 50019 Florence, Italy

[⊥]Department of Biology, McMaster University, Hamilton, Ontario L8S 4L8, Canada

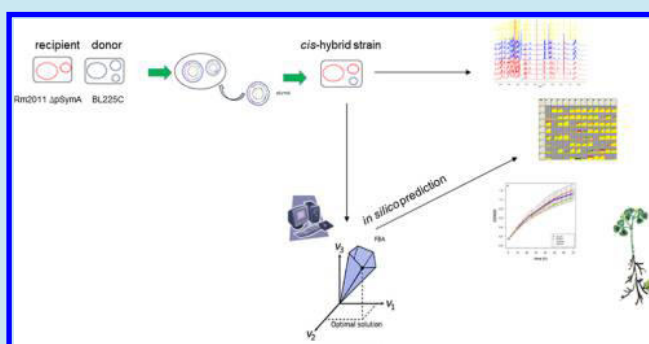
[#]CERM and Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy

[∇]European Laboratory for Non-Linear Spectroscopy, LENS, 50019 Sesto Fiorentino, Italy

Supporting Information

ABSTRACT: Many bacteria, often associated with eukaryotic hosts and of relevance for biotechnological applications, harbor a multipartite genome composed of more than one replicon. Biotechnologically relevant phenotypes are often encoded by genes residing on the secondary replicons. A synthetic biology approach to developing enhanced strains for biotechnological purposes could therefore involve merging pieces or entire replicons from multiple strains into a single genome. Here we report the creation of a genomic hybrid strain in a model multipartite genome species, the plant-symbiotic bacterium *Sinorhizobium meliloti*. We term this strain as *cis*-hybrid, since it is produced by genomic material coming from the same species' pangenome. In particular, we moved the secondary replicon pSymA (accounting for nearly 20% of total genome content) from a donor *S. meliloti* strain to an acceptor strain. The *cis*-hybrid strain was screened for a panel of complex phenotypes (carbon/nitrogen utilization phenotypes, intra- and extracellular metabolomes, symbiosis, and various microbiological tests). Additionally, metabolic network reconstruction and constraint-based modeling were employed for *in silico* prediction of metabolic flux reorganization. Phenotypes of the *cis*-hybrid strain were in good agreement with those of both parental strains. Interestingly, the symbiotic phenotype showed a marked cultivar-specific improvement with the *cis*-hybrid strains compared to both parental strains. These results provide a proof-of-principle for the feasibility of genome-wide replicon-based remodelling of bacterial strains for improved biotechnological applications in precision agriculture.

KEYWORDS: replicon independence, genome coadaptation, experimental transplantation, accessory genome, *Sinorhizobium meliloti*



Interest in large-scale genome modification and synthetic bacterial chromosome construction has strongly increased over the past decade (for instance, see the literature^{1,2}) with a goal of engineering bacterial strains with new or improved traits. However, phenotypes are often the result of the coordinated function of many genes acting together in a defined genome architecture.³ Hence, the ability to predict the phenotypic outcomes of large-scale genome modification requires a precise understanding of the genetic and regulatory interactions between each gene or gene product in the genome. As such, there is a need for integrated approaches, combining experimental evidence with computational-based methods, to interpret and potentially predict the outcomes of genome-wide DNA manipulations.

In this context, multipartite (or divided) genomes (*i.e.*, genomes possessing more than one informational molecule) are particularly interesting. The genome of bacteria with a multipartite structure is typically composed of a principal chromosome that encodes the core housekeeping and metabolic genes essential for cellular life, and one (or more) secondary replicons (termed chromids and megaplasmids). More than 10% of the presently sequenced bacterial genomes are characterized by the presence of a multipartite architecture.^{4,5} The secondary replicons can account for up to half of the total genome size, and their level of integration into cellular

Received: April 10, 2018

Published: September 17, 2018

regulatory and metabolic networks is variable.^{6–9} In some cases, strong replicon-centric transcriptional networks have been suggested.^{10–14} The apparently functional modularity of secondary replicons is particularly attractive from both ecological and biotechnological viewpoints. Indeed, secondary replicons might act as plug-and-play functional modules, potentially allowing the recipient strain to obtain previously untapped genetic information.¹⁵ This, in turn, might allow the emergence of novel phenotypic features leading, for example, to the colonization of a new ecological niche.¹⁶ Moreover, such modularity paves the way for large-scale, genome-wide manipulations of bacterial strains with a multipartite genome structure, by synthetically merging complex biotechnologically important traits in the same strain.¹⁷

However, it remains unclear to what extent complex phenotypes can be directly transferred into a recipient strain, as secondary replicons are in part coadapted to the host genome, for example, through regulatory interaction and/or inter-replicon metabolic cross-talk.^{10,18–21} In the past years, there have been examples of conjugal transfer of large plasmids in some bacterial species, including rhizobia and agrobacteria.^{22–34} However, especially for the very large (>1 Mb) megaplasmids of rhizobia that have been transferred between species, the transfer was accompanied by only partial transmission of the plasmid phenotypic features. However, interspecies transfer events are conceptually distinct from the proposal to create *cis*-hybrid strains through intraspecies (interstrain) replicon transfers. We are aware of only one study that deeply examined the phenotypic consequences of replacing a large (>800 kb) native secondary replicon with a homologous replicon of closely related strains. In that study, the third replicon of *Burkholderia cepacia* complex strains was mobilized and the effects on various phenotypes including virulence was examined.³⁵ It was found that in some cases, phenotypes were dependent solely on the secondary replicon, whereas in other cases, the phenotypes depended on genetic/regulatory interactions with the other replicons.³⁵ However, additional studies are required to examine the generalizability and the controllability of those observations.

To further test the feasibility, the stability, and the predictability of secondary replicon shuffling on the phenotype(s) of the cell, here we have performed experimental and *in silico* replicon transplantation between two bacterial strains. We used the symbiotic nitrogen-fixing bacterium *Sinorhizobium meliloti* as a model, considering that it is characterized by a well-studied multireplicon genome structure.^{16,36} Additionally, *S. meliloti* represents a highly valuable microorganism in agriculture, as its symbiosis with crops like alfalfa is estimated to be worth more than \$70 million/year in the United States.³⁷ The genome of the mostly commonly studied *S. meliloti* strains (the two very closely related strains Rm1021 and Rm2011) are composed by a chromosome (~3.7 Mb), and by two secondary replicons: a chromid (~1.7 Mb, called pSymB) and a megaplasmid (~1.4 Mb, called pSymA). The pSymA megaplasmid determines many of the biotechnologically relevant functions, as it harbors most of the key genes involved in plant symbiotic colonization and in nitrogen-fixation.³⁸ The *S. meliloti* large replicons have recently been proposed as scaffolds for novel shuttle vectors for synthetic biology.³⁹ Furthermore, previously performed genome reduction experiments have led to the complete removal of one or both of the two secondary replicons,^{16,40} and an *in silico* genome-scale metabolic model has been

reconstructed,⁴¹ paving the way for massive genome-scale remodelling of *S. meliloti*.

Here, we constructed a hybrid strain containing the chromosome and the chromid of the laboratory *S. meliloti* Rm2011 strain with the pSymA replicon from the wild isolate *S. meliloti* BL225C. The genome of BL225C is 290 kbps larger than that of Rm2011;^{36,42} moreover, up to 1,583 genes are differentially present in the genomes of these strains.^{43,44} Furthermore, the BL225C strain has been shown to have several interesting biotechnological features, including plant growth promotion, and nodulation efficiency.^{36,45} The majority of the genetic differences between Rm2011 and BL225C are associated with the homologous pSymA and pSINMEB01 homologous megaplasmids (as shown in Supplemental Figure S1); 836 of the 1,583 variable genes are located on this replicon.⁴³ We therefore expect that creating a hybrid strain between Rm2011 and BL225C, by moving the pSymA-equivalent from BL225C to the Rm2011 derivative lacking pSymA, will provide a good testing ground to examine (i) the feasibility of large replicon shuffling between strains, and (ii) the stability and predictability of the phenotypes linked to such replicons. We term this novel hybrid strain as *cis*-hybrid since it derives from *cis*-genetic manipulation and contains genetic material from the pangenome pool of the same species (in contrast to a *trans*-genetic strain that would contain genes from a distinct species). *cis*-Hybrid strains could be an important way to promote environmental-friendly and regulatory compliant biotechnology and synthetic biology in bacterial species of interest in agricultural and environmental microbiology.¹⁷

RESULTS AND DISCUSSION

We report here the creation of a *cis*-hybrid *S. meliloti* strain, where the symbiotic megaplasmid pSINMEB01, was transferred from the natural strain BL225C to the laboratory strain Rm2011. The pSINMEB01 megaplasmid is homologous to pSymA, and is ~1.6 Mb in size, accounting for nearly 23% of total genome. *In silico* metabolic reconstruction and a large set of phenotypic tests, including Nuclear Magnetic Resonance (NMR)-based metabolomic profiling, Phenotype Microarray, and symbiotic assays with different host plant cultivars have been performed.

Experimental Creation of a *cis*-Hybrid Strain. Starting with the derivative of the *S. meliloti* Rm2011 strain that lacks the pSymA replicon, herein referred to as Δ pSymA,⁴⁶ we produced a *cis*-hybrid strain that contains the Rm2011 chromosome and pSymB chromid, and the pSymA replicon from a genetically and phenotypically distinct *S. meliloti* strain, BL225C (all strains used in this work are listed in Table 1).^{36,45} The *cis*-hybrid strain was produced through a series of conjugations as described in the Methods section. Briefly, a plasmid for overexpressing *rctB*⁴⁷ was transferred to BL225C; as RctB is a negative regulator of RctA,⁴⁷ which in turn is a negative regulator of the pSymA conjugal genes,^{48,49} this step was necessary to promote pSymA transfer without mutating the replicon. Concurrently, a plasmid carrying an antibiotic resistance gene marker (gentamycin, plasmid pMP760S⁵⁰) was transferred to the Δ pSymA strain to allow for the use of the gentamycin resistance marker in the selection of *cis*-hybrid transconjugants in the next step. Finally, a mating mixture of the Δ pSymA acceptor strain and the BL225C donor strain was prepared, and *cis*-hybrid transconjugants were isolated on a medium selective for the gain of the pSymA replicon (see Methods). Correct construction of the *cis*-hybrid strain was

Table 1. Strains and Plasmids Used in This Study^a

strain or plasmid	description	reference
<i>Sinorhizobium meliloti</i>		
Rm2011	Wild type SU47 derivative; Sm ^R	Sallet <i>et al.</i> , 2013 ⁴²
BL225C	Wild isolate from <i>Medicago sativa</i> in Lodi (Italy)	Galardini <i>et al.</i> , 2011 ³⁶
RmP3498	Rm2011 ΔpSymAB+B with engA and tRNA into the chromosome; Sm ^R Sp ^R	diCenzo <i>et al.</i> , 2014 ¹⁶
BM 826	RmP3498 with pMp7605; Sm ^R Sp ^R Gm ^R	This study
BM 806	BM 826 with pSymA from BL225C; Sm ^R Sp ^R Gm ^R	This study
BM 848	BL225C with pTE3rctB; Tc ^R	This study
<i>Escherichia coli</i>		
MT616	Helper strain carrying pRK600 that has the RK2 <i>tra</i> genes; Cm ^R	Finan <i>et al.</i> , 1986 ³¹
Plasmids		
pMp7605	Broad host range vector constitutively expressing the <i>mCherry</i> gene; Gm ^R	Lagendijk <i>et al.</i> , 2010 ⁵⁰
pTE3rctB	Broad host range vector overexpressing the <i>Rhizobium etli rctB</i> gene; Tc ^R	Nogales <i>et al.</i> , 2013 ⁴⁷

^aCode of strains and plasmid is reported. A succinct description of the main phenotypic features is shown; Sm^R, streptomycin resistance, Sp^R, spectinomycin resistance, Gm^R gentamycin resistance, Tc^R tetracycline resistance, Cm^R, chloramphenicol resistance.

initially confirmed through PCR amplification on specific unique genes on the Rm2011 chromosome and pSymB, and the pSymA homologue of BL225C (pSINMEB01) (see Table S1). Subsequently, whole genome sequencing (Figure 1, Supplemental Figure S1) confirmed the complete transfer of pSymA, and the banding pattern observed in pulse-field gel electrophoresis (PFGE) (Supplemental Figure S2) was consistent with

pSINMEB01 being present as an independent replicon (*i.e.*, not integrated into the chromosome or pSymB).

In Silico Metabolic Network Reconstruction. In addition to the experimental creation of the *cis*-hybrid strain, we attempted to predict its metabolic outcomes by generating a new metabolic model which includes the genomic features present in the *cis*-hybrid strain. The curated iGD1575 reconstruction (herein referred to as the Rm2011 reconstruction) was used to represent metabolism of *S. meliloti* Rm2011,⁴⁶ although iGD1575 is based on the strain Rm1021, the genomic content of these strains is 99,9% identical, with the exception of numerous SNPs⁴² that are not considered in the process of metabolic reconstruction. Next, our recently described pipeline⁵¹ was used to build a representation of BL225C based on a draft reconstruction built with the KBase Web server and enhanced based on the iGD1575 model. An *in silico* representation of the *cis*-hybrid strain was then built by removing all pSymA genes (and dependent reactions) from the Rm2011 model, followed by the addition of all pSymA (pSINMEB01) genes (and associated reactions) from the BL225C model using our published pipeline.⁵¹ Despite there being numerous (47 to 143 gene) differences in the gene content of the metabolic reconstructions, the Rm2011 model differed from the BL225C and the *cis*-hybrid models by no more than a half dozen reactions (Table 2). The low reaction variability between models may (i) reflects the difficulty in predicting the function of the *S. meliloti* variable gene content, (ii) suggests the presence of nonorthologous genes encoding proteins catalyzing the same reaction(s), and/or (iii) indicates that few metabolic features are dependent on the accessory gene set. Not surprisingly, given the near identical reaction content of the reconstructions, the outputs of flux balance analysis simulations for the different reconstructions

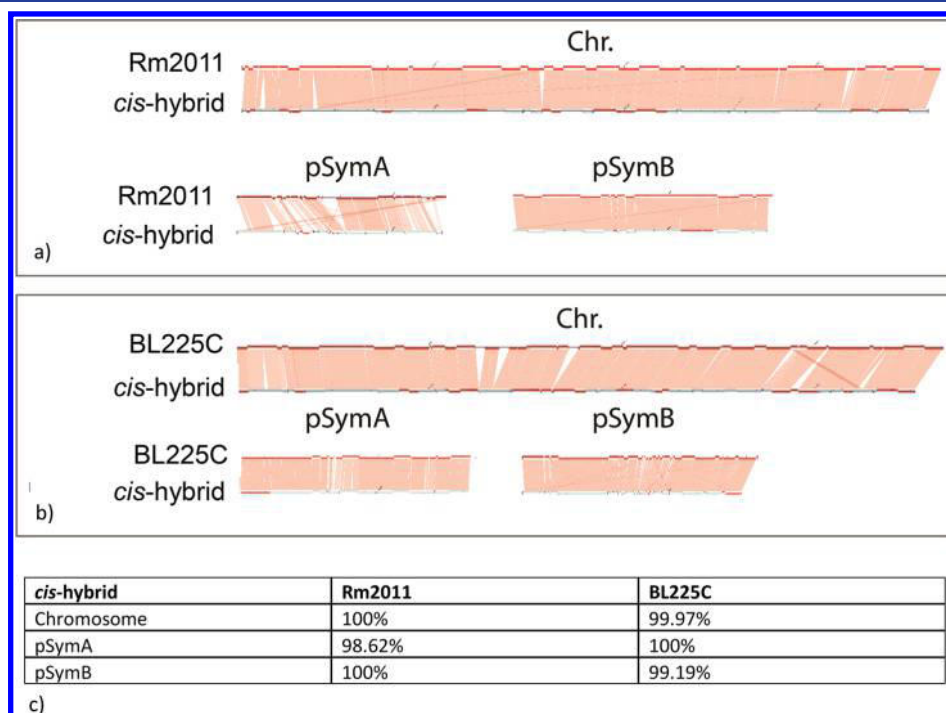


Figure 1. Confirmation of the genome structure of the *cis*-hybrid strain. (a) Comparison of *cis*-hybrid strain genome sequence with Rm2011 chromosome and pSymB, and with pSymA of the donor strain (BL225C) (contigs alignment performed with Contiguator). Apparent indels in the alignment of the *cis*-hybrid assembly with the Rm2011 genome are artifacts of the assembly, as revealed by directly mapping sequencing reads to the reference Rm2011 genome (Figure S1). (b) Percentage of identity of each replicon which composed the multipartite genome of *cis*-hybrid strain with those of the donor strains (Rm2011 and BL225C).

Table 2. Comparison of *S. meliloti* Metabolic Network Reconstructions^a

strain	genes			reactions		
	BL225C	<i>cis</i> -hybrid	Δ pSymA	BL225C	<i>cis</i> -hybrid	Δ pSymA
Rm2011	1525/52/91	1551/26/76	1336/241/0	1821/6/6	1823/4/6	1755/72/0
BL225C	–	1598/18/29	1308/308/28	–	1827/0/2	1753/74/2
<i>cis</i> -hybrid	–	–	1336/291/0	–	–	1755/74/0

^aThe gene and reaction content of the four *S. meliloti* metabolic reconstructions used in this work are shown. For each cell, the values are a comparison of the strain indicated on the left with the strain indicated along the top. Three values are provided in each cell, and these correspond to the following. The first value is the number of genes or reactions in common between the models. The second value is the number of genes or reactions present in the reconstruction on the left but not in the one along the top. The third value is the number of genes or reactions present in the reconstruction along the top but not in the one on the left.

were nearly identical (data not shown); therefore, we do not describe these results further.

Metabolic Phenotypes and Profiles of the *cis*-Hybrid Strain. To confirm that the process of replicon swapping did not result in any unintended metabolic perturbations, the *cis*-hybrid strain and the parental strains were characterized using Phenotype MicroArray experiments to test metabolic capacity, and using ¹H nuclear magnetic resonance (NMR) to compare their metabolomic profiles.

Phenotype MicroArray experiments were performed to test the substrate utilization abilities (as presence of an oxidative metabolism, hence growth) of the *cis*-hybrid strain, as well as the parental and wild type strains, with 192 different carbon sources and 96 different nitrogen sources. Previous work has shown that the pSymA megaplasmid has little contribution to the metabolic capacity of *S. meliloti*.^{16,40} Consequently, we expected a small number of differences in the *cis*-hybrid strain. A previous Phenotype MicroArray experiment identified only four carbon substrates (3-methyl glucose, D-gluconic acid, D-ribose-1,4-lactone, and β -D-allose) and one nitrogen source (cytosine) that appeared dependent on pSymA in Rm2011.¹⁶ 3-methyl glucose was the only carbon source of these four that supported growth of Rm2011 in the current study. The contribution of pSymA to metabolism of BL225C has not been elucidated. Consistent with the past results, only minor changes in the substrate utilization abilities were observed following the introduction of the pSymA of BL225C (pSINMEB01) into the Δ pSymA strain (Figure 2, Table S2). Importantly, the Δ pSymA strain did not grow with 3-methyl glucose as a carbon source and cytosine as a nitrogen source. Both these metabolic abilities were restored in the *cis*-hybrid strain. These results confirm the expression of metabolic functions the pSymA homologue in the *cis*-hybrid strain.

Concerning the differences between the *cis*-hybrid strain and the two parental ones, Rm2011 and BL225C, the *cis*-hybrid strain showed a higher activity index (a metrics of substrate utilization⁵²) with respect to both the parental strains on 22 out of 192 (11.4%) carbon sources tested, and a lower activity index in 29 of 192 (15.1%) carbon sources (Supplemental Table S2a). For the nitrogen sources utilization (Supplementary Table S2b), the *cis*-hybrid strain showed higher activity index values with respect to the parental strains in 6 out of 96 (6.2%) tested compounds, and a lower activity index in 27 of the 96 (28.1%) compounds. Most of the compounds on which the *cis*-hybrid strain displayed different activity index with respect to both parental strains (as for instance formic acid, propionic acids, pyruvic acid, L-alanine, L-serine *etc.*) are the same shown previously accounting for metabolic differences by *S. meliloti* wild-type strains, including Rm1021 and BL225C.⁴⁵

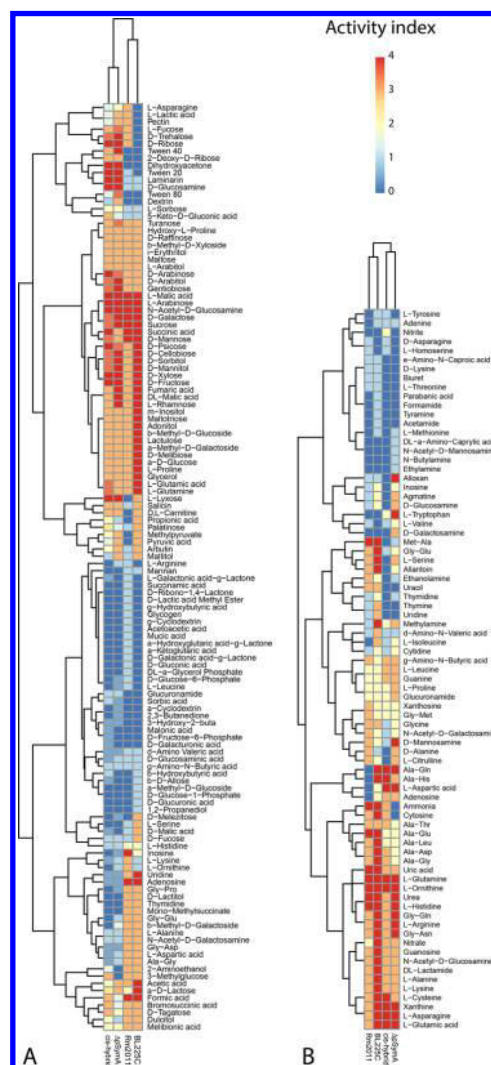


Figure 2. Metabolic phenotype of the *cis*-hybrid strain. Heatmap of Phenotype Microarray profiles of the growth on different carbon and nitrogen sources for Rm2011, BL225C, *cis*-hybrid and Δ pSymA strains. (a) Heatmap with Euclidean clustering; (b) values of pairwise Euclidean distances.

We further inspected if the changes in metabolic abilities were related to genetic differences in the megaplasmids. This analysis was performed with DuctApe dape module.⁵² In 87 of 192 carbon sources, the *cis*-hybrid strain showed different metabolic activity with respect to the sole parental Rm2011, with whom it shares the chromosomal and chromid background. Among these, the *cis*-hybrid activity index is lower in 48 and higher in 39 of the 87 carbon sources. Only in few cases

the observed differences mirror the replicon-localization of the metabolic pathways. Indeed, a small number of genes involved in the metabolic pathways for the carbon sources are located on pSymA (for instance, SM2011_a0233 of Rm2011 and SinmeB_5530 of BL225C involved in D-trehalose metabolism, SM2011_a0796, SM2011_a1844, SM2011_a2213 of Rm2011 and SinmeB_5059, SinmeB_5059, SinmeB_6004 of BL225C involved in acetic acid metabolism, SM2011_a0878 of Rm2011 and SinmeB_6071 of BL225C involved in D-fructose-6-phosphate metabolism, SM2011_a0398 and SM2011_a0306 of Rm2011 and SinmeB_6485 and SinmeB_6542 of BL225C in histidine metabolism). Only few of such genes are present in one megaplasmid only (for instance, the Rm2011 aminotransferase related with L-alanine metabolism SM2011_a1495 and the ethanolamine ammonia-lyase subunit EutB (SinmeB_5765) involved in 2-aminoethanol metabolism, and the glucoamylase SinmeB_5936 of dextrin metabolism, present only in BL225C). The same trend was observed for the nitrogen sources: in 48 of the 96 tested, the *cis*-hybrid strain showed different metabolic activity with respect to the parental Rm2011 only, in 31 cases its activity index is lower and in 17 it is higher respect to the reference strain. Also in this case, only few genes involved in the nitrogen sources metabolism are present in one of megaplasmids only (for instance, SinmeB_5853, involved in L-asparagine metabolism and SinmeB_5765 involved in agmatina metabolism, on pSymA homologue of BL225C). Interestingly, in some cases one gene present on pSymA of BL225C (SinmeB_6469) takes part to metabolic reactions which in Rm2011 request only genes on the chromosome (as in the case of guanosine, inosine and xanthosine). In this situation, the additional participation of a gene on BL225C pSymA-homologue (pSINMEB01), which is necessary in the metabolism of the strain, can determine an increase or a decrease of metabolism in a different chromosomal background. In conclusion, the DuctApe dape analysis showed that only part of the observed differences between the *cis*-hybrid strain and the parental ones can be related with genetics differences (as genes loss or acquisition) at the pSymA level. In agreement with such analysis is interesting to note that the substrate utilization abilities of the *cis*-hybrid strain were difficultly predictable with the metabolic modeling also. This is perhaps not surprising given that less than 1% of reactions differed between the *in silico* metabolic reconstructions of the strains. Consequently, we can speculate that the differences in Phenotype Microarray attributed of the *cis*-hybrid strain could be associated with regulatory differences, changes in gene expression levels and/or with alleles with different functionality, all differences which are not considered in the metabolic modeling.

To more in deep investigate the possible metabolic alterations following pSymA transplantation, a metabolomic analysis through NMR was performed. Using an untargeted approach, both cellular lysates and spent growth media were analyzed to identify the fingerprint of the endo- and exometabolomes of the two parental strains, the *cis*-hybrid strain, and the Δ pSymA recipient.

Principal component analysis (PCA) was used to generate an initial overview of the metabolome differences among the four strains (Figure 3a,b), followed by PCA-Canonical Analysis (CA) to obtain the best discrimination among the strains by maximizing the differences among their metabolomic profiles (Figure 3c,d). In both the PCA and PCA-CA score plots (Figure 3), the *cis*-hybrid strain clustered very close to both the

Δ pSymA recipient strain and to the parental strain Rm2011, whereas the parental donor strain BL225C clustered separately. These results are consistent with previous data indicating that pSymA has little contribution to the metabolome,⁷ proteome,⁸ or transcriptome⁵³ of *S. meliloti* Rm2011 in laboratory conditions. Importantly, these results confirmed that the synthetic large-scale horizontal gene transfer performed here to produce the *cis*-hybrid strain did not result in a major perturbation of the cellular metabolism.

In addition to the multivariate analysis of the metabolic NMR fingerprints described above, the signals of 25 and 19 metabolites were unambiguously assigned and integrated in the ¹H NMR spectra of the cell lysates and growth media, respectively (Figure S3). The metabolites that are characterized by statistically significant differences in concentration levels in at least one strain with respect to the two other strains are indicated in Supplementary Table S3 and are also reported in Supplemental Figure S4. Validating the ability of this approach to detect metabolic differences between the strains, it was noted that the Δ pSymA strain exported cytosine unlike the wild type Rm2011 or the *cis*-hybrid strain, consistent with the inability of this strain to catabolize cytosine as shown by the Phenotype MicroArray data.

Overall, the data reported here are consistent with (i) at least some of the genes of the homologous pSymA being properly expressed, and their gene products function, and (ii) that the replicon swapping procedure did not result in any major, unintentional metabolic consequences.

Assessment of the Phenotypes of the *cis*-Hybrid Strain. Growth Profiles in Synthetic Laboratory Media. Growth profiles of the *cis*-hybrid and parental strains in complex (TY) and defined (M9-succinate) media are reported in Figure 4. In the complex TY medium (Figure 4a), the growth of the *cis*-hybrid strain was impaired relative to the recipient (Δ pSymA), and to the Rm2011 and BL225C parental strains. In other words, gain of the BL225C pSymA replicon by the Δ pSymA strain resulted in a decrease in the growth rate in TY medium. Although we cannot provide a definitive explanation for this phenomenon, it may be that the simultaneous gain of hundreds of new genes not integrated into cellular networks imposes a high metabolic cost to the cell, resulting in impaired fitness. In contrast, little to no difference was observed in the growth rate of any of the strains in the defined M9-succinate medium (Figure 4b). The lack of an observable growth impairment of the *cis*-hybrid strain in the M9-succinate medium may be due to the masked general decrease in growth rate of all strains in this medium. Moreover, the similarity of the growth profiles of all strains in the minimal medium suggest that, at least in artificial laboratory conditions, the main growth characteristics of these strains are primarily dependent on the core, not accessory, genome.

Growth Using Root Exudate as a Nutrient Source. Root exudates can be considered a proxy of the nutritional conditions of the plant rhizosphere and contain numerous compounds that can be used by bacteria as carbon and nitrogen sources.⁵⁴ We therefore evaluated the ability of the four strains to grow on M9 mineral medium supplemented with root exudates of *Medicago sativa*, a *S. meliloti* symbiotic partner. None of the strains were able to grow when the root exudate was used as the sole carbon source (data not shown); this was likely due to a very low amount of carbon sources contained in the root exudate unable to support the growth. In contrast, all strains can utilize the root exudate as the sole nitrogen source

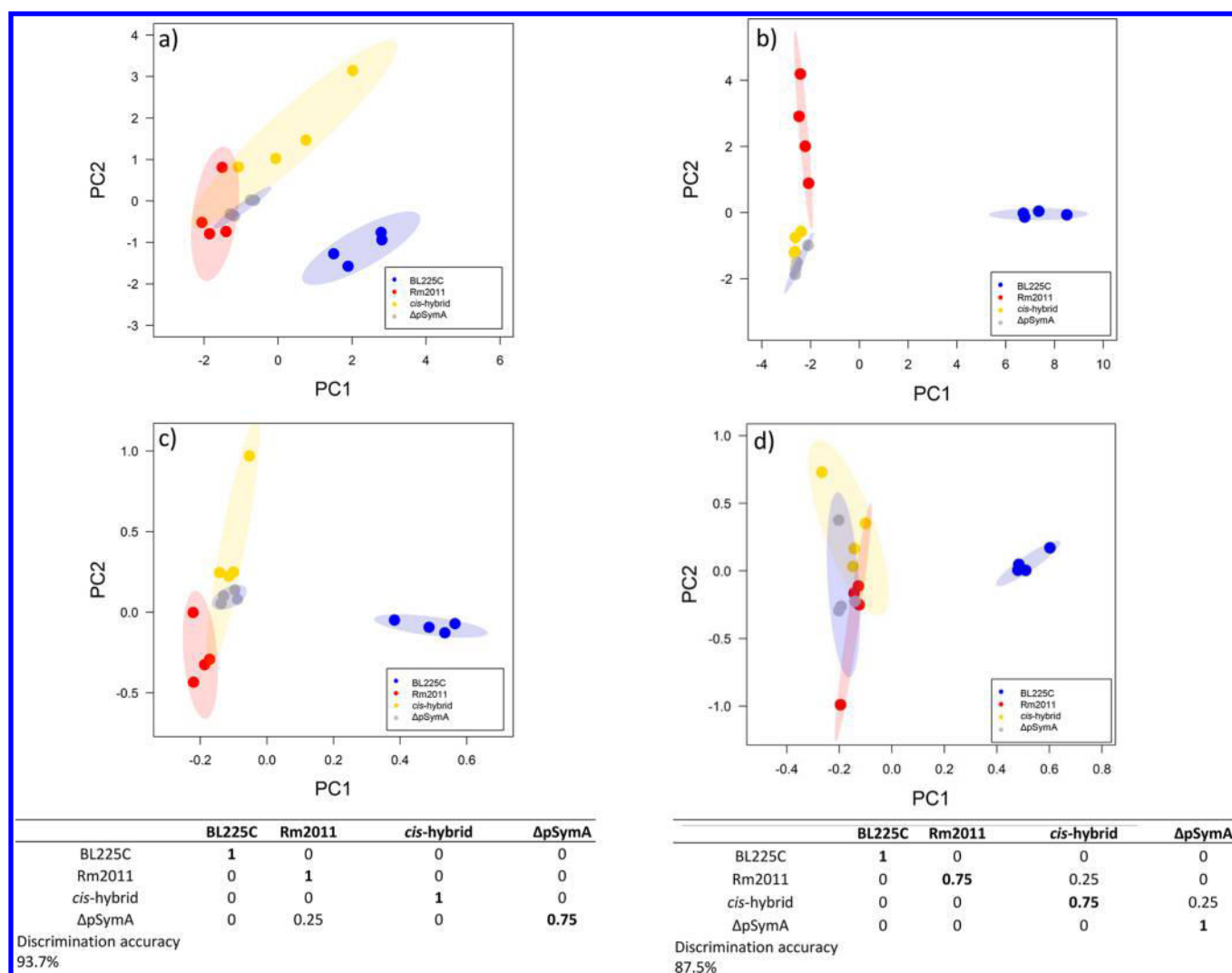


Figure 3. ^1H nuclear magnetic resonance (NMR)-based metabolomic profiles of cellular lysates and growth media of Rm2011, BL225C, *cis*-hybrid and Δ pSymA strains. Score plot of PCA (a,b) and PCA-CA (c,d) analysis of cell lysates (a,c) growing media (b,d). The confusion matrices and the discrimination accuracy values for PCA-CA analysis are also reported. Ellipses in the score plots illustrate the 95% confidence level.

when provided succinate as a carbon source to the M9 medium, as confirmed in the control the nitrogen-free M9 medium (Supplementary Figure S5). Interestingly, different growth kinetics were observed (Figure 4c). In particular, BL225C displayed the highest growth among all four strains when grown with root exudates as a sole nitrogen source. Plating for viable colony forming units confirmed the differences in the final population densities (data not shown). As the robust growth of BL225C with root exudates did not transfer to the *cis*-hybrid strain, it is likely that this phenotype is primarily dependent on the chromosome and/or pSymB of BL225C, as was suggested by previous studies.^{16,43,46,55} This observation would further suggest that the adaptation of the tested strains to growth in the rhizosphere occurred prior to the gain of pSymA and symbiotic abilities, consistent with recent work indicating that the majority of *S. meliloti* rhizosphere growth-promoting genes are chromosomally encoded.⁵⁶ Finally, considering that there are relatively few differences in the nitrogen metabolic capacity of Rm2011 and BL225C,⁴⁵ and that FBA (flux balance analysis) simulations for the metabolic model reconstructions were nearly identical (data not shown), we hypothesize that the

growth differences observed between these strains are primarily related to regulatory differences, and less so to differences in metabolic genes.

Biofilm Formation. Biofilm formation is a key factor in root colonization and plant invasion for many Proteobacteria.⁵⁷ In particular, optimal colonization of the roots by rhizobia can influence nodule formation efficiency and competitiveness.^{58,59} The formation of biofilm is a complex phenomenon and is determined by a series of regulatory mechanism involving many genes.⁵⁹ Consequently, to evaluate if the large genomic manipulation of the *cis*-hybrid strain resulted in a possible impairment of this process, biofilm formation was measured for the *cis*-hybrid strain and the parental strains in two growth conditions (the complex TY medium and the defined Rhizobium defined medium (RDM) medium). Biofilm production (estimated as the total biofilm-to-biomass ratio) by the *cis*-hybrid was similar to the parental strains in TY medium, and was higher than the parental strains in minimal RDM medium ($p < 0.005$; Figure S6). This result supports the robustness of this phenotype over the genomic manipulation performed. The higher production of biofilm in RDM by the *cis*-hybrid may additionally suggest that some still unknown

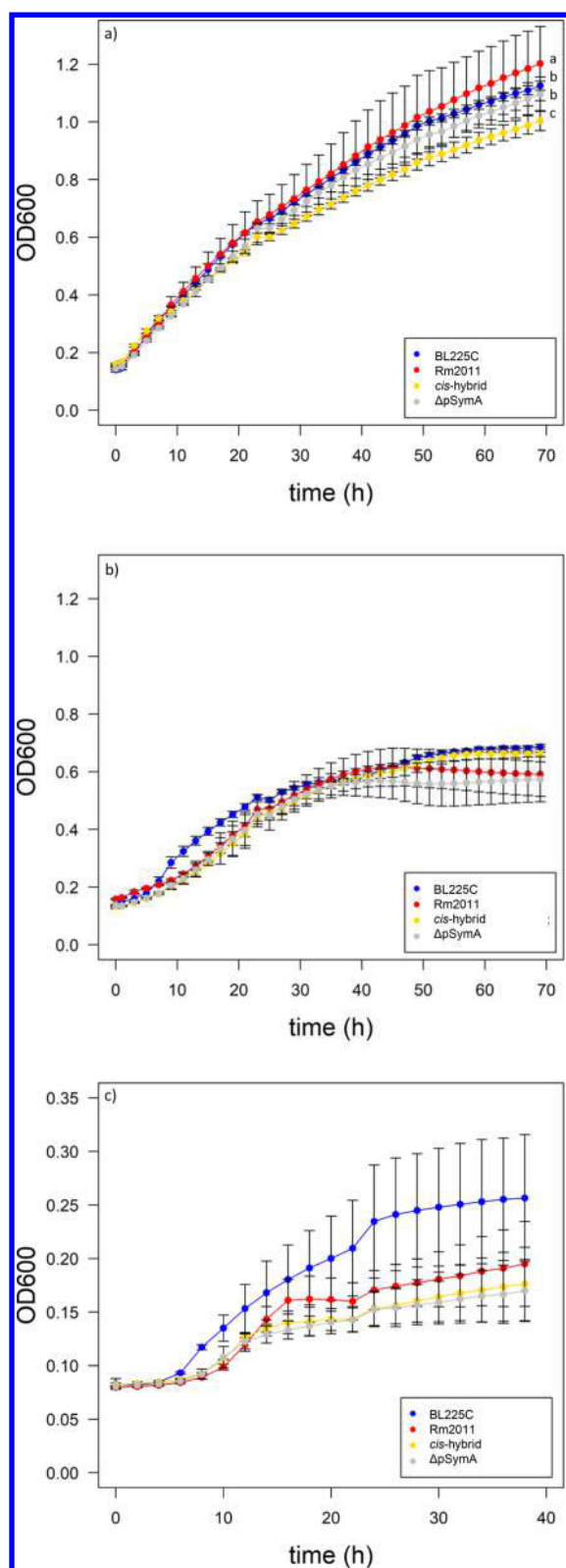


Figure 4. Growth phenotypes of the *cis*-hybrid strain. The growth of *S. meliloti* was examined in TY complex medium (a) M9 minimal medium (b) and (c) M9 + succinate and root exudates as sole N source. Data points represent averages from quadruplicate measurements. The letters on the curves represent the statistically significant differences among the strains growth ($p < 0.005$, Tukey *post hoc* contrasts).

regulatory mechanisms may have been modified. This difference between TY and RDM is also observed for the Δ pSymA recipient strain, which produced a higher ($p < 0.005$) level of biofilm with respect to the other three strains in the complex TY medium (Figure S6a), but a lower or similar level in the defined medium RDM (Figure S6b), as already described for the deletion of the common *nod* genes (harbored by pSymA) and of the entire pSymA.⁶⁰ The different phenotype observed on TY and RDM media further supports the suggestion that during growth in complex medium, there may be a still unknown pSymA-mediated negative regulation of biofilm formation.

Symbiotic Phenotypes of the *cis*-Hybrid Strain. Many of the key genes required for symbiotic abilities (e.g., nodule formation and nitrogen-fixation) are present on pSymA of *S. meliloti* Rm2011^{38,61} and the homologous megaplasmid pSINMEB01 of BL225C.³⁶ These replicons additionally contain nonessential genes that promote improved symbiotic abilities.⁶² While many of the symbiotic genes are conserved between these strains, relevant differences between pSymA and pSINMEB01 are present. The 482 genes exclusive to pSINMEB01 encode symbiotic (e.g., *nws*, *hemA* homologue, C P450³⁶) and nonsymbiotic functions (e.g., *acdS*).⁶³ For these reasons, this replicon swapping study was initiated in large part to evaluate whether swapping the symbiotic megaplasmid could promote differential symbiotic abilities.

To test the robustness of symbiotic abilities following replicon transplantation, *in vitro* symbiotic assays were performed on a panel of alfalfa cultivars, as alfalfa is the main host legume of *S. meliloti*⁶⁴ (Figure 5). In particular, the *cis*-hybrid strain and the two parental strains (Rm2011, BL225C) were tested in combination with six alfalfa cultivars (Table S4). These cultivars belong to the species *Medicago sativa*, *Medicago x varia*, and *Medicago falcata*, and they are representative of the variability of cultivars and germplasm mainly used as crops in Europe. Moreover, BL225C was originally isolated on the *M. sativa* cultivar “Lodi” at the CREA-FLC institute (Italy) during a long-course experiment.⁶⁵ The percentage of nodulated plants (Figure 5a), the number of nodules per plant (Figure 5b), the plant aerial part length (Figure 5c), the shoot dry weight (Figure 5d), and nodule colonization abilities (Supplemental Figure S7) were recorded using standard procedures.^{45,66} Not surprisingly, for each strain there was high variability in the symbiotic phenotypes observed with the different cultivars. The symbiotic interaction is a multistep developmental process which involves a tight exchange of signals between the bacterium and the plant root at both rhizospheric and endophytic levels.^{17,67} Earlier works have demonstrated strain and cultivar specificities in this process, which results in *S. meliloti* strains displaying differential symbiotic effectiveness with various plant genotypes.^{65,68,69}

The *cis*-hybrid strain performed very poorly in symbiosis with some cultivars, such as in the cultivars “Prosementi”, “Camporegio” and “Verbena”. In particular, the number of nodules per plant and the length of the aerial part were lower (Figure 5b,c, Table S5) ($p < 0.005$), indicating that the pSymA and pSINMEB01 megaplasmids are not always interchangeable. This could reflect the importance of the genomic context of the symbiotic megaplasmid and hypothetically the importance of inter-replicon regulatory networks.^{10,70,71} Alternatively, pSINMEB01 may lack important (but still unknown) symbiotic genes whose function may be replaced by chromosomal genes in BL225C but not by chromosomal genes in Rm2011.

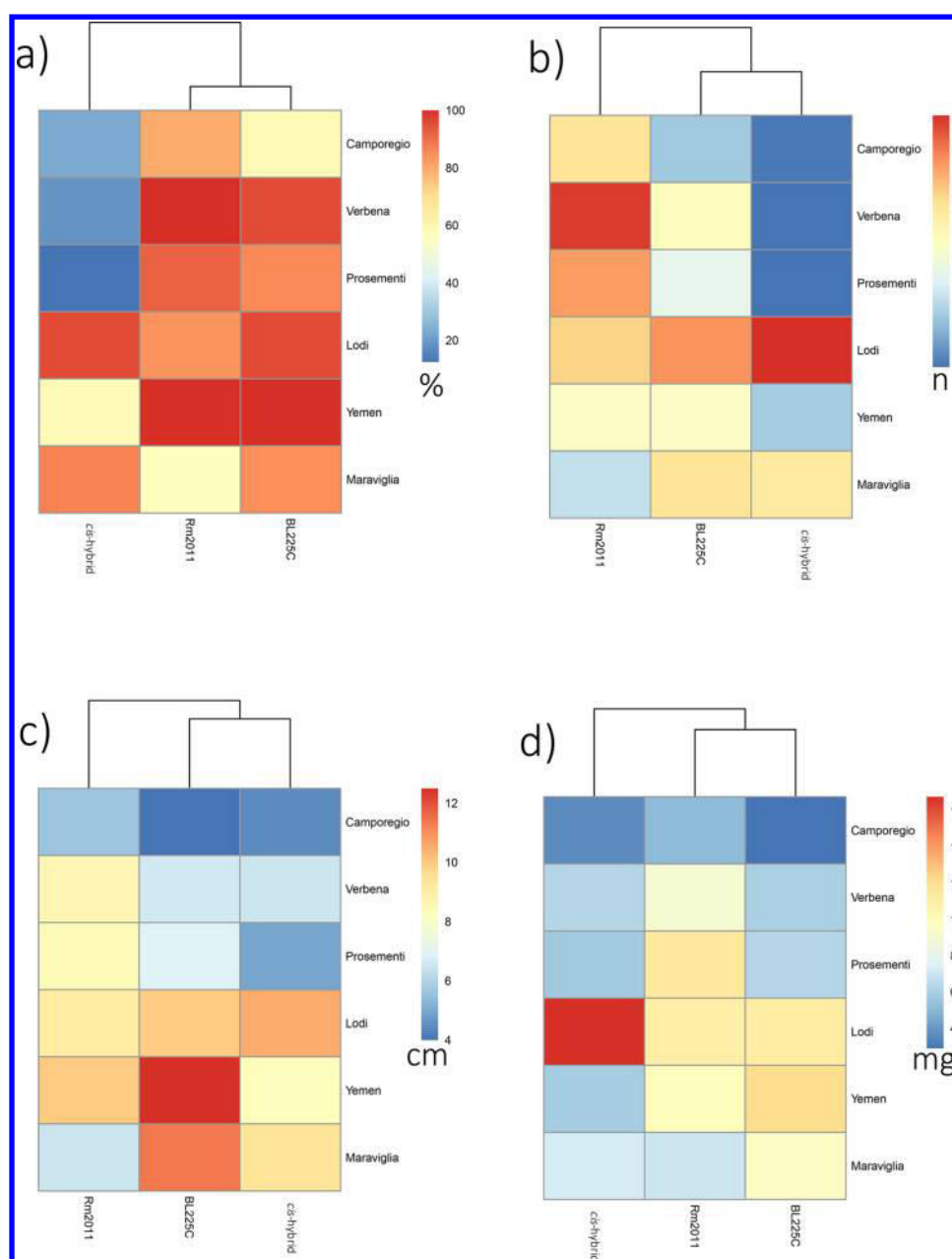


Figure 5. Symbiotic capabilities of the *cis*-hybrid strain. Heatmaps of symbiotic performances profiles for Rm2011, BL225C and *cis*-hybrid strains in a panel of six alfalfa cultivars; (a) percentage of nodulated plants, (b) number of nodules per plant, (c) plant aerial part length (cm) and (d) the shoot dry weight (mg).

Strikingly, the *cis*-hybrid strain displayed clearly improved symbiotic capabilities during symbiosis with the cultivar “Lodi” compared to both Rm2011 and BL225C (Figure 5). This was true for several key measures of symbiosis, including nodule number per plant, shoot dry weight, and length of the aerial part of the plant. The competitive abilities of the *cis*-hybrid strains, assessed through a coinoculation experiment, were similar to those of the two parental strains (Rm2011 and BL225C) (Figure S8).

These data suggest the presence of nonlinear and genomic context dependent genic interactions in the establishment of symbiotic abilities. Such interactions may resemble (at the logic level) those present in some eukaryotic genomes that result in the so-called “hybrid vigor”, *i.e.*, the tendency for hybrids to be superior to the parental genotypes.⁷² However,

since hybrid vigor is related to heterozygosity, in our case we may speculate that strain-by-strain variability of regulons,¹⁰ as well as the metabolic redundancy of *S. meliloti* genome^{5,73,74} (which could in some way mimic the presence of multiple alleles) could contribute to the increase in the observed symbiosis-related phenotypes.

Summing up, these data highlight the potential of a large-scale genome manipulation approach to obtain highly effective, and cultivar specific, rhizobial strains. This provides a rational basis for the use of similar approaches in the development of elite bioinoculants for use in precision agriculture.^{17,75}

CONCLUSIONS

The work presented here provides a proof-of-principle for the feasibility of using a large-scale genome manipulation approach

that makes use of the species' pangenome (*i.e.*, the extended gene set present in a group of microbial strains belonging to the same species⁷⁶) to produce daughter strains with improved biotechnologically relevant (*i.e.*, nitrogen fixing symbiosis) characteristics.¹⁷ In the current work, the large-scale genome manipulation was based on the transplantation of the primary symbiotic megaplasmid of a bacterial multipartite genome, a genome organization commonly found in the rhizobia. Although an entire replicon accounting for more than 20% of the total genome content was replaced with a homologous replicon of a closely related strain resulting in the gain of 482 new genes (in addition to numerous SNPs) and the loss of 354 genes, most of the core metabolic phenotypes appeared largely resilient to modification with this approach. However, other phenotypes, particularly complex (*i.e.*, multigenic) phenotypes such as the symbiotic phenotypes, gave interesting features that support the validity of this approach to improve biotechnologically relevant properties.

METHODS

Microbiological and Genetic Methods. Strains and plasmids used in this study are described in Table 1. Conjugation between *E. coli* and *S. meliloti* were performed as described in the literature.⁷⁷ All growth media (LB, LBmc, TY, M9, RDM) and antibiotic concentrations were as previously described.^{16,73,78}

***cis*-Hybrid Strain Construction.** First, a triparental mating between the wild type strain BL225C (the future donor), the helper strain *E. coli* MT616 (carrying pRK600 that has the RK2 *tra* genes),³¹ and *E. coli* with the pTE3rctB vector (replicative plasmid overexpressing the *R. elii* *rctB* gene and carrying a tetracycline resistance marker)⁴⁷ was performed to create the BM848 (BL225C-rctB) strain. Second, a biparental mating between *S. meliloti* Rm3498 (Δ pSymA)¹⁶ and an *E. coli* S17-1 strain carrying the pMp7605 vector (carrying a gentamycin resistance marker)⁵⁰ was performed to generate the strain BM826. Lastly, the *cis*-hybrid strain BM806 was created through a biparental mating between the strain BM848 (BL225C-rctB) as the donor and the strain BM826 (Δ pSymA + pMp7605) as the acceptor. Selection for the *cis*-hybrid transconjugant strain (which had the pSymA replicon of the donor strain) was performed on M9 medium containing 1 mM MgSO₄, 0.25 mM CaCl₂, 0.001 mg/mL biotin, 42 μ M CoCl₂, 76 μ M FeCl₂, 10 mM trigonelline, streptomycin, and gentamycin. Streptomycin and gentamycin were used to select for the recipient strain, while the presence of trigonelline as the sole carbon source selected for the gain of pSINMEB01, as the trigonelline catabolic genes are located on pSymA/pSINMEB01.⁷⁹

Validation of the Transplanted Strain. Pulsed-Field Gel Electrophoresis (PFGE) was performed to verify the successful uptake of pSymA *via* restriction digestion of genomic DNA with *PmeI*. The applied PFGE protocol was modified from Herschleb *et al.* 2007⁸⁰ and Mavingui *et al.* 2002,⁸¹ and a protocol from Sharon Long's research group (Stanford University, available at <http://cmgm.stanford.edu/biology/long/files/protocols/Purification%20of%20S%20meliloti.pdf>). *S. meliloti* cultures were grown to an OD₆₀₀ of 1.0 in TY medium supplemented with suitable antibiotics and harvested by centrifugation (3000g, 15 min, 4 °C). All following steps were carried out either on ice or at 4 °C. Sedimented cells were washed with TE buffer (10 mM Tris-HCl, 1 mM EDTA) supplemented with 0.1% (w/v) *N*-Lauroylsarcosine, and a second time with TE buffer. Washed cell pellets were then resuspended in TE buffer and mixed (1:1) with 1.6% (w/v) low-melt agarose (50 °C),

thereby resulting in a final concentration of $\sim 8 \times 10^8$ cells/mL. Two hundred μ L of each suspension was casted into a moistened mold and gelatinized at 4 °C. The resulting agar plugs were subsequently incubated at 37 °C for 3 h in lysis buffer (6 mM Tris-HCl, 1 M NaCl, 100 mM EDTA, 0.5% (w/v) Brij-58, 0.2% (w/v) Sodium deoxycholate, 0.5% (w/v) *N*-Lauroylsarcosine) supplemented with 1.5 mg/mL lysozyme (SERVA Electrophoresis GmbH, Germany). Treated agar plugs were then washed in H₂O, followed by incubation at 50 °C for 48 h in Proteinase K buffer (100 mM EDTA, 10 mM Tris-HCl, 1% (w/v) *N*-Lauroylsarcosine, 0.2% (w/v) Sodium deoxycholate, pH 8.0) supplemented with 1 mg/mL Proteinase K (AppliChem GmbH, Germany). Finally, agar plugs were sequentially washed in four steps, 1 h per wash. After incubation in washing buffer (10 mM Tris-HCl, 50 mM EDTA), plugs were washed in washing buffer supplemented with 1 mM Phenylmethylsulfonyl fluoride, then in washing buffer, and finally in 0.1 \times concentrated washing buffer.

For restriction digestion with *PmeI* (New England Biolabs, USA), the prepared agar plugs were incubated in 1 mL of restriction enzyme buffer (1 \times concentrated) for 1 h with gentle agitation at room temperature. Then, the plugs were transferred into 300 μ L of fresh enzyme buffer supplemented with *PmeI* (50 units per 100 μ L agar plug). Restriction digestions were incubated overnight at 37 °C. After overnight incubation, agar plugs were washed in 1 \times washing buffer for 1 h. For PFGE analysis, 1/8th of each agar plug was used. PFGE was performed using the Rotaphor System 6.0 (Analytik Jena, Germany) following the manufacturer's instructions. Separation of DNA fragments was achieved using a 0.5% agarose gel (Pulse Field Certified Agarose, Bio-Rad, USA) and 0.5 \times TBE buffer (44.5 mM Tris-HCl, 44.5 mM boric acid, 1 mM EDTA). The following settings were applied. Step 1, 18 h, 130–100 V (logarithmic decrease), angle: 130°–110° (logarithmic decrease), interval: 50–175 s (logarithmic increase). Step 2, 18 h, 130–80 V (logarithmic decrease), angle: 110°, interval: 175–500 s (logarithmic increase). Step 3, 40 h, 80–50 V (logarithmic decrease), angle: 106°, interval: 500–2000 s (logarithmic increase). Buffer temperature was adjusted to 12 °C.

For whole genome sequencing, a Nextera XT DNA library was constructed⁸² and sequenced using the Illumina MiSeq platform which generated 2,504,130 paired-end reads. After trimming, assembly was performed with SPAdes 3.9.0,⁸³ which produced 399 contigs. Contigs were aligned against the genomes of *S. meliloti* 2011 and BL225C with MeDuSa.⁸⁴ A further check with raw sequence reads was performed with QualiMap Software, using default parameters.⁸⁵ The assembly has been deposited to the GenBank database under the BioProject ID PRJNA434498.

Finally, several PCR primer pairs for amplification of unique genes of Rm2011 and BL225C (Supplementary Table S3), selected based on a comparative genome analysis with Roary,⁸⁶ were routinely used to ensure the correct identification of strains during all experiments.

Growth Curves. Growth curves were initiated by diluting overnight cultures to an OD₆₀₀ of 0.1 in TY medium or in M9 medium supplemented with succinate as a carbon source. Incubation was performed in 150 μ L volumes in a 96 well microtiter plate. The microplates were incubated without shaking at 30 °C and growth was measured with a microplate reader (Tecan Infinite 200 PRO, Tecan, Switzerland).

Growth with Root Exudate. The ability to colonize plant roots was tested using growth on root exudate as a

metabolic proxy for colonization. Root exudate were produced from seedlings of *M. sativa* (cv. Maraviglia), as previously described.⁶³ Strains were grown on TY plates, following which a single colony was resuspended in 0.9% NaCl solution to a final OD₆₀₀ of 0.5 (1×10^9 CFU/mL). Then, each microplate well was inoculated with 75 μ L of either M9 without a carbon source or a nitrogen-free M9 composition with succinate as a carbon source, 20 μ L of root exudate, and 5 μ L of the culture. The microplates were incubated without shaking at 30 °C and the growth was measured on a microplate reader (Tecan Infinite 200 PRO, Tecan, Switzerland). At the end of the incubation period, aliquots from each well were diluted and viable titers of *S. meliloti* cells were estimated after incubation on TY plates at 30 °C.

Plant Symbiotic Assays. Symbiotic assays were performed in microcosm conditions in plastic pots containing a 1:1 mixture of sterile vermiculite and perlite, supplemented with 200 mL of Fahraeus N-free liquid plant growth medium⁸⁷ *S. meliloti* strains were grown in liquid TY medium at 30 °C for 48 h. Cultures were then washed three times in 0.9% NaCl solution and resuspended to an OD₆₀₀ of 1.0. Approximately 1×10^7 cells were added to each pot, corresponding to $\sim 4 \times 10^4$ cells/cm³. Washed cell-suspensions were then directly spread over the roots of one-week old seedlings that were directly germinated in the pots, and grown in a growth chamber maintained at 26 °C with a 16 h photoperiod (100 microeinstein/m²/s) for 5 weeks. Nodule counts were performed after the 5 weeks, then the shoots dried at 50 °C for 7 days. Competition assays and the estimations of number of bacterial genome copies per nodule (determined with qPCR) were done as previously reported.⁶⁶ The alfalfa cultivars (*M. sativa*, *M. falcata*, *Medicago x varia*) used and their main features are reported in Supplemental Table S4.

Biofilm Assays. Strains were inoculated in 5 mL of tryptone-yeast extract (TY) medium and in 5 mL of RDM⁷⁸ and grown for 24 h with shaking. After growth, cultures were diluted to an OD₆₀₀ of 0.02 in fresh TY and RDM medium, respectively, and 100 μ L of the diluted culture was inoculated into a microtiter plate. The plates were incubated at 30 °C for 48 h, after which the OD₆₀₀ was measured to determine the cell biomass. Each well was then stained with 20 μ L of crystal violet solution for 10 min. The medium containing the planktonic cells was gently removed and the microtiter plate wells were washed three times with 200 μ L of PBS (0.1 M, pH 7.4) buffer and allowed to dry for 15 min. The crystal violet in each well was then solubilized by adding 100 μ L of 95% EtOH and incubating for 15 min at room temperature as described in.⁸⁸ The plate was then read at 560 nm using a microtiter plate reader (Tecan Infinite 200 PRO, Tecan, Switzerland).

Phenotype Microarray. Phenotype MicroArray experiments using Biolog plates PM1 (carbon sources), PM2A (carbon sources), and PM3 (nitrogen sources) were performed largely as described previously.⁴⁵ All bacterial strains used in this study (parental and *cis*-hybrid) are listed in Table 1. Data analysis was performed with DuctApe.⁵² Activity index (AV) values were calculated following subtraction of the blank well from the experimental wells. Growth with each compound was evaluated with AV values from 0 (no growth) to 4 (maximal growth), after elbow test calculation (Table S3c,d).

NMR Metabolomics of the Cell Lysates and Media. Overnight cultures were washed, resuspended, and diluted in 100 mL of fresh M9 medium (41 mM Na₂HPO₄, 22 mM KH₂PO₄, 8.6 mM NaCl, 18.7 mM NH₄Cl, 4.1 μ M biotin,

42 nM CoCl₂, 1 mM MgSO₄, 0.25 mM CaCl₂, 38 μ M FeCl₃, 5 μ M thiamine-HCl, 10 mM succinate).¹⁶ For cell lysates, when cultures reached an OD₆₀₀ of 1, 50 mL of each culture was pelleted by centrifuging for 25 min at 15 000g. For the media, 1 mL of the supernatant of each culture was collected. For cell lysate analysis, each pellet was resuspended in 500 μ L of PBS, and sonicated for 20 min with cycles of 1 s of activity and 9 s of rest (292.5 W, 13 mm tip), with contemporary cooling on ice. After cell lysis, the samples were centrifuged for 25 min at 4 °C at 8000g. For each strain, four independent experiments were performed. NMR samples were prepared in 5.00 mm NMR tubes (Bruker BioSpin) with 55 μ L of a ²H₂O solution containing 10 mM sodium trimethylsilyl[2,2,3,3-²H₄] propionate (TMSP) and 500 μ L of sample.

¹H NMR spectra were acquired for both the cell lysates and the growth media. High reproducibility between samples was seen (Supplemental Figure S3), as expected based on previous studies with eukaryotic cells.^{89,90} NMR spectra were recorded using a Bruker 900 MHz spectrometer (Bruker BioSpin) equipped with a CP TCI ¹H/¹³C/¹⁵N probe. Before measurement, samples were kept for 5 min inside the NMR probe head for temperature equilibration at 300 K. ¹H NMR spectra were acquired with water peak suppression and a standard Carr–Purcell–Meiboom–Gill (CPMG) sequence (cpmgrp; Bruker BioSpin srl), using 192 or 256 scans (for cell lysates and growing media, respectively) over a spectral region of 18 kHz, 110 K points, an acquisition time of 3.07 s, and a relaxation delay of 4 s. This pulse sequence⁹¹ was used to impose a T₂ filter that allows the selective observation of small molecular weight components in solutions containing macromolecules.

The raw data were multiplied by a 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions and calibrated (chemical shift was referenced to the doublet of alanine at 1.48 ppm for cell lysates, and to the singlet of TMSP at 0.00 ppm for growth media) using TopSpin 3.5 (Bruker BioSpin srl). Multivariate and univariate analyses were performed on the obtained data using R software. For multivariate analysis, each spectrum in the region 10–0.2 ppm was segmented into 0.02 ppm chemical shift bins, and the corresponding spectral areas were integrated using the AMIX software (Bruker BioSpin). Binning is a way to reduce the number of total variables and to compensate for small shifts in the signals, making the analyses more robust and reproducible. The area of each bin was normalized to the total spectral area, calculated with exclusion of the water region (4.50–5.15 ppm), in order to correct the data for possible differences in the cell count of each of the NMR samples.

Unsupervised PCA was used to obtain a preliminary overview of the data (visualization in a reduced space, cluster detection, screening for outliers). CA was used in combination with PCA to increase supervised data reduction and classification. Accuracy, specificity, and sensitivity were estimated according to standard definitions. The global accuracy for classification was assessed by means of a leave-one-out cross-validation scheme. The metabolites, whose peaks in the spectra were well-defined and resolved, were assigned and their levels analyzed. The assignment procedure was performed using an internal NMR spectral library of pure organic compounds, public databases such as the *E. coli* Metabolome Database⁹² storing reference NMR spectra of metabolites, and spiking NMR experiments.⁹³ The relative concentrations of the various metabolites were calculated by integrating the corresponding

signals in the spectra,⁹⁴ using a homemade program for signal deconvolution. The nonparametric Wilcoxon–Mann–Whitney test was used for the determination of the meaningful metabolites: a *p*-value of 0.05 was considered statistically significant. The molecule 1,4-dioxane was used as a standard to perform the quantitative NMR analysis with the aim of obtaining the absolute concentrations (μM) of the analyzed metabolites.

NMR data were uploaded on the MetaboLights database (www.ebi.ac.uk/metabolights) with the accession number MTBLS576.

Generation of the Metabolic Models. The manually curated iGD1575 reconstruction of *S. meliloti* Rm1021⁴¹ was modified to expand the composition of the biomass reaction through the inclusion of an additional 31 compounds, including vitamins, coenzymes, and ions at trace concentrations as described elsewhere (Table S6).⁷⁴ Although iGD1575 is based on *S. meliloti* Rm1021, it is expected to accurately represent Rm2011 metabolism as these two strains are derived from the same field isolate (SU47) and have nearly identical gene contents,⁴² while there are numerous SNPs between the strains, SNPs are not considered during the process of metabolic reconstruction.

All other metabolic models were constructed using our recently published protocols for template-assisted metabolic reconstruction and assembly of hybrid bacterial models.⁵¹ Briefly, a draft metabolic reconstructions of *S. meliloti* BL225C was produced using the KBase Web server (www.kbase.us) with gap filling. The draft model was enhanced using the curated Rm1021 model as a template according to,⁵¹ using orthologous gene sets between BL225C and Rm1021 produced with InParanoid.⁹⁵ Additionally, an appropriate “protein synthesis” reaction was manually added to the model. Finally, replicon transplantation between the BL225C model and the Rm1021 model was performed as described recently,⁵¹ making use of the InParanoid generated orthology data and the information contained within each model. All metabolic reconstructions used in this work are provided in Supplementary File S1 in COBRA format within a MATLAB MAT-file. The enhancement and transplantation pipeline is available at <https://github.com/TVignolini/replicon-swap>.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00158.

Tables S1, S3–S6; Figures S1–S8 (PDF)

Table S2 (XLSX)

File S1: Supporting Data (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: marco.fondi@unifi.it.

*E-mail: alessio.mengoni@unifi.it.

ORCID

Anke Becker: 0000-0003-4561-9184

Marco Fondi: 0000-0001-9291-5467

Paola Turano: 0000-0002-7683-8614

Alessio Mengoni: 0000-0002-1265-8251

Author Contributions

A. Checcucci created the strains, performed microbiological analyses. G. diCenzo contributed in metabolic model creation

and performed computation analyses on the metabolic modeling. V. Ghini, P. Turano, C. Luchinat performed NMR analyses and contributed in NMR spectra interpretation. V. Ghini contributed in preparing illustrations. A. Becker and J. Döhlemann contributed PFGE analysis and interpretation. T. Vignolini and M. Fondi contributed the first draft of metabolic model and preliminary computational simulations. G. diCenzo performed computational simulations. F. Decorosi and C. Viti contributed in Phenotype Microarray analysis and interpretation. A. Checcucci and C. Fagorzi contributed to *in vitro* symbiotic assays. M. Bazzicalupo and T. Finan provided data interpretation. A. Mengoni, M. Fondi, G. diCenzo, A. Checcucci conceived the work. A. Checcucci, A. Mengoni, V. Ghini, M. Fondi, G. diCenzo prepared the manuscript. All authors have read and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Dr. Carla Scotti (CREA-FLC, Lodi, Italy) for kindly providing seeds of *M. sativa* cultivars and to Gabriele Brazzini for assistance in symbiotic assays. NMR spectra were acquired and analyzed at CERM, Core Centre of Instruct-ERIC, an ESFRI Landmark, supported by national member subscriptions. This work was supported by the University of Florence, project “Dinamiche dell’evoluzione dei genomi batterici: l’evoluzione del genoma multipartito e la suddivisione in moduli funzionali”, call “PROGETTI STRATEGICI DI ATENE ANNO 2014” to AM. AC was supported by Fondazione Buzzati-Traverso. GCD was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a PDF fellowship, VG was supported by Fondazione Umberto Veronesi. Work in the TMF lab is supported by NSERC. AB and JD were supported by the German Research Foundation (TRR 174).

■ REFERENCES

- (1) Lau, Y. H., Stirling, F., Kuo, J., Karrenbelt, M. A., Chan, Y. A., Riesselman, A., and Gibson, D. G. (2017) Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.* 45, 6971.
- (2) Smith, H. O., Hutchison, C. A., Pfannkoch, C., and Venter, J. C. (2003) Generating a synthetic genome by whole genome assembly: ϕ X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* 100 (26), 15440–15445.
- (3) Burton, R. S., Rawson, P. D., and Edmands, S. (1999) Genetic architecture of physiological phenotypes: empirical evidence for coadapted gene complexes. *Am. Zool.* 39 (2), 451–462.
- (4) Harrison, P. W., Lower, R. P. J., Kim, N. K. D., and Young, J. P. W. (2010) Introducing the bacterial ‘chromid’: not a chromosome, not a plasmid. *Trends Microbiol.* 18 (4), 141–148.
- (5) diCenzo, G. C., and Finan, T. M. (2017) The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiol. Mol. Biol. Rev.* 81 (3), x.
- (6) Agnoli, K., Schwager, S., Uehlinger, S., Vergunst, A., Viteri, D. F., Nguyen, D. T., and Eberl, L. (2012) Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Mol. Microbiol.* 83 (2), 362–378.
- (7) Fei, F., Bowdish, D. M., McCarry, B. E., and Finan, T. M. (2016) Effects of synthetic large-scale genome reduction on metabolism and metabolic preferences in a nutritionally complex environment. *Metabolomics* 12 (2), 23.
- (8) Chen, H., Higgins, J., Oresnik, I. J., Hynes, M. F., Natera, S., Djordjevic, M. A., and Rolfe, B. G. (2000) Proteome analysis demonstrates complex replicon and luteolin interactions in pSymA-

cured derivatives of *Sinorhizobium meliloti* strain 2011. *Electrophoresis* 21, 3833–3842.

(9) González, V., Santamaría, R. I., Bustos, P., Hernández-González, I., Medrano-Soto, A., Moreno-Hagelsieb, G., Janga, S. C., Ramírez, M. A., Jiménez-Jacinto, V., and Collado-Vides, J. (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci. U. S. A.* 103 (10), 3834–3839.

(10) Galardini, M., Brilli, M., Spini, G., Rossi, M., Roncaglia, B., Bani, A., Chianciani, M., Moretto, M., Engelen, K., Bacci, G., Pini, F., Biondi, E. G., Bazzicalupo, M., and Mengoni, A. (2015) Evolution of Intra-specific Regulatory Networks in a Multipartite Bacterial Genome. *PLoS Comput. Biol.* 11 (9), e1004478.

(11) Xu, Q., Dziejman, M., and Mekalanos, J. J. (2003) Exposing the third chromosome of *Burkholderia cepacia* complex strains as a virulence plasmid. *Proc. Natl. Acad. Sci. U. S. A.* 100 (3), 1286–1291.

(12) Yoder-Himes, D. R., Konstantinidis, K. T., and Tiedje, J. M. (2010) Identification of potential therapeutic targets for *Burkholderia cenocepacia* by comparative transcriptomics. *PLoS One* 5 (1), e8724.

(13) Ramachandran, V. K., East, A. K., Karunakaran, R., Downie, J. A., and Poole, P. S. (2011) Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol.* 12, R106.

(14) López-Guerrero, M. G., Ormeño-Orrillo, E., Acosta, J. L., Mendoza-Vargas, A., Rogel, M. A., Ramírez, M. A., and Martínez-Romero, E. (2012) Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. *Plasmid* 68 (3), 149–158.

(15) Young, J. P. W. (2016) Bacteria Are Smartphones and Mobile Genes Are Apps. *Trends Microbiol.* 24 (12), 931–932.

(16) diCenzo, G. C., MacLean, A. M., Milunovic, B., Golding, G. B., and T, F. (2014) Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. *PLoS Genet.* 10 (10), e1004742.

(17) Checucci, A., diCenzo, G. C., Bazzicalupo, M., and Mengoni, A. (2017) Trade, diplomacy & warfare. The quest for elite rhizobia inoculant strains. *Front. Microbiol.*, DOI: 10.3389/fmicb.2017.02207.

(18) Ramachandran, R., Jha, J., Paulsson, J., and Chatteraj, D. (2017) Random versus cell cycle-regulated replication initiation in Bacteria: insights from studying *Vibrio cholerae* chromosome 2. *Microbiol. Mol. Biol. Rev.*, DOI: 10.1128/MMBR.00033-16.

(19) Villaseñor, T., Brom, S., Dávalos, A., Lozano, L., Romero, D., and Santos, AG-DL. (2011) Housekeeping genes essential for pantothenate biosynthesis are plasmid-encoded in *Rhizobium etli* and *Rhizobium leguminosarum*. *BMC Microbiol.* 11, 66.

(20) Pini, F., De Nisco, N. J., Ferri, L., Penterman, J., Fioravanti, A., Brilli, M., Mengoni, A., Bazzicalupo, M., Viollier, P. H., Walker, G. C., and Biondi, E. G. (2015) Cell Cycle Control by the Master Regulator CtrA in *Sinorhizobium meliloti*. *PLoS Genet.* 11 (5), e1005232.

(21) Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A., Gwinn, M. L., Dodson, R. J., and Gill, S. R. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406, 477–483.

(22) Rao, J. R., Fenton, M., and Jarvis, B. D. W. (1994) Symbiotic plasmid transfer in *Rhizobium leguminosarum* biovar trifolii and competition between the inoculant strain ICMP2163 and trans-conjugant soil bacteria. *Soil Biol. Biochem.* 26 (3), 339–351.

(23) Novikova, N., and Safronova, V. (1992) Transconjugants of *Agrobacterium radiobacter* harbouring sym genes of *Rhizobium galegae* can form an effective symbiosis with *Medicago sativa*. *FEMS Microbiol. Lett.* 93 (3), 261–268.

(24) Martínez, E., Palacios, R., and Sanchez, F. (1987) Nitrogen-fixing nodules induced by *Agrobacterium tumefaciens* harboring *Rhizobium phaseoli* plasmids. *J. Bacteriol.* 169 (6), 2828–2834.

(25) Hooykaas, P. J. J., Snijderwint, F. G. M., and Schilperoord, R. A. (1982) Identification of the Sym plasmid of *Rhizobium leguminosarum* strain 1001 and its transfer to and expression in other rhizobia and *Agrobacterium tumefaciens*. *Plasmid* 8 (1), 73–82.

(26) Abe, K., Miyake, K., Nakamura, M., Kojima, K., Ferri, S., Ikebukuro, K., and Sode, K. (2014) Engineering of a green-light

inducible gene expression system in *Synechocystis* sp. PCC6803. *Microbiol. Microb. Biotechnol.* 7 (2), 177–183.

(27) Nakatsukasa, H. H., Uchiyumi, T., Kucho, K. I., Suzuki, A., Higashi, S., and Abe, M. (2008) Transposon mediation allows a symbiotic plasmid of *Rhizobium leguminosarum* bv. trifolii to become a symbiosis island in *Agrobacterium* and *Rhizobium*. *J. Gen. Appl. Microbiol.* 54, 107–118.

(28) Wong, C. H., Pankhurst, C. E., Kondorosi, A., and Broughton, W. J. (1983) Morphology of root nodules and nodule-like structures formed by *Rhizobium* and *Agrobacterium* strains containing a *Rhizobium meliloti* megaplasmid. *J. Cell Biol.* 97, 787–794.

(29) Hirsch, A. M., Wilson, K. J., Jones, J. D., Bang, M., Walker, V. V., and Ausubel, F. M. (1984) *Rhizobium meliloti* nodulation genes allow *Agrobacterium tumefaciens* and *Escherichia coli* to form pseudonodules on alfalfa. *J. Bacteriol.* 158, 1133–1143.

(30) Truchet, G., Rosenberg, C., Vasse, J., Julliot, J. S., Camut, S., and Denarie, J. (1984) Transfer of *Rhizobium meliloti* pSym genes into *Agrobacterium tumefaciens*: host-specific nodulation by atypical infection. *J. Bacteriol.* 157 (1), 134–142.

(31) Finan, Kunkel, De Vos, and Signer (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J. Bacteriol.* 167 (1), 66–72.

(32) Hynes, M. F., Simon, R., Müller, P., Niehaus, K., Labes, M., and Pühler, A. (1986) The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. *Molecular and General Genetics* MGG 202 (3), 356–362.

(33) Muro-Pastor, A. M., Kuritz, T., Flores, E., Herrero, A., and Wolk, C. P. (1994) Transfer of a genetic marker from a megaplasmid of *Anabaena* sp. strain PCC 7120 to a megaplasmid of a different *Anabaena* strain. *J. Bacteriol.* 176 (4), 1093–1098.

(34) Romanchuk, A., Jones, C. D., Karkare, K., Moore, A., Smith, B. A., Jones, C., and Baltrus, D. A. (2014) Bigger is not always better: transmission and fitness burden of ~1MB *Pseudomonas syringae* megaplasmid pMPPla107. *Plasmid* 73, 16–25.

(35) Agnoli, K., Freitag, R., Gomes, M. C., Jenul, C., Suppiger, A., Mannweiler, O., and Eberl, L. (2017) Use of Synthetic Hybrid Strains To Determine the Role of Replicon 3 in Virulence of the *Burkholderia cepacia* Complex. *Appl. Environ. Microbiol.*, DOI: 10.1128/AEM.00461-17.

(36) Galardini, M., Mengoni, A., Brilli, M., Pini, F., Fioravanti, A., Lucas, S., Lapidus, A., Cheng, J. F., Goodwin, L., Pitluck, S., Land, M., Hauser, L., Woyke, T., Mikhailova, N., Ivanova, N., Daligault, H., Bruce, D., Detter, C., Tapia, R., Han, C., Teshima, H., Mocali, S., Bazzicalupo, M., and EG, B. (2011) Exploring the symbiotic pangenome of the nitrogen-fixing bacterium *Sinorhizobium meliloti*. *BMC Genomics* 12 (1), 235.

(37) Vance, C. P. (2001) Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiol.* 127 (2), 390–397.

(38) Barnett, M. J., Fisher, R. F., Jones, T., Komp, C., Abola, A. P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J., Gurjal, M., Hong, A., Huizar, L., Hyman, R. W., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Palm, C., Peck, M. C., Surzycki, R., Wells, D. H., Yeh, K. C., Davis, R. W., Federspiel, N. A., and Long, S. R. (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc. Natl. Acad. Sci. U. S. A.* 98 (17), 9883–8.

(39) Döhlemann, J., Wagner, M., Happel, C., Carrillo, M., Sobetzko, P., Erb, T. J., and Becker, A. (2017) A family of single copy repABC-type shuttle vectors stably maintained in the alpha-proteobacterium *Sinorhizobium meliloti*. *ACS Synth. Biol.* 6 (6), 968–984.

(40) Oresnik, Liu, Yost, and Hynes (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. *Journal of bacteriology* 182 (12), 3582–3586.

(41) diCenzo, G. C., Zamani, M., Milunovic, B., and Finan, T. M. (2016) Genomic resources for identification of the minimal N₂-fixing symbiotic genome. *Environ. Microbiol.* 18 (8), 2534–2547.

(42) Sallet, E., Roux, B., Sauviac, L., Jardinaud, M. F. O., Carrère, S., Faraut, T., and Bruand, C. (2013) Next-generation annotation of

prokaryotic genomes with EuGene-P: application to *Sinorhizobium meliloti* 2011. *DNA Res.* 20 (4), 339–354.

(43) Galardini, M., Pini, F., Bazzicalupo, M., Biondi, E. G., and Mengoni, A. (2013) Replicon-Dependent Bacterial Genome Evolution: The Case of *Sinorhizobium meliloti*. *Genome Biol. Evol.* 5 (3), 542–558.

(44) Giuntini, E., Mengoni, A., De Filippo, C., Cavalieri, D., Aubin-Horth, N., Landry, C. R., Becker, A., and Bazzicalupo, M. (2005) Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. *BMC Genomics* 6, 158.

(45) Biondi, E. G., Tatti, E., Comparini, D., Giuntini, E., Mocali, S., Giovannetti, L., Bazzicalupo, M., Mengoni, A., and Viti, C. (2009) Metabolic capacity of *Sinorhizobium (Ensifer) meliloti* strains as determined by phenotype microarray analysis. *Appl. Environ. Microbiol.* 75 (16), 5396–5404.

(46) diCenzo, Checcucci, Bazzicalupo, Mengoni, Viti, Dziewit, Finan, Galardini, and Fondi (2016) Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* 7, 12219.

(47) Nogales, J., Blanca-Ordóñez, H., Olivares, J., and Sanjuán, J. (2013) Conjugal transfer of the *Sinorhizobium meliloti* 1021 symbiotic plasmid is governed through the concerted action of one- and two-component signal transduction regulators. *Environ. Microbiol.* 15, 811–821.

(48) Pérez-Mendoza, D., Sepúlveda, E., Pando, V., Munoz, S., Nogales, J., Olivares, J., and Sanjuán, J. (2005) Identification of the rctA gene, which is required for repression of conjugative transfer of rhizobial symbiotic megaplasmids. *J. Bacteriol.* 187 (21), 7341–7350.

(49) Blanca-Ordóñez, H., Oliva-García, J. J., Pérez-Mendoza, D., Soto, M. J., Olivares, J., Sanjuán, J., and Nogales, J. (2010) pSymA-dependent mobilization of the *Sinorhizobium meliloti* pSymB megaplasmid. *Journal of bacteriology* 192 (23), 6309–6312.

(50) Lagendijk, E. L., Validov, S., Lamers, G. E., De Weert, S., and Bloembergen, G. V. (2010) Genetic tools for tagging Gram-negative bacteria with mCherry for visualization in vitro and in natural habitats, biofilm and pathogenicity studies. *FEMS Microbiol. Lett.* 305 (1), 81–90.

(51) Vignolini, T., Mengoni, A., and Fondi, M. (2018) Template-assisted metabolic reconstruction and assembly of hybrid bacterial models. In *Metabolic Network Reconstruction and Modeling: Methods and Protocols*, Methods in Molecular Biology Series, pp 177–196, Humana Press, New York, NY.

(52) Galardini, Mengoni, Biondi, Semeraro, Florio, Bazzicalupo, Benedetti, and Mocali (2014) DuctApe: A suite for the analysis and correlation of genomic and OmniLog Phenotype Microarray data. *Genomics* 103, 1.

(53) diCenzo, G. C., Wellappili, D., Golding, G. B., and Finan, T. M. (2018) Inter-replicon Gene Flow Contributes to Transcriptional Integration in the *Sinorhizobium meliloti* Multipartite Genome. *G3: Genes, Genomes, Genet.* 8, g3–300405.

(54) Venturi, V., and Keel, C. (2016) Signaling in the rhizosphere. *Trends Plant Sci.* 21 (3), 187–198.

(55) Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B., and Puhler, A. (2001) The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *Proc. Natl. Acad. Sci. U. S. A.* 98 (17), 9889–94.

(56) Salas, M. E., Lozano, M. J., López, J. L., Draghi, W., Serrania, J., Torres Tejerizo, G. A., and Parisi, G. (2017) Specificity traits consistent with legume-rhizobia coevolution displayed by *Ensifer meliloti* rhizosphere colonization. *Environ. Microbiol.* 19, 3423.

(57) Ramey, B. E., Koutsoudis, M., von Bodman, S. B., and Fuqua, C. (2004) Biofilm formation in plant-microbe associations. *Curr. Opin. Microbiol.* 7 (6), 602–609.

(58) Amaya-Gómez, C. V., H. A. M., and Soto, M. J. (2015) Biofilm formation assessment in *Sinorhizobium meliloti* reveals interlinked control with surface motility. *BMC Microbiol.* 15, 58.

(59) Calatrava-Morales, N., McIntosh, M., and Soto, M. J. (2018) Regulation Mediated by N-Acyl Homoserine Lactone Quorum Sensing Signals in the Rhizobium-Legume Symbiosis. *Genes* 9 (5), 263.

(60) Fujishige, N. A., Rinauldi, L., Giordano, W., and Hirsch, A. M. (2006) Superficial liaisons: colonization of roots and abiotic surfaces by rhizobia. In *Biology of Plant–Microbe Interactions*, in Proceedings of the 12th International Congress on Molecular Plant–Microbe Interactions (Sánchez, F., Quinto, C., López-Lara, I. M., and Geiger, O., Eds) Vol. 5, pp 292–299, ISMPMI Press, St. Paul, MN.

(61) Rosenberg, C., Boistard, P., Dénarié, J., and Casse-Delbart, F. (1981) Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol. Gen. Genet.* 184 (2), 326–333.

(62) Pobigaylo, N., Szymczak, S., Nattkemper, T. W., and Becker, A. (2008) Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. *Mol. Plant-Microbe Interact.* 21, 219–231.

(63) Checcucci, A., Azzarello, E., Bazzicalupo, M., De Carlo, A., Emiliani, G., Mancuso, S., and Mengoni, A. (2017) Role and regulation of ACC deaminase gene in *Sinorhizobium meliloti*: is it a symbiotic, rhizospheric or endophytic gene? *Front. Genet.*, DOI: 10.3389/fgene.2017.00006.

(64) Sprent, J. I. (2001) *Nodulation in Legumes*, Royal Botanic Gardens, London.

(65) Carelli, M., Gnocchi, S., Fancelli, S., Mengoni, A., Paffetti, D., Scotti, C., and Bazzicalupo, M. (2000) Genetic diversity and dynamics of *Sinorhizobium meliloti* populations nodulating different alfalfa varieties in Italian soils. *Appl. Environ. Microbiol.* 66, 4785–4789.

(66) Checcucci, A., Azzarello, E., Bazzicalupo, M., Galardini, M., Lagomarsino, A., Mancuso, S., Marti, L., Marzano, M. C., Mocali, S., Squartini, A., Zanardo, M., and Mengoni, A. (2016) Mixed nodule infection in *Sinorhizobium meliloti* - *Medicago sativa* symbiosis suggest the presence of cheating behavior. *Front. Plant Sci.* 7, 835.

(67) Poole, P., Ramachandran, V., and Terpolilli, J. (2018) Rhizobia: from saprophytes to endosymbionts. *Nat. Rev. Microbiol.* 16, 291.

(68) Burghardt, L. T., Guhlin, J., Chun, C. L., Liu, J., Sadowsky, M. J., Stupar, R. M., and Tiffin, P. (2017) Transcriptomic basis of genome by genome variation in a legume-hizobia mutualism. *Mol. Ecol.* 26, 6122.

(69) Paffetti, D., Daguin, F., Fancelli, S., Gnocchi, S., Lippi, F., Scotti, C., and Bazzicalupo, M. (1998) Influence of plant genotype on the selection of nodulating *Sinorhizobium meliloti* strains by *Medicago sativa*. *Antonie van Leeuwenhoek* 73 (1), 3–8.

(70) Bobik, C., Meilhoc, E., and Batut, J. (2006) FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J. Bacteriol.* 188 (13), 4890–902.

(71) Barnett, M. J., Toman, C. J., Fisher, R. F., and Long, S. R. (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proc. Natl. Acad. Sci. U. S. A.* 101 (47), 16636–41.

(72) Chen, Z. J. (2010) Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* 15 (2), 57–71.

(73) diCenzo, G. C., and F, T. M. (2015) Genetic redundancy is prevalent within the 6.7 Mb *Sinorhizobium meliloti* genome. *Mol. Genet. Genomics* 290 (4), 1345–1356.

(74) diCenzo, G. C., Benedict, A. B., Fondi, M., Walker, G. C., Finan, T. M., Mengoni, A., and Griffiths, J. S. (2017) Robustness encoded across essential and accessory replicons in an ecologically versatile bacterium. *PLoS Genet.* 14 (4), e1007357.

(75) Parnell, J. J., Berka, R., Young, H. A., Sturino, J. M., Kang, Y., Barnhart, D. M., and DiLeo, M. V. (2016) From the lab to the farm: an industrial perspective of plant beneficial microorganisms. *Front. Plant Sci.* 7, 1110.

(76) Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005) The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15 (6), 589–594.

(77) Pini, F., Frage, B., Ferri, L., De Nisco, N. J., Mohapatra, S. S., Taddei, L., and Biondi, E. G. (2013) The DivJ, CbrA and PleC system

controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. *Mol. Microbiol.* 90 (1), 54–71.

(78) Vincent, J. M. (1970) *A Manual for the Practical Study of the Root-Nodule Bacteria*, Blackwell Scientific Publications.

(79) Boivin, C., Barran, L. R., Malpica, C. A., and Rosenberg, C. (1991) Genetic analysis of a region of the *Rhizobium meliloti* pSym plasmid specifying catabolism of trigonelline, a secondary metabolite present in legumes. *J. Bacteriol.* 173 (9), 2809–2817.

(80) Herschleb, J., Ananiev, G., and Schwartz, D. C. (2007) Pulsed-field gel electrophoresis. *Nat. Protoc.* 2 (3), 677.

(81) Mavingui, P., Flores, M., Guo, X., Dávila, G., Perret, X., Broughton, W. J., and Palacios, R. (2002) Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. *Journal of bacteriology* 184 (1), 171–176.

(82) Adessi, A., Spini, G., Presta, L., Mengoni, A., Viti, C., Giovannetti, L., and De Philippis, R. (2016) Draft genome sequence and overview of the purple non sulfur bacterium *Rhodopseudomonas palustris* 42OL. *Stand. Genomic Sci.* 11 (1), 24.

(83) Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068.

(84) Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M. F., Lió, P., Crescenzi, P., Fani, R., and Fondi, M. (2015) MeDuSa: a multi-draft based scaffold. *Bioinformatics* 31 (15), 2443–2451.

(85) García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., and Conesa, A. (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28 (20), 2678–2679.

(86) Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., and Parkhill, J. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31 (22), 3691–3693.

(87) Fahraeus, G. (1957) The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique. *Microbiology* 16 (2), 374–381.

(88) Rinaudi, L. V., and González, J. E. (2009) The low-molecular-weight fraction of exopolysaccharide II from *Sinorhizobium meliloti* is a crucial determinant of biofilm formation. *J. Bacteriol.* 191, 7216–7224.

(89) Bernacchioni, C., Ghini, V., Cencetti, F., Japtok, L., Donati, C., Bruni, P., and Turano, P. (2017) NMR metabolomics highlights sphingosine kinase-1 as a new molecular switch in the orchestration of aberrant metabolic phenotype in cancer cells. *Mol. Oncol.* 11 (5), 517–533.

(90) Ghini, V., Di Nunzio, M., Tenori, L., Valli, V., Danesi, F., Capozzi, F., and Bordoni, A. (2017) Evidence of a DHA Signature in the Lipidome and Metabolome of Human Hepatocytes. *Int. J. Mol. Sci.* 18 (2), 359.

(91) Carr, H. Y., and Purcell, E. M. (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* 94, 630–638.

(92) Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., and Wishart, D. S. (2016) ECMDDB 2.0: A richer resource for understanding the biochemistry of *E. coli*. *Nucleic Acids Res.* 44 (D1), D495–D501.

(93) Psychogios, N., Hau, D. D., Peng, J., et al. (2011) The human serum metabolome. *PLoS One* 6, e16957.

(94) Wishart, D. S. (2008) Quantitative metabolomics using NMR. *TrAC, Trends Anal. Chem.* 27, 228–237.

(95) Sonnhammer, E. L., and Östlund, G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* 43 (D1), D234–D239.

Article

Bacterial Communities from Extreme Environments: Vulcano Island

Camilla Fagorzi ^{1,†}, Sara Del Duca ^{1,†}, Stefania Venturi ^{2,3}, Carolina Chiellini ⁴, Giovanni Bacci ¹, Renato Fani ¹ and Franco Tassi ^{2,3,*}

¹ Department of Biology, University of Florence, 50019 Sesto Fiorentino, Italy

² Department of Earth Sciences, University of Florence, 50121 Florence, Italy

³ National Research Council of Italy (CNR), Institute of Geosciences and Earth Resources (IGG), 50121 Florence, Italy

⁴ Department of Agriculture Food and Environment, University of Pisa, 56124 Pisa, Italy

* Correspondence: franco.tassi@unifi.it

† These authors contributed equally to this work.

Received: 7 June 2019; Accepted: 16 August 2019; Published: 20 August 2019



Abstract: Although volcanoes represent extreme environments for life, they harbour bacterial communities. Vulcano Island (Aeolian Islands, Sicily) presents an intense fumarolic activity and widespread soil degassing, fed by variable amounts of magmatic gases (dominant at La Fossa Crater) and hydrothermal fluids (dominant at Levante Bay). The aim of this study is to analyse the microbial communities from the different environments of Vulcano Island and to evaluate their possible correlation with the composition of the gas emissions. Microbial analyses were carried out on soils and pioneer plants from both La Fossa Crater and Levante Bay. Total DNA has been extracted from all the samples and sequenced through Illumina MiSeq platform. The analysis of microbiome composition and the gases sampled in the same sites could suggest a possible correlation between the two parameters. We can suggest that the ability of different bacterial genera/species to survive in the same area might be due to the selection of particular genetic traits allowing the survival of these microorganisms. On the other side, the finding that microbial communities inhabiting different sites exhibiting different emission profiles are similar might be explained on the basis of a possible sharing of metabolic abilities related to the gas composition.

Keywords: bacterial communities; extreme environments; interstitial soil gases

1. Introduction

In ecology, the term “extreme” commonly refers to unfavourable environmental factors that depress the ability of organisms to function [1]. Understanding the mechanisms underlying the adaptation of microbes to extreme environments is of fundamental importance to deeply investigate processes that led to the evolution of Earth [2]; for example, it has been shown that microbial communities are mainly shaped by environmental conditions and the microbes inhabiting extreme habitats evolve faster than those populating benign environments [1]. Bacterial communities from natural extreme environments represent not only a gene reservoir for potential biotechnological applications (e.g., the discovery of the *Taq* polymerase, isolated from the thermophilic bacteria *Thermus aquaticus*), but they can be used as a model system to explore relationships between diversity and environmental factors [3]. Adaptive features of these extremophiles permit them to survive under such extreme and hostile environmental conditions [4].

In this context, volcanic environments arouse interest in the scientific community, as they are considered analogous to some of the earliest environments on Earth [5]. In the last decades, several

studies focused on understanding the distribution of microorganisms in volcanic environments [6], such as Antarctic and Icelandic volcanoes. These studies were motivated by the desire to elucidate how the geochemically extreme conditions of such environments can influence microbial diversity both on the surface and in the subsurface of the Earth. Interest in the microbial diversity of geothermal and volcanic soils has been growing, and high diversity, including novel and rare bacteria, has been detected so far [7].

The Aeolian volcanic archipelago is situated north of Sicily (Italy), along the southern margin of the Tyrrhenian Sea and is underlain by a 15 to 20 km thick continental crust. Lipari, Vulcano, and Salina are three summits of a volcanic pile located along a regional crustal lineament crossing the north-eastern coast of Sicily and continuing as far as the Aeolian Islands. Each island in the Aeolian arc has had a different tectonic and magmatic history and their activity is not always correlated in time [8]. As usually occurs, shallow hydrothermal systems are associated with the Aeolian volcanoes [9]. Vulcano and the other Aeolian Islands have been investigated for the microbial populations [10], leading to the isolation of novel halotolerant and thermophilic species [11–14]. The thermal springs in the bay Baia di Levante of Vulcano Island (Aeolian Islands) host dozens of aerobic and anaerobic, thermophilic, and hyperthermophilic microorganisms belonging to *Bacteria* and *Archaea* domains [9].

Despite the abundance of researches about the microbial communities from these environments, from a scientific viewpoint, very little is known about the possible correlation between interstitial soil gases and soil bacteria.

Due to their detectability in almost every environment and the high information content, volatile (organic) compounds (VOCs) have been widely studied in relation to different organisms, functions, and interactions [15]. VOCs produced by bacteria have been correlated with inhibition of laccase activity [16], alteration of nitrification processes [17], antagonistic interactions with other species [18], and plant growth promotion [19]. Asensio et al. found microorganism, roots, and physio-chemical factors to be involved in VOCs emissions [20].

It is known that geothermal fluxes shape microbial soil communities [21] and that bacteria produce volatile compounds [22], but there is an almost complete lack of information about the possibility that bacteria may actually participate in the production of these interstitial gases.

Volcanoes are extreme environments for plants too. The role of bacteria associated with plants represents, at the moment, a very interesting topic. Symbiotic bacteria can establish a relationship with the host plant that gives benefits to both parts [23]. In an extreme environment, it is reliable to suppose that the role of bacteria for the survival of the plant in harsh conditions (lack of organic matter, high or low values of pH, presence of high concentration of metals and heavy metals) might be of major importance [24]. Microbial communities associated with plants in diverse conditions, including extreme environments, have been characterized for their potential biotechnological applications in industry, medicine, and agriculture [25]. Plants belonging to the *Poa* genus are pioneer plants widely found in volcanic environments [26–29], thus they result in interesting models to study the endophytic communities in extreme environments.

In this study, we aim to investigate the diversity of bacterial communities of Vulcano Island, Aeolian Islands (Italy). Physicochemical characteristics of fluids from the different sampling areas analysed in this study were investigated, in order to establish a possible correlation of these data with the composition of bacterial communities, and to evaluate the possibility that bacteria may be involved in the production of these interstitial soil gases.

2. Materials and Methods

2.1. Study Area: Vulcano Island

Vulcano is the southernmost island of the Aeolian archipelago (Sicily, southern Italy), a volcanic arc consisting of seven islands and nine sea mountains with an extension of 22 km². The base of the volcanic cone leans on the Calabro-Peloritano block at an average depth of 1000 m below sea level,

for a diameter of about 15 km [30]. The evolution of the island is linked to a trans-tensional tectonics, related to the subduction process [31].

Sampling Areas

Samples for microbiological and geochemical analyses were mostly collected from two areas marked by intense exhalative activity: La Fossa Crater and the Levante Bay, along the eastern side of the isthmus connecting Vulcano Island to the Vulcanello peninsula.

The north-western sector of the crater rim of La Fossa Crater is currently affected by intense fumarolic degassing, with outlet temperatures ranging from the boiling point of water up to 450 °C. The fumarolic fluids have a typical magmatic composition, with dry gas fraction dominated by CO₂ and with relevant concentrations of HCl, SO₂, H₂S, HF, and CO [32,33]. The mineralogy of the crater fumaroles is dominated by sulfides and sulfosalts, deposited due to a sublimate formation under high-temperature reducing conditions [34–36]. Notable quantities of ammonium minerals were also recognized [37]. Rapidly changing temperatures, magmatic fluids input, soil permeability, etc., offer a multitude of environment in a very narrow area. On the volcanic edifice, pioneer plants belonging to the *Poa* genus start to grow at the altitude of nearly 360 m above the sea level.

The Levante Bay area is characterized by the presence of low-temperature fumaroles ($T < 100$ °C), both subaerial and submerged, that originated by boiling of a shallow geothermal aquifer heated by the uprising hot magmatic fluids. The discharged fluids have a typical hydrothermal composition, being characterized by higher contents of CH₄ and H₂S relative to the crater fumaroles. Pioneer plants are present in Levante Bay, in high CO₂ fluxes areas, and in high salinity environments. This site and, in particular, the submarine gas vents were selected as natural laboratories for research studies regarding, e.g., (i) ocean acidification [38] and (ii) thermophilic bacteria [11].

2.2. Interstitial Soil Gas Sampling and Analyses

Sampling was performed on May 2017 in different sites of Vulcano Island, Aeolian Islands (Sicily, Italy). In this work, analyses were concentrated at La Fossa Crater (38°24'15" N 14°57'42" E) and Levante Bay (38°25'02" N 14°57'34" E).

Interstitial soil gases were collected from 8 sampling sites, as follows: (i) 3 sites on the northern flank of La Fossa Crater at different distances from fumarolic emissions (CSCS, CSF56, CSF94), (ii) 1 site on the western flank of the volcanic edifice characterized by the presence of a pioneer plant (CPBS), and (iii) 4 sites in the Levante Bay (LSV11, LSV13, LSV14, LPBS), selected according to in situ soil CO₂ flux measurements. Soil CO₂ fluxes were measured according to the accumulation chamber (AC) method [39], using a cylindrical chamber (basal area 200 cm² and inner volume 3060 cm³) and a Licor[®] (NE, USA) Li-820 Infra-Red (IR) spectrophotometer, as described in Venturi et al. [40].

Gas samples were collected by inserting a stainless-steel probe within the soil down to the sampling depth (5, 10, or 30 cm) connected, through silicon tubing and a PTFE three-way valve, to a 60 mL plastic syringe in order to pump the gas into (i) a 12 mL glass vial equipped with a silicon rubber membrane and (ii) a 1 L Supelco's (Darmstadt, Germany) Tedlar[®] gas sampling bag. The detailed description of the sampling equipment and procedure is reported in previously published literature [40–43]. After sampling, soil temperature was measured with a thermocouple inserted within the soil down to the sampling depth.

The chemical analyses of major gas constituents were performed on gas samples stored in the glass vials using a Shimadzu (Duisburg, Germany) 15A gas chromatograph (GC), equipped with a thermal conductivity detector (TCD) and either a 10 m long 5A Molecular Sieve column (for N₂, O₂, Ar and H₂) or a 3 m long column filled with 80/100 mesh Porapak Q (for CO₂ and H₂S). The analyses of VOCs were performed using (i) a Shimadzu 14A GC equipped with a flame ionization detector (FID) and a 10 m long stainless-steel column filled with 23% SP 1700 on Chromosorb PAW (80/100 mesh) for the analysis of light hydrocarbons (C₁–C₃), and (ii) a Thermo Trace Ultra GC coupled with a Thermo

DSQ Quadrupole Mass Spectrometer (MS) for the analysis of C₄₊ VOCs. The analytical procedure and instrumental specifications are described in Venturi et al. [43].

The analysis of carbon isotopic composition of CO₂ and CH₄ (expressed as δ¹³CO₂ and δ¹³CH₄, respectively, in ‰ vs. V-PDB) was performed on gas samples stored in the Tedlar® bags by using a Cavity Ring-Down Picarro G2201-i spectrometer after proper gas sample treatments, as described in Venturi et al. [43].

2.3. Biological Sample Collection

Samples were collected into 50 mL sterile falcon tubes. Then, 20 g of soil samples were taken in the same sites where interstitial soil gases were measured, immediately before the gas sampling.

A pioneer plant, belonging to the *Poa* genus, present both on the western flank of the volcanic crater and in Levante Bay, was included in the study. Each plant sample analysed was composed of a pool of three plants collected in the same site. Plants were divided into two main compartments: Aerial part and roots. Moreover, rhizospheric and bulk soils were collected from each site, resulting in four different samples for each plant (Levante Bay: LPA, LPR, LPRS, LPBS; La Fossa Crater: CPA, CPR, CPRS, CPBS).

Table 1 reports the list of samples collected from each site.

Table 1. List of samples collected on Vulcano Island.

No.	Code	Site	Sample
1	CPA	La Fossa Crater	Plant Aerial Part
2	CPR	La Fossa Crater	Plant Roots
3	CPRS	La Fossa Crater	Plant Rhizospheric Soil
4	CPBS	La Fossa Crater	Plant Bulk Soil
5	LPA	Levante Bay	Plant Aerial Part
6	LPR	Levante Bay	Plant Roots
7	LPRS	Levante Bay	Plant Rhizospheric Soil
8	LPBS	Levante Bay	Plant Bulk Soil
9	CSF94	La Fossa Crater	Soil, 94 °C
10	CSF56	La Fossa Crater	Soil, 56 °C
11	CSCS	La Fossa Crater	Soil from Cap Surface
12	LSV11	Levante Bay	Soil, site V11
13	LSV13	Levante Bay	Soil, site V13
14	LSV14	Levante Bay	Soil, site V14

2.4. Extraction of Genomic DNA and Next Generation Sequencing

DNA extraction was performed from each sample by using the PowerLyzer® PowerSoil® DNA Isolation Kit (MO BIO laboratories, Inc., Carlsbad, CA, USA) following the manufacturer's instruction. Concentration and purity of extracted DNA were checked by 0.8% agarose gel electrophoresis. Bacterial 16S rRNA gene contains conserved sequences and nine hypervariable regions, named V1–V9 [44,45]. Hypervariable regions are used as molecular markers for bacterial identification in HTS analysis [46]. In particular, in this study, we sequenced the V3–V4 regions using the primer 341F and 805R [47] according to the protocol reported in the 16S Metagenomic Sequencing Library Preparation protocol from Illumina [48]. Library preparation and demultiplexing were performed following Illumina's standard pipeline [49]. Libraries were sequenced in a single run using Illumina MiSeq technology with pair-end sequencing strategy with MiSeq Reagent Kit v3. PCR amplification, library construction, and sequencing were performed by an external company (IGA Technology Services, Udine, Italy). Sequence files were deposited in the NCBI sequence read archive (SRA) under the accession PRJNA508599.

2.5. Sequence Analysis

Primers were removed from all the sequences using Cutadapt [50], version 2.4. Sequences were clustered following the DADA2 pipeline (version 1.8) described at <https://benjjneb.github.io/dada2/tutorial.html> [51], using the R software version 3.4.3 [52]. Sequences were trimmed, quality checked (Figure S1), denoised, and then forward and reverse reads were merged. An amplicon sequence variant (ASV) table was obtained, a higher-resolution analogue of the traditional OTU table [53]. Chimeras were removed; then the taxonomy was assigned to the output sequences through the Silva database (version 128) [54]. Tables produced by the DADA2 pipeline were imported into phyloseq (through the phyloseq R package version 1.22.3), for further analyses of microbiome data. Representative sequences that were not assigned to domain Bacteria were removed from the dataset (Tables S1 and S2).

2.6. Statistical Testing

All statistical analyses were performed into the R environment, version 3.4.3 [52,55]. Good's coverage estimator was calculated using the formula: $1-(n/N) \cdot 100$, where n is the number of sequences found only one time in a specimen (singletons) and N is the total number of sequences assigned to an ASV in that specimen [56].

Bacterial diversity was estimated using the vegan package (version 2.5–4, Dixon 2003). The function 'diversity' was used to compute the Shannon index (H), whereas species evenness (J) was estimated as a function of the Shannon diversity and the number of ASVs detected in the sample (S), according to the Pielou's formula $J = H/\log(S)$ [57]. Differences in bacterial diversity between the different sampling sites were tested using the Student's t -test (' $t.test$ ' R function), and the possible correlation between samples' Shannon diversity and environmental parameters was calculated with the ' $cor.test$ ' function through the Pearson method.

Beta-diversity was analysed with the Bray–Curtis dissimilarity, using the 'vegdist' function of the vegan package. Correspondence analysis (CA) and canonical correspondence analysis (CCA) were performed using the 'cca' function of the vegan package.

Different community structures were tested using permutational multivariate analysis of variance (' $adonis2$ ' function of the vegan R package) with 1000 permutations.

The possible correlation between taxa distribution and gas composition of soil samples was evaluated with a Mantel test of correlation (' $mantel$ ' function of vegan package with Bray–Curtis distance).

The composition of all interstitial gases detected was fitted onto ordination analyses previously produced using the 'envfit' function of the vegan package.

2.7. Isolation of Culturable Bacterial Strains

Isolation of bacteria was performed on Tryptic Soy Agar (TSA) medium using as starting material 1 g of soil (for samples CSCS, CSF56, CSF94, CPBS, CPRS, LSV11, LSV13, LSV14, LPBS, LPRS), and 0.5 g of surface sterilized plant tissue (samples CPA, CPR, LPA, LPR) [58].

Samples were serial diluted and plated on TSA medium, in triplicate. Bacterial plate counts were carried out after 48 h incubation at 30 °C. For each sample, a variable number of colonies (based on the different number of isolates obtained from the samples) was isolated on TSA medium and stored at –80 °C in 20% glycerol for further analysis.

Cell lysates of bacterial isolates were prepared by processing with thermal lysis (95 °C for 10 min, 0 °C for 5 min) an isolated colony dissolved in 20 µL of distilled sterile water.

2.8. Random Amplified Polymorphic DNA (RAPD) Analysis

Random amplification of DNA fragments [59] was performed on DNA samples obtained from thermal lysis. Reactions were performed in a 25 µL total volume, as described in Chiellini et al. [58] using primer 1253 (5'-GTT TCCGCC-3') [58,60]. Amplicons were analysed by 2% agarose gel

electrophoresis. For each biological sample, isolates showing the same RAPD profile were grouped together into the same haplotype.

2.9. Taxonomical Characterization of Culturable Bacterial Strains

One strain for each RAPD haplotype was chosen for the amplification of the 16S rRNA gene, following the protocol described in Chiellini et al. [58]. Amplification of 16S rRNA genes was performed in a final volume of 20 μ L, with DreamTaq DNA Polymerase reagents (ThermoFisher Scientific, Waltham, MA, USA), 0.5 μ M of P0 and P6 primers (P0, 5'-GAGAGTTTGATCCTGGCTCAG; P6 5'-CTACGGCTACCTTGTACGA) [61] and 1 μ L cell lysate was used as template [61]. Amplification conditions were the following: 90'' denaturation at 95 °C, 30 cycles of 30'' at 95 °C, 30'' at 50 °C, and 1' at 72 °C, followed by a final extension of 10' at 72 °C. Sequencing of the 16S rRNA amplicon was performed by an external company (IGA Technology Services) with primer P0. Taxonomic affiliation of the 16S rRNA gene sequences was performed using the Classifier tool of Ribosomal Database Project (RDP) [62]. Each sequence obtained was submitted to GenBank with the accession numbers from MK249884 to MK249979.

3. Results

3.1. Geochemical Features of Interstitial Soil Gases

3.1.1. La Fossa Crater

Sampling sites at La Fossa Crater were characterized by widely varying soil temperatures at 5 cm depth, sharply decreasing as the distance from the fumarolic emission increased, i.e., 94 °C at less than 1 m (CSF94), 56 °C at about 10 m (CSF56), and 31 °C at about 50 m (CSCS). Similarly, the chemical composition of soil gases changed as a function of the distance from the fumarolic vent. Although N₂ was the dominant gas constituent (from 507 to 877 mmol/mol), CO₂, H₂S, and H₂ progressively increased approaching the fumarolic vent, while N₂, Ar, and O₂ showed an opposite behaviour (Table S3). The isotopic composition of CO₂ was slightly heavier than that of the fumarolic gas (−0.86‰ vs. V-PDB; data not reported), with $\delta^{13}\text{C}_{\text{CO}_2}$ values progressively increasing from 1.11‰ to 2.39‰ vs. V-PDB moving away from the fumarolic vent. Considering that the isotopic signature of CO₂ in the fumarolic gases implies an origin related to mixing of magmatic and carbonate sources [32], the $\delta^{13}\text{C}_{\text{CO}_2}$ values of the interstitial gases, which are enriched in ¹³C, were likely produced by partial consumption of volcanogenic CO₂ occurring in the soil due to microbial activity [43]. CH₄ showed similar concentrations in CSF94 and CSF56 (6.2 and 7.1 μ mol/mol, respectively) and markedly decreased in CSCS (2.3 μ mol/mol), while the $\delta^{13}\text{C}_{\text{CH}_4}$ values markedly increased from −47.2‰ and −44.8‰ vs. V-PDB in CSF94 and CSF56 to −5.5‰ vs. V-PDB in CSCS. Such a wide range of $\delta^{13}\text{C}_{\text{CH}_4}$ values was likely caused by oxidation, at different degrees, of volcanogenic CH₄ [43], the latter likely being produced by thermogenic processes acting on pre-existing organic matter at relatively high temperatures typical of a volcanic environment [63].

The organic fraction of interstitial soil gases was largely dominated by alkanes (C₂ to C₁₁ compounds, mainly represented by ethane), with relative abundances ranging from 61% to 79% of Σ VOCs (Table S3), followed by aromatics, which were characterized by relative abundances varying from 12% to 26% of Σ VOCs and decreasing as the distance from the fumarolic emission increased. A similar trend was observed for cycloalkanes and S-bearing compounds ($\leq 2.96\%$ and $\leq 8.75\%$ of Σ VOCs), whilst the relative abundance of O-bearing species sharply increased as the distance from the fumarolic vent increased (from 0.26% to 7.40% of Σ VOCs). Alkenes (only represented by iso-butene) were present at relative abundances varying in a narrow range, from 1.06% to 1.99% of Σ VOCs. Overall, Σ VOCs concentrations ranged from 0.15 to 0.53 μ mol/mol and decreased moving away from the fumarolic emission.

3.1.2. Levante Bay

Sampling sites at Levante Bay were characterized by soil temperatures ranging from 30.9 to 37.5 °C and strongly different soil CO₂ fluxes (hereafter, ΦCO_2), with values of 6.5, 53, and 346 g m⁻² day⁻¹ in LSV13, LSV11, and LSV14, respectively (Table S3). Similarly, the chemical composition of interstitial soil gases varied widely among the selected sites. N₂ was the dominant gas component (from 537 to 963 mmol/mol), but the CO₂/N₂ ratio progressively increased with increasing ΦCO_2 values (from 0.02 to 0.77), the CO₂ concentrations varying from 15 to 411 mmol/mol. H₂S and H₂ were only detected in LSV11 and LSV14, where O₂ and Ar concentrations were from 36 to 44 mmol/mol and around 7 mmol/mol, respectively, and CH₄ ranged from 1.7 to 1.9 $\mu\text{mol/mol}$. Differently, LSV13 was characterized by lower O₂ (8.7 mmol/mol) and higher Ar (13 mmol/mol) and CH₄ (5.1 $\mu\text{mol/mol}$) contents (Table S3). The isotopic compositions of CO₂ and CH₄ (Table S3) were largely different among the selected sites, with heavier values in sites characterized by higher ΦCO_2 (from -3.33‰ to -1.15‰ vs. V-PDB and from 4.30‰ to 6.47‰ vs. V-PDB, respectively) and lighter values in LSV13 (-8.13‰ and -46.2‰ vs. V-PDB, respectively). Methane concentration was 840 $\mu\text{mol/mol}$, i.e., two orders of magnitude higher than those of the interstitial soil gases. The isotopic compositions of CO₂ and CH₄ in this gas sample were -2.21‰ and -16.2‰ vs. PDB, respectively.

Alkanes dominated the organic fraction of interstitial soil gases (Table S3), being represented by C₂ to C₁₁ compounds with overall relative abundances $\geq 74\%$ of ΣVOCs in high ΦCO_2 sampling sites and by C₂ to C₆ compounds with overall relative abundance of 51% of ΣVOCs in LSV13. Ethane was the most abundant alkane, followed by propane. Aromatics (dominated by benzene) represented the second most abundant organic functional group, with relative abundances $\leq 19\%$ of ΣVOCs in LSV11 and LSV14 and of 43% in LSV13. Iso-butene was constantly present at ca. 2% of ΣVOCs , whereas cycloalkanes and S-bearing species were only detected in LSV11 and LSV14 (up to 1.37 and 3.5% of ΣVOCs). On the other hand, O-bearing compounds were present at relative abundances $\leq 0.37\%$ of ΣVOCs in LSV11 and LSV14 and of 3.39% in V13. Overall, ΣVOCs concentrations ranged from 0.12 (in LSV13) to 1.48 (in LSV11) $\mu\text{mol/mol}$.

3.1.3. Pioneer Plant Sites

The chemical composition of interstitial soil gases from the two pioneer plants bulk soil (temperatures around 30 °C) was largely dominated by N₂ (up to 965 mmol/mol). No relevant difference was observed in terms of inorganic constituents between CPBS and LPBS, being characterized by CO₂ concentrations of 16 and 21 mmol/mol, respectively, and Ar and O₂ of 14 and 15 mmol/mol and 5.2 and 3.1 mmol/mol, respectively. H₂S and H₂ were not detected (Table S3). LPBS, characterized by a low ΦCO_2 value (2.6 g m⁻² day⁻¹), was slightly enriched in CH₄ (5.7 $\mu\text{mol/mol}$) with respect to CPBS (2.8 $\mu\text{mol/mol}$). Although the $\delta^{13}\text{CH}_4$ values were similar in the two sites (-45.4‰ and -47.7‰ vs. V-PDB in CPBS and LPBS, respectively), the $\delta^{13}\text{CO}_2$ values strongly differed, the interstitial soil gas from CPBS being much heavier (8.60‰ vs. V-PDB) with respect to that from LPBS (-15.5‰ vs. V-PDB) (Table S3).

The chemical composition of the organic fraction of the two interstitial soil gases was different in terms of relative abundances of alkanes and aromatics. The former (C₂ to C₆ compounds) represented 63% and 58% of ΣVOCs in CPBS and LPBS, respectively, whereas the latter accounted for 24% and 37% of ΣVOCs in the two sites, respectively (Table S3). Cycloalkanes and S-bearing species were not detected, whereas iso-butene was present at ca. 1.6% ΣVOCs in both sites. A sharp difference was observed in the relative abundances of O-bearing compounds, accounting for 12% and 3.8% of ΣVOCs in CPBS and LPBS, respectively. Overall, ΣVOCs concentrations were 0.10 and 0.14 $\mu\text{mol/mol}$ in CPBS and LPBS, respectively (Table S3).

An overall view of the interstitial gases composition of every sample site is shown in Figure 1.

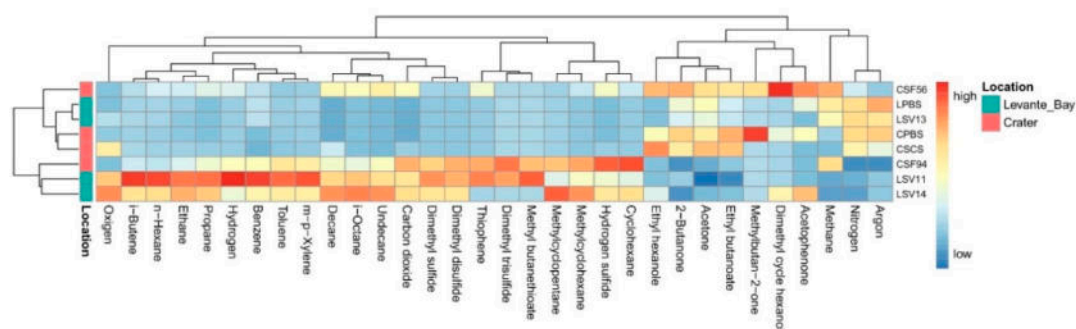


Figure 1. Heatmap showing the scaled measures of detected interstitial soil gases.

3.2. Molecular Characterization of Bacterial Communities

Bacterial communities of soil and plant tissues samples collected in both emissive and non-emissive areas of Vulcano Island were examined through NGS analysis. Sequencing yielded 2,360,693 paired sequences (2×300 bp). More than 60% of the initial pairs were correctly merged (1,446,377 sequences) with a mean of 103,313 sequences *per* sample. Quality filtering steps produced 975,139 high-quality sequences that were correctly mapped into 5073 ASVs with an average of 69,653 sequences *per* sample. Samples reported a Good's coverage estimator ranging from 99% to 100% indicating that roughly 1% of the reads in a given sample came from ASVs that appear only once in that sample [64].

Representative sequences for each ASV were correctly classified into 520 bacterial genera belonging to 23 phyla according to the Silva database (see Materials and Methods), with a different distribution across samples. The analysis of the taxonomic composition revealed that more than 83% of the ASVs were classified into five phyla: *Proteobacteria* (32.22%), *Planctomycetes* (16.44%), *Actinobacteria* (13.56%), *Firmicutes* (11.93%), and *Bacteroidetes* (8.95%).

Comparison between La Fossa Crater and Levante Bay Soil and Plant Samples

Bacterial communities of soil samples from La Fossa Crater (at different soil temperatures) and Levante Bay (at different soil CO₂ fluxes) were analysed, and data obtained are shown in Figure 2.

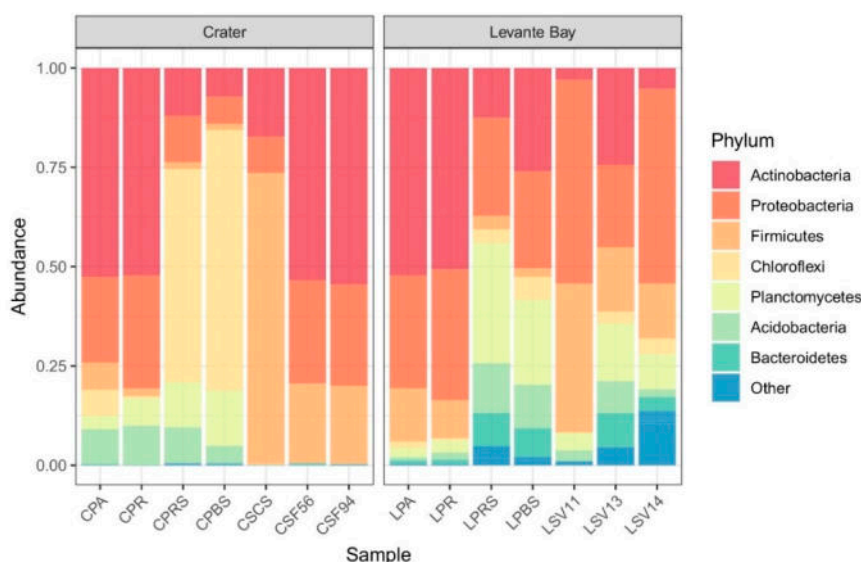


Figure 2. Bar plot showing the relative abundances of bacterial phyla in each sample (from Crater, on the left, and Levante Bay, on the right). Amplicon sequence variants (ASVs) representing <3% of the whole community were reported as “Other”.

Exploring the different communities at the phylum level, the Crater soil samples showed the presence of three main phyla (*Actinobacteria*, *Proteobacteria*, and *Firmicutes*) that, in the samples CSCS, CSF56, and CSF94, comprise more of the 99% of the detected ASVs. Both CSF94 and CSF56 showed a frequency of over 50% in *Actinobacteria* (54.5% and 53.5%, respectively), of 25.6% and 25.9% in *Proteobacteria*, and of 19.7% and 20% in *Firmicutes*. The sample CSCS showed frequencies of 73.3% in *Firmicutes*, 17.3% in *Actinobacteria*, and 9.1% in *Proteobacteria*. Conversely, the sample CPBS presented a frequency in *Chloroflexi* equal to 65.6%, and in *Planctomycetes* of 13.9%, followed by *Actinobacteria* (7.3%), *Proteobacteria* (6.8%), *Acidobacteria* (4.3%), and *Firmicutes* (1.6%).

In Levante Bay soil samples, the main phyla detected were *Proteobacteria* (51.3%, 20.7%, 49%, and 24.4%), *Firmicutes* (37.5%, 16.2%, 13.9%, and 2.2%), *Planctomycetes* (4.4%, 14.4%, 8.9%, and 21.4%), and *Actinobacteria* (2.9%, 24.4%, 5.2%, and 26%) in LSV11, LSV13, LSV14, and LPBS, respectively.

Bacterial communities of *Poa* plant aerial part, roots, rhizosphere, and bulk soil from both Levante Bay and La Fossa Crater were analysed too (Figure 2). All the samples presented a high frequency in *Actinobacteria* and *Proteobacteria*, except for the samples CPBS and CPRS in which *Chloroflexi* are predominant (65.6% and 53.7%, respectively).

Levante Bay samples showed a higher biodiversity than Crater samples (Figure 3a,b), and the Student's t-test between the Shannon indices and evenness of the two sampling sites resulted in statistically significant *p*-values (*p*-value = 0.0033 and 0.0028, respectively).

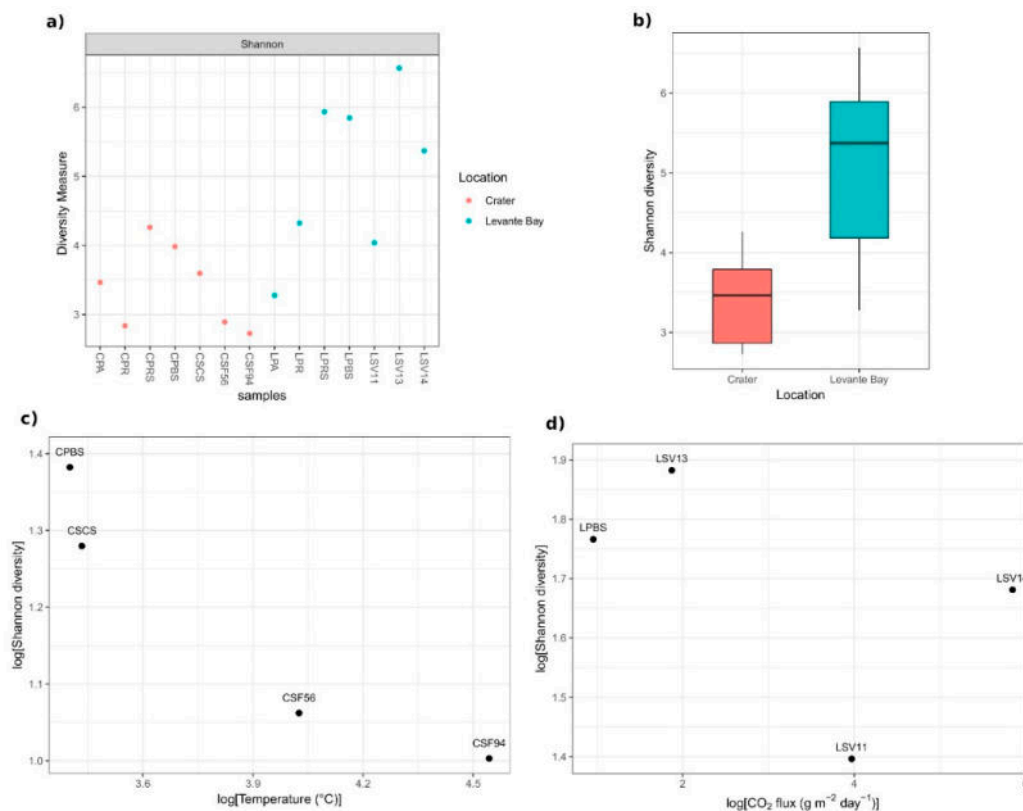


Figure 3. Shannon diversity of samples bacterial communities. (a) Samples' Shannon diversity indices, calculated on raw data; (b) box-whiskers plot of the indices based on the different sampling area: Ends of the whiskers represent the lowest and the highest point still within 1.5 of the inter quartile range (IQR); (c) relation between Crater samples soil temperatures and Shannon diversity indices (log transformation of the data has been performed); (d) absence of correlation between Levante Bay samples CO₂ fluxes and Shannon diversity indices (log transformation of the data has been performed).

A reduction in the Shannon diversity of the Crater soil samples was observed as the distance of the sampling site to the fumarolic emission decreases (so, as the soil temperature enhancing) (Pearson's

$r = -0.94$, p -value = 0.06) (Figure 3c). The analysis of the Levante Bay soil samples did not reveal a linear correlation between the diversity and the soil CO₂ fluxes (Pearson's $r = -0.50$, p -value = 0.50) (Figure 3d).

Correspondence analysis (CA) showed a high similarity between the Crater soil samples CSF94, CSF56, and CSCS, while Levante Bay soil bacterial communities of the different samples resulted in being diverse from each other (Figure 4a). Different community structures were tested using permutational multivariate analysis of variance, resulting in a coefficient of determination $R^2 = 0.24$ and in a non-statistically significant p -value (p -value = 0.09).

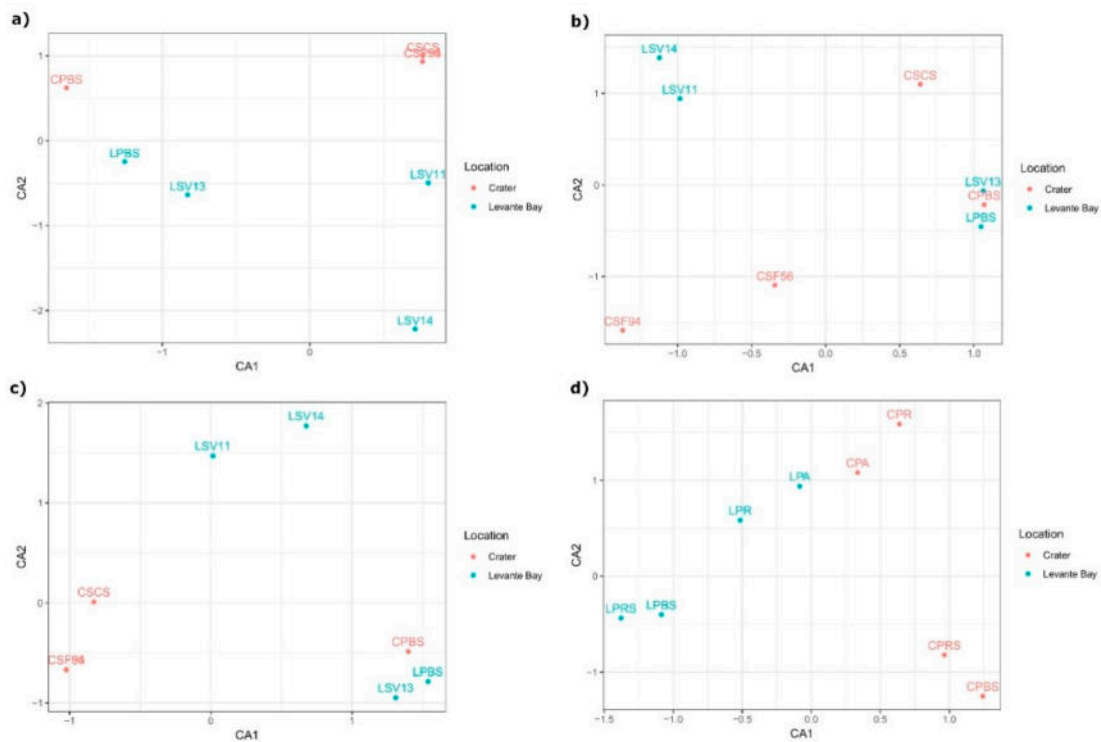


Figure 4. Correspondence analysis (CA) ordination analysis of samples based on (a) soil microbiome taxonomic composition; (b) soil gas emissions; (c) soil microbiome taxonomic composition collapsed at the genus level; (d) plant microbiome taxonomic composition.

To evaluate whether the differences between the two sampling sites, in terms of composition of the bacterial community, can be possibly correlated to the gas emissions present in these areas, a CA analysis was performed on the samples on the basis of the gas emissions (Figure 4b). In this case, the samples LSV11 and LSV14 appeared to be similar to each other, as did LPBS, CPBS, and LSV13, whereas the other samples were separated. Permutational multivariate analysis of variance resulted in a p -value = 0.97, showing no interstitial gases differences between the sampling sites.

Subsequently, a CA analysis was performed on the microbiome taxa composition collapsed at the genus level (Figure 4c): Both the CA analysis and the CCA analysis constrained on the interstitial gases composition (Figure S2) resulted in the same spatial distribution of the samples, which appeared to be comparable to the distribution of the samples based on the gases composition: In particular, samples LSV13, LPBS, and CPBS, and LV11 and LSV14 formed two clusters that are conserved in both the distributions. However, a contrasting pattern was observed for the samples CSF94, CSF56, and CSCS, when comparing microbiome and gases ordinations. Indeed, a Mantel test showed an overall non-significant correlation between taxa and gases composition (p -value > 0.1).

To assess the possibility that one (or more) gases could affect the taxonomic distribution, the fitting of the interstitial gases composition into taxa ordination analysis was performed, resulting in

significant p -values for decane ($p = 0.009$), argon ($p = 0.018$), i-octane ($p = 0.029$), and undecane ($p = 0.044$) (Figure S3).

Beta diversity between plant samples was analysed through CA analysis (Figure 4d) and revealed a marked separation between Levante Bay and La Fossa Crater samples, on the CA1. Permutational multivariate analysis of variance between the two sites gave back a $R^2 = 0.23$ and a p -value = 0.047, resulting in a statistically significant difference between the plants bacterial communities of La Fossa Crater and Levante Bay.

3.3. Culturable Bacterial Community Analysis

Due to the very scarce amount of microbiological studies concerning the cultivable microbiome of Vulcano Island, we performed the isolation of cultivable bacteria from the samples collected in this study. The analysis of the titer of cultivable bacteria in the samples revealed that they contained from 5×10^2 to 1.5×10^5 CFU/g (Table 2).

Table 2. Table counts of cultivable bacteria in the collected samples.

No.	Code	Site	Sample	Bacterial Plate Count (CFU/g)	Number of Isolated Strains
1	CPA	La Fossa Crater	Plant Aerial Part	$1.5 \pm 0.7 \times 10^4$	14
2	CPR	La Fossa Crater	Plant Roots	$3.8 \pm 0.6 \times 10^4$	5
3	CPRS	La Fossa Crater	Plant Rhizospheric Soil	$2.7 \pm 2.5 \times 10^4$	11
4	CPBS	La Fossa Crater	Plant Bulk Soil	$9.5 \pm 0.7 \times 10^3$	14
5	LPA	Levante Bay	Plant Aerial Part	$1.03 \pm 0.09 \times 10^3$	16
6	LPR	Levante Bay	Plant Roots	$4.9 \pm 0.9 \times 10^3$	16
7	LPRS	Levante Bay	Plant Rhizospheric Soil	1.1×10^4	9
8	LPBS	Levante Bay	Plant Bulk Soil	$6.1 \pm 1.4 \times 10^4$	14
9	CSF94	La Fossa Crater	Soil, 94 °C	0	0
10	CSF56	La Fossa Crater	Soil, 56 °C	0	0
11	CSCS	La Fossa Crater	Soil	$5 \pm 0.007 \times 10^2$	2
12	LSV11	Levante Bay	Soil	$1 \pm 0.7 \times 10^4$	15
13	LSV13	Levante Bay	Soil	$1.2 \pm 0.2 \times 10^5$	46
14	LSV14	Levante Bay	Soil	$1.4 \pm 1.6 \times 10^5$	32

All the isolates were screened through RAPD analysis in order to split them into groups, on the basis of their RAPD profile (hereinafter haplotype), assuming that isolates exhibiting the same haplotype correspond to the same strain. For one strain for each haplotype, a 16S rRNA gene sequencing was performed, in order to taxonomically affiliate the strain. Table S4 reports the list of strains isolated from the samples collected in this study.

4. Discussion

The comprehension of the mechanisms that may shape bacterial communities living in extreme environments is crucial from both evolutionary and ecological viewpoint. In this context, Vulcano Island (Aeolian Islands, Sicily, Italy) represents an interesting field of study for analysing bacterial communities of extreme environments, since it offers a multitude of different situations in a very narrow area. Even though previous studies have been carried out on microbial communities of Vulcano Island, very little is known about the possible correlation between interstitial soil gases and soil bacteria. Thus, in this work, we aimed to analyse bacterial communities from different substrates of Vulcano Island and explore the possibility that these bacteria may participate in the production of interstitial gases.

4.1. Bacterial Communities Study

A wide number of studies on extreme environments microbial communities has been carried out in the last decades. Chan et al. performed a similar bacterial communities' study on Malaysian circumneutral hot springs, finding that *Firmicutes* and *Proteobacteria* dominated this environment.

Medrano-Santillana et al. found that the bacterial communities in volcanic soils were characterized by the prevalence of *Firmicutes*, *Proteobacteria*, and *Actinobacteria* [65,66]. The same phyla were disclosed in this study, along with others, such as *Planctomycetes* and *Bacteroidetes*, in agreement with other studies showing that members of these phyla have been found in a wide variety of environments, confirming their strong versatility.

4.1.1. Soil Samples Analysis

Analysing the composition of the bacterial communities of La Fossa Crater and Levante Bay samples, it was shown that Crater samples exhibited a lower biodiversity with respect to Levante Bay ones (Figure 3a,b). Considering the Shannon diversity of the bacterial communities of La Fossa Crater and Levante Bay soil samples, it was shown that Crater samples exhibited a negative trend between the Shannon diversity index and the proximity to the fumarolic emission, in agreement with previous studies showing that extreme environments are less diverse than other ones [67]. Moreover, Crater samples, with the exception of CPBS sample, showed a major taxonomic uniformity compared to Levante Bay ones, confirming that more emissive environments select only few taxa able to survive under these conditions (Figure 4a).

The correlation between the soil bacterial composition and interstitial soil gases composition was analysed. The ordination analyses resulted in different spatial distribution of samples between the taxa composition CA plot and the gases CA plot (Figure 4a,b). These data did not highlight the existence of a direct influence of bacteria on the interstitial gases' composition, or *vice versa*. However, bacterial communities of Levante Bay samples differentiate from each other on the basis of the sampling sites (they present differences in the relative abundances of the main phyla detected, that are *Proteobacteria*, *Firmicutes*, *Planctomycetes*, and *Actinobacteria*), but belong to the same clusters in the interstitial gases ordination analysis; therefore, we can hypothesize that, in spite of their taxonomical diversity, they could share metabolic abilities related to the gas composition. On the other hand, we can also speculate that sites with similar bacterial communities' composition and different interstitial gases (i.e., Crater samples) could host different genetic repertoire responsible for different functions within the same taxonomic unit.

Performing a CA analysis on the taxa composition collapsed at the genus level (Figure 4c); the distribution of samples obtained appeared to have some similarities to the distribution based on interstitial gases composition. This could suggest the possibility of a direct relation between some soil bacterial genera and interstitial soil gases. However, a Mantel test did not support statistical correlation, possibly in relation to the relatively low number of samples analysed. Anyway, we can hypothesize that some genetic traits, possibly related to the presence of some interstitial gases, could be shared among the same genus.

Further, the fitting of the interstitial gases' composition into the taxa ordination analysis was performed to evaluate the possibility that one (or more) gases could affect the taxonomic distribution. The analysis suggested that decane, argon, i-octane, and undecane could be related with the different composition of the microbial communities; however, data obtained from this study do not allow to understand whether the gases affect the taxa distribution, or if bacteria are involved in the production of such gases. Moreover, since these results do not represent a cause-effect relation, it is not possible to know which group of bacteria is responsible for such evidence.

4.1.2. Plant Samples Analysis

It is known that microbial communities associated with plants can establish relationships with the host, giving benefits to both parts [23]. Accordingly, in an extreme environment, it is plausible that the role of bacteria could be of major importance for the survival of the plant [24].

In this study, we choose a pioneer plant belonging to the *Poa* genus, since they are widely distributed in volcanic environments, as a model to study the bacterial communities associated with it.

The bacterial communities of plants tissues and the surrounding soil from both Levante Bay and La Fossa Crater presented the predominance of *Actinobacteria* and *Proteobacteria*, except for two Crater soil samples (bulk soil and rhizospheric soil) in which *Chloroflexi* are the most abundant phyla detected (Figure 2). In a previous study [68], *Chloroflexi* phylum has been found as dominant group in the Atacama Desert, the oldest and driest desert located in South America. So, it is possible that the aridity might represent a driving force for the development of such communities. At the same time, the absence of this phylum in the Crater soil samples CSCS, CSF56, and CSF94 could be due to the presence of higher fumarolic emissions in these sites.

Evaluating the distances between the samples, ordination analysis resulted in a marked separation between Levante Bay and Crater samples showing that, despite the differences between the plant compartments analysed, the composition of bacterial communities between the two sampling areas is more different than between the plant compartments of the same area (Figure 4d).

4.2. Cultivable Bacteria

Due to the very scarce amount of microbiological studies concerning the cultivable microbiome of Vulcano Island, we also isolated cultivable bacteria from the samples for further analyses and future perspectives. About the soil samples, the vital titer resulted in a difference between sampling areas: In particular, samples collected in La Fossa Crater showed a lower amount of cultivable bacteria, with a total absence in the samples CSF94 and CSF56, collected at the temperatures of 94 °C and 56 °C, respectively, and in proximity of the fumarolic emission. Concerning the plants samples, there are no obvious differences between the sampling sites.

Analysing the 16S rRNA gene of the isolates (Table S4), a prevalence of the *Bacillus* genus in Levante Bay soil and plant tissue samples was observed, while in Crater samples, a higher diversity of cultivable bacteria was detected, with isolates mainly belonging to *Arthrobacter* and *Staphylococcus* genera, and the almost total absence of *Bacillus* genus.

5. Conclusions

Vulcano Island is an interesting place harbouring different geochemical situations able to shape bacterial communities. In this context, our work represents one of the first geo-microbiological approaches for the study of volcanic environments. Data obtained on the composition of the bacterial communities from different sites collapsed at the genus level with different emission profiles suggested the possibility of a correlation between the two parameters. It is plausible to imagine that the ability of different bacterial genera/species to survive in the same area might be due to the selection of particular genetic traits, allowing the survival of these microorganisms. On the other side, the finding that microbial communities inhabiting different sites exhibiting different emission profiles are similar might be explained on the basis of a possible sharing of metabolic abilities related to the gas composition.

Moreover, analysing the gases individually, a possible correlation between decane, argon, i-octane, and undecane and the different composition of the microbial communities emerged, even though it is not possible to know which group of bacteria is responsible for such evidence.

Additional analyses are still required, involving a major number of samples and a targeted analysis on archaea communities, combined with gas composition profiling.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1424-2818/11/8/140/s1>, Figure S1: Plots of the estimated error rates of the amplicon dataset, Table S1: Number of reads throughout the filtering steps, Table S2: Relative abundances of reads through the filtering steps, Table S3: Geochemical data of the sampling sites, Table S4: List of the isolated strains and their taxonomic characterization, Figure S2: CCA ordination analysis of soil microbiome taxonomic composition collapsed at the genus level, constrained on the interstitial gases composition, Figure S3: Plot of the gases fitted on the taxa distribution.

Author Contributions: Conceptualization, C.F., S.D.D., S.V., C.C., R.F. and F.T.; data curation, C.F. and S.D.D.; formal analysis, S.D.D. and G.B.; investigation, C.F., S.V. and C.C.; project administration, R.F. and F.T.; supervision, G.B., R.F. and F.T.; writing—original draft, C.F., S.D.D. and S.V.; writing—review and editing, C.F., S.D.D., S.V., C.C., G.B., R.F. and F.T.

Funding: This research received no external funding.

Conflicts of Interest: Funders had no role in the design of the study, in the collection, analyses, or interpretation of data, and in the writing of the manuscript, or in the decision to publish the results.

References

1. Li, S.J.; Hua, Z.S.; Huang, L.N.; Li, J.; Shi, S.H.; Chen, L.X.; Kuang, J.L.; Liu, J.; Hu, M.; Shu, W.S. Microbial communities evolve faster in extreme environments. *Sci. Rep.* **2014**, *4*, 1–9. [[CrossRef](#)] [[PubMed](#)]
2. Chiellini, C.; Miceli, E.; Bacci, G.; Fagorzi, C.; Coppini, E.; Fibbi, D.; Bianconi, G.; Mengoni, A.; Canganella, F.; Fani, R. Spatial structuring of bacterial communities in epilithic biofilms in the Acquarossa river (Italy). *FEMS Microbiol. Ecol.* **2018**. [[CrossRef](#)] [[PubMed](#)]
3. Canfora, L.; Lo Papa, G.; Vittori Antisari, L.; Bazan, G.; Dazzi, C.; Benedetti, A. Spatial microbial community structure and biodiversity analysis in “extreme” hypersaline soils of a semiarid Mediterranean area. *Appl. Soil Ecol.* **2015**. [[CrossRef](#)]
4. Yadav, A.N.; Verma, P.; Kumar, M.; Pal, K.K.; Dey, R.; Gupta, A.; Padaria, J.C.; Gujar, G.T.; Kumar, S.; Suman, A.; et al. Diversity and phylogenetic profiling of niche-specific Bacilli from extreme environments of India. *Ann. Microbiol.* **2015**, *65*, 611–629. [[CrossRef](#)]
5. Van Kranendonk, M.J.; Pirajno, F. Geochemistry of metabasalts and hydrothermal alteration zones associated with c. 3.45 Ga chert and barite deposits: Implications for the geological setting of the Warrawoona Group, Pilbara Craton, Australia. *Geochem. Explor. Environ. Anal.* **2005**, *4*, 253–278. [[CrossRef](#)]
6. Herrera, A.; Cockell, C.S. Exploring microbial diversity in volcanic environments: A review of methods in DNA extraction. *J. Microbiol. Methods* **2007**, *70*, 1–12. [[CrossRef](#)]
7. Stott, M.B.; Crowe, M.A.; Mountain, B.W.; Smirnova, A.V.; Hou, S.; Alam, M.; Dunfield, P.F. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environ. Microbiol.* **2008**, *10*, 2030–2041. [[CrossRef](#)]
8. Crisci, G.M.; De Rosa, R.; Esperan, S.; Mazzuoli, R.; Sonnino, M.; Scienze, D.; Calabria, U. Temporal evolution of a three component system: The island of Lipari (Aeolian Arc, southern Italy) GM. *Bull. Volcanol.* **1991**, *53*, 207–221. [[CrossRef](#)]
9. Maugeri, T.L.; Lentini, V.; Gugliandolo, C.; Italiano, F.; Cousin, S.; Stackebrandt, E. Bacterial and archaeal populations at two shallow hydrothermal vents off Panarea Island (Eolian Islands, Italy). *Extremophiles* **2009**, *13*, 199–212. [[CrossRef](#)]
10. Amend, J.P.; Rogers, K.L.; Shock, E.L.; Gurrieri, S.; Inguaggiato, S. Energetics of chemolithoautotrophy in the hydrothermal system of Vulcano Island, southern Italy. *Geobiology* **2003**, *1*, 37–58. [[CrossRef](#)]
11. Caccamo, D.; Gugliandolo, C.; Stackebrandt, E.; Maugeri, T.L. *Bacillus vulcani* sp. nov., a novel thermophilic species isolated from a shallow marine hydrothermal vent. *Int. J. Syst. Evol. Microbiol.* **2000**, *50*, 2009–2012. [[CrossRef](#)] [[PubMed](#)]
12. Maugeri, T.L.; Gugliandolo, C.; Caccamo, D.; Stackebrandt, E. A polyphasic taxonomic study of thermophilic bacilli from shallow, marine vents. *Syst. Appl. Microbiol.* **2001**, *24*, 572–587. [[CrossRef](#)] [[PubMed](#)]
13. Maugeri, T.L.; Gugliandolo, C.; Caccamo, D.; Stackebrandt, E. Three novel halotolerant and thermophilic Geobacillus strains from shallow marine vents. *Syst. Appl. Microbiol.* **2002**, *25*, 450–455. [[CrossRef](#)] [[PubMed](#)]
14. Gugliandolo, C.; Maugeri, T.L.; Caccamo, D.; Stackebrandt, E. *Bacillus aeolius* sp. nov. a novel thermophilic, halophilic marine Bacillus species from Eolian Islands (Italy). *Syst. Appl. Microbiol.* **2003**, *26*, 172–176. [[CrossRef](#)]
15. Insam, H.; Seewald, M.S.A. Volatile organic compounds (VOCs) in soils. *Biol. Fertil. Soils* **2010**, *46*, 199–213. [[CrossRef](#)]
16. Mackie, A.E.; Wheatley, R.E. Effects and incidence of volatile organic compound interactions between soil bacterial and fungal isolates. *Soil Biol. Biochem.* **1999**. [[CrossRef](#)]
17. Bending, G.D.; Lincoln, S.D. Inhibition of soil nitrifying bacteria communities and their activities by glucosinolate hydrolysis products. *Soil Biol. Biochem.* **2000**. [[CrossRef](#)]
18. Orlandini, V.; Maida, I.; Fondi, M.; Perrin, E.; Papaleo, M.C.; Bosi, E.; de Pascale, D.; Tutino, M.L.; Michaud, L.; Lo Giudice, A.; et al. Genomic analysis of three sponge-associated *Arthrobacter* Antarctic strains, inhibiting the growth of *Burkholderia cepacia* complex bacteria by synthesizing volatile organic compounds. *Microbiol. Res.* **2014**, *169*, 593–601. [[CrossRef](#)] [[PubMed](#)]

19. Fincheira, P.; Quiroz, A. Microbial volatiles as plant growth inducers. *Microbiol. Res.* **2018**, *208*, 63–75. [[CrossRef](#)]
20. Asensio, D.; Peñuelas, J.; Filella, I.; Llusà, J. On-line screening of soil VOCs exchange responses to moisture, temperature and root presence. *Plant Soil* **2007**, *291*, 249–261. [[CrossRef](#)]
21. Gagliano, A.L.; Tagliavia, M.; D'Alessandro, W.; Franzetti, A.; Parello, F.; Quatrini, P. So close, so different: Geothermal flux shapes divergent soil microbial communities at neighbouring sites. *Geobiology* **2016**. [[CrossRef](#)] [[PubMed](#)]
22. Kai, M.; Hausteine, M.; Molina, F.; Petri, A.; Scholz, B.; Piechulla, B. Bacterial volatiles and their action potential. *Appl. Microbiol. Biotechnol.* **2009**, *81*, 1001–1012. [[CrossRef](#)] [[PubMed](#)]
23. Rosenblueth, M.; Martínez-Romero, E. Bacterial Endophytes and Their Interactions with Hosts. *Mol. Plant-Microbe Interact.* **2007**, *19*, 827–837. [[CrossRef](#)] [[PubMed](#)]
24. Kong, Z.; Glick, B.R. *The Role of Plant Growth-Promoting Bacteria in Metal Phytoremediation*, 1st ed.; Elsevier Ltd.: Amsterdam, The Netherlands, 2017; Volume 71, ISBN 0065-2911.
25. Verma, P.; Yadav, A.N.; Kumar, V.; Singh, D.P.; Saxena, A.K. Beneficial plant-microbes interactions: Biodiversity of microbes from diverse extreme environments and its impact for crop improvement. In *Plant-Microbe Interactions in Agro-Ecological Perspectives*; Springer: Singapore, 2017; ISBN 9789811065934.
26. Fridriksson, S. Plant Colonization of a Volcanic Island, Surtsey, Iceland. *Arct. Alp. Res.* **2006**, *19*, 425. [[CrossRef](#)]
27. Korablev, A.P.; Smirnov, V.E.; Neshataeva, V.Y.; Khanina, L.G. Plant Life-Forms and Environmental Filtering during Primary Succession on Loose Volcanic Substrata (Kamchatka, Russia). *Biol. Bull.* **2018**, *45*, 255–264. [[CrossRef](#)]
28. Moral, R.; Wood, D.M. Early primary succession on the volcano Mount St. Helens. *J. Veg. Sci.* **2006**. [[CrossRef](#)]
29. Clarkson, B.D.; Clarkson, B.R.; Juvik, J.O. Pattern and process of vegetation change (succession) on two northern New Zealand island volcanoes. *Surtsey Res.* **2015**, 45–48.
30. Gabbianelli, G. Submarine morphology and tectonics of Vulcano (Aeolian Islands, Southern Tyrrhenian Sea). *Acta Vulcanol.* **1991**, *1*, 135–141.
31. Calanchi, N.; Peccerillo, A.; Tranne, C.A.; Lucchini, F.; Rossi, P.L.; Kempton, P.; Barbieri, M.; Wu, T.W. Petrology and geochemistry of volcanic rocks from the island of Panarea: Implications for mantle evolution beneath the Aeolian island arc (Southern Tyrrhenian sea). *J. Volcanol. Geotherm. Res.* **2002**, *115*, 367–395. [[CrossRef](#)]
32. Capasso, G.; Favara, R.; Inguaggiato, S. Chemical features and isotopic composition of gaseous manifestations on Vulcano Island, Aeolian Islands, Italy: An interpretative model of fluid circulation. *Geochim. Cosmochim. Acta* **1997**. [[CrossRef](#)]
33. Inguaggiato, S.; Mazot, A.; Diliberto, I.S.; Inguaggiato, C.; Madonia, P.; Rouwet, D.; Vita, F. Total CO₂ output from Vulcano island (Aeolian Islands, Italy). *Geochem. Geophys. Geosyst.* **2012**, *13*, 1–19. [[CrossRef](#)]
34. Garavelli, A.; L'Aviano, R.; Vurro, F. Sublimate deposition from hydrothermal fluids at the Fossa. *Eur. J. Mineral.* **1997**, *9*, 423–432. [[CrossRef](#)]
35. Cheynet, B.; Dall'Aglio, M.; Garavelli, A.; Grasso, M.F.; Vurro, F. Trace elements from fumaroles at Vulcano Island (Italy): Rates of transport and a thermochemical model. *J. Volcanol. Geotherm. Res.* **2000**. [[CrossRef](#)]
36. Pinto, D.; Garavelli, A.; Mitolo, D. Baliczunicite, Bi₂O(SO₄)₂, a new fumarole mineral from La Fossa crater, Vulcano, Aeolian Islands, Italy. *Mineral. Mag.* **2014**. [[CrossRef](#)]
37. Demartin, F.; Campostrini, I.; Castellano, C.; Gramaccioli, C.M.; Russo, M. D'ansite-(Mn), Na₂₁Mn²⁺(SO₄)₁₀Cl₃ and d'ansite-(Fe), Na₂₁Fe²⁺(SO₄)₁₀Cl₃, two new minerals from volcanic fumaroles. *Mineral. Mag.* **2012**. [[CrossRef](#)]
38. Boatta, F.; D'Alessandro, W.; Gagliano, A.L.; Liotta, M.; Milazzo, M.; Rodolfo-Metalpa, R.; Hall-Spencer, J.M.; Parello, F. Geochemical survey of Levante Bay, Vulcano Island (Italy), a natural laboratory for the study of ocean acidification. *Mar. Pollut. Bull.* **2013**, *73*, 485–494. [[CrossRef](#)] [[PubMed](#)]
39. Chiodini, G.; Cioni, R.; Guidi, M.; Raco, B.; Marini, L.; Chiodini, G.; Cioni, R.; Guidi, M.; Raco, B.; Marini, L. Soil CO₂ flux measurements in volcanic and geothermal areas. *Appl. Geochem.* **1998**, *13*, 543–552. [[CrossRef](#)]

40. Venturi, S.; Tassi, F.; Cabassi, J.; Vaselli, O.; Minardi, I.; Neri, S.; Caponi, C.; Capasso, G.; Di Martino, R.M.R.; Ricci, A.; et al. A multi-instrumental geochemical approach to assess the environmental impact of CO₂-rich gas emissions in a densely populated area: The case of Cava dei Selci (Latium, Italy). *Appl. Geochem.* **2019**, *101*, 109–126. [[CrossRef](#)]
41. Tassi, F.; Venturi, S.; Cabassi, J.; Capecchiacci, F.; Nisi, B.; Vaselli, O. Volatile organic compounds (VOCs) in soil gases from Solfatara crater (Campi Flegrei, Southern Italy): Geogenic source(s) vs. biogeochemical processes. *Appl. Geochem.* **2015**, *56*, 37–49. [[CrossRef](#)]
42. Tassi, F.; Venturi, S.; Cabassi, J.; Vaselli, O.; Gelli, I.; Cinti, D.; Capecchiacci, F. Organic Geochemistry Biodegradation of CO₂, CH₄ and volatile organic compounds (VOCs) in soil gas from the Vicano—Cimino hydrothermal system (central Italy). *Org. Geochem.* **2015**, *86*, 81–93. [[CrossRef](#)]
43. Venturi, S.; Tassi, F.; Magi, F.; Cabassi, J.; Ricci, A.; Capecchiacci, F.; Caponi, C.; Nisi, B.; Vaselli, O. Carbon isotopic signature of interstitial soil gases reveals the potential role of ecosystems in mitigating geogenic greenhouse gas emissions: Case studies from hydrothermal systems in Italy. *Sci. Total Environ.* **2019**, *655*, 887–898. [[CrossRef](#)]
44. Chakravorty, S.; Helb, D.; Burday, M.; Connell, N.; Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **2007**. [[CrossRef](#)]
45. Petrosino, J.F.; Highlander, S.; Luna, R.A.; Gibbs, R.A.; Versalovic, J. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* **2009**, *55*, 856–866. [[CrossRef](#)]
46. Huse, S.M.; Dethlefsen, L.; Huber, J.A.; Welch, D.M.; Relman, D.A.; Sogin, M.L. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* **2008**. [[CrossRef](#)]
47. Herlemann, D.P.R.; Labrenz, M.; Jürgens, K.; Bertilsson, S.; Waniek, J.J.; Andersson, A.F. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **2011**. [[CrossRef](#)]
48. Kozich, J.J.; Westcott, S.L.; Baxter, N.T.; Highlander, S.K.; Schloss, P.D. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Appl. Environ. Microbiol.* **2013**. [[CrossRef](#)]
49. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Huntley, J.; Fierer, N.; Owens, S.M.; Betley, J.; Fraser, L.; Bauer, M.; et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **2012**. [[CrossRef](#)]
50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **1994**, *17*, 10–12. [[CrossRef](#)]
51. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Dada, S.P.H. High resolution sample inference from Illumina amplicon data. *Nat. Methods* **2016**, *13*, 581–583. [[CrossRef](#)]
52. R Core Team. R: A Language and Environment for Statistical Computing. 2014. Available online: <http://www.R-project.org/> (accessed on 20 August 2019).
53. Callahan, B.J.; McMurdie, P.J.; Holmes, S.P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **2017**. [[CrossRef](#)]
54. McMurdie, P.J.; Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [[CrossRef](#)]
55. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**. [[CrossRef](#)]
56. Good, I.J. The Population Frequencies of Species and the Estimation of Population Parameters. *Biometrika* **1953**, *40*, 237–264. [[CrossRef](#)]
57. Hill, M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **1973**. [[CrossRef](#)]
58. Chiellini, C.; Maida, I.; Emiliani, G.; Mengoni, A.; Mocali, S.; Fabiani, A.; Biffi, S.; Maggini, V.; Gori, L.; Vannacci, A.; et al. Endophytic and rhizospheric bacterial communities isolated from the medicinal plants *echinacea purpurea* and *echinacea angustifolia*. *Int. Microbiol.* **2015**, *17*, 165–174. [[CrossRef](#)]
59. Williams, J.G.K.; Kubelik, A.R.; Livak, K.J.; Rafalski, J.A.; Tingey, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **1990**. [[CrossRef](#)]
60. Mori, E.; Liò, P.; Daly, S.; Damiani, G.; Perito, B.; Fani, R. Molecular nature of RAPD markers from *Haemophilus influenzae* Rd genome. *Res. Microbiol.* **1999**. [[CrossRef](#)]
61. Di Cello, F.; Fani, R. A molecular strategy for the study of natural bacterial communities by PCR-based techniques. *Minerva Biotechnol.* **1996**, *8*, 126–134.

62. Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **2014**. [[CrossRef](#)]
63. Capaccioni, B.; Tassi, F.; Vaselli, O. Organic and inorganic geochemistry of low temperature gas discharges at the Baia di Levante beach, Vulcano Island, Italy. *J. Volcanol. Geotherm. Res.* **2001**, *108*, 173–185. [[CrossRef](#)]
64. Bacci, G.; Cerri, M.; Lastrucci, L.; Ferranti, F.; Ferri, V.; Foggi, B.; Gigante, D.; Venanzoni, R.; Viciani, D.; Mengoni, A.; et al. Applying predictive models to decipher rhizobacterial modifications in common reed die-back affected populations. *Sci. Total Environ.* **2018**. [[CrossRef](#)]
65. Chan, C.S.; Chan, K.G.; Ee, R.; Hong, K.W.; Urbietta, M.S.; Donati, E.R.; Shamsir, M.S.; Goh, K.M. Effects of physiochemical factors on prokaryotic Biodiversity in Malaysian circumneutral hot springs. *Front. Microbiol.* **2017**. [[CrossRef](#)]
66. Medrano-Santillana, M.; Souza-Brito, E.M.; Duran, R.; Gutierrez-Corona, F.; Reyna-López, G.E. Bacterial diversity in fumarole environments of the Paricutín volcano, Michoacán (Mexico). *Extremophiles* **2017**. [[CrossRef](#)]
67. Lanzén, A. Analysis of Sequencing Data in Environmental Genomics Exploring the Diversity of the Microbial Biosphere. Ph.D. Thesis, University of Bergen, Bergen, Norway, 2013.
68. Neilson, J.W.; Quade, J.; Ortiz, M.; Nelson, W.M.; Legatzki, A.; Tian, F.; LaComb, M.; Betancourt, J.L.; Wing, R.A.; Soderlund, C.A.; et al. Life at the hyperarid margin: Novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* **2012**, *16*, 553–566. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

RESEARCH ARTICLE

Spatial structuring of bacterial communities in epilithic biofilms in the Acquarossa river (Italy)

Carolina Chiellini¹, Elisangela Miceli¹, Giovanni Bacci¹, Camilla Fagorzi¹, Ester Coppini², Donatella Fibbi², Giovanna Bianconi³, Alessio Mengoni¹, Francesco Canganella³ and Renato Fani^{1,*}

¹Department of Biology, University of Florence, Via Madonna del Piano 6, 50019 Sesto Fiorentino (FI), Italy, ²G.I.D.A. S.p.A., Via di Baciavalle 36, 59100 Prato (PO), Italy and ³Department of Biological, Agrofood and Forestry Sciences, University of Tuscia, Via San Camillo de Lellis snc, I-01100, Viterbo, Italy

*Corresponding author: Renato Fani, Department of Biology, University of Florence, via Madonna del Piano 6, 50019 Sesto Fiorentino (FI), Italy. Tel: 055 4574742; E-mail: renato.fani@unifi.it

One sentence summary: Structure of bacterial communities in black and red epilithic biofilm in the Acquarossa river (Viterbo, Italy) has been investigated with both culture dependent and independent approaches.

Editor: Anna Roman

ABSTRACT

Epilithic river biofilms characterize the rock surfaces along the Acquarossa river (Viterbo, Italy); they are in part red and in part black colored, maintaining a well-defined borderline. This peculiarity has raised questions about the biotic and abiotic phenomena that might avoid the mixing of the two biofilms. In this study, the structuring of bacterial communities in black and red epilithic biofilm in the Acquarossa river has been investigated with both culture dependent and independent approaches. Data obtained highlighted a (very) different taxonomic composition of black and red epilithons bacterial communities, dominated by *Acinetobacter* sp. and iron-oxidizing bacteria, respectively. The chemical characterization of both river water and biofilms revealed a substantial heavy metals pollution of the environment; heavy metals were also differentially accumulated in red and black epilithons. Overall, our data revealed that the structuring of red and black epilithons might be affected mainly by the antagonistic interactions exhibited by bacterial genera dominating the two biofilms. These findings suggest that biotic factors might be responsible for the structuring of natural bacterial communities, suggesting that there is a selection of populations at very small scale, and that different populations might compete for different niches.

Keywords: biofilm; epilithon; antagonistic interactions; *Pseudomonas*; *Acinetobacter*

INTRODUCTION

Understanding the ecological processes that lead to the structure and the function of microbial communities in the environment is a field that raised great interest in recent years, due to the crucial roles that they play in natural ecosystems, human health and industrial biotechnology (Kastman *et al.* 2016; Widder *et al.* 2016). Interestingly, biotic interactions among microbes

have been widely described as a force driving the structuring of environmental communities (Maida *et al.* 2016 and references therein). The assemblage of strains in a microbial community is the result of many factors such as random drift, selection by abiotic conditions, and biotic interactions (Stoodley *et al.* 2002). Abiotic conditions have largely been studied in the past years (e.g. Mathur *et al.* 2007; Rubin and Leff 2007); they vary depending on many factors (e.g. environmental conditions) and can be

Received: 23 October 2017; Accepted: 7 September 2018

© FEMS 2018. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Table 1. Basic physico-chemical variables and concentration of nitrogen and sulfur species as well as heavy metals in the Acquarossa river water.

Parameter	Acquarossa (river water)
pH	6.50
Conductivity ($\mu\text{S}/\text{cm } 20^\circ\text{C}$)	843
Ammonium (NH_4^+) (mg/l)	0.1
Nitrites (NO_2) (mg/l)	<0.05
Nitrates (NO_3) (mg/l)	<1
Sulfates (mg/l)	<40
Sulfides (mg/l)	<0.1
COD (mg/l O_2)	<10
Cd (mg/l)	<0.01
Ni (mg/l)	<0.05
Cu (mg/l)	<0.05
As (mg/l)	<0.02
Zn (mg/l)	<0.01
Fe tot (mg/l)	1.788
Fe ²⁺ (mg/l)	0.733
Fe ³⁺ (mg/l)	1.05

manipulated at lab scale to understand their influence on bacterial community composition (Stubbenieck and Straight 2016). On the other hand, the complexity of biotic interactions, which play a major role by altering the structure and the degree of organization of complex communities, are challenging to investigate at lab scale and difficult to understand (Moënne-Loccoz *et al.* 2015; Battin *et al.* 2016). Epilithic biofilms are microbial communities (Battin *et al.* 2007) whose structure and composition have been investigated (e.g. Kobayashi *et al.* 2009; Ledger and Hildrew 1998) leading to the suggestion that structure and composition of river epilithic biofilms can vary in response to many factors, such as anthropogenic nutrient and organic matter (Kobayashi *et al.* 2009). Geographical factors, related for example to the altitudinal gradient, play also an important role in the composition and diversity of epilithic communities (Bartrons, Catalan and Casamayor 2012; Besemer *et al.* 2013; Wilhelm *et al.* 2015). Moreover, the interaction between prokaryotic and eukaryotic microbial communities in epilithons has been recently investigated (Zancarini *et al.* 2017).

Epilithic biofilm of the Acquarossa river (Viterbo) are particularly intriguing in terms of spatial structuring. Interestingly, they form two physically separated colored biofilms, that are in part red and in part black colored. Black and red epilithons can be found on the rocks very close to each other, but they do not blend together, maintaining a well-defined borderline. This peculiarity has raised questions about the biotic and abiotic phenomena that might avoid the mixing of the two biofilms. The Acquarossa site is characterized by the presence of an ancient Etruscan village (625–550 B.C.), historically known for its metallurgic activity (Harrison, Cattani and Turfa 2010), which caused a contamination due to the spread of significant amounts of undesirable heavy metals, especially arsenic, into the environment (Hook 2007). The river is also characterized by a high iron concentration, conferring a red color to the water and giving to the river its name Acquarossa, that in Italian stands for 'red water'. From a scientific viewpoint, there is an almost complete lack of information on this site, especially concerning the biological and environmental features; indeed, most of the available literature focuses on its archaeological and historical characteristics (e.g. Staccioli 1976; Meyers, 2003, 2013).

The few available information on the Acquarossa site are mainly focused on the biological aspects and characteristics of the whole environment; the lack of information on the microbiological aspects of the river arose great interest in this unexplored site.

For these reasons, the goal of this study is to assess the characterization of the bacterial communities inhabiting this environment, focusing on the abiotic and biotic factors that may drive the structuring of black and red epilithic biofilms. To this purpose, both a culture dependent and a culture independent approach have been applied to determine the structure and composition of bacterial communities of red and black epilithon. Physico-chemical parameters were also measured in the river water and in the two biofilms. Moreover, phenotypic characterization of the main bacterial genera detected through cultivation has been assessed, focusing on resistance patterns towards heavy metals and antibiotics, and on the antagonistic interactions between the two communities.

MATERIALS & METHODS

Site and sample collection

Samples were collected on July 2016 near the Etruscan city of 'Acquarossa', 42°28'47.0"N 12°07'17.3"E. The site of naturalistic and archaeological interest is located halfway between Viterbo and the town of Grotte Santo Stefano (Viterbo, Italy). Fig. 1a shows a detail of the Acquarossa river course. Fig. 1b and c shows details of the red and black epilithic biofilms in the sampling site. Several physico-chemical variables were measured, namely: pH, conductivity, concentrations of ammonium (NH_4^+), nitrites (NO_2), nitrates (NO_3), Chemical Oxygen Demand (COD), sulfates, sulfides, total iron content, Fe²⁺, Fe³⁺ and heavy metals (Cu, Zn, As, Cd, Ni).

The concentrations of the following heavy metals were measured in black and red biofilms: Cu, Zn, As, Cd, Ni and total iron content.

All the analyzes mentioned above were performed by an external service. Determination of the concentration of heavy metals was performed according to UNI EN 13 657:2004 and UNI EN ISO 11 885:2009 methods (Table 1). Fe²⁺ and Fe³⁺ were measured following the Standard Methods '3500-Fe B. Phenanthroline Method' (Copyright 1999 by American Public Health Association, American Water Works Association, Water Environment Federation).

Six samples, three replicates from the red epilithon and three replicates from the black one, were collected for both culture dependent analysis, and for High-Throughput Sequencing (HTS) analysis.

Extraction of genomic DNA, HTS, sequence analysis and statistical testing

DNA extraction was performed from each sample by using PowerLyzer® PowerSoil® DNA Isolation Kit (MO BIO laboratories, Inc., Carlsbad, California, USA) following the manufacturer's instruction. Concentration and purity of extracted DNA were checked by 0.8% agarose gel electrophoresis. Bacterial 16S rRNA gene contains conserved sequences and nine hypervariable regions (V1–V9), whose lengths range from approximately 50–100 bases (e.g. Chakravorty *et al.* 2007; Petrosino *et al.* 2009). Hypervariable regions are used as molecular markers for bacterial identification in HTS analysis (e.g. Huse *et al.*



Figure 1. The Acquarossa river (A) and details of the red (B) and black (C) epilithic biofilms.

2008). For this reason, on the extracted DNAs the bacterial V4 region of 16S rRNA genes was amplified with specific primers (515F, TGYCAGCMGCCGCGGTAA; 806R, GGAACACNVGGGTWCTAAT, Caporaso et al. 2011) using the protocol reported in the 16S Metagenomic Sequencing Library Preparation protocol from Illumina (Illumina 2013). Library preparation and demultiplexing have been performed following Illumina's standard pipeline (Caporaso et al. 2012). Libraries were sequenced in a single run using Illumina MiSeq technology with pair-end sequencing strategy with MiSeq Reagent Kit v3. PCR amplification, library construction and sequencing were performed by an external company (IGA Technology Services, Udine, Italy). Sequences were merged and clustered using the UPARSE pipeline (version 10.0.240) (Edgar 2013). Reads were merged with the 'fastq_mergepairs' command setting the identity threshold to 80% to account for long overlaps ($2 \times 300\text{bp}$, V3 – V4). Low complexity reads that may have been generated during sequencing were detected and removed using the 'filter_low' command before quality assessment. Merged reads were then quality checked using StreamingTrim 1.0 with a quality cut-off of 18 Phred (Bacci et al. 2014). An additional quality filtering step was performed using the 'fastq_filter' command of the UPARSE pipeline with a maximum expected error threshold of 1. Obtained sequences were truncated to a fixed length of 420 in order to remove PCR primers and retain only amplicons with the expected length. Finally, sequences were clustered into Operational Taxonomic Units (OTUs) by using the 'cluster_otus' command and representative sequences were taxonomically classified using the 'sintax' command along with the RDP training set (version 16). Representative sequences that were not assigned to Domain Bacteria were removed from the dataset. The OTU table was produced with the 'otutab' command.

The following analyzes were performed by using the R software version 3.4.4 (R Core Team 2014; <http://www.R-project.org>). Assigned OTUs were normalized by cumulative sum scaling normalization implemented in the metagenomeSeq package (version 1.16) (Paulson et al. 2013). Differences in OTU abundances between red and black epilithon were assessed using a zero-inflated log-normal model through the 'fitFeatureModel' function whereas different community structures were tested using permutational multivariate analysis of variance ('adonis2' function of the vegan R package, version 2.4, Dixon 2003) with 1000 permutations. Taxonomic units reporting adjusted *P* values lower than 0.05 were considered as characteristic of a biofilm type and were analyzed separately. Normalized counts were used for principal component analysis (PCA) after counts standardization ('decostand' function of the vegan package version 2.4). Environmental factors (black and red epilithon) were fitted onto the ordination analysis using the 'envfit' function of the vegan package (version 2.4). Bacterial diversity was estimated using R (vegan package version 2.4, Dixon 2003). The function was used to compute the Shannon index (*H*) (Hill 1973) whereas

species evenness (*J*) (Hill 1973) was estimated as a function of the Shannon diversity and the number of OTUs detected (*S*), according to the Pielou's formula $J = H / \log(S)$. Differences in bacterial diversity between black and red epilithon were tested using Student's *t*-test ('t.test' R function). Correlation coefficients between culture dependent and culture independent methods were computed correcting the number of reads assigned to each genus by the average number of 16S rRNA gene copies as reported in the rrnDB (Stoddard et al. 2014). Raw data have been submitted to NCBI SRA archive under the BioProject accession PRJNA412007.

Isolation, taxonomical characterization of culturable bacterial strains and statistical analysis

Bacterial plate counts were assessed as described in Chiellini et al (2014) in Tryptic Soy Agar (TSA) medium using 1 g of biofilm of each sample. Bacterial plate counts were carried out after 48 h incubation at 21°C, and the *t*-test was applied using PAST3 software (Hammer, Harper and Ryan 2001) to check for any difference in the enumeration of culturable bacterial community between red and black epilithon. For each sample, 12–59 colonies were isolated on solid TSA medium and stored at –80°C in glycerol (20% final concentration) for further analysis. The variable number of isolates was determined by the number of isolated colonies that could be recovered during plate count enumeration.

Cell lysates of bacterial isolates were prepared by processing an isolated colony dissolved in 30 µl of distilled sterile water with thermal lysis (95°C for 10 min, followed by cooling on ice for 5 min). For each isolated strain, the 16S rRNA gene was amplified following the protocol described in Chiellini et al. (2014). Briefly, PCR amplification of 16S rRNA genes was carried out in 20-µl reactions using DreamTaq DNA Polymerase reagents (ThermoFisher Scientific, Waltham, MA, USA) at the concentrations suggested by the company, and 0.5 µM of primers P0 (5'-GAGAGTTTGATCCTGGCTCAG) and P6 (5'-CTACGGCTACCTTGTTCAG) (Di Cello and Fani 1996); 1 µl of cell lysate was used as template. Amplification conditions were the following: 90-s denaturation at 95°C, 30 cycles of 30 s at 95°C, 30 s at 50°C and 1 min at 72°C, followed by a final extension of 10 min at 72°C. Direct sequencing of the amplified 16S rRNA genes was performed with primer P0 by an external company (IGA Technology Services, Udine, Italy). Each 16S rRNA gene sequence was submitted to GenBank and assigned an accession number from MF964679 to MF964869. Taxonomic affiliation of the 16S rRNA gene sequences was attributed using the 'classifier' tool of the Ribosomal Database Project—RDP (Cole et al. 2014). Alpha diversity indices (Shannon, Evenness and Chao1) were calculated using PAST3 software (Hammer, Harper and Ryan 2001) on the taxonomic composition of the cultivable fraction of bacteria

from black and red epilithon (genus level). BioEdit Software (Hall 1999) was used to align the obtained sequences together with high-quality sequences of closely related Type Strains that were previously downloaded from the RDP database. MEGA5 Software (Tamura et al. 2011) was used for phylogenetic tree construction, by using the Neighbor Joining algorithm; the robustness of the inferred trees was evaluated by 1000 bootstrap resampling.

Random amplified polymorphic DNA analysis.

Random amplification of DNA fragments (Williams et al. 1990) was independently performed on *Pseudomonas* sp. and *Acinetobacter* sp. bacterial strains. Reactions were performed in a 25- μ l total volume, as described in Chiellini et al (2014) using primer 1253 (5'-GTT TCCGCC-3') (Mori et al. 1999). For each genus, strains showing the same Random amplified polymorphic DNA (RAPD) profile were grouped together into the same haplotype. Alpha diversity indices (Shannon, Evenness and Chao1) were calculated using PAST3 software (Hammer, Harper and Ryan 2001) on the results obtained from haplotype attribution, considering the number of strains attributed to each detected haplotype.

Phenotypic characterization of *Acinetobacter* and *Pseudomonas* spp. bacterial strains: resistance to antibiotics and heavy metals, and antagonistic interactions

The resistance patterns to six antibiotics and to six different heavy metals were obtained through the broth microdilution methods in Muller Hinton Broth (MHB), according to the protocols of Clinical and Laboratory Standards Institute (Jorgensen 1993). Briefly, 96-well plates were used for the experiments. Each well contained 10 μ l of bacterial inoculum and 90 μ l of MHB medium enriched with each tested substance at different concentrations, in order to obtain the requested final concentration in each well. The bacterial inoculum was prepared by dissolving in liquid MHB the isolated bacterial colonies after 48 h growth at 21°C in solid medium; when the suspension was adjusted to 0.5 McFarland (corresponding to 1×10^8 CFU/mL), a 1:20 dilution in MHB medium was then performed, in order to yield 5×10^6 CFU/mL in a volume of 10 μ l. The final test concentration of bacteria was approximately 5×10^5 CFU/mL.

Six antibiotics belonging to different classes and having different cellular targets were chosen and tested at different concentrations: Tetracycline (0.5–1.25–2.5–5–12.5–25 μ g/ml), Chloramphenicol (1–2.5–5–10–25–50 μ g/ml), Rifampicin (5–10–25–50–100 μ g/ml), Ciprofloxacin (0.5–1–2.5–5–10 μ g/ml), Streptomycin and Kanamycin (0.5–1–2.5–5–10–50 μ g/ml).

Six different heavy metals were chosen and tested at the following concentrations: Cu (CuCl₂ 1–2.5–5–10 mM), Ni (NiCl₂ 5–10–15–25 mM), Zn (ZnSO₄ 5–10–15–25 mM), Cd (Cd(NO₃)₂ 1–2.5–5–10 mM), As (V) (KH₂AsO₄ 0.5–1–2.5–5–7.5–10–12.5–25 mM) and As (III) (NaAsO₂ 0.5–1–2.5–5–10–15) mM. Results from Broth microdilution methods applied for antibiotics and heavy metals were validated by using TECAN microplate reader (Tecan, Durham, USA) at 600 nm wavelength, after 48 h incubation at 21°C.

Antagonistic interactions among 36 bacterial strains were tested by the cross-streak method (Arasu, Veeramuthu and Savarimuthu 2013; Thirumurugan and Vijayakumar 2015) using a 36 \times 36 (total 1296) array of tests, as described in Maida et al (2016). All tests were performed in TSA medium. Antagonism assays were performed both with petri dishes with or without a septum separating the two hemi-cycles. In petri dishes without

Table 2. Concentration (mg/kg) of heavy metals in red and black epilithic biofilms collected from the Acquarossa river.

Heavy metal (mg/kg)	Red epilithon	Black epilithon
Cd	<1	<1
Ni	<5	54
Cu	<5	17
As	7	17
Zn	14	108
Fe tot	54 314	11 668

septum, the inhibitory activity might be due to both the presence of volatile and of molecules diffusing in the agar medium; in petri dishes containing a central septum, the inhibition is due to the production of volatile compounds only, since the septum physically separates the two chambers hosting the tester and the target strains, respectively.

Statistical analysis of phenotypic results

Phenotypic characterization results were organized in a table with different colors corresponding to different levels of growth/inhibition. For both heavy metals and antibiotic resistance pattern experiments, a positive control consisting on the bacterial inoculum in MHB medium was assessed to verify the growth of each strain in the absence of heavy metal or antibiotic. A negative control, consisting of MHB medium, was also set up in triplicate in each microplate, to verify the absence of any external contamination. The OD₆₀₀ value measured after 48 h incubation at 21°C for the positive control, was considered as the 100% value. All the other measured values corresponding to the different concentrations of each tested substance, were reported as percentage of growth in proportion to the positive control. The data matrix containing the data of growth expressed as % on the positive control has been used for statistical analysis (UPGMA with Bray Curtis distance measure, and PCA) with PAST3 software (Hammer, Harper and Ryan 2001), as described in Mengoni et al (2014).

The results from the cross-streak inhibition assay were organized in two distinct inhibition matrices representing the test performed with both petri dishes with or without septum; in each matrix, rows stand for a target strains and each column stands for a tester strain, as described in Maida et al (2016). The inhibition values corresponding to three different inhibition levels observed during the cross-streak experiment are complete (red), strong (orange), weak (yellow) and absence (white) of inhibition.

RESULTS

Physico-chemical characterization of river water and epilithic biofilms

The values of the measured water variables are resumed in Table 1; the concentration of heavy metals tested in the biofilms is shown in Table 2. The pH of the river water was 6.5, and the conductivity was 1133 μ S/cm. Data obtained revealed that water samples collected in the Acquarossa river were enriched in iron (total content 1.788 mg/l). The Fe²⁺ and Fe³⁺ content in the Acquarossa river water was 0.733 and 1.05 mg/l, respectively. The total iron content within epilithic biofilm matrix was about four orders of magnitude higher in respect to river water content. Moreover, total iron content of red epilithic biofilm was

Table 3. Mean values \pm SEM of richness and diversity indices calculated for both type of biofilms ($n = 3$). P values resulting from Student t-test are also shown.

Index	Mean \pm SEM		P
	Red epilithon	Black epilithon	
Shannon	3.86 \pm 0.121	3.73 \pm 0.666	0.86
Simpson	0.95 \pm 0.001	0.87 \pm 0.070	0.34
Richness	784.67 \pm 183.550	870.33 \pm 217.428	0.78
Evenness	0.59 \pm 0.007	0.56 \pm 0.094	0.79

five times higher than black epilithic biofilm. Quantification of Cd, Ni, Cu, As and Zn revealed that these heavy metals accumulated in the epilithic biofilms. Particularly, the heavy metals content within the biofilm matrix was almost 100 times higher for Cd, 100–1000 times higher for Ni, Cu and As, and up to 10^4 times higher for Zn respect to the river water.

HTS analysis

Sequencing yielded 6 804 040 paired sequences (2×300 bp) with a mean of 567 003 sequences per pair. More than 50% of the initial pairs were correctly merged (1 822 737 sequences) with a mean of 151 894 sequence per sample. Quality filtering step produced 1 238 851 high-quality sequences that were clustered into 1850 OTUs. Representative sequences for each OTU were correctly classified into 276 genera, 212 families, 114 orders, 84 classes and 35 phyla according to the RDP database (see Materials and Methods). Sequences were correctly mapped to OTUs with 80 262 sequences per sample on average.

Samples from red and black epilithon showed similar degree of bacterial diversity according to the main indices computed (Table 3). Indeed, no significant differences were found between the two biofilms studied (Student's t-test). The taxonomic distribution of bacterial phyla in each biofilm, obtained through HTS of the 16S rRNA gene, is reported in Fig. 2. Both epilithons were characterized by a dominance of *Proteobacteria* (ranging from 41% to 89%), followed by *Bacteroidetes* and *Acidobacteria* (ranging from 0.5% to 20% and from 2% to 37%, respectively). Replicates from red epilithon appeared to be more homogeneous in respect to replicates from black epilithon. This finding was also confirmed by a principal component analysis (PCA, Fig. 3) revealing that black epilithon samples were more different than those coming from the red one, spanning a higher range of the plot. However, both multivariate analysis of variance and environmental fitting on PCA did not support a statistically significant difference in the whole community assemblage ($P = 0.1$; environmental fitting, $R^2 = 0.8$; $P = 0.1$).

At a finer scale, 14 OTUs were more represented in red epilithon samples whereas only one was more abundant in black epilithon (zero-inflated log-normal model, $p < 0.05$, adjusted using false discovery rate correction) (Fig. 4). Interestingly, while *Gammaproteobacteria* appeared the most abundant class in red epilithon (Fig. 2), at OTU level, there was a dominance by members of *Betaproteobacteria*, as members of *Gallionellaceae* family (15%), *Sideroxydans* species (11%) and *Gallionella* species (11%). On the contrary *Acinetobacter* was the dominant genus in black epilithon.

Culturable bacterial community analysis and taxonomic diversity

Analysis of the concentration of heterotrophic cultivable bacteria in red and black epilithic biofilm samples revealed that they contained from 5.9×10^4 to 7.5×10^7 CFU/g (Table 4). No significant differences between the two epilithons were detected ($p > 0.5$).

To check the structure and the composition of the cultivable bacterial communities isolated from the two epilithons, we used a two-step strategy. First, we amplified the 16S rRNA gene from bacterial isolates randomly chosen coming from the two epilithons. Once we obtained a phylogenetic affiliation, a RAPD analysis was carried out to check if the different bacterial isolates corresponded to the same or different strains.

Phylogenetic analysis of 16S rRNA gene sequences from the bacterial isolates obtained (Fig. 5) revealed that the black epilithon was dominated by bacteria affiliated to *Acinetobacter* sp. (56.6% vs 6.5% of red epilithon), whereas *Pseudomonas* sp. was prevalent in red epilithon (53.2% vs 8.5% of black epilithon). *Curvobacterium* sp. and *Sphingobacterium* sp. were detected in black epilithon samples, but with low percentages (7.8% and 2.3% respectively). *Bacillus* sp. and *Aeromonas* sp. were two other bacterial genera contributing to the differentiation of the two kinds of biofilm; *Bacillus* sp. was mostly present in black epilithon (12.4%) with respect to red epilithon (3.2%) while *Aeromonas* sp. showed a greater abundance in red epilithon (16.1%) with respect to black epilithon (1.6%). Even though the Chao1 index was higher in black epilithon than in the red one (11.33 vs 9), and the Evenness value was higher in red epilithon (0.5313) than in the black one (0.4212), the differences were not statistically significant ($p = 0.07$ in both cases). On the contrary, Shannon index was significantly different in the two epilithons ($p = 0.006$), showing a value higher in red than in black epilithon (1.565 vs 1.533, respectively).

A good correlation between the taxonomic composition (at the genus level) detected with both culture dependent and culture independent methods was found (Fig. S1, Supporting Information). *Pseudomonas* and *Acinetobacter* were among the most abundant genera detected in both black and red epilithons, with the former exhibiting a higher presence in the cultured isolates than in HTS analysis.

Due to the dominance of cultivable *Acinetobacter* or *Pseudomonas* in the two epilithons, we further inspected the genetic structure and the degree of strain sharing of the *Acinetobacter* and *Pseudomonas* populations associated to the two epilithons using the RAPD technique. The analysis was performed on a total of 77 and 44 *Acinetobacter* sp. and *Pseudomonas* sp. isolates, respectively. Each RAPD fingerprinting obtained was compared with the other ones to split the bacterial isolates into groups (or haplotypes), very likely corresponding to different strains. Data revealed that the 77 *Acinetobacter* and the 44 *Pseudomonas* strains were split into 12 and 19 RAPD haplotypes, respectively, and the two epilithons harbored distinct populations of *Pseudomonas* and *Acinetobacter* (Tables S1a and b, Supporting Information). Moreover, the higher values of richness and diversity indices calculated for the cultivable fraction of bacterial communities in red and black biofilms suggested that the genetic diversity of *Pseudomonas* community was higher than that of the *Acinetobacter* one (Table S2, Supporting Information). Very interestingly, the two epilithons did not share any *Pseudomonas* and/or *Acinetobacter* strain, suggesting the existence of a hidden/unknown factor driving the structuring of these communities. The phylogenetic analysis on *Acinetobacter* sp. (Fig. S2, Supporting Information)

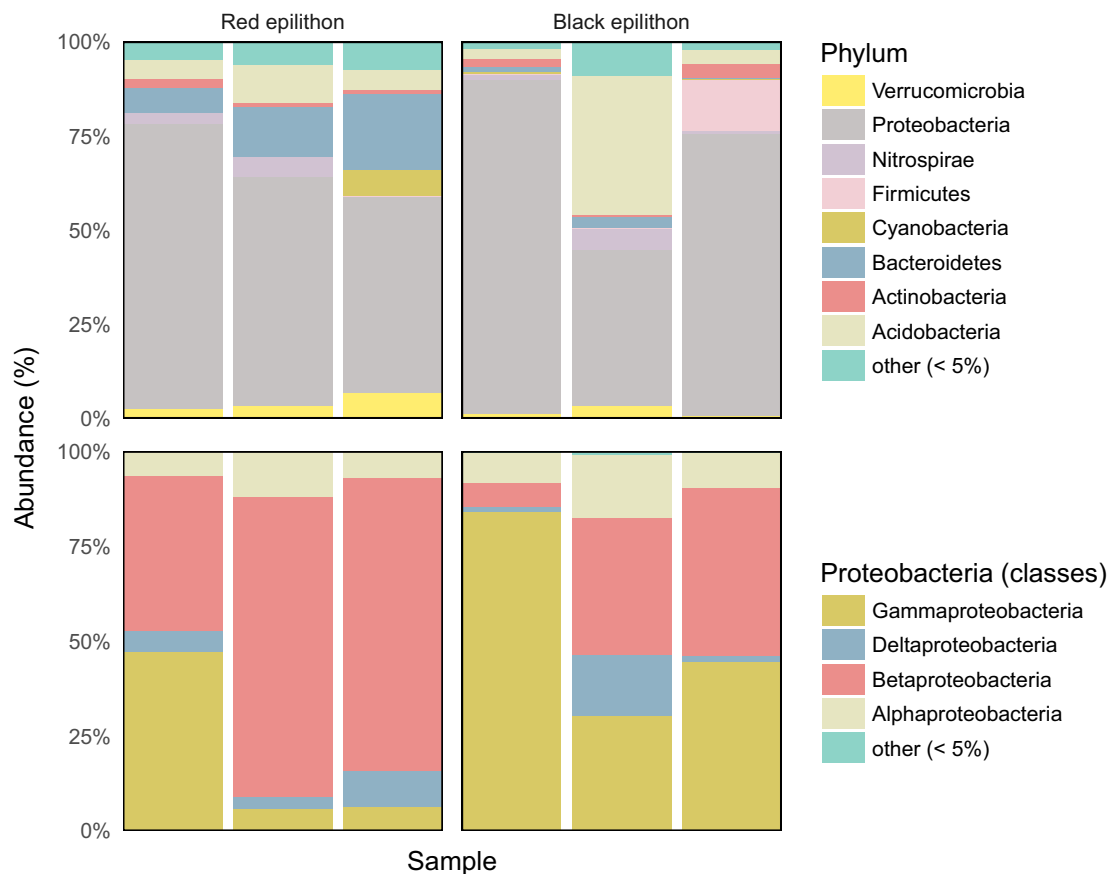


Figure 2. Affiliation of 16S rRNA sequences obtained from red (left panels) and black (right panels) epilithic biofilms. Phyla with relative abundance lower than 0.5% were grouped together and labelled as 'Other taxa'.

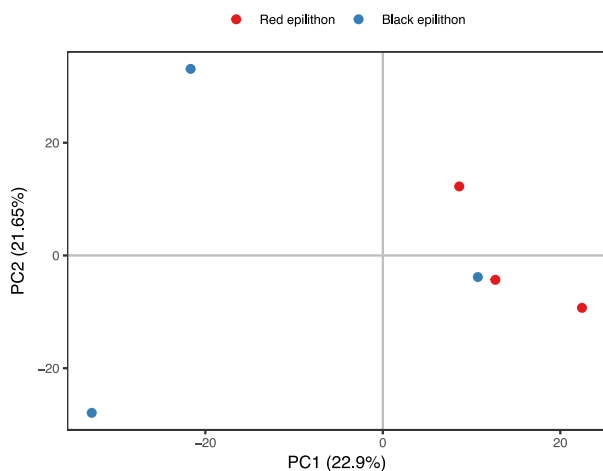


Figure 3. PCA analysis of HTS data on samples collected in red and black epilithons.

and *Pseudomonas* sp. strains (Fig. S3, Supporting Information), confirmed the separation of the strains in the different haplotypes detected by RAPD analysis. Indeed, data revealed that isolates exhibiting the same RAPD profile, clustered together in the phylogenetic tree. Moreover, the phylogenetic analysis confirmed the higher genetic diversity of *Pseudomonas* in respect to *Acinetobacter* strains.

Phenotypic characterization of *Pseudomonas* and *Acinetobacter* community

In order to obtain insights into the (different) phenotypic features of *Acinetobacter* and *Pseudomonas* strains, some phenotypic tests (resistance to heavy metals and antibiotics, antagonistic interactions) were carried out on a panel of 13 *Acinetobacter* sp. strains (3 from red epilithon and 10 from black epilithon) and 23 *Pseudomonas* sp. strains (8 from black epilithon and 15 from red epilithon), representative of all the identified RAPD haplotypes.

The *Pseudomonas* and *Acinetobacter* strains exhibited different resistance patterns to antibiotics (Table S3, Supporting Information); this was particularly evident for Streptomycin and Rifampicin. Indeed, most *Pseudomonas* strains grew at Streptomycin concentrations up to 10–50 $\mu\text{g/ml}$, while most *Acinetobacter* strains grew up to 5–10 $\mu\text{g/ml}$. Similarly, *Pseudomonas* strains were able to grow at Rifampicin concentrations up to 5–10 $\mu\text{g/ml}$, while *Acinetobacter* strains were sensitive to all the tested concentrations of this antibiotic.

Concerning heavy metals, *Pseudomonas* strains were more sensitive to Arsenate than *Acinetobacter* ones; on the contrary, *Pseudomonas* could grow on Cadmium (up to 5 mM), while *Acinetobacter* were sensitive to all the Cadmium tested concentrations. When analyzed separately, heavy metals resistance patterns did not reveal any strong separation between *Acinetobacter* and *Pseudomonas* communities (Fig. 6a); conversely, antibiotic resistance patterns allowed a strong and complete separation between the two distinct communities (Fig. 6b).

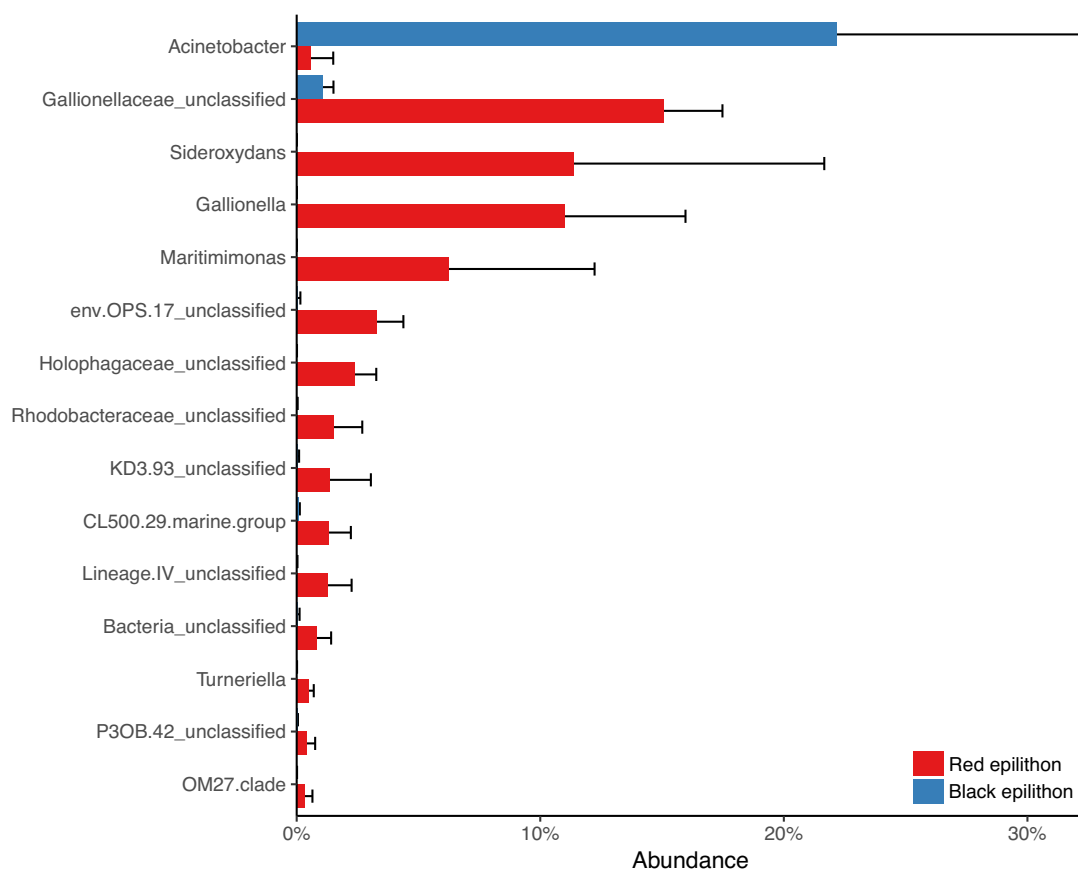


Figure 4. Top—HTS analysis results of bacterial phyla distribution within samples collected in red and black epilithon. Phyla with a relative abundance lower than 0.5% were pooled together and labelled as 'Other'. Bottom—Distribution of classes belonging to Proteobacteria phylum.

Table 4. Viable counts of heterotrophic bacteria in red and black epilithic biofilms.

Sample ID	Type of biofilm	Bacterial plate count (CFU/g)	Number of isolated strains
1	Black	$2.08 \pm 0.09 \times 10^7$	63
5	Black	1.65×10^5	35
6	Black	$2.63 \pm 0.02 \times 10^6$	106
2	Red	$7.53 \pm 3.5 \times 10^7$	38
3	Red	$1.73 \pm 0.4 \times 10^6$	39
4	Red	$5.88 \pm 4.4 \times 10^4$	12

Antagonistic interactions experiments carried out between *Acinetobacter* sp. and *Pseudomonas* sp. and between strains belonging to the same genus (*Acinetobacter* vs *Acinetobacter* and *Pseudomonas* vs *Pseudomonas*) using single-chambered petri dishes highlighted a strong inhibitory activity of *Pseudomonas* sp. strains against the *Acinetobacter* strains and a moderate inhibitory activity between *Pseudomonas* sp. strains (Table 5, upper part). Moreover, since it is known that bacterial strains can produce volatile organic compounds that may inhibit the growth of other bacteria (Papaleo et al. 2013), the same experiments were carried out using two-chambered petri dishes with a central septum physically separating the tester strain from the target ones. Data obtained are shown in Table 5 (lower part) whose analysis revealed that the *Pseudomonas* sp. strains were

less active against *Acinetobacter* strains; moreover, the inhibition among *Pseudomonas* strains disappeared completely. On the other side, *Acinetobacter* sp. strains exhibited an almost complete absence of inhibition activity against *Pseudomonas*, either with single-chambered or two-chambered petri dishes, with the only exception represented by *Acinetobacter* strain 1.7, which showed a slightly moderate inhibitory activity vs *Pseudomonas* strains when grown on single-chambered petri dishes.

DISCUSSION

The Acquarossa river (Viterbo, Italy) represents a still unexplored site of naturalistic interest. To the best of authors' knowledge this is the first study exploring the structure and complexity of bacterial communities of rock biofilms along the river course, focusing on both chemical and microbiological aspects. Data obtained in this work shed some light on the role that biotic factors might play in driving the structuring of the bacterial communities of rock biofilms.

On the basis of HTS data, iron-oxidizing bacteria, mainly related to *Sideroxydans* sp. and *Gallionellaceae* represent an important fraction of the entire microbiota of the red-colored biofilms. It is known that freshwater iron-oxidizing bacteria are a group of bacteria associated to aqueous environments containing appreciable concentrations of Fe(II) (Emerson, Fleming and McBeth 2010; Emerson et al. 2013). The analysis of Fe²⁺ and Fe³⁺ content in the Acquarossa river water revealed that almost 60% of the total iron content (1.05/1.788 mg/l) is in the oxidized Fe³⁺

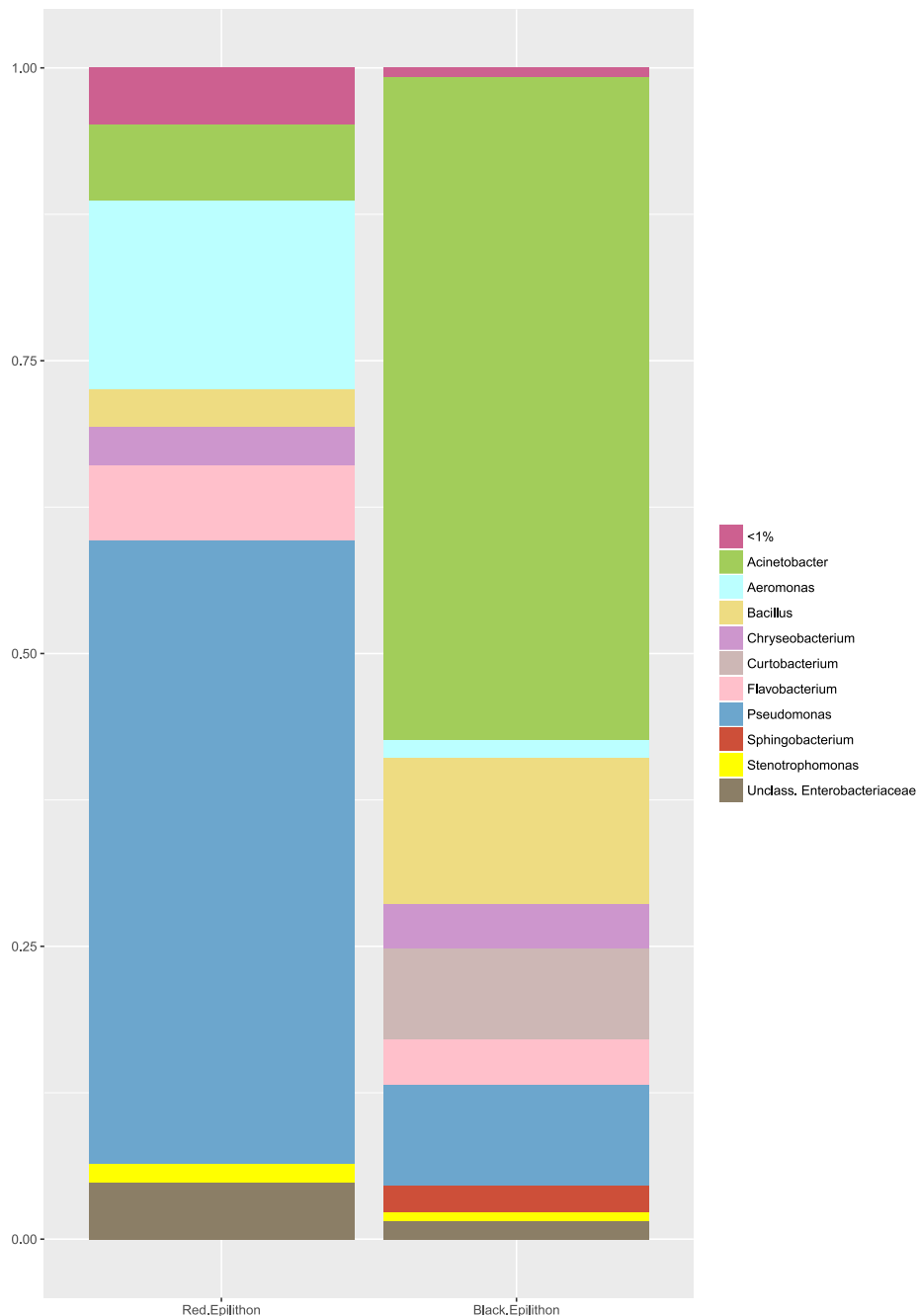


Figure 5. Bacterial genera distribution in black and red epilithons obtained by analysing culturable bacterial communities.

form, while the 40% is in the form of Fe^{2+} . The presence of high amounts of Fe^{2+} in the river is favoured by pH (about 6.5 in the flowing water) (Hem and Cropper 1960) thus sustaining the activity of iron-oxidizing bacteria (Emerson, Fleming and McBeth 2010) and deposition of iron hydroxides in biofilms dominated by these microorganisms. Therefore, the red color shown by this biofilm might be related to the presence of iron hydroxides that accumulates within the biofilm matrix (presumably due to the formation of Fe_2O_3) (Fig. 1c).

The color of the black epilithon might be due to the presence of other compounds in the river water. One possible explanation might be related to the presence of iron sulfides, which are known to form a black colored precipitate (Berner

1964) and that can be the result of the combination of Fe^{2+} with S (Table 1) by sulfate-reducing bacteria in the biofilm matrix. HTS data revealed the presence of sulfate-reducing bacteria affiliated to *Nitrospirae* and *Deltaproteobacteria* both in red and black epilithic biofilms (Fig. 2). Among them, the analysis of OTUs at genus level revealed the presence of sulfate reducing *Desulfobulbus* and *Desulfuromonas*. However, the multivariate analysis of variance did not show any significant difference in the whole community assemblage between the two epilithons ($p = 0.1$), and as a consequence, there are not significant differences in sulfate-reducing bacteria distribution within the two biofilms. Hence, it might be possible that other bacterial and/or archaeal groups not analyzed in this work could carry out the chemical reaction leading to the production of black precipitates characterizing this biofilm.

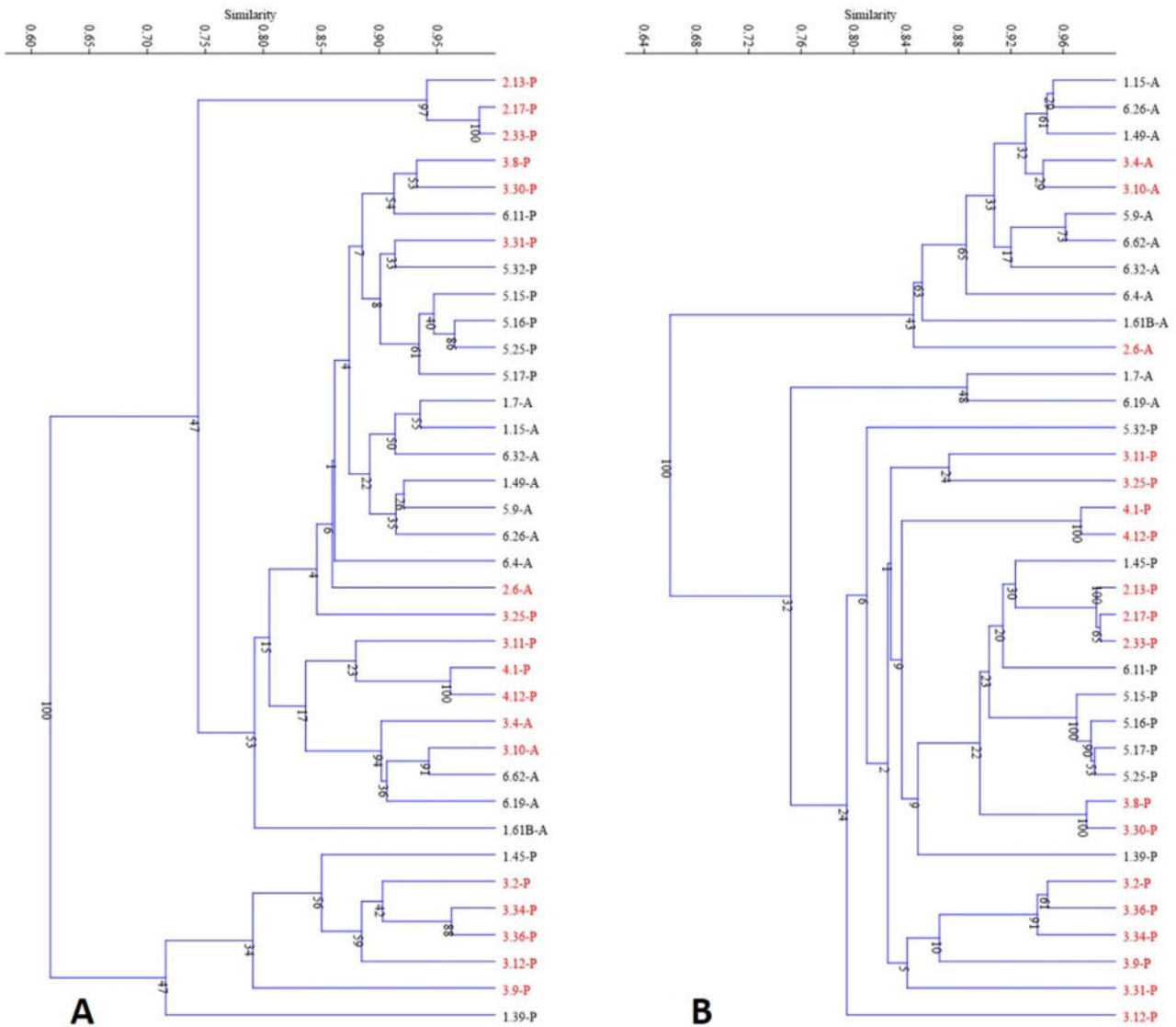


Figure 6. UPGMA analysis (Bray Curtis distance) performed on the heavy metal data matrix (A) and on the antibiotic data matrix (B) obtained by means of the broth microdilution methods.

The prevalence of *Proteobacteria* in red and black epilithons might be expected since *Proteobacteria* dominance in epilithic biofilm has been already reported from different environments (e.g. Bartrons, Catalan and Casamayor 2012; Ragon et al. 2012). Moreover, it has been reported that a limited number of dominant bacterial phyla (<10) prevails in rock biofilm bacterial communities (McNamara and Mitchell 2005). The bacterial communities in black and red epilithons were dominated by *Acinetobacter* and unclassified *Gallionellaceae*, *Sideroxydans* sp. and *Gallionella* sp., respectively. The HTS data partially agree with those obtained using culture-dependent approach: indeed, strains affiliated to *Acinetobacter* spp. were more abundant in black epilithon samples, whereas *Pseudomonas* sp. prevailed in red epilithon ones (Fig. 5). This discrepancy in the red epilithon might be explained as bacteria belonging to the genera *Gallionella* and *Sideroxydans* don't grow on the TSA medium, hence, they might have not been isolated by the culture conditions used in this work. Indeed, they are considered 'iron bacteria' because they are able to oxidize iron

(Emerson, Fleming and McBeth 2010) and live in environments with high iron concentration, preferentially at the oxic-anoxic interfaces.

Bacterial plate counts gave comparable results in both red and black epilithon samples revealing that there was not a significant difference between the two biofilms. This result suggest that the fractions of cultivable and non-cultivable bacterial communities are comparable in the two environments and do not contribute to the differences between the red and black epilithon.

The complete absence of RAPD haplotype sharing between red and black epilithon both for *Acinetobacter* and *Pseudomonas*, suggests that the two communities are genetically differentiated at the strain level. These findings are in agreement with previous data suggesting that when two communities co-exist in the same environment and are spatially and physically in contact with each other, there could be some selective pressure that create and/or maintain a barrier to the free intermix of bacterial strains (Chiellini et al. 2014 and references therein). This selective pressure might likely be due to biotic conditions such as the

two biofilm types. This is in agreement with data obtained with other reports concerning the structuring on bacterial communities isolated from different compartments (Mengoni et al. 2014). *Pseudomonas* spp. population showed higher resistance levels towards the tested antibiotics in respect to the *Acinetobacter* spp. one. The comparison of our results with other published data is not easy to perform due to the scarcity of similar studies at environmental scale. One example can be given by *Acinetobacter* sp. and *Pseudomonas* sp. strains associated with fish and water from Congonhas river, that revealed no significant differences in resistance patterns towards Streptomycin at concentration of 10 µg/ml (Sousa and Silva-Souza 2001).

The RAPD and phylogenetic analysis performed for the two bacterial groups suggested that *Pseudomonas* constitutes a panmictic population and dominates the red epilithon, while *Acinetobacter* is a clonal population that dominates the black epilithon. Bacteria are naturally organized in clonal populations due to binary fission; however, the occurrence of horizontal gene transfer might shift the genetic structure of populations to panmictic (Vogel et al. 2003).

Moreover, it is known that in natural communities the release of toxic/antagonistic compounds synthesized by some bacteria can prevent the invasion from other (micro)organisms by creating an inhospitable zone for competitors (Stubbenieck and Straight 2016; Stubbenieck, Vargas-Bautista and Straight 2016). The competition for space and resources and the production of molecules with inhibiting potential against other organisms represent an advantage for the colonization of a niche by bacteria, as previously suggested by the few studies on this issue (Pérez-Gutiérrez et al. 2013).

It has also been reported that antagonistic interactions between bacterial strains belonging to the same or to different taxa might play a role in driving the structuring of microbial communities (Maida et al. 2016). Cross-streaking experiments (Table 5) performed in this work demonstrated that *Pseudomonas* sp. strains are able to inhibit the growth of *Acinetobacter* sp. strains by synthesizing diffusible (and very likely not volatile) antibiotic compounds. The ability of *Pseudomonas* sp. strains in synthesizing antibiotic compounds has been previously documented (e.g. Raaijmakers, Weller and Thomashow 1997; Haas and Keel 2003). This finding strongly supports the importance of the antagonistic interaction (biotic factor) as a driving force for the assembly of microbial communities in those environments that are spatially structured, such as biofilms (Pérez-Gutiérrez et al. 2013). This might suggest that the molecule(s) synthesized by *Pseudomonas* strains and able to inhibit the growth of *Acinetobacter* can diffuse within the biofilm and might be responsible for the structuring of the two distinct biofilms characterizing the black and the red epilithons. Indeed, diffusion has been previously demonstrated as being the prevalent transport process for molecules in biofilms (Stewart 2003). Considering that the black and red epilithons never mix but they are spatially located close to each other, 'diffusion' might be the main mechanism, if not the only one, acting in these environments, since molecules can easily diffuse/flow between cells; indeed, in aquatic environment—such as the Acquarossa river—molecules produced by bacteria are carried off the cell and there is a low probability that they might reach neighbor bacteria (Watnick and Kolter 2000).

CONCLUSION

The structuring of bacterial communities in black and red epilithic rock biofilms in Acquarossa river (Viterbo, Italy) has

been investigated by means of cultivation and unculturable dependent approaches. In both cases, differences in taxonomic composition of black and red epilithon were highlighted. While black epilithon is dominated by *Acinetobacter* sp., bacteria inhabiting red epilithon are most iron-oxidizing strains. The proportion of culturable and non-culturable fractions of the community, as well as the resistance patterns towards heavy metals, do not seem to affect the differential structuring of the communities. On the other side, antibiotic resistance patterns and, in a larger proportion, the antagonistic interactions between the dominant bacterial genera seem to affect the whole structuring of red and black epilithic biofilms. These findings enforce the role of biotic factors as responsible for the structuring of natural bacterial communities. Overall, the study suggests that there is a selection of population at very small scale, and that different population might compete for different niches.

SUPPLEMENTARY DATA

Supplementary data are available at [FEMSEC](https://www.femsec.org) online.

ACKNOWLEDGEMENTS

The authors acknowledge Renato Zompanti for his great technical and logistic support to explore the Acquarossa site. This work was financially supported by G.I.D.A. S.p.A. Authors are very grateful to the two anonymous reviewers for the suggestions to improve the manuscript.

Conflict of interest. None declared.

REFERENCES

- Arasu MV, Veeramuthu D, Savarimuthu I. Antibacterial and antifungal activities of polyketide metabolite from marine *Streptomyces* sp. AP-123 and its cytotoxic effect. *Chemosphere* 2013;**90**:479–87.
- Bacci G, Bazzicalupo M, Benedetti A et al. StreamingTrim 1.0: a Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies. *Mol Ecol Resour* 2014;**14**:426–34.
- Bartrons M, Catalan J, Casamayor EO. High bacterial diversity in epilithic biofilms of oligotrophic mountain lakes. *Microb Ecol* 2012;**64**:860–9.
- Battin TJ, Sloan WT, Kjelleberg S et al. Microbial landscapes: new paths to biofilm research. *Nat Rev Microbiol* 2007;**5**:76.
- Battin TJ, Besemer K, Bengtsson MM et al. The ecology and biogeochemistry of stream biofilms. *Nat Rev Microbiol* 2016;**14**:251–63.
- Berner RA. Iron sulfides formed from aqueous solution at low temperatures and atmospheric pressure. *J Geol* 1964;**72**:293–306.
- Besemer K, Singer G, Quince C et al. Headwaters are critical reservoirs of microbial diversity for fluvial networks. *P Biol Sci* 2013;**280**:20131760.
- Caporaso JG, Lauber CL, Walters WA et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *P Natl Acad Sci USA* 2011;**108**:4516–22.
- Caporaso JG, Lauber CL, Walters WA et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 2012;**6**:1621–4.
- Chakravorty S, Helb D, Burday M et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Meth* 2007;**69**:330–9.

- Chiellini C, Maida I, Emiliani G et al. Endophytic and rhizospheric bacterial communities isolated from the medicinal plants *Echinacea purpurea* and *Echinacea angustifolia*. *Int Microbiol* 2014;17:165–74.
- Cole JR, Q Wang JA, Fish B et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014;42:D633–42.
- Di Cello F, Fani R. A molecular strategy for the study of natural bacterial communities by PCR-based techniques. *Minerva Biotechnol* 1996;8:126–34.
- Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci* 2003;14:927–30.
- Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat methods* 2013;10:996–8.
- Emerson D, Fleming EJ, McBeth JM. Iron-oxidizing bacteria: an environmental and genomic perspective. *Annu Rev Microbiol* 2010;64:561–83.
- Emerson D, Field EK, Chertkov O et al. Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. *Front Microbiol* 2013;4:254.
- Haas D, Keel C. Regulation of antibiotic production in root-colonizing *Pseudomonas* spp. and relevance for biological control of plant disease. *Ann Rev Phytopathol* 2003;41:117–53.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95 /98/NT. *Nucl Acid S* 1999;41:95–98.
- Hammer Ø, Harper DAT, Ryan PD. PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 2001;4:1–9.
- Harrison AP, Cattani I, Turfa JM. Metallurgy, environmental pollution and the decline of Etruscan civilisation. *Environ Sci Poll Res* 2010;17:165–80.
- Hem JD, Cropper WH. Survey of ferrous-ferrous chemical equilibria and redox potentials. In: *Chemistry of Iron in Natural Water, Geological Survey Water-supply Paper 1459*. Washington: United States Government Printing Office, 1960, 1459.
- Hill MO. Diversity and evenness: a unifying notation and its consequences. *Ecology* 1973;54:427–32.
- Hook D. The composition and technology of selected Bronze Age and Early Iron Age copper alloy artefacts from Italy. In: Bietti Sestieri AM, Macnamara E (eds). *Prehistoric Metal Artefacts From Italy (3500–720 BC) in the British Museum (British Museum Research Publication 159)*. London: British Museum, 2007;308–23.
- Huse SM, Dethlefsen L, Huber JA et al. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLOS GENET* 2008;4:e1000255.
- Illumina. 16S Metagenomic sequencing library preparation. In: *Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System*. Illumina, 2013, 1–28. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
- Jorgensen HJ, NCCLS Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. Approved standard, NCCLS-M7. *Infect Dis Clin N Am* 1993; 7:393–409.
- Kastman EK, Kamelamela N, Norville JW et al. Biotic interactions shape the ecological distributions of *Staphylococcus* species. *MBio* 2016;7:e01157–16.
- Kobayashi Y, Kim C, Yoshimizu C et al. Longitudinal changes in bacterial community composition in river epilithic biofilms: influence of nutrients and organic matter. *Aquat Microb Ecol* 2009;54:135–52.
- Ledger ME, Hildrew AG. Temporal and spatial variation in the epilithic biofilm of an acid stream. *Freshwater Biol* 1998;40:655–70.
- Maida I, Chiellini C, Mengoni A et al. Antagonistic interactions between endophytic cultivable bacterial communities isolated from the medicinal plant *Echinacea purpurea*. *Environ Microbiol* 2016;18:2357–65.
- Mathur J, Bizzoco RW, Ellis DG et al. Effects of abiotic factors on the phylogenetic diversity of bacterial communities in acidic thermal springs. *Appl Environ Microb* 2007;73:2612–23.
- McNamara CJ, Mitchell R. Microbial deterioration of historic stone. *Front Ecol Environ* 2005;3:445–51.
- Mengoni A, Maida I, Chiellini C et al. Antibiotic resistance differentiates *Echinacea purpurea* endophytic bacterial communities with respect to plant organs. *Res Microbiol* 2014;165:686–94.
- Meyers GE. Etrusco-Italic monumental architectural space from the Iron Age to the Archaic period: an examination of approach and access. *Doctoral dissertation*. <https://repositories.lib.utexas.edu/handle/2152/784>, 2003.
- Meyers GE. Approaching monumental architecture: mechanics and movement in Archaic Etruscan palaces. *Pap Brit Sch Rome* 2013;81:39–66.
- Moënne-Loccoz Y, Mavingui P, Combes C et al. Microorganisms and biotic interactions. In: *Environmental Microbiology: Fundamentals and Applications*. Netherlands: Springer, 2015, 395–444.
- Mori E, Liò P, Daly S et al. Molecular nature of RAPD markers amplified from *Haemophilus influenzae* Rd genome. *Res Microbiol* 1999;150:83–93.
- Papaleo MC, Romoli R, Bartolucci G et al. Bioactive volatile organic compounds from Antarctic (sponges) bacteria. *New Biotechnol* 2013;30:824–38.
- Paulson JN, Stine OC, Bravo HC et al. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* 2013;10:1200–2.
- Pérez-Gutiérrez RA, López-Ramírez V, Islas A et al. Antagonism influences assembly of a *Bacillus* guild in a local community and is depicted as a food-chain network. *ISME J* 2013;7:487.
- Petrosino JF, Highlander S, Luna RA et al. Metagenomic pyrosequencing and microbial identification. *Clin Chem*; 2009;55:856–66.
- R Core Team. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*. Vienna, Austria: R Core Team, 2014.
- Raaijmakers JM, Weller DM, Thomashow LS. Frequency of antibiotic-producing *Pseudomonas* spp. in natural environments. *Appl Environ Microb* 1997;63:881–7.
- Ragon M, Fontaine MC, Moreira D et al. Different biogeographic patterns of prokaryotes and microbial eukaryotes in epilithic biofilms. *Mol Ecol* 2012;21:3852–68.
- Rubin MA, Leff LG. Nutrients and other abiotic factors affecting bacterial communities in an Ohio River (USA). *Microb Ecol* 2007;54:374–83.
- Sousa JAD, Silva-Souza ÂT. Bacterial community associated with fish and water from Congonhas River, Sertaneja, Paraná, Brazil. *Braz Arch Biol Techn* 2001;44:373–81.
- Staccioli RA. Considerazioni sui complessi monumentali di Murlo e di Acquarossa. *Publications de l'École française de Rome* 1976;27:961–72.
- Stewart PS. Diffusion in biofilms. *J Bacteriol* 2003;185:1485–91.
- Stoddard SF, Smith BJ, Hein R et al. rrn DB: improved tools for interpreting rRNA gene abundance in bacteria and archaea

- and a new foundation for future development. *Nucleic Acids Res* 2014;**43**:D593–8.
- Stoodley P, Sauer K, Davies DG et al. Biofilms as complex differentiated communities. *Ann Rev Microbiol* 2002;**56**:187–209.
- Stubbendieck RM, Vargas-Bautista C, Straight PD. Bacterial communities: interactions to scale. *Front Microbiol* 2016;**7**:1234.
- Stubbendieck RM, Straight PD. Multifaceted interfaces of bacterial competition. *J Bacteriol* 2016;**198**:2145–55.
- Tamura K, Peterson D, Peterson N et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;**28**:2731–9.
- Thirumurugan D, Vijayakumar R. Characterization and structure elucidation of antibacterial compound of *Streptomyces* sp. ECR77 isolated from East Coast of India. *Curr Microbiol* 2015;**70**:1–11.
- Vogel J, Normand P, Thioulouse J et al. Relationship between spatial and genetic distance in *Agrobacterium* spp. in 1 cubic centimeter of soil. *Appl Environ Microb* 2003;**69**:1482–7.
- Watnick P, Kolter R. Biofilm, city of microbes. *J Bacteriol* 2000;**182**:2675–9.
- Widder S, Allen RJ, Pfeiffer T et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J* 2016;**10**:2557.
- Wilhelm L, Besemer K, Fragner L et al. Altitudinal patterns of diversity and functional traits of metabolically active microorganisms in stream biofilms. *ISME J* 2015;**9**:2454–64.
- Williams JG, Kubelik AR, Livak KJ et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 1990;**18**:6531–5.
- Zancarini A, Echenique-Subiabre I, Debroas D et al. Deciphering biodiversity and interactions between bacteria and microeukaryotes within epilithic biofilms from the Loue River, France. *Sci Rep* 2017;**7**:4344.



Temporal Evolution of Bacterial Endophytes Associated to the Roots of *Phragmites australis* Exploited in Phytodepuration of Wastewater

Alberto Vassallo¹, Elisangela Miceli¹, Camilla Fagorzi¹, Lara Mitia Castronovo¹, Sara Del Duca¹, Sofia Chioccioli¹, Silvia Venditto¹, Ester Coppini², Donatella Fibbi² and Renato Fani^{1*}

¹ Department of Biology, University of Florence, Sesto Fiorentino, Italy, ² G.I.D.A. SpA, Prato, Italy

OPEN ACCESS

Edited by:

Margherita Sosio,
Naicons Srl, Italy

Reviewed by:

Gaurav Saxena,
Jawaharlal Nehru University, India
Shaohua Chen,
Institute of Urban Environment (CAS),
China

*Correspondence:

Renato Fani
renato.fani@unifi.it

Specialty section:

This article was submitted to
Microbiotechnology,
a section of the journal
Frontiers in Microbiology

Received: 27 March 2020

Accepted: 25 June 2020

Published: 17 July 2020

Citation:

Vassallo A, Miceli E, Fagorzi C,
Castronovo LM, Del Duca S,
Chioccioli S, Venditto S, Coppini E,
Fibbi D and Fani R (2020) Temporal
Evolution of Bacterial Endophytes
Associated to the Roots
of *Phragmites australis* Exploited
in Phytodepuration of Wastewater.
Front. Microbiol. 11:1652.
doi: 10.3389/fmicb.2020.01652

Improvement of industrial productions through more environment-friendly processes is a hot topic. In particular, land and marine environment pollution is a main concern, considering that recalcitrant compounds can be spread and persist for a long time. In this context, an efficient and cost-effective treatment of wastewater derived from industrial applications is crucial. Phytodepuration has been considered as a possible solution and it is based on the use of plants and their associated microorganisms to remove and/or transform pollutants. In this work we investigated the culturable microbiota of *Phragmites australis* roots, sampled from the constructed wetlands (CWs) pilot plant in the G.I.D.A. SpA wastewater treatment plant (WWTP) of Calice (Prato, Tuscany, Italy) before and after the CW activation in order to check how the influx of wastewater might affect the resident bacterial community. *P. australis* specimens were sampled and a panel of 294 culturable bacteria were isolated and characterized. This allowed to identify the dynamics of the microbiota composition triggered by the presence of wastewater. 27 out of 37 bacterial genera detected were exclusively associated to wastewater, and *Pseudomonas* was constantly the most represented genus. Moreover, isolates were assayed for their resistance against eight different antibiotics and synthetic wastewater (SWW). Data obtained revealed the presence of resistant phenotypes, including multi-drug resistant bacteria, and a general trend regarding the temporal evolution of resistance patterns: indeed, a direct correlation linking the appearance of antibiotic- and SWW-resistance with the time of exposure to wastewater was observed. In particular, nine isolates showed an interesting behavior since their growth was positively affected by the highest concentrations of SWW. Noteworthy, this study is among the few investigating the *P. australis* microbiota prior to the plant activation.

Keywords: *Phragmites australis*, phytodepuration, wastewater, endophytes, antibiotic resistance, metal resistance

INTRODUCTION

Plants and microorganisms have been living in association for a very long time. In fact, arbuscular mycorrhizal mutualism is believed to have had a key importance in the terrestrialization process and in the evolution and diversification of plant phototrophs (Selosse and Le Tacon, 1998; Heckman et al., 2001). Different microorganisms (bacteria and fungi) can establish (more or less) deep associations with plants; some of them exhibit an endophytic lifestyle, in that they colonize plant tissues internally, although a more specific definition of endophytes states that they are organisms which, at some moment of their life cycle, colonize the internal plant tissues without causing any type of harm to the host (Patriquin and Döbereiner, 1978). Potential endophytes often inhabit the surrounding soil, especially rhizosphere, from where they can enter plant tissues switching to an endophytic lifestyle. They may thus enter plant tissues through wounds, germinating radicles, emergence points of lateral roots or root elongation and differentiation zones (Reinhold-Hurek et al., 2006; Sturz et al., 2010). Once inside, bacteria adapt to different environmental conditions (e.g., pH, osmotic pressure, carbon source, and availability of oxygen) and overcome plant defense responses (Zeidler et al., 2004).

The plant host and the bacterial endophytes create a mutualistic interaction, with bacteria gaining nutrients and a niche to colonize (Sturz et al., 2010). Even though the exact role of endophytes within plant tissues has not been fully understood yet, it is well-established that in many cases endophytes are beneficial to plants (Schlaeppli and Bulgarelli, 2015; Wani et al., 2015). The most common functions observed for bacterial endophytes are (i) uptake of nutrients (e.g., N, P, S, Mg, Fe, and Ca; Duijff et al., 1999; Çakmakçı et al., 2006), (ii) biosynthesis of phytohormones promoting plant growth (Spaepen et al., 2007), (iii) 1-aminocyclopropane-1-carboxylate deaminase activity (ACC; Glick et al., 2007), (iv) nitrogen fixation (Doty et al., 2009), (v) prevention of pathogenic infections (Weller, 2007; Pérez-García et al., 2011), (vi) acceleration of seedling emergence (Hardoim et al., 2008), and (vii) tolerance to pollution and stresses (Ryan et al., 2008; Lugtenberg and Kamilova, 2014).

In the context of the present work, particularly important is the ability of plant-associated bacteria to increase tolerance to pollution and/or increase the ability of plants to detoxify polluted environments. Environmental pollution, especially water pollution, represents a concern of considerable prominence in the current society. In this regard, phytodepuration is the overarching term for a group of technologies that utilizes plants and the associated rhizospheric microorganisms to remove and/or transform contaminants leached from soils/sediments and from used water streams (He et al., 2017; Saxena et al., 2020). It represents an environmental-friendly and a valuable solution for environmental cleanup, in particular for wastewater treatment, and it is popular because of its cost effectiveness, aesthetic advantages, and long-term applicability (Puvanakrishnan et al., 2019). In the present manuscript the term “phytodepuration” has been used to indicate specifically the remediation process regarding water and wastewater, rather than “phytoremediation,”

TABLE 1 | Antibiotics used in this work.

Antibiotic	Class	Target
Ampicillin	Penicillins	Cell wall synthesis: inhibitor of D-Ala-D-Ala carboxypeptidase
Chloramphenicol	Phenicols	Ribosome: inhibitor of peptidyl transferase activity of 23S rRNA
Ciprofloxacin	Fluoroquinolones	Topoisomerases
Kanamycin	Aminoglycosides	Ribosome: inhibitor of 30S ribosomal subunit
Rifampicin	Ansamycins	DNA-dependent RNA polymerase
Streptomycin	Aminoglycosides	Ribosome: inhibitor of 30S ribosomal subunit
Tetracycline	Tetracyclines	Ribosome: it blocks the binding of aminoacyl-tRNAs
Trimethoprim	Diaminopyrimidines	DNA replication: inhibitor of dihydrofolate reductase

TABLE 2 | Composition of synthetic wastewaters (SWWs).

Compound	1X SWW	2X SWW	3X SWW
H ₃ BO ₃	20	40	60
FeCl ₂ · 4H ₂ O	15	30	45
Na ₂ SeO ₃	0.03	0.06	0.09
NaCl	5,000	10,000	15,000

Concentrations are expressed as mg/L.

which has a more general meaning, encompassing applications regarding, for example, soil remediation.

The constructed wetlands (CW) are engineered systems designed to mimic the self-purification processes of natural wetlands. For decades, CW have been successfully used for treating wastewater of different origins and have been identified as a sustainable wastewater management option worldwide (Wang et al., 2017), demonstrating their ability to eliminate diffuse pollutants from urban, rural, and industrial emissions. In literature, the effectiveness of the use of CW in the treatment of sewage containing heavy metals and high salinity is reported (Vymazal, 2011). This process is due to the interaction between plants, microorganisms, soil, and polluting substances (Zhou et al., 2013).

In CW, the rhizosphere is the mainly involved plant compartment, where multiple different physiochemical and biological processes occur (Stottmeister et al., 2003). The common reed *Phragmites australis* is one of the most employed plant species, because of its ability to flourish in marshy areas and swamps and the high detoxification and phytodepuration potential. Moreover, it is widely used to treat industrial wastewater containing heavy metals (Zhang et al., 2017). One peculiar characteristic of *P. australis* is that its internal environment is characterized by a relatively constant osmotic gradient determined by the downward transportation of Na⁺ from stems to roots (Vasquez et al., 2006). For this reason, *P. australis* is also well-adapted to salty ecosystems. In CW, vegetation is responsible for only a small amount of pollutant removal (0.02%; Zhang et al., 2017), while its main function is

TABLE 3 | Bacterial counts in roots of *Phragmites australis* collected during the five samplings and meteorological conditions registered monthly.

	Samplings				
	1st (March 2017)	2nd (July 2017)	3rd (November 2017)	4th (June 2018)	5th (December 2018)
SFS-v (CFU/g)	4×10^6	1×10^7	1×10^6	3×10^6	6×10^5
SFS-h (CFU/g)	5×10^6	1×10^7	5×10^6	2×10^7	1×10^6
Average air temperature (°C)	13.8	26.4	10.6	23.5	7.4
Total precipitations (mm)	58.8	2.0	108.2	48.6	42.6

Bacterial counts are expressed as colony forming units per gram (CFU/g) of roots, air temperature in Celsius degrees, total precipitations in millimeters.

to provide additional oxygen and organic matter for microbial growth (Zhou et al., 2013). Indeed, microorganisms have been described as the main actors of pollutant removal in CW (Zhang et al., 2017). Phytodepuration has proved to effectively remove or neutralize hazardous environmental contaminants and it is predicted to have a growing application in the next years. However, this process presents some limitations, such as the toxic effects of pollutants on the growth and health of the plants (Glick, 2003). In fact, plant biomass is critical for phytodepuration (Germaine et al., 2010) and even hyperaccumulator plants, which can accumulate concentrations of toxic elements up to 100-fold higher than other plant species, usually exhibit a reduced growth. Also, phytodepuration may determine the accumulation of contaminants in plant tissues, which, in turn, is responsible for ecological and airborne exposure issues (Ho et al., 2012). In this scenario, rhizobacteria and endophytic bacteria can aid plants by supporting their growth (Tesar et al., 2002; Shaw and Burns, 2004; Chaudhry et al., 2005), reducing phytotoxicity effects, increasing pollutant uptake and removal (Glick and Stearns, 2011), reducing the release of toxic compounds into the atmosphere (Barac et al., 2004), removing contaminants and/or accumulating heavy metals (Germaine et al., 2010; Ho et al., 2012).

The experimental plant of Calice (Prato, Italy), managed by G.I.D.A. SpA, has therefore set itself as a goal to verify the action of this association in tertiary treatment of landfill leachate (LFL; Coppini et al., 2019).

The aim of this work was to characterize the cultivable bacterial communities associated to the roots of *P. australis* plants in Calice CW and to analyze their temporal dynamics before and after the activation of the plant for 22 months. This allowed the assessment of wastewater influx effect in shaping the composition of pre-existing bacterial communities. Moreover, bacteria isolated from roots were tested for their ability to grow in the presence of synthetic wastewater (SWW), along with their resistance against a panel of antibiotics commonly used to treat infections in humans. To the best of our knowledge, this work is among the few taking in consideration the bacterial composition of endophytes before the activation of CW, and likely the first regarding this issue in *P. australis*.

MATERIALS AND METHODS

Site Description

P. australis plants were obtained from the CWs pilot plant managed by G.I.D.A. SpA and located at Calice Wastewater

Treatment Plant (WWTP) in Prato, Italy. The CW of Calice was designed for the tertiary treatment of LFL. This CW is located downstream of a membrane bio-reactor (MBR) designed to pretreat a mixture of LFLs before their discharging in the main line of a full-scale WWTP, which treats both urban and industrial wastewater.

Constructed wetlands medium, used as substrate for the growth of *P. australis*, consists of four layers of gravel and sand; proceeding from the top to the bottom they are (thickness of layers and diameter range of particles are reported in brackets, respectively): gravel (20 cm; 5–10 mm) – sand (60 cm; 0.1–0.4 mm and 0.02–0.1 mm) – gravel (10 cm; 5–10 mm) – gravel (10 cm; 40–70 mm). CW implant was designed with two parallel lines, named “Line A” and “Line B,” respectively, with a total surface area of 1,620 m². Each line is a two-stage subsurface flow system (SFS), consisting of a vertical system (SFS-v) followed by a horizontal one (SFS-h). The SFS-v of Line A is subdivided into four parallel separated tanks (SFS-v1, SFS-v2, SFS-v3 e SFS-v4), while the SFS-v of Line B is composed by two tanks (SFS-v5 e SFS-v6). Furthermore, both SFS-h lines are composed by three tanks, each one receiving the same hydraulic load. The maximum hydraulic load supplied to the entire system was 95 m³/day corresponding to a 1.9-day Hydraulic Retention Time for the horizontal stage (Coppini et al., 2019).

Sampling and Isolation of Bacteria

Samples from the roots of *P. australis* were collected using sterile plastic bags and immediately transported to the laboratory for the subsequent processing. All procedures described hereinafter were carried out under sterile conditions to avoid external contaminations. Samples of three different specimens of *P. australis* growing in three different tanks were grouped and pooled before starting any procedure. Two pools were obtained from both SFS-v and SFS-h for each sampling campaign. 1 g of fresh tissue from each pool was surface-sterilized with 1% v/v HClO solution at room temperature to remove epiphytic bacteria and then washed three times with sterile water. Aliquots of 100 µL of water from the last wash were plated in triplicate as sterility controls. Subsequently, samples were homogeneously potted in a sterile mortar with the addition of 2 mL of 0.9% w/v NaCl sterile solution. Serial dilutions of tissue extracts were plated in triplicate on trypticase soy agar (TSA) medium (Biolife) and incubated at 30°C for 48 h. The total number of aerobic heterotrophic fast-growing bacteria of each sample was expressed as colony forming units per gram of roots (CFU/g), and it was determined as an average of three replicates. Isolated bacteria

TABLE 4 | Number of bacterial isolates grouped for genus and sampling.

	Samplings					Total	%
	1st	2nd	3rd	4th	5th		
<i>Achromobacter</i>	-	1	3	4	4	12	4.08
<i>Acinetobacter</i>	-	4	-	-	-	4	1.36
<i>Aeromonas</i>	-	-	-	1	-	1	0.34
<i>Agrobacterium</i>	-	2	1	9	1	13	4.42
<i>Arthrobacter</i>	-	-	-	1	-	1	0.34
<i>Bacillus</i>	9	2	9	-	4	24	8.16
<i>Buttiauxella</i>	1	-	-	-	-	1	0.34
<i>Chryseobacterium</i>	-	-	-	-	3	3	1.02
<i>Comamonas</i>	-	1	-	-	-	1	0.34
<i>Devosia</i>	-	-	-	1	-	1	0.34
<i>Enterobacter</i>	-	-	2	-	-	2	0.68
<i>Flavobacterium</i>	1	-	-	2	2	5	1.70
<i>Halomonas</i>	-	1	6	-	-	7	2.38
<i>Idiomarina</i>	-	1	-	-	-	1	0.34
<i>Isoptericola</i>	-	-	1	-	-	1	0.34
<i>Janthinobacterium</i>	1	-	-	-	-	1	0.34
<i>Lelliottia</i>	2	-	-	-	-	2	0.68
<i>Lysobacter</i>	-	-	-	1	-	1	0.34
<i>Microbacterium</i>	-	2	1	1	-	4	1.36
<i>Micrococcus</i>	-	-	-	1	-	1	0.34
<i>Ochrobactrum</i>	-	2	-	-	1	3	1.02
<i>Paenibacillus</i>	-	-	-	-	5	5	1.70
<i>Pannonibacter</i>	-	1	2	3	-	6	2.04
<i>Pantoea</i>	7	-	-	-	2	9	3.06
<i>Paracoccus</i>	-	1	-	-	-	1	0.34
<i>Pectobacterium</i>	1	-	-	-	-	1	0.34
<i>Planococcus</i>	-	-	-	-	1	1	0.34
<i>Pseudomonas</i>	43	15	12	21	19	110	37.41
<i>Pseudoxanthomonas</i>	-	2	1	-	-	3	1.02
<i>Rheinheimera</i>	-	7	2	1	-	10	3.40
<i>Rhizobium</i>	-	-	2	13	1	16	5.44
<i>Shinella</i>	-	-	-	2	-	2	0.68
<i>Sphingobium</i>	-	-	-	1	-	1	0.34
<i>Staphylococcus</i>	1	2	-	-	-	3	1.02
<i>Stenotrophomonas</i>	1	-	8	1	10	20	6.80
<i>Thalassospira</i>	-	4	7	5	-	16	5.44
<i>Vibrio</i>	-	1	-	-	-	1	0.34
Total number of isolates	67	49	57	68	53	294	
Total number of genera	10	17	14	17	12	37	

–: absence of isolates.

were name-coded according to the portion of the CW from whom they were isolated (V for the SFS-v and H for the SFS-h, respectively) and the pool of origin collected during the five samplings (1–2, 3–4, 5–6, 7–8, and 9–10 for the first, second, third, fourth, and fifth sampling, respectively).

Amplification and Sequencing of 16S rRNA Gene

Polymerase chain reaction (PCR) were performed to amplify the 16S rRNA coding gene. 2 μ L of colony thermal lysate

were used as template for a PCR in 1X DreamTaq Buffer (Thermo Scientific) containing 200 μ M of each dNTPs, 0.2 μ M of primers P0 (5'-GAGAGTTTGATCCTGGCTCAG-3') and P6 (5'-CTACGGCTACCTTGTACGA-3'; Di Cello and Fani, 1996), 2 U of DreamTaq DNA Polymerase (Thermo Scientific) in a final volume of 25 μ l. The PCR cycling for 16S rRNA gene amplification was 95°C for 3 min followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min, then a final extension at 72°C for 10 min. A Bio-Rad T100 thermal cycler was used. Sequencing of 16S rRNA gene was performed by IGA Technology Services Srl (Udine, Italy).

Taxonomic and Phylogenetic Analyses

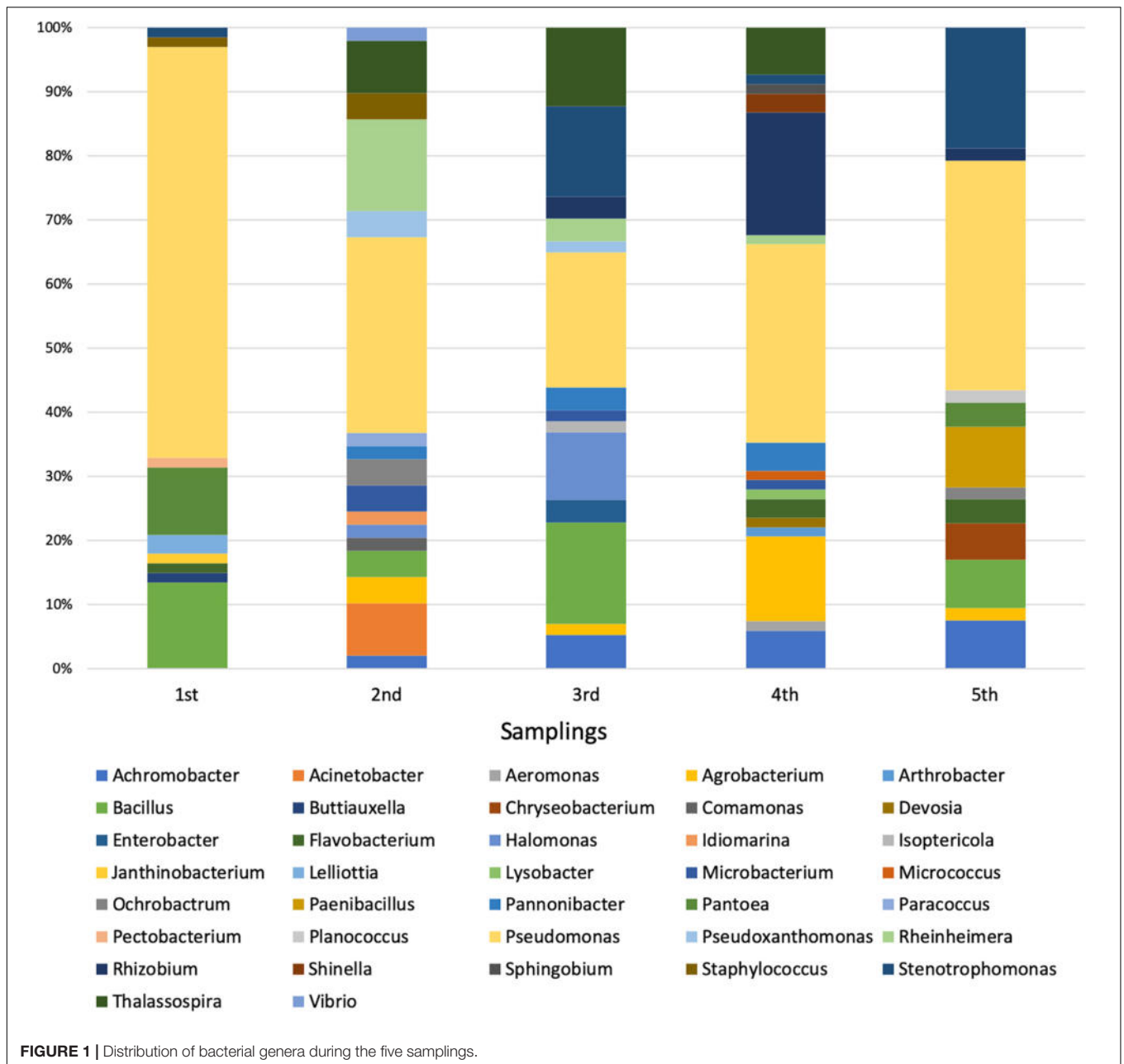
Taxonomic affiliation of isolates was determined through the alignment of sequences to those of type strains downloaded from the ribosomal database project (RDP; Cole et al., 2014) using BioEdit (Hall, 1999). The obtained alignment was then used to build a phylogenetic tree through MEGA7 (Kumar et al., 2016), applying the Neighbor-Joining algorithm with a 1000-bootstrap resampling.

Antibiotic Resistance Assay

Isolates were tested for their resistance against eight antibiotics (i.e., rifampicin, ampicillin, kanamycin, tetracycline, chloramphenicol, streptomycin, trimethoprim and ciprofloxacin) at six different concentrations (i.e., 1 – 10 – 25 – 50 – 100 – 150 μ g/mL; **Table 1**). Bacteria were firstly grown overnight on TSA (Biolife) at 30°C, then a single colony was resuspended in 100 μ L of 0.9 w/v NaCl sterile solution. The obtained suspensions were streaked on Mueller–Hinton II Agar (Biolife) plates supplemented with the tested antibiotics. Bacteria were also cultivated on the same medium in the absence of antibiotics, using these cultures as control to evaluate the presence of growth inhibition in presence of antibiotics. All plates were incubated at 30°C and growth performances were evaluated after 48 h. The minimal inhibitory concentration (MIC) value for each antibiotic was considered as the lowest concentration of the compound that inhibited visible growth of the tested isolate.

Growth in Presence of Synthetic Wastewater

Growth of strains isolated from roots of *P. australis* in presence of SWW was assayed through the broth microdilution methods (Wiegand et al., 2008) using trypticase soy broth (TSB) medium (Biolife). The bacterial inoculum for the experiment was prepared by dissolving an isolated bacterial colony in 10 ml of TSB medium after 24 h-growth at 30°C on TSA. The inoculum was incubated overnight at 30°C under shaking. Upon incubation, absorbance at 600 nm was measured and adjusted to 0.1. The experiment was performed using 96-well plates. Each well contained 10 μ L of bacterial inoculum, 80 μ L of TSB medium and 10 μ L of 10X, 20X, and 30X SWW, to reach the final concentration of 1X, 2X, and 3X, respectively. The composition of SWWs used for this assay is shown in **Table 2**. Growth performances in presence of SWW were evaluated calculating the ratio between the OD₆₀₀ of cultures in presence of SWW (herein after indicated



as OD_{600SWW} and OD_{600} of controls (i.e., bacteria grown in TSB lacking SWW). Bacterial isolates were considered sensitive to SWW when this parameter assumed values <0.7 , while they were evaluated as resistant when it was >1.3 .

RESULTS

Bacterial Counts

P. australis plants were sampled from the CW in Calice during a period of 22 months, spanning from March 2017 to December 2018; 5 samplings were conducted, with the first one (i.e., March 2017) performed before the activation of the CW (Table 3). This

experimental strategy allowed us to compare the composition of the cultivable bacterial community associated to the *P. australis* roots before and after the beginning of wastewater influx. The titer of viable bacteria associated to roots was determined as described in section “Materials and Methods”. Data obtained revealed that there were no great differences between the CFU counts in SFS-v and SFS-h, exception for the third and fourth samplings in which CFU values were higher in SFS-h (Table 3).

In general, bacterial load was quite constant during the experiment and fluctuations might be related to different factors, such as wastewater composition, frequency of raining, and/or seasonal variations. Indeed, it is likely that the weather exerted a main effect on bacterial growth since the highest bacterial loads

TABLE 5 | Frequencies of minimal inhibitory concentration (MIC) values among bacterial isolates.

	MIC (µg/mL)	Samplings				
		1	2	3	4	5
Rifampicin	1	12	18	26	9	8
	10	26	28	9	21	20
	25	23	–	9	22	13
	50	–	–	11	1	5
	100	–	–	–	3	4
	150	–	–	–	–	2
Ampicillin	> 150	–	–	–	–	–
	1	3	7	18	5	2
	10	1	8	1	1	3
	25	4	3	4	–	2
	50	2	7	6	5	5
	100	10	6	1	9	5
Kanamycin	150	1	4	2	3	5
	> 150	39	8	23	33	30
	1	1	–	–	2	3
	10	50	25	31	16	27
	25	2	7	5	7	4
	50	3	3	2	14	1
Tetracycline	100	1	1	1	5	–
	150	1	2	1	–	–
	> 150	1	6	15	12	17
	1	30	43	37	33	12
	10	28	4	8	19	31
	25	–	–	8	3	1
Chloramphenicol	50	–	–	1	–	–
	100	–	–	–	–	8
	150	–	–	–	–	–
	> 150	–	–	–	1	–
	1	20	14	9	15	4
	10	8	10	17	5	9
Streptomycin	25	4	7	14	13	14
	50	4	7	8	6	14
	100	16	6	–	–	6
	150	2	2	2	–	1
	> 150	7	–	4	17	4
	1	3	12	2	–	2
Trimethoprim	10	32	13	24	16	10
	25	10	9	9	4	12
	50	10	3	1	5	–
	100	4	–	4	2	8
	150	1	–	–	9	–
	> 150	3	7	14	20	20
Ciprofloxacin	1	12	15	8	–	5
	10	–	5	9	7	7
	25	1	3	2	2	7
	50	1	6	3	5	2
	100	2	7	15	4	5
	150	7	5	5	3	6
	> 150	42	5	12	35	20
	1	65	43	40	48	39
	10	–	3	14	8	5
	25	–	–	–	–	8
	50	–	–	–	–	–
	100	–	–	–	–	–
	150	–	–	–	–	–
	> 150	–	–	–	–	–

–: absence of isolates showing a specific MIC value.

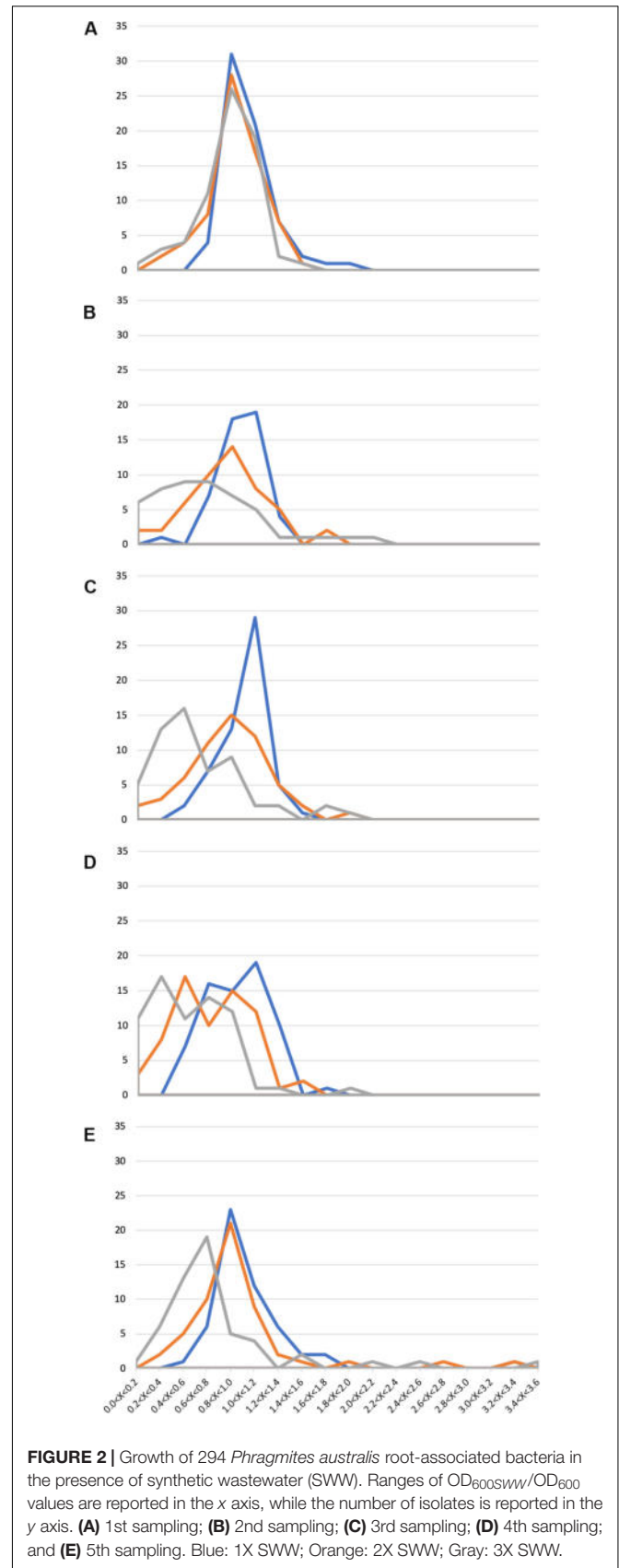


FIGURE 2 | Growth of 294 *Phragmites australis* root-associated bacteria in the presence of synthetic wastewater (SWW). Ranges of OD_{600SWW}/OD₆₀₀ values are reported in the x axis, while the number of isolates is reported in the y axis. **(A)** 1st sampling; **(B)** 2nd sampling; **(C)** 3rd sampling; **(D)** 4th sampling; and **(E)** 5th sampling. Blue: 1X SWW; Orange: 2X SWW; Gray: 3X SWW.

were observed during summer (i.e., second and fourth samplings, respectively), when the higher temperatures probably facilitated bacterial growth and the poor precipitations probably caused a higher concentration of wastewater.

Taxonomic Affiliation of Cultivable Bacteria

A total of 294 isolates (67, 49, 57, 68, and 53 from the first, second, third, fourth, and fifth sampling, respectively) were isolated from the *P. australis* roots. The attention was focused on bacteria isolated from this plant compartment because it has been reported that it is primarily involved in the depuration process (Riva et al., 2020). Each of the 294 isolates underwent a taxonomic characterization; to this purpose the amplification, sequencing, and analysis of 16S rRNA coding gene(s) were performed as described in section “Materials and Methods.”

Each sequence was submitted to Genbank and was assigned the accession number reported in **Supplementary Table S1**. The comparative analysis of each sequence with those available in databases allowed to split the 294 isolates into 37 different genera (**Supplementary Figures S1–S30**). The analysis revealed that 254 isolates were Gram-negative while 40 were Gram-positive bacteria. Moreover, a total of four different phyla were represented, with 246 belonging to *Proteobacteria* (59 *Alphaproteobacteria*, 14 *Betaproteobacteria*, and 173 *Gammaproteobacteria*), 8 to *Bacteroidetes* (all belonging to *Flavobacteriia* class), 33 to *Firmicutes* (all belonging to *Bacilli* class), and 7 to *Actinobacteria* (all belonging to *Actinobacteria* class). The most represented genus was *Pseudomonas*, whose members accounted for 37% of all isolates, as shown in **Table 4**. The abundance of *Pseudomonas* was not directly related to the activation of the CW, because it was the most represented genus even before the influx of wastewater (**Figure 1**).

Among the 37 genera, only 4 were exclusively present during the first sampling (i.e., *Buttiauxella*, *Janthinobacterium*, *Lelliottia*, and *Pectobacterium*), while 27 started being present from the second one on (i.e., *Achromobacter*, *Acinetobacter*, *Aeromonas*, *Agrobacterium*, *Arthrobacter*, *Chryseobacterium*, *Comamonas*, *Devosia*, *Enterobacter*, *Halomonas*, *Idiomarina*, *Isoptericola*, *Lysobacter*, *Microbacterium*, *Micrococcus*, *Ochrobactrum*, *Paenibacillus*, *Pannonibacter*, *Paracoccus*, *Planococcus*, *Pseudoxanthomonas*, *Rheinheimera*, *Rhizobium*, *Shinella*, *Sphingobium*, *Thalassospira*, and *Vibrio*). Hence, it is possible that the bacteria belonging to these 27 genera might derive from the wastewater, although it cannot be established whether they were present in wastewater with either urban or industrial origin. In addition to this, we cannot *a priori* exclude the possibility that they were already present in the pre-existing community (even though in low percentage) and that the presence of the wastewater might have exerted a selective pressure favoring their reproduction. In most cases, the phylogenetic trees showed a narrow taxonomic distribution of isolates, which clustered together in the same branch. For example, all *Acinetobacter* strains were phylogenetically close to *A. haemolyticus* (**Supplementary Figure S1**), all *Achromobacter* isolates formed a distinct cluster and were close

to *A. spanius* (**Supplementary Figure S5**), all *Chryseobacterium* were related to *C. indoltheticum* (**Supplementary Figure S7**), all *Paenibacillus* belonged to the same cluster and were close to *P. tundrae* (**Supplementary Figure S19**), all *Pannonibacter* were affiliated to *P. phragmitetus* (**Supplementary Figure S24**), and all *Thalassospira* isolates formed a separate branch in the phylogenetic tree (**Supplementary Figure S29**). Moreover, *Rheinheimera* isolates distributed in close branches which included *R. aquimaris*, *R. pacifica*, and *R. nanhaiensis* (**Supplementary Figure S22**). On the contrary, a higher phylogenetic diversity was observed in the case of *Bacillus* (**Supplementary Figure S4**), *Enterobacteriales* (**Supplementary Figure S10**), *Pseudomonas* (**Supplementary Figure S20**), *Rhizobiales* (**Supplementary Figure S23**), and *Stenotrophomonas* (**Supplementary Figure S28**). However, with the exception of the bacteria belonging to the *Enterobacteriales* order, in the case of these genera the formation of distinct clusters was observed.

Resistance Against Antibiotics

All 294 isolates were tested for their resistance against a panel of eight antibiotics used to treat human infections as described in section “Materials and Methods.” These compounds were chosen because they are representatives of diverse antibiotic classes and they are directed toward different cellular targets (**Table 1**). Data obtained are shown in **Table 5** and **Supplementary Table S2** and revealed that, overall, the most effective antibiotics were rifampicin, tetracycline and, above all, ciprofloxacin. On the contrary, the more tolerated antibiotic was ampicillin, especially in the case of *Pseudomonas* and *Stenotrophomonas* (**Supplementary Table S2**). Although resistant bacteria were isolated since the first sampling, a correlation between the time of exposure to the wastewater (i.e., earlier vs later samplings) and the increase of MIC values was observed for almost all antibiotics (**Table 5** and **Supplementary Table S2**).

On one hand, tests with rifampicin, ciprofloxacin and tetracycline showed a progressive increase in the number of isolates having the highest MIC values going from the first to the fifth sampling. On the other one, although MIC values were not determined in the assayed conditions because isolates were able to grow even in presence of the highest concentration of antibiotic, in the case of kanamycin and streptomycin the number of isolates with MIC > 150 µg/mL increased during time. Finally, clear trends were not detected using ampicillin, chloramphenicol and trimethoprim: indeed, there were bacteria able to survive in the presence of the highest concentration since the first sampling and, also, the frequency of resistant isolates was not subjected to temporal variations.

According to the MIC breakpoints provided by the European Committee on Antimicrobial Susceptibility Testing (EUCAST – Breakpoint tables for interpretation of MICs and zone diameters; Version 10.0, 2020¹), relatively to the antibiotics assayed in this work and limiting to the species reported by EUCAST, six isolates could be defined as multi-drug resistant strains, since they were resistant against at least three different antibiotics (**Supplementary Table S2**). In detail, two *Lelliottia* (V2R14 and

¹<http://www.eucast.org>

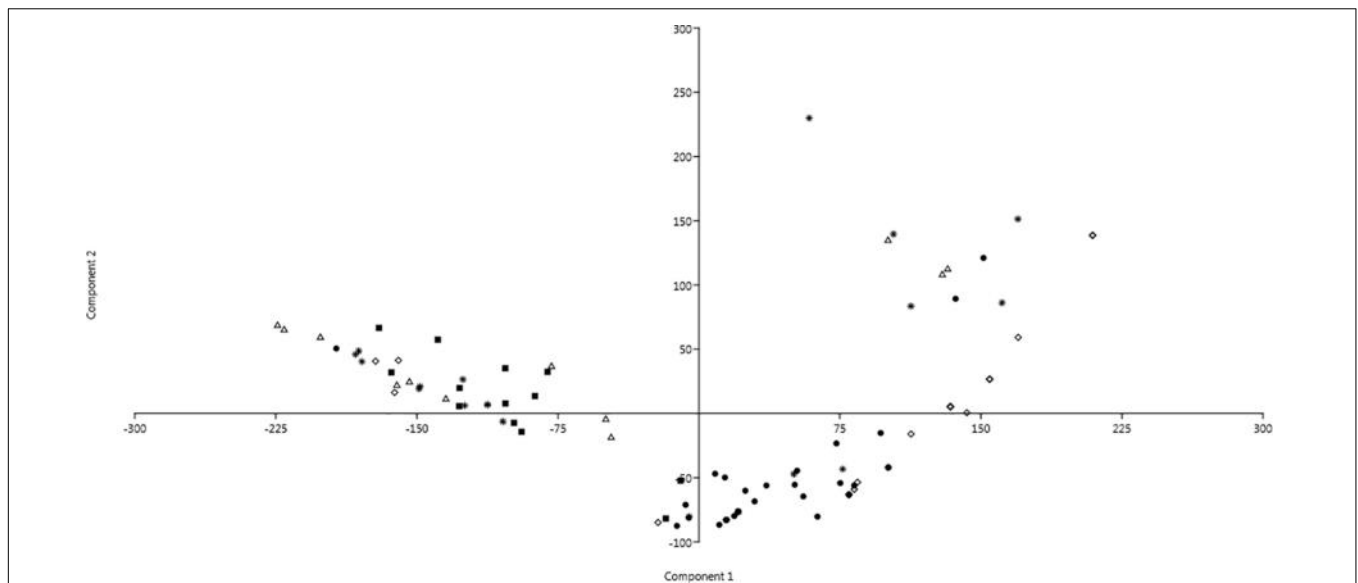


FIGURE 3 | Principal component analysis (PCA) showing profiles of resistance against antibiotics and SWW of isolates belonging to the genus *Pseudomonas*. Dot: 1st sampling; Filled square: 2nd sampling; Triangle: 3rd sampling; Diamond: 4th sampling; and Star: 5th sampling.

H1R21) and two *Enterobacter* (H5R6 and H5R7) isolates were resistant to ampicillin, chloramphenicol, and ciprofloxacin; lastly, the two *Pantoea* isolates H9R2 and H9R15 that were resistant to ampicillin, chloramphenicol, trimethoprim, and ciprofloxacin.

Growth in the Presence of SWW

The 294 isolated *P. australis* root-associated endophytes were also tested for their ability to grow in the presence of SWW containing B, Fe, and Se since these elements are critical for the WWTP studied in this work. Selection of bacteria able to grow efficiently in the presence of these compounds is of relevant interest because CW might be enriched with these more tolerant microorganisms, which, in turn, might increase the pollutant removal efficiency in wastewater. All bacterial isolates were assayed for their growth in TSB medium supplemented with three different concentration of SWW: 1X (i.e., a mix of H_3BO_3 , $FeCl_2 \cdot 4H_2O$, Na_2SeO_3 , and NaCl at the maximum concentrations allowed by law for sewer emission), and 2X and 3X in which TSB medium was supplemented with two- and threefold higher concentrations of 1X SWW, respectively.

In general, the presence of 1X SWW did not alter the growth of isolates, indicating that these endophytes can tolerate the presence of the tested compounds (Figure 2 and Supplementary Table S3). The analysis of data shown in Supplementary Table S3 and Figure 2 also revealed that 211 bacterial isolates were able to grow efficiently also in the presence of either 2X SWW or 3X SWW. Interestingly, eight isolates (i.e., V2R8, H3R17, H4R18, V6R1, V5R1, V8R24, H9R1, and H10R8), belonging to the genera *Bacillus*, *Planococcus*, *Pseudomonas*, and *Rheinheimera*, showed a positive correlation between growth and concentration of SWW: indeed, the higher the SWW concentration, the higher the growth of these bacteria. This finding suggests that these bacteria could represent good candidates for future

applications and for improvements of phytodepuration efficiency and pollutant removal. Moreover, the isolate H9R16 deserves further investigations, since it showed the highest growth increase (about 350%) in presence of SWW. The analysis of the 16S rDNA phylogenetic tree revealed that it joined bacteria belonging to the *Bacillus gibsonii* species (Supplementary Figure S4), alkaliphilic bacteria exploited for production of alkaline proteases (Martinez et al., 2013; Deng et al., 2014).

Similarly to the case of antibiotic resistance, also in this assay a correlation between the time of exposure to wastewater in the CW and the appearance of more resistant isolates was highlighted. Indeed, a progressive increase of the number of isolates showing an augmented growth (measured as OD_{600SWW}/OD_{600} ratio as described in section “Materials and Methods”) was observed from the first to the last sampling. For instance, several isolates with an $OD_{600SWW}/OD_{600} > 2.4$ (Figure 2E) were detected only during the last sampling.

Changes of resistance profiles along time, considering those against either antibiotics or SWW, were particularly clear in the case of isolates belonging to the genus *Pseudomonas*. As shown in Figure 3 by the mean of principal component analysis (PCA) performed with the PAST4 software (Hammer et al., 2001), the formation of two different main clusters was observed with the main part of isolates from the first sampling clustering independently from those isolated from all the other four samplings.

DISCUSSION

The aim of this work was the analysis of the composition, the phenotypic characterization, and the temporal dynamics of the cultivable microbiota associated to the roots of *P. australis*

grown in the CW of Calice (Prato, Italy) before and after the activation of the CW.

The composition of root cultivable microbiota was determined through five samplings spanning from March 2017 to December 2018. We focused on cultivable bacterial communities because their isolation and characterization might permit the identification of strains particularly resistant to the antibiotics and/or to the compounds present in the wastewater. Hence, these strains could be used to construct a “synthetic” consortium that, in turn, might be exploited in pilot-experiments with the goal of increasing the phytodepuration efficiency of the plant. The taxonomic analysis was performed on 294 cultivable bacteria through the analysis of the 16S rRNA genes. Even though we are completely aware that the number of isolates could not be representative of the entire community, the analysis performed gave useful hints on the effect of the activation of CW on bacterial community composition. Data obtained revealed that the wastewater income exerted a shaping effect on the bacterial composition. Overall, 37 bacterial genera were disclosed and six of them were detected both before and after the activation of CW. Moreover, bacteria belonging to 27 different genera were detected only after the activation of the CW, while, on the contrary, 4 genera that were present at the beginning were not found in the following samplings. The most represented genus in all five samplings was *Pseudomonas*, which accounted for the 37% of all isolates.

As it might be expected, the effect of the wastewater income was exerted not only at the taxonomic level, but also at the phenotypic one. Each of the 294 bacterial isolates was assayed for its resistance against a panel of eight antibiotics belonging to different chemical classes and acting toward different cellular targets. This analysis was performed since it is known that WWTP are reservoirs of antibiotic resistance genes and/or resistant bacteria. Data obtained revealed the presence of several resistant (or multi-resistant) strains and, mostly important, that the number of antibiotic resistant isolates and the degree of antibiotic resistance increased over time, from the first to the last sampling. This strongly suggests that the wastewater income might generate a selective pressure favoring the growth of those isolates intrinsically resistant to antibiotics, even though it cannot be *a priori* excluded the possibility of horizontal gene transfer (HGT) events and/or the acquisition of resistance through mutations in chromosomal genes.

All the 294 isolates were assayed also for their ability to grow in SWW containing three different concentrations of B, Se, Fe, and NaCl. Analogously to the antibiotic resistance pattern, we detected a similar correlation between SWW exposure and the ability to grow in its presence. Interestingly, among isolates able to grow in the presence of these compounds, nine of them showed an increasing growth at the highest concentrations of SWW: these isolates will deserve a specific focus to identify the molecular mechanisms behind this intriguing behavior. Moreover, a PCA carried out on *Pseudomonas* strains, isolated from all the five samplings, furtherly suggested that the main event changing the resistance patterns against antibiotics and SWW was the activation of the plant (i.e., when conveyance of the permeate into the tanks occurred). The characterization of heavy metal

resistant strains may be crucial to better understand the diffusion of antibiotic resistance genes in the environment. As a matter of fact, it has been reported that the occurrence of multiple heavy metal resistance markers is associated with the onset of antibiotic resistance (Wales and Davies, 2015; Wu et al., 2018; Zhu et al., 2019). This might be due to the co-localization of resistance genes against antibiotics and heavy metals in the same mobile genetic element(s) and, as a consequence, the accumulation of heavy metals in the environment can cause the selection of antibiotic resistant species. So, the dissemination of these heavy metal resistance genes represents an issue that should not be underestimated. Moreover, monitoring the presence of bacteria resistant to antibiotics and/or heavy metals specifically in WWTPs should be considered as a priority to contrast the spreading of multi-drug resistant (MDR) pathogens. Indeed, WWTPs represent hotspots for HGT events, because of the mixing of bacteria from diverse sources (e.g., households, hospitals, industries, etc.), the high bacterial densities, stressful conditions triggering SOS responses and presence of antibiotics at sublethal concentrations (Karkman et al., 2018). It must be also considered that although HGT occurring in WWTPs might not directly regard human pathogens, these could acquire resistance markers from harmless bacteria which act as vectors as soon as effluent is released in the environment (Manaia, 2017).

To deeply characterize this phenomenon, future work could take advantage of emulsion, paired isolation and concatenation PCR (epicPCR) as previously reported (Spencer et al., 2016; Hultman et al., 2018). Indeed, this kind of analysis could allow the “tagging” of resistance genes with phylogenetic markers, such as 16S rRNA gene, helping to compare these pairs in wastewater entering the WWTP and in effluents. However, this would be limited to target resistance markers with known sequence and for whom it is thereof possible to design specific primers.

CONCLUSION

The experimental approach used in this work revealed that the cultivable bacterial community existing prior to the plant activation underwent fluctuations in terms of both taxonomy and resistance to antibiotics and SWW compounds. As it might be expected, the influx of wastewater exerted a selective pressure on the resident bacterial community, selecting and/or bringing bacterial strains progressively more resistant to SWW and/or antibiotics. We are completely aware that the analysis of the entire community (both cultivable and uncultivable) might give more detailed insights into the composition of the total community. In spite of this, only the selection of particular cultivable strains, i.e., more resistant to SWW and antibiotics, can permit the formulation of a synthetic bacterial community to improve the phytodepuration properties of *P. australis*.

DATA AVAILABILITY STATEMENT

The datasets generated in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/genbank/>, MK110895, MK110946, MK110896, MK110920, MK110921, MK110947, MK110948, MK110922, MK110897, MK110949, MK110898, MK110950, MK110899, MK110959, MK110923, MK110925, MK110924, MK110960, MK110926, MK110945, MK110957, MK110927, MK110928, MK110929, MK110930, MK110931, MK110932, MK110900, MK110901, MK110902, MK110933, MK110934, MK110935, MK110958, MK110936, MK110937, MK110938, MK110939, MK110940, MK110941, MK110942, MK110943, MK110903, MK110904, MK110905, MK110906, MK110907, MK110908, MK110909, MK110910, MK110911, MK110912, MK110951, MK110913, MK110914, MK110915, MK110952, MK110916, MK110953, MK110917, MK110954, MK110918, MK110944, MK110961, MK110919, MK110955, MK110956, MK134509, MK134489, MK134488, MK134487, MK134486, MK134554, MK134508, MK134547, MK134496, MK138850, MK134502, MK134511, MK134551, MK134555, MK134559, MK134558, MK134549, MK138851, MK134553, MK134542, MK134557, MK134546, MK134544, MK134543, MK134541, MK134540, MK134510, MK134497, MK134490, MK134495, MK134494, MK134493, MK134499, MK134505, MK134552, MK134556, MK134545, MK134548, MK134500, MT165525, MK134507, MK134504, MK134503, MK134550, MK134501, MK134498, MK134506, MK134492, MK134491, MK134518, MK130934, MK130935, MK134524, MK134539, MK134538, MK134515, MK138852, MK134526, MK130907, MK130906, MK130937, MK134534, MK134533, MK138853, MK130915, MK130913, MK130910, MK130921, MK130917, MK134514, MK130914, MK134516, MK130911, MK134532, MK134531, MK134530, MK134528, MK134485, MK138854, MK130912, MK130908, MK130920, MK130932, MK130933, MK130936, MK134513, MK130922, MK130919, MK134521, MK134512, MK134537, MK134536, MK134535, MK134522, MK134519, MK134517, MK134529, MK134525, MK134523, MK134520, MK134527, MK130916, MK130931, MK130909, MK130923, MK130918, MK130945, MK130957, MK130901, MK130905, MK133358, MK138868, MK138869, MK138870, MK138872, MK138874, MK138875, MK130924, MK130928, MK138881, MK138862, MK130903, MK138867, MK138878, MK138861, MK130939, MK138863, MK130949, MK130953, MK130904, MK138876, MK138879, MK130926, MK130927, MK130929, MK138880, MK130930, MK138882, MK138883, MK138884, MK138885, MK138886, MK138887, MK130940, MK130941, MK130943,

MK138889, MK130944, MK130902, MK130900, MK130946, MK138877, MK138855, MK138856, MK138857, MK138858, MK138859, MK138860, MK138864, MK138865, MK138866, MK138871, MK138873, MK130925, MK130942, MK130948, MK130950, MK130955, MK138888, MK130951, MK130952, MK130954, MK130956, MK130947, MT165526, MT165527, MT165547, MT165551, MT165553, MT165552, MT165528, MT165529, MT165530, MT165531, MT165532, MT165533, MT165534, MT165557, MT165554, MT165535, MT165536, MT165555, MT165558, MT165559, MT165561, MT165548, MT165562, MT165560, MT165549, MT165563, MT165564, MT165537, MT165538, MT165565, MT165539, MT165540, MT165566, MT165570, MT165569, MT165567, MT165571, MT165568, MT165578, MT165550, MT165541, MT165542, MT165543, MT165572, MT165544, MT165545, MT165573, MT165546, MT165574, MT165556, MT165575, MT165576, and MT165577.

AUTHOR CONTRIBUTIONS

RF, EC, and DF conceived the project. AV, EM, CF, and RF designed the experiments. AV, EM, CF, SV, SD, LC, and SC performed the experiments. RF supervised the experiments. AV, EM, and RF analyzed the results. AV wrote the original draft of the manuscript. AV, SD, LC, SC, CF, EM, EC, and RF reviewed and edited the manuscript. All authors read and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01652/full#supplementary-material>

FIGURE S1–S30 | Phylogenetic trees.

TABLE S1 | List of bacteria isolated from roots of *Phragmites australis* and used in this work. Accession numbers of 16S rRNA gene partial sequences are reported.

TABLE S2 | MIC ($\mu\text{g/ml}$) values of bacterial isolates described in this work.

TABLE S3 | Growth of 294 bacterial isolates associated to *Phragmites australis* roots in TSB in presence of three different concentrations of synthetic wastewater (SWW).

REFERENCES

- Barac, T., Taghavi, S., Borremans, B., Provoost, A., Oeyen, L., Colpaert, J. V., et al. (2004). Engineered endophytic bacteria improve phytoremediation of water-soluble, volatile, organic pollutants. *Nat. Biotechnol.* 22, 583–588. doi: 10.1038/nbt960
- Çakmakçı, R., Dönmez, F., Aydın, A., and Şahin, F. (2006). Growth promotion of plants by plant growth-promoting rhizobacteria under greenhouse and two different field soil conditions. *Soil Biol. Biochem.* 38, 1482–1487. doi: 10.1016/j.soilbio.2005.09.019
- Chaudhry, Q., Blom-Zandstra, M., Gupta, S., and Joner, E. J. (2005). Utilising the synergy between plants and rhizosphere microorganisms to enhance breakdown of organic pollutants in the environment. *Environ. Sci. Pollut. Res. Int.* 12, 34–48. doi: 10.1065/espr2004.08.213
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244
- Coppini, E., Palli, L., Antal, A., Del Bubba, M., Miceli, E., Fani, R., et al. (2019). Design and start-up of a constructed wetland as tertiary treatment for landfill leachates. *Water Sci. Technol.* 79, 145–155. doi: 10.2166/wst.2019.030
- Deng, A., Zhang, G., Shi, N., Wu, J., Lu, F., and Wen, T. (2014). Secretory expression, functional characterization, and molecular genetic analysis of novel

- halo-solvent-tolerant protease from *Bacillus gibsonii*. *J. Microbiol. Biotechnol.* 24, 197–208. doi: 10.4014/jmb.1308.08094
- Di Cello, F., and Fani, R. (1996). A molecular strategy for the study of natural bacterial communities by PCR-based techniques. *Minerva Biotechnol.* 8, 126–134.
- Doty, S. L., Oakley, B., Xin, G., Kang, J. W., Singleton, G., Khan, Z., et al. (2009). Diazotrophic endophytes of native black cottonwood and willow. *Symbiosis* 47, 23–33. doi: 10.1007/BF03179967
- Duijff, B. J., Recorbet, G., Bakker, P. A., Loper, J. E., and Lemanceau, P. (1999). Microbial antagonism at the root level is involved in the suppression of fusarium wilt by the combination of nonpathogenic *Fusarium oxysporum* Fo47 and *Pseudomonas putida* WCS358. *Phytopathology* 89, 1073–1079. doi: 10.1094/PHYTO.1999.89.11.1073
- Germaine, K. J., Liu, X., Cabello, G. G., Hogan, J. P., Ryan, D., and Dowling, D. N. (2010). Bacterial endophyte-enhanced phytoremediation of the organochlorine herbicide 2,4-dichlorophenoxyacetic acid. *FEMS Microbiol. Ecol.* 57, 302–310. doi: 10.1111/j.1574-6941.2006.00121.x
- Glick, B. R. (2003). Phytoremediation: synergistic use of plants and bacteria to clean up the environment. *Biotechnol. Adv.* 21, 383–393. doi: 10.1016/s0734-9750(03)00055-7
- Glick, B. R., and Stearns, J. C. (2011). Making phytoremediation work better: maximizing a plant's growth potential in the midst of adversity. *Int. J. Phytoremediation* 13(Suppl. 1), 4–16. doi: 10.1080/15226514.2011.568533
- Glick, B. R., Todorovic, B., Czarny, J., Cheng, Z., and Mcconkey, B. (2007). Promotion of plant growth by bacterial ACC deaminase. *CRC Crit. Rev. Plant Sci.* 26, 227–242. doi: 10.1080/07352680701572966
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/ NT. *Nucleic Acids Symp. Ser.* 41, 95–98. doi: 10.12691/ajmr-2-6-8
- Hammer, O., Harper, D. A. T., and Ryan, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Paleontol. Electron.* 4, 1–9.
- Haridoim, P. R., van Overbeek, L. S., and Elsas, J. D. (2008). Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* 16, 463–471. doi: 10.1016/j.tim.2008.07.008
- He, Y., Langenhoff, A. A. M., Sutton, N. B., Rijnaarts, H. H. M., Blokland, M. H., Chen, F., et al. (2017). Metabolism of ibuprofen by *Phragmites australis*: uptake and phytodegradation. *Environ. Sci. Technol.* 51, 4576–4584. doi: 10.1021/acs.est.7b00458
- Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L., and Hedges, S. B. (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science* 293, 1129–1133. doi: 10.1126/science.1061457
- Ho, Y. N., Mathew, D. C., Hsiao, S. C., Shih, C. H., Chien, M. F., Chiang, H. M., et al. (2012). Selection and application of endophytic bacterium *Achromobacter xylosoxidans* Strain F3B for improving phytoremediation of phenolic pollutants. *J. Hazard. Mater.* 21, 43–49. doi: 10.1016/j.jhazmat.2012.03.035
- Hultman, J., Tamminen, M., Pärnänen, K., Cairns, J., Karkman, A., and Virta, M. (2018). Host range of antibiotic resistance genes in wastewater treatment plant influent and effluent. *FEMS Microbiol. Ecol.* 94:fy038. doi: 10.1093/femsec/fy038
- Karkman, A., Do, T. T., Walsh, F., and Virta, M. P. J. (2018). Antibiotic-resistance genes in waste water. *Trends Microbiol.* 26, 220–228. doi: 10.1016/j.tim.2017.09.005
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lugtenberg, B., and Kamilova, F. (2014). Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* 63, 541–556. doi: 10.1146/annurev.micro.62.081307.162918
- Manaia, C. M. (2017). Assessing the risk of antibiotic resistance transmission from the environment to humans: non-direct proportionality between abundance and risk. *Trends Microbiol.* 25, 173–181. doi: 10.1016/j.tim.2016.11.014
- Martinez, R., Jakob, F., Tu, R., Siegert, P., Maurer, K. H., and Schwaneberg, U. (2013). Increasing activity and thermal resistance of *Bacillus gibsonii* alkaline protease (BgAP) by directed evolution. *Biotechnol. Bioeng.* 10, 711–720. doi: 10.1002/bit.24766
- Patriquin, D. G., and Döbereiner, J. (1978). Light microscopy observations of tetrazolium-reducing bacteria in the endorhizosphere of maize and other grasses in Brazil. *Can. J. Microbiol.* 24, 734–742. doi: 10.1139/m78-122
- Pérez-García, A., Romero, D., and de Vicente, A. (2011). Plant protection and growth stimulation by microorganisms: biotechnological applications of *Bacilli* in agriculture. *Curr. Opin. Biotechnol.* 22, 187–193. doi: 10.1016/j.copbio.2010.12.003
- Puvanakrishnan, R., Sivasubramanian, S., and Hemalatha, T. (2019). *Microbes and Enzymes: Basics and Applied*. Chennai: MJP Publisher.
- Reinhold-Hurek, B., Maes, T., Gemmer, S., Van Montagu, M., and Hurek, T. (2006). An endoglucanase is involved in infection of rice roots by the not-cellulose-metabolizing *Endophyte azoarcus* sp. strain BH72. *Mol. Plant Microbe Interact.* 19, 181–188. doi: 10.1094/MPMI-19-0181
- Riva, V., Riva, F., Vergani, L., Crotti, E., Borin, S., and Mapelli, F. (2020). Microbial assisted phytodepuration for water reclamation: environmental benefits and threats. *Chemosphere* 241:124843. doi: 10.1016/j.chemosphere.2019.124843
- Ryan, R. P., Germaine, K., Franks, A., Ryan, D. J., and Dowling, D. N. (2008). Bacterial endophytes: recent developments and applications. *FEMS Microbiol. Lett.* 278, 1–9. doi: 10.1111/j.1574-6968.2007.00918.x
- Saxena, G., Purchase, D., Mulla, S. I., Saratale, G. D., and Bharagava, R. N. (2020). Phytoremediation of heavy metal-contaminated sites: eco-environmental concerns, field studies, sustainability issues, and future prospects. *Rev. Environ. Contam. Toxicol.* 249, 71–131. doi: 10.1007/398_2019_24
- Schlaeppli, K., and Bulgarelli, D. (2015). The plant microbiome at work. *Mol. Plant Microbe Interact.* 28, 212–217. doi: 10.1094/MPMI-10-14-0334-FI
- Selosse, M. A., and Le Tacon, F. (1998). The land flora: a phototroph-fungus partnership? *Trends Ecol. Evol.* 13, 15–20. doi: 10.1016/S0169-5347(97)01230-5
- Shaw, L. J., and Burns, R. G. (2004). Enhanced mineralization of [U-14C] 2,4-dichlorophenoxyacetic acid in soil from the rhizosphere of trifolium pratense. *Appl. Environ. Microbiol.* 70, 4766–4774. doi: 10.1128/AEM.70.8.4766-4774.2004
- Spaepen, S., Vanderleyden, J., and Remans, R. (2007). Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol. Rev.* 31, 425–448. doi: 10.1111/j.1574-6976.2007.00072.x
- Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., et al. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* 10, 427–436. doi: 10.1038/ismej.2015.124
- Stottmeister, U., Wiessner, A., Kusch, P., Kappelmeyer, U., Kästner, M., Bederski, O., et al. (2003). Effects of plants and microorganisms in constructed wetlands for wastewater treatment. *Biotechnol. Adv.* 22, 93–117. doi: 10.1016/j.biotechadv.2003.08.010
- Sturz, A. V., Christie, B. R., and Nowak, J. (2010). Bacterial endophytes: potential role in developing sustainable systems of crop production. *CRC Crit. Rev. Plant Sci.* 19, 1–30. doi: 10.1080/07352680091139169
- Tesar, M., Reichenauer, T. G., and Sessitsch, A. (2002). Bacterial rhizosphere populations of black poplar and herbal plants to be used for phytoremediation of diesel fuel. *Soil Biol. Biochem.* 34, 1883–1892. doi: 10.1016/S0038-0717(02)00202-X
- Vasquez, E. A., Glenn, E. P., Guntenspergen, G. R., Brown, J. J., and Nelson, S. G. (2006). Salt tolerance and osmotic adjustment of *Spartina alterniflora* (Poaceae) and the invasive *M. Haplotype* of *Phragmites australis* (Poaceae) along a salinity gradient. *Am. J. Bot.* 93, 1784–1790. doi: 10.3732/ajb.93.12.1784
- Vymazal, J. (2011). Constructed wetlands for wastewater treatment: five decades of experience. *Environ. Sci. Technol.* 45, 61–69. doi: 10.1021/es101403q
- Wales, A. D., and Davies, R. H. (2015). Co-selection of resistance to antibiotics, biocides and heavy metals, and its relevance to foodborne pathogens. *Antibiotics* 4, 567–604. doi: 10.3390/antibiotics4040567
- Wang, M., Zhang, D. Q., Dong, J. W., and Tan, S. K. (2017). Constructed wetlands for wastewater treatment in cold climate – A review. *J. Environ. Sci.* 57, 293–311. doi: 10.1016/j.jes.2016.12.019
- Wani, Z. A., Ashraf, N., Mohiuddin, T., and Riyaz-Ul-Hassan, S. (2015). Plant-endophyte symbiosis, an ecological perspective. *Appl. Microbiol. Biotechnol.* 99, 2955–2965. doi: 10.1007/s00253-015-6487-3
- Weller, D. M. (2007). *Pseudomonas* biocontrol agents of soilborne pathogens: looking back over 30 years. *Phytopathology* 97, 250–256. doi: 10.1094/PHYTO-97-2-0250

- Wiegand, I., Hilpert, K., and Hancock, R. E. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* 3, 163–175. doi: 10.1038/nprot.2007.521
- Wu, C., Lin, C., Zhu, X., Liu, H., Zhou, W., Lu, J., et al. (2018). The β -Lactamase Gene Profile And A Plasmid-Carrying Multiple Heavy Metal Resistance Genes of *Enterobacter cloacae*. *Int. J. Genomics* 2018:4989602. doi: 10.1155/2018/4989602
- Zeidler, D., Zähringer, U., Gerber, I., Dubery, I., Hartung, T., Bors, W., et al. (2004). Innate immunity in *Arabidopsis thaliana*: lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15811–15816. doi: 10.1073/pnas.0404536101
- Zhang, C., Wang, B., Dai, X., Li, S., Lu, G., and Zhou, Y. (2017). Structure and function of the bacterial communities during rhizoremediation of hexachlorobenzene in Constructed Wetlands. *Environ. Sci. Pollut. Res. Int.* 24, 11483–11492. doi: 10.1007/s11356-017-8463-1
- Zhou, Y., Tigane, T., Li, X., Truu, M., Truu, J., and Mander, U. (2013). Hexachlorobenzene dechlorination in constructed wetland mesocosms. *Water Res.* 47, 102–110. doi: 10.1016/j.watres.2012.09.030
- Zhu, Y., Zhang, W., Schwarz, S., Wang, C., Liu, W., Chen, F., et al. (2019). Characterization of a blaIMP-4-carrying plasmid from *Enterobacter cloacae* of swine origin. *J. Antimicrob. Chemother.* 74, 1799–1806. doi: 10.1093/jac/dkz107

Conflict of Interest: EC and DF were employed by G.I.D.A. SpA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Vassallo, Miceli, Fagorzi, Castronovo, Del Duca, Chioccioli, Venditto, Coppini, Fibbi and Fani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments

Elena Perrin^{1,7}, Veronica Ghini^{2,7}, Michele Giovannini¹, Francesca Di Patti^{3,4}, Barbara Cardazzo⁵, Lisa Carraro⁵, Camilla Fagorzi¹, Paola Turano⁶, Renato Fani¹ & Marco Fondi^{1,4}✉

It is commonly thought that when multiple carbon sources are available, bacteria metabolize them either sequentially (diauxic growth) or simultaneously (co-utilization). However, this view is mainly based on analyses in relatively simple laboratory settings. Here we show that a heterotrophic marine bacterium, *Pseudoalteromonas haloplanktis*, can use both strategies simultaneously when multiple possible nutrients are provided in the same growth experiment. The order of nutrient uptake is partially determined by the biomass yield that can be achieved when the same compounds are provided as single carbon sources. Using transcriptomics and time-resolved intracellular ¹H-¹³C NMR, we reveal specific pathways for utilization of various amino acids. Finally, theoretical modelling indicates that this metabolic phenotype, combining diauxie and co-utilization of substrates, is compatible with a tight regulation that allows the modulation of assimilatory pathways.

¹Department of Biology, University of Florence, Florence, Italy. ²Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Florence, Italy. ³Department of Physics and Astronomy, University of Florence, Florence, Italy. ⁴CSDC, University of Florence, Florence, Italy. ⁵Department of Comparative Biomedicine and Food Science, University of Padova, Padova, Italy. ⁶Center of Magnetic Resonance (CERM), University of Florence, Florence, Italy. ⁷These authors contributed equally: Elena Perrin, Veronica Ghini. ✉email: marco.fondi@unifi.it

Microorganisms must quickly and efficiently adapt to a variety of possible fluctuations in the surrounding environment. When considering changes in the pool of available nutrients, this is usually achieved by a tight regulation of their metabolic phenotypes by sensing the availability of specific compounds, synthesizing the enzymes required for their catabolism and repressing them after specific metabolites are depleted¹. The spectrum of possible bacterial metabolic adaptation strategies can be observed, for example, when growing cells in a medium containing a simple mixture of carbon sources. In this situation bacterial may exhibit different patterns including diauxic growth², simultaneous consumption³, and bistable growth^{4,5}. Further, nutrients concentration and growth medium composition are known to affect other important cellular features such as motility and cell adhesion^{6,7} and biofilm formation^{8,9}. Typically, these phenomena have been studied (both theoretically and experimentally) in model organisms (e.g., *Escherichia coli* and *Lactococcus lactis*)¹⁰, grown on defined media containing simple mixtures of 2/3 carbohydrates, e.g., glucose and lactose^{11–14}. In natural conditions, however, bacteria rarely encounter simple combinations of exploitable carbon/energy sources. Rather, complex mixtures of nutrients are common and often colonized by actively growing bacteria. This is the case, for example, of intracellular pathogens that are commonly faced with a diverse set of host nutrients in infected tissues. In these cases, bacteria have been shown to adapt to this situation by the simultaneous exploitation of plastic and flexible nutrient utilization strategies^{15–19}.

Do the same models developed for simplified conditions hold also in real-case scenarios? At present, we witness a knowledge gap concerning the study of these processes in experimental settings that do not involve model organisms and/or defined media and we lack a sound theoretical understanding of the mechanisms driving nutrients assimilation strategies in conditions that are closer to the ones found in natural settings.

Bacterial exploitation of nutrient patches is made up of (at least) two different stages, i.e., physical interaction followed by carbon sources metabolic degradation. The capability of bacteria to interact with transient nutrient sources is well documented and has revealed their high efficiency in exploiting transient nutrient patches^{20,21}. Little is known, instead, on the molecular aspects regulating and influencing bacterial productivity once micro-scale nutrient hot spots are colonized. At this stage, i.e., when cells start to feed on the available carbon source(s), other cellular mechanisms need to be involved to ensure a systematic exploitation of the resource. Indeed, as nutrient patches are likely composed of complex nutrient mixtures (that may include carbohydrates, amino acids, lipids and nucleic acids) bacteria need to dynamically activate specific degradation pathways according to the kind and concentration of external nutrients. In other words, a continuous and flexible genetic reprogramming needs to be active to ensure that the preferred compound(s) are sequentially or simultaneously taken up from the external environment and properly metabolized. Up to now, this latter aspect has been mostly overlooked despite it might be central in the understanding of micro-scale nutrients dynamics.

In this regard, the marine environment represents a paradigmatic example of the challenges encountered by microorganisms when it comes to the efficient (and rapid) exploitation of complex nutritional inputs. Such a habitat is thought to be characterized by a low average nutrient level (e.g., the concentration of amino acids is in the range of $\sim 10^{-9}$ M) and nutrients in general appear and disappear in a sporadic fashion, demanding a precise chemical response, a fast swimming speed, and ability to localize and exploit a nutrient patch once it is found⁶. These are the conditions that are commonly faced by marine heterotrophic bacteria,

i.e., those microorganisms relying on the assimilation of external biomass for both energy generation and nutrition. Their metabolism is pivotal for the maintenance and the correct balance of oceanic biogeochemical cycles as they are central to the so-called microbial loop, i.e., the trophic pathway of the marine food web responsible for the microbial assimilation of dissolved organic matter, by transforming phytoplankton-derived organic matter and fuelling the entire ocean biogeochemical nutrient cycle. In this regard, the metabolism of *Pseudoalteromonas haloplanktis* TAC125 (PhTAC125), a heterotrophic marine bacterium isolated from Antarctica, has recently gained a certain attention due to its potential biotechnological exploitation²², its capability to synthesize anti-biofilm compounds²³, the necessity to set up efficient culture conditions^{24,25} and the metabolic reprogramming during growth in complex environments²⁶. In particular, the analysis of its growth phenotype in an amino acid rich medium has shown the presence of metabolic switches among different groups of amino acids²⁵, although nothing could be said about the molecular mechanisms underlying such phenotype and the possible regulation involved. Using constraint-based metabolic modelling we attempted to provide a systems level scheme of PhTAC125 metabolic re-wiring as a consequence of carbon source switching in such a nutritionally complex medium. Our simulations highlighted an efficient reprogramming of PhTAC125 metabolic machinery to quickly adapt to a nutritionally unstable environment, compatible with adaptation to fast growth in a highly competitive environment²⁶.

Here we have investigated the global regulation of a marine heterotrophic bacterium when grown in both a complex and a defined rich medium (i.e., including multiple possible carbon sources) using and integrating a set of complementary-omics techniques (i.e., transcriptomics and ¹H and ¹³C NMR metabolomics) with measured growth parameters. We show that the two main nutritional strategies commonly observed (co-utilization and sequential uptake of multiple substrates) can coexist in the same growth experiment, leading to an efficient exploitation of the available carbon sources. We also developed two theoretical models accounting for nutrients switching in a nutritionally rich environment in the presence and absence of cell regulation acting at the level of resource allocation in the synthesis of nutrient assimilation pathways. We show that a model taking into consideration an overall regulatory control on the sequence of nutrients uptake produces a better fit with available experimental data with respect to a purely Michaelis–Menten kinetic model.

Results

Global regulation of a triauxic growth. *P. haloplanktis* TAC125 (hereinafter PhTAC125) cells were grown in shaken flasks in a complex medium composed of Schatz salts²⁷ and peptone as their C source. Optical density (OD) was measured every hour and cellular RNA was sampled in five different time points of their growth (Fig. 1a). To increase the time points for a better growth rate estimation, we also used an interpolation technique on the data generated at this stage. Details on the specific interpolation approach used are reported in “Methods” and the resulting plot is reported in Supplementary Fig. 1. The growth curve displays a triauxic pattern (Fig. 1a). An initial growth phase (growth rate of 0.023 h^{-1} , 0.021 considering interpolated data) is interrupted by a lag phase between min. 180 and min. 240; afterwards, cells start growing over but such growth is interrupted by another lag phase between min. 280 and min. 340. Cells then started growing again, until the end of the experiment (growth rate of 0.004 h^{-1} , 0.006 h^{-1} considering interpolated data). The average growth rate across all the time points was estimated to be 0.01 h^{-1} . To identify transcriptional changes during cell growth total RNA was extracted and

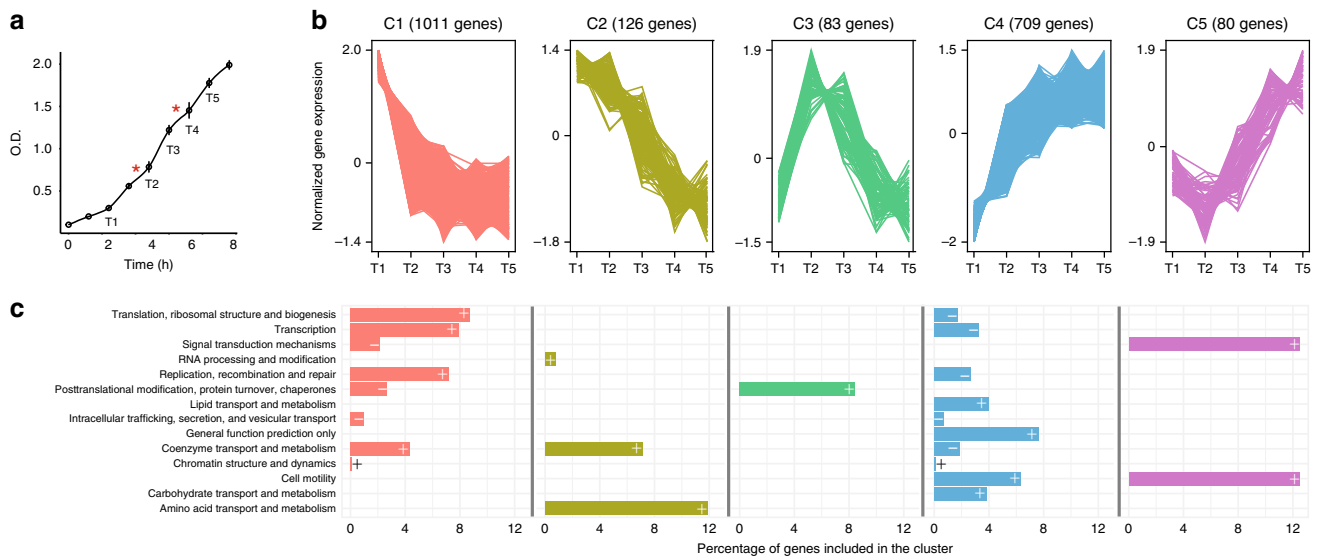


Fig. 1 Transcriptomic data of the multi-auxic growth. **a** PhTAC125 growth curves in peptone. Asterisks represent hypothetical switch points and T1–T5 indicate the sampled time points. **b** Clustering of PhTAC125 genes according to their expression values (normalized according to *clust* algorithm, see “Methods”) throughout the growth curve. **c** COG functional categories that are significantly over- (“+” sign) or under- (“–” sign) represented inside each cluster. Source data are provided as a Source Data file.

sequenced (two biological and two technical replicates for a total of 20 samples) using Illumina MiSeq (Genomix4Life, Naples, Italy). The main features of the 20 sequenced samples are reported in Supplementary Table 1. We clustered the genes according to their expression during the growth and identified six major trends (clusters C1–C6, Fig. 1b). Overall, we were able to cluster 2045 genes out of the 3448 encoded by the PhTAC125 genome (roughly 60%). We then performed a functional annotation and a functional enrichment analysis for the genes embedded in each cluster. One of these clusters (C6) did not include any significantly enriched functional category and thus it was discarded. Cluster C1 includes genes that display a decrease in their expression between the first two time points (T1 and T2) and a constant (low) expression across the rest of the growth curve.

Over-represented genes embedded in this cluster included those involved in basic housekeeping functions such as translation, DNA replication and transcription (Fig. 1c). Genes embedded in Cluster C2 displayed a decreasing trend throughout the growth curve and mainly included genes involved in RNA processing, metabolism of coenzymes and amino acids transport and metabolism. The expression of the 83 genes included in Cluster C3 was characterized by an abrupt increase between T1 and T2 and then an overall decrease until the end of the curve. This cluster significantly included genes involved in post-translational modification, protein turnover and chaperons. Clusters C4 and C5 included genes whose expression tended to increase in the later stages of the growth; over-represented genes in C4 mainly belonged to lipid metabolism, cell motility and amino acids transport and metabolism. The expression of genes included in C5 decreased during the first stages of the growth and is then increased for the rest part of the curve. The cluster of genes included those involved in signal transduction mechanisms and cell motility.

Whole-genome transcriptomics data depict a scenario in which PhTAC125 is active and fast-growing mainly during the first stages of the curve, as reflected by the relatively high expression of translation, transcription replication and coenzyme metabolism genes. Genes embedded in these categories are under-represented among those increasing their expression in the last stages of the growth (Fig. 1c) and over-represented among those with high

expression values in the first stages of the growth. Metabolically, PhTAC125 cells seem to rely more on amino acids metabolism in the initial stages of their growth, consistently with their progressive exhaustion in the medium. The last part of the growth experiment was also characterized by an increase in gene expression of cell motility-related genes (over-represented in C4 and C5). Finally, genes generally related to post-processing mechanisms peak their expression at T2.

A non-*E. coli*-like regulatory response to nutrients exhaustion.

The triauxic growth curve reported in Fig. 1a (and in Supplementary Fig. 1 using interpolated data) suggests the presence of a dynamic control on the adjustment of cell physiology. Here we sought to quantify the regulatory effort required to growing cells for modulating such cellular response. We focused on transcriptional factors (TFs) and two-component response systems (TCRSs) and analyzed differentially expressed genes among three points of PhTAC125 growth curve, namely T1 vs. T3 and T3 vs. T5. These points should capture PhTAC125 cells during exponential growth after the first growth lag (T1), in-between the two growth lags (T3) and after the final growth lag but before getting to plateau (T5).

First, we checked whether PhTAC125 regulation system somehow resembled the model scheme of the known overall metabolic regulation (i.e., the one characterized in *E. coli*). Of the 81 transcription factors known to directly or indirectly control central metabolic enzymes²⁸, we found a reliable homologue (E -value $< 1e^{-20}$) only for 34 of them (Supplementary Note 1 and Supplementary Data 1). PhTAC125, for example, lacks key players in bacterial diauxic shifts as the major global regulator of catabolite-sensitive operons (when complexed to cAMP) *crp* and the genes responsible for the synthesis of cyclic AMP (adenylate cyclase, *cyaA*). Among the 34 global regulators identified, only ten (roughly 25% of the shared ones and 12% of the entire *E. coli* set) displayed a significantly altered expression following the first transition (T1–T3) and none of them was differentially expressed following the second one (T3–T5). Details on the shared, differentially expressed TFs are provided in Supplementary Table 2.

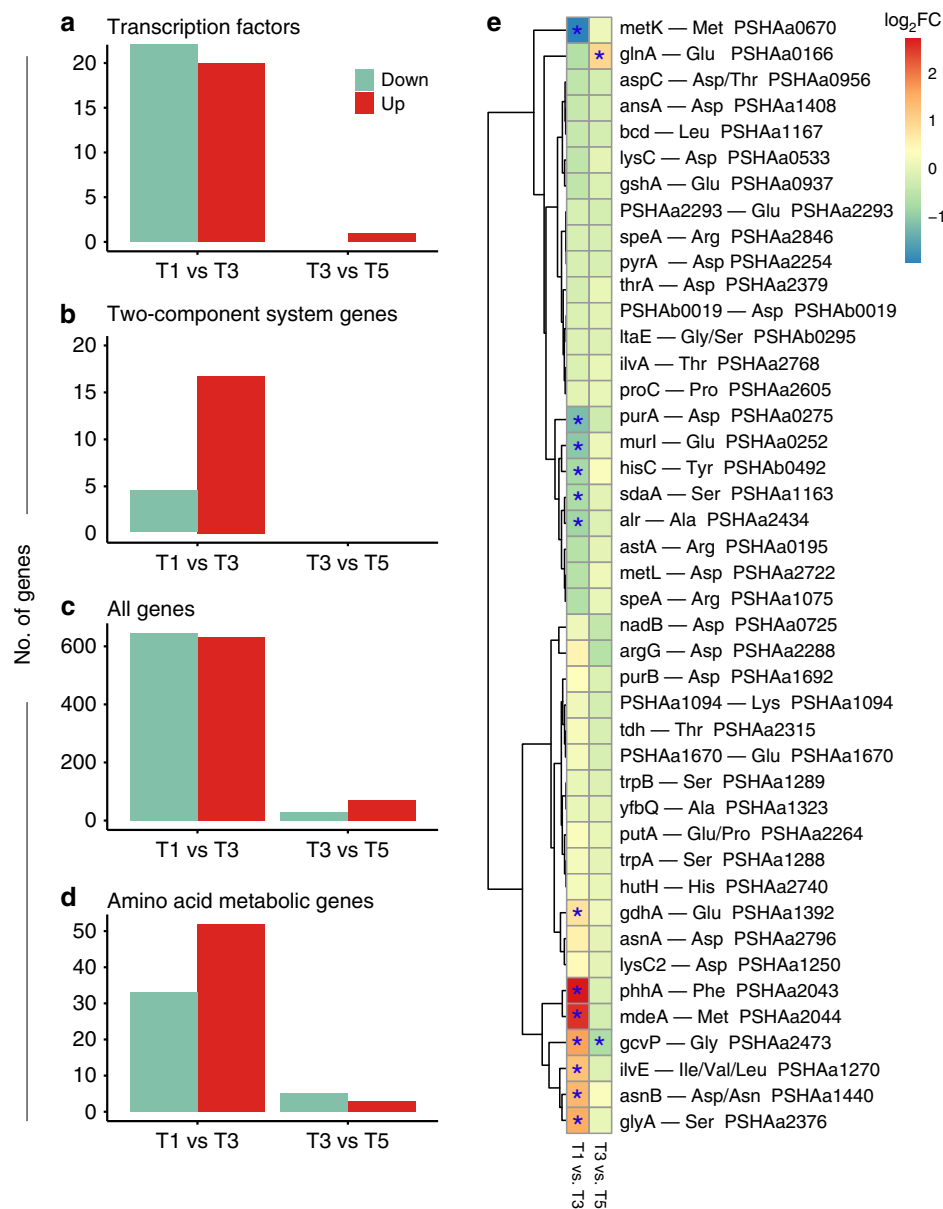


Fig. 2 Differentially expressed genes. a–d Number of up- and downregulated genes in the two contrasts considered, for different gene categories. **e** Fold change of each gene responsible for the first degradative step of each amino acid. Differentially expressed genes are marked with an asterisk. Source data are provided as a Source Data file.

A similar situation was observed for eight selected sigma-factors that control gene expression globally²⁸. In this case, as expected, an ortholog was found for each of them, but only two of them showed an altered expression following T1–T3 transition and the expression of none of them was significantly altered following T3–T5 one. The two genes displaying a significant change in gene expression were *rpoS* and *rpoD*. RpoS is the primary regulator of stationary phase genes, whereas RpoD is the primary sigma factor during exponential growth. Expectedly, the first resulted to be upregulated following the T1–T3 transition, whereas the second was downregulated (Supplementary Table 3).

With the exception of RpoS and RpoD, whose expression is in line with the global control of exponential vs. stationary phases, it appears that growth lags are regulated by mechanisms that poorly overlap with our current knowledge.

For this reason, we evaluated the expression of the entire repertoire of PhTAC125 TFs across the two points that involved the ceasing of cellular growth in our experiment, i.e., T1–T3 and

T3–T5. Overall, we identified 41 differentially expressed TFs, 22 downregulated and 19 upregulated (Fig. 2b, Supplementary Fig. 2 and Supplementary Data 2) following the first growth interruption. The second growth lag was characterized by the significant change in expression (upregulation) of just one TF. Together with TFs, TCRSs are a basic stimulus-response coupling mechanism to sense and react to changes in environmental conditions, e.g., nutrient concentration. We identified differentially expressed TCRSs in the two selected contrasts. Overall, we found 21 TCRSs-related genes that were differentially expressed in T1 vs. T3 and none in the T3 vs. T5 transition (Fig. 2c and Supplementary Data 3).

We conclude that the two growth lags observed in the curve apparently point to different reprogramming efforts that, in turn, may underpin distinct nutrients uptake strategies. The first growth interruption seems to have a deeper impact on the entire metabolic system, whereas the second could imply a fine tuning of the catabolic machinery. This is further confirmed by the

overall number of DEGs across the selected contrasts (Fig. 2a), 1280 between T1 and T3 (633 and 647 up- and downregulated genes, respectively) and only 101 between T3 and T5.

Amino acid assimilation pathways and their (dis)regulation.

Previous experiments have shown that PhTAC125 displays a coordinated sequence of amino acids degradation when grown in medium embedding complex mixtures of such molecules (i.e., peptone or casamino acid-based media, Supplementary Note 2)²⁵. In other words, some amino acids are preferred over others and are metabolized early in PhTAC125 growth curve. This switching among nutrients suggests that an active and modulated reprogramming occurs during PhTAC125 growth in a nutritionally complex environment. Similarly, differentially expressed amino acid metabolic genes (hereinafter AA-genes) are unevenly distributed among the two contrasts considered (Fig. 2d). T1 vs. T3 displays a higher number of DEGs (85, 33 downregulated and 55 upregulated) with respect to T3 vs. T5 (8, 5 downregulated and 3 upregulated). Considering the amino acid assimilation pathways of differentially expressed AA-genes in the T1 vs. T3 contrast (Supplementary Fig. 3), we did not observe a clear functional bias towards specific routes. Almost all the pathways are represented, both in terms of up- and downregulated genes. Similarly, the switch between T3 and T5 included downregulated genes involved in a broad spectrum of metabolic pathways including Val, Leu and Ile degradation, Tyr metabolism and Ala, Asp and Glu metabolism (one gene for each pathway) and upregulated genes in Gly, Ser and Thr metabolism (1 gene), Lys biosynthesis (1 gene) and Arg biosynthesis (3 genes).

To unravel the faith of each amino acid inside the cell, we analyzed the expression of the genes involved in all their possible first assimilatory step (Fig. 2e and Supplementary Data 4). We found 13 DEGs in the T1–T3 contrasts and 2 DEGs in the T3–T5 contrast. The first set comprised seven upregulated and six downregulated genes; upregulated genes were involved in Glu, Phe, Met, Gly Ile/Val/Leu, Asn/Asp and Ser degradation, whereas downregulated genes were responsible for the first assimilatory step of Met, Asp, Glu, Tyr, Ser and Ala. DEGs identified in the second contrast included genes involved in Glu and Gly degradation. Taking the DEGs indicated above as a proxy for the entire assimilatory process of the corresponding amino acids, we noticed a good overlap with available PhTAC125 physiological data²⁵.

DEGs analysis also allowed the identification of the major amino acid entry points into PhTAC125 metabolism. We counted, for example, 12 alternative possibilities steps to metabolize Asp in PhTAC125 (Fig. 2e), but the expression of only two of them (*purA* and *asnB*) appeared to be significantly modified during PhTAC125 growth. Similarly, six alternative steps can convert Glu to other cellular intermediates following its uptake. At T1, only two of these genes showed an altered expression level, suggesting that these may represent the most relevant players in Glu assimilation and usage. Nearly the same holds for Ser, with five distinct entry points and only two of them being differentially regulated.

Overall, we have identified possible key players, both in the switch among the set of metabolized amino acids, and in the entrance of amino acids into PhTAC125 entire metabolic network. However, nutrients switching requires an efficient genetic regulation to ensure that each catabolic pathway is active at the right moment, allowing a correct proteome allocation. For this reason, we analyzed the co-expression of genes belonging to the same metabolic pathway (Supplementary Note 3) and identified an overall dis-regulation of such genes (average Fisher's Z transformation average of Pearson correlation coefficient 0.49,

Fig. 3a, b, Supplementary Fig. 4 and Supplementary Table 4). Focusing on the known regulons including AA-genes (Supplementary Table 5), we noticed that nearly half of them (three out of seven) displayed a relatively low (0.47, ArgR) or almost absent (0.26 and 0.11, MetJ and TyrR1, respectively) correlation among the expression values of the corresponding genes (Table 1). Figure 3c–e summarizes the details of the correlation existing among each gene of each pathway. In the case of ArgR regulon, for example, the major contribution to the low intra-regulon correlation is due to *astA* and *astD* (PSHAa0195 and 0196, respectively), showing an almost opposite expression pattern compared with the other ArgR regulated genes, especially with PSHAa2287-91 (Fig. 3C). *astA* and *astD* are involved in the conversion of Arg to Glu, whereas ArgHA, B, C, F, G (encoded by PSHAa2287-91, respectively) are involved in the synthesis of Arg from Glu through the formation of citrulline and fumarate.

Concerning MetJ regulon, we noticed a group of genes (including PSHAa2222, PSHAa2223, PSHAa0287 and PSHAa2292) whose expression values are negatively correlated with those of genes PSHAa2274-76 and PSHAa1226 (Fig. 3d). This group of co-regulated genes include those involved in the conversion of homocysteine to Met (PSHAa2222 and PSHAa2223), an L-alanine-DL-glutamate epimerase (PSHAa0287) and a Methylthioribulose-1-phosphate dehydratase involved in the Met salvage pathway (PSHAa2292). Finally, as for TyrR1 regulon, PSHAa2042-43, coding for 4a-hydroxytetrahydrobiopterin dehydratase and phenylalanine-4-hydroxylase are negatively correlated to the other genes in the same regulon (Fig. 3e). PSHAa2043 encodes *phhA* the gene responsible for the synthesis of Tyr from Phe, whereas PSHAa2042 (*phhB*) encodes a Pterin-4- α -carbinolamine dehydratase responsible for the conversion of 4a-hydroxytetrahydrobiopterin to dihydrobiopterin.

Upstream of most of the genes belonging to the three regulons considered, we were able to identify a conserved motif for each regulon (Fig. 3f–h), partially overlapping with their known TF binding site. Finally, a closer inspection to the metabolic steps encoded by the differentially regulated genes of these regulons revealed that they usually belong to different and symmetric regions of the same metabolic pathway. PSHAa0195 and PSHAa0196 respectively encodes for *astA* and *astB*, responsible for the first steps of the route leading to the formation of Glu from Arg. The other genes of the ArgR regulon are mostly involved in the production of Arg starting from Glu (Fig. 3i). Similarly, *phhA* (encoded by PSHAa2043) is involved in the formation of Tyr (from Phe), whereas all the other genes are responsible for the formation of Tyr from a set of different precursors (e.g., prephenate) (Fig. 3k). In the case of MetR regulon, among the genes that could be reliably assigned to the methionine metabolic pathway, one of the two group of co-regulated genes belong to the upper part of the pathway (upstream the main product methionine), whereas members the other one are in its close proximity (PSHAa2222) or involved in the methionine salvage pathway (PSHAa2292), the set of reactions responsible for the recycle of the thiomethyl group of S-adenosylmethionine from methylthioadenosine (Fig. 3j).

In a previous work²⁶, we have simulated the growth of PhTAC125 in a nutritionally complex environment (peptone) and derived the overall metabolic reprogramming occurring during growth in a rich undefined medium using constraint-based metabolic modelling (i.e., flux balance analysis, FBA) (Supplementary Note 4). We observed a good correlation between the measured changes in the expression of AA gene regulons and the predicted changes in metabolic fluxes of their encoded reactions using FBA (Pearson's product-moment correlation = 0.89, p value = 0.019). See Supplementary Table 6 for further details.

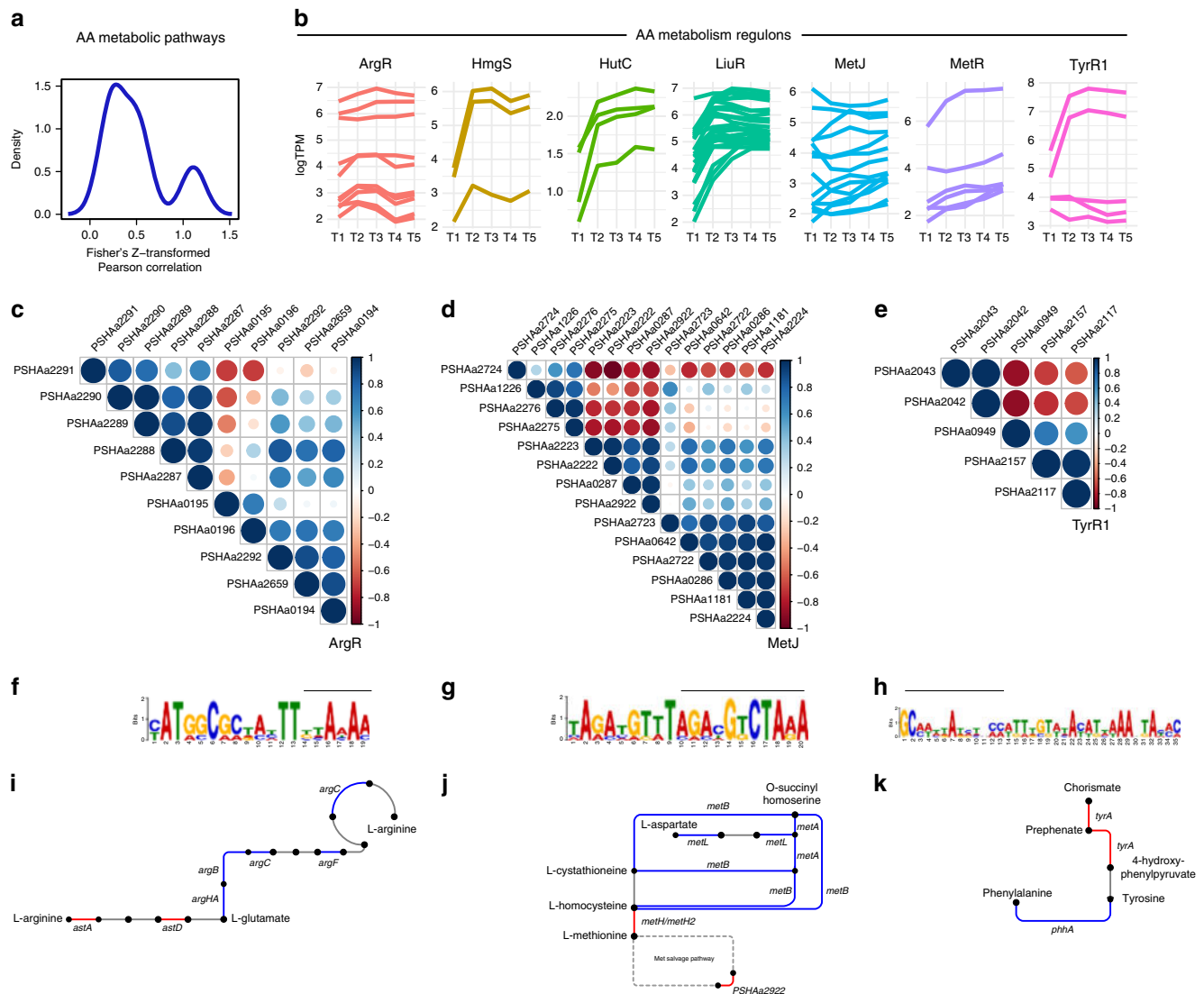


Fig. 3 Deregulation of amino acid metabolic pathways. **a** Cumulative density plot of Fisher's Z transformation average of Pearson correlation expressing the co-regulation of genes belonging to the same pathway. **b** Log₂-transformed TPM values of the genes belonging to the same regulon. **c–e** Graphic representation of the Pearson correlation matrix for the genes belonging to the regulons showing the lowest average correlation, Arginine metabolism/biosynthesis, Methionine biosynthesis and Aromatic amino acids metabolism. Locus tags in bold indicate those of the regulator of the corresponding regulon. **f–h** Show the conserved upstream motif found for each gene of the considered regulon. The bar above each conserved motif indicates the overlap existing with the known TF binding site according to the RegPrecise database. Finally, in (**i–k**) the reactions encoded by the genes included in the same regulon are schematically represented. Red and blue lines schematically represent the correlation coefficient of Fig. 3c, d. Source data are provided as a Source Data file.

Taken together these results suggest that (i) the two growth lags observed (Fig. 1a) may be the same phenotypic representation of two different cellular states (i.e., assimilation strategies) and that (ii) a rather complex genetic regulation is at work to ensure a correct decision-making process in nutritionally dynamic environments. In the next sections these two aspects will be elucidated using controlled growth conditions and a combination of NMR experiments and theoretical modelling.

Combination of simultaneous and sequential amino acids uptake. Up to now, we have analyzed the behaviour of bacterial cells in a complex medium, using gene expression as a proxy for amino acids assimilation pathways. The medium used (peptone) is a complex mixture of nutrients whose exact composition is unknown.

Accordingly, it is not possible to conclude that the observed growth features (i.e., triaxial growth) are due to the exhaustion of certain preferred amino acids in the medium. For this reason, we assembled a medium including 19 amino acids, 0.2 mM each (named 19 AA medium, cysteine was not included in the list because of difficulties in its unambiguous quantitation during the experiments due to its spontaneous oxidation, as also reported in ref. 29) and determined the kinetics of their usage during PhTAC125 growth by analyzing the growing media using ¹H NMR (Supplementary Note 5 and Supplementary Fig. 5). Data obtained revealed that an important fraction of all the provided amino acids (16 out of 19) are consumed in the first 7 h of the growth. Afterwards, the remaining three amino acids (His, Met and Trp) are (slowly) metabolized (Fig. 4a). Clustering the amino acids assimilation profiles allowed a clearer visualization of the order in which amino acids are used by PhTAC125 during its

growth (Fig. 4b). This analysis divided the set of metabolized compounds into four main, non-overlapping clusters. Gln, Glu and Arg are the first amino acids to be consumed in the medium. Their concentration reaches (negligible) values close to 0.01 mM after 4.5 h of growth, remaining constant afterwards. The second set of amino acids is composed of Asn, Asp, Leu and Pro. Their degradation starts with a small delay with respect to the one of the first cluster, and they are completely removed from the

medium only between 6 and 6.5 h. The third cluster includes nine amino acids (Fig. 4b). Their consumption is rather slow in the first 3 h of growth; afterwards, it accelerates leading to negligible concentration of the corresponding amino acids at 7.5 h. The concentration of amino acids belonging to the fourth cluster remains overall constant for the first 6 h of growth. After that moment, corresponding to the point in which all the other amino acids are consumed, it starts decreasing. Importantly, this pattern of amino acids assimilation results in a triauxic growth curve (Fig. 4c). Indeed, (short) growth lag phases are observed after 4 and 6 h of growth, in correspondence with the major transition in amino acids assimilation pattern.

Overall, this behaviour highlights a balanced mix between simultaneous and sequential uptake of nutrients. Amino acids belonging to the same group (Fig. 4d) are simultaneously metabolized by the cells but the assimilation of different groups occurs with different dynamics and is responsible for growth lags in the curve. Finally, a typical diauxic nutrient shift is observed when all the main (preferred) sources are exhausted and the degradation of the other (previously ignored) compounds begins. Available growth phenotypes^{24,30} (Supplementary Note 6) seem to suggest that the order in which nutrients are used by PhTaC125 during the growth depends both on the final biomass and on the specific growth rate achievable when grown with amino acids as sole carbon sources (Supplementary Fig. 6A, B). Also, the order of amino acids uptake, partially reflects their entry point into the TCA cycle (Supplementary Fig. 7).

Table 1 Fisher's Z transformation average of Pearson correlation of the same-regulon genes.

Regulon	Process	Fisher's Z transformation of Pearson correlation coefficient
HutC	His degradation	2.55
HmgS	Tyr degradation	1.92
MetR	Met biosynthesis	1.22
LiuR	Branched-chain amino acid degradation	0.97
ArgR	Arg biosynthesis/ degradation	0.47
MetJ	Met metabolism, Met degradation	0.26
TyrR1	Aromatic amino acids metabolism	0.11

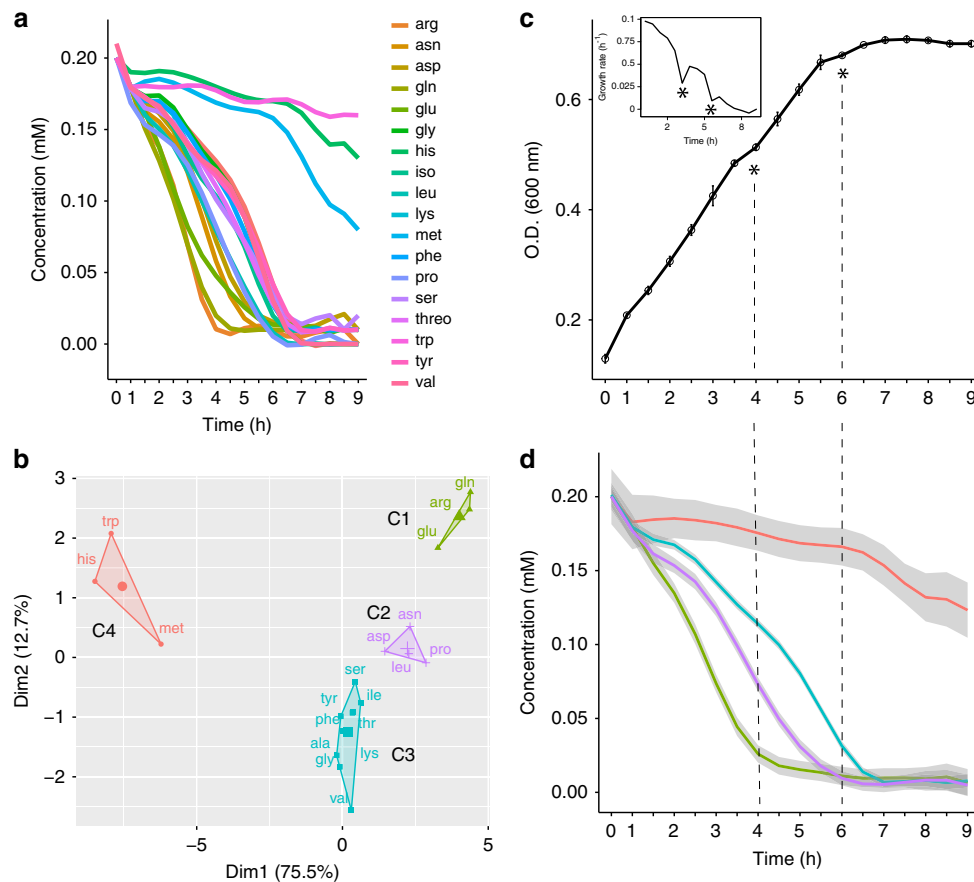


Fig. 4 Amino acids assimilation profiles. **a** Degradation dynamics for each of the 19 amino acids included in the defined AA medium. **b** Clustering of time-resolved concentration values for the 19 amino acids analyzed. Error bars represent SD of two different cell cultures in two independent experiments. **c** Growth curve of the 19 amino acids experiment. Asterisks indicate growth lags. **d** Degradation dynamics for each of the four identified clusters of amino acids included in the defined AA medium. Colour codes as in **b**. Grey shaded area includes the 95% confidence of the linear regression (coloured) line over the concentrations of the amino acids belonging to the same group. Source data are provided as a Source Data file.

Despite the trend seems to be quite clear, a certain variability was observed in the results shown in Supplementary Fig. 6A, B. This is the case, for example, of Leu in cluster C2 that is preferred to Ala (C3) despite the growth rate of PhTAC125 growth on Leu is about twofold lower than that on Ala. This might be accounted for by promiscuous uptake of nutrients. Indeed, the broad spectrum of some amino acids transporters is well described in bacteria^{31,32}. Thus, some amino acids might be taken up from the medium not as the result of an active cellular control over the most efficient carbon sources, but as the result of the broad-spectrum activity of a membrane transporter.

An additional explanation to the composition of the clusters might involve the overlap among the catabolic pathways of the substrate that are co-consumed (i.e., belonging to the same cluster, Supplementary Note 7). Indeed, we noticed that C1 amino acids (Arg, Glu and Gln) share part of their catabolic pathways and are all converted to alpha-ketoglutarate before entering the TCA cycle. Two (out of four, i.e., Asp and Asn) amino acids of cluster C2 are converted to oxaloacetate before entering the TCA cycle. Most of the C3 amino acids (six out of nine, i.e., Ala, Ser, Gly, Thr, Phe and Tyr) are converted to pyruvate, thus not being directly catabolized into one of the TCA cycle intermediates. Other three C3 amino acids (Thr, Ile and Val) can be catabolized to form pyruvate and a TCA intermediate (i.e., Succinyl-CoA). Finally, C4 includes those amino acids that are negligibly used during PhTAC125 growth, so no degradation pathways overlap is required to explain their inclusion in the same group (Supplementary Fig. 7).

In order to check whether the sequential or co-consumption of substrates also depends on their own concentrations, we performed a set of additional experiments aimed at evaluating the effect of higher concentrations of nutrients on the pattern of nutrients assimilation (Supplementary Note 8). We both tested the effects of (i) providing all the 19 amino acids at concentrations five and ten times higher (1 and 2 mM, respectively) than those used in the 19 AA experiment described above and (ii) increasing the concentration of late-metabolized amino acids (His, Met, Trp, 2 mM) on the timing of the last metabolic switch observed during PhTAC125 growth. The results of these analyses are reported in (Supplementary Figs. 8 and 9) and revealed that metabolic phenotype identified in the original 19 AA experiment is poorly affected by the concentration of the available carbon sources. Also, we compared the phenotype observed in PhTAC125 with the one of a model organism (*E. coli*, Supplementary Note 9), when grown in the same nutritional environment (19 AA medium, 0.2 mM each). A poor overlap was observed between the specific response to nutrients switching of the two microorganisms (Supplementary Figs. 10 and 11).

The different fate of the catabolized amino acids. Next, we investigated the different fate of amino acids belonging to the same cluster once entered inside PhTAC125 cells. To this aim, we used uniformly labelled ¹³C amino acids (Glu, Asp, Ala and Met, belonging to C1, C2, C3 and C4, respectively) and followed the path of their labelled carbon atoms inside the cells by NMR. More in detail, for each labelled amino acid used, we prepared four different parallel cultures, each containing 18 amino acids plus the ¹³C labelled one as the only carbon source for PhTAC125. These four cultures were run in parallel and each of them was sacrificed at a different time point. Specifically, we analyzed four time points, i.e., early and late exponential growth (3 and 6 h) and early and late stationary phase (8 h and 30 min, 24 h). For each time point we analyzed both growing media and the cell lysates by acquiring mono-dimensional ¹H NMR spectra and bidimensional ¹H-¹³C HSQC NMR spectra. A scheme of the structure of

this experiment is reported in Supplementary Fig. 12. This experiment allowed us to study the metabolic fluxes in a time-dependent fashion and provided hints on the fate of the metabolized amino acids inside the cell (Fig. 5).

First, by using ¹H NMR spectra of growing media, we confirmed the same growth dynamics for all the replicates of the experiment and the same overall growth features observed in the unlabelled 19 AA experiment (compare Fig. 5a, c). Similarly, amino acids were consumed in the same order and with the same overall rates previously observed (compare Figs. 5b and 4a).

A principal component analysis (PCA) performed on the ¹H spectra of cell lysates clustered the samples according to their sampling time, thus confirming the consistency and the high reproducibility among the different replicates (Supplementary Fig. 13), besides the occurrence of significantly different metabolic profiles of PhTAC125 cells along the four sampled time points.

¹H-¹³C NMR spectra of cell lysates were used to have an overview of the fate of the metabolized amino acids inside the cells. Each of the four selected amino acids showed a specific path of labelled carbon atoms inside the cells, as revealed by the PCA reported in Fig. 4c (¹H-¹³C HSQC NMR spectra). While the spectra of Asp and Glu almost overlap, the spectra of Ala and Met are very well separated. This trend probably reflects the fact that both Glu and Asp are used to directly feed the TCA cycle, whereas Met and Ala are used to feed different pathways inside the cell (see below).

In particular, we were able to show that Ala is mostly used to feed many important pathways inside the cell, namely glycolysis/gluconeogenesis, nucleotide precursors metabolism and (partly) TCA cycle. At T3 we found ¹³C coming from Ala degradation in most of the key compounds that are the input/output of the aforementioned pathways, i.e., pyruvate, phosphoenolpyruvate (PEP), fatty acids and AXP/GXP (Fig. 5d, Supplementary Figs. 14 and 15). Oxaloacetate is the only TCA cycle intermediate that displays Ala labelling at T3 (Fig. 5d and Supplementary Fig. 16). At T3, we found Asp-derived ¹³C labelling on all the TCA cycle intermediates identified and on purine biosynthesis intermediates, AXP and GXP (Fig. 5d, Supplementary Figs. 15 and 16). At later growth stages Asp labelling appeared also on PEP and NAD precursors (Fig. 5d, Supplementary Figs. 14 and 17). Glu-derived ¹³C labelling appears in all the TCA intermediates analyzed (with the exception of oxaloacetate) starting from T3 (Fig. 5d and Supplementary Fig. 16), suggesting that Glu is readily redirected towards the TCA upon its uptake. Interestingly, Glu seems also to be used as a substrate for NAD (and precursors) biosynthesis from the early stage of growth (Fig. 5D and Supplementary Fig. 17). Finally, this experiment confirmed that Met is incorporated into the PhTAC125 metabolic network at a later stage of its growth, in that no identified compounds was labelled with ¹³C of Met at T3. Met labelled ¹³C appear to be included into homocysteine starting from T6. After 24 h, we found Met labelled ¹³C on a compound whose NMR pattern corresponds to that of trimethylamine (Fig. 5d and Supplementary Fig. 18). Interestingly, in a previous work³³, we had characterized PhTAC125 as a methylamine producer and also showed that adding Met to the growth medium was pivotal to allow PhTAC125 to produce this compound (and inhibit the growth of human opportunistic pathogens). The metabolic flux analysis performed here suggests a link between the production of methylamine from the degradation of Met, possibly through the formation of trimethylamine.

Overall, using ¹³C-labelled precursors provided evidence that the different amino acids used by PhTAC125 during its growth have different and complementary roles within the cell. Ala is readily converted into pyruvate and this is used to feed both gluconeogenic pathways and (possibly through the formation of

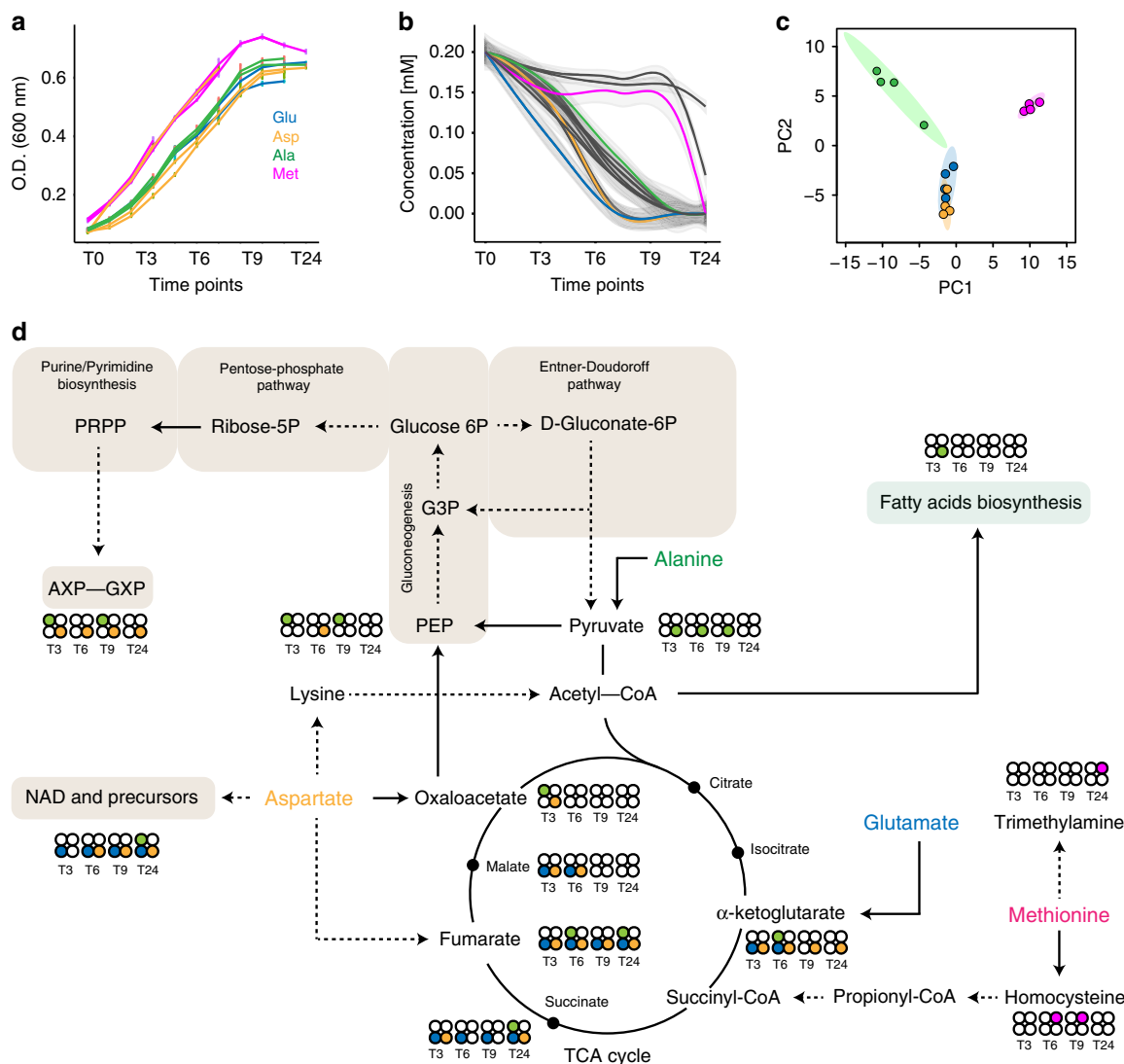


Fig. 5 The fate of catabolized amino acids. **a** PhTAC125 growth curves across all replicates in the 19 AA medium with one of them being ¹³C labelled. Each colour indicates which was the ¹³C-labelled amino acids for each replicate. Error bars represent SD of four different cell cultures. **b** Extracellular concentration of all the amino acids in time. Colour codes as in **a**. **c** PCA on the ¹³C spectra. Colour codes as in **a**. Grey shaded area includes the 95% confidence of the linear regression (coloured) line over the concentrations of the amino acids belonging to the same group. **d** Central PhTAC125 metabolic network, highlighting the possible entry points of the labelled amino acids. Circles close to compounds indicate the ¹³C signal recovered at the different time points (T3–T24) and coming from different amino acids (colour codes as in **a**) corresponding. Source data are provided as a Source Data file.

acetyl-CoA) fatty acids synthesis. Only at later growth stages, TCA intermediates started to display Ala-derived labelling. Conversely, Glu is immediately used to feed the TCA cycle as all its intermediates but one (oxaloacetate) carry Glu labelling at T3. No other compounds in the upper part of Fig. 5d displayed Glu-derived ¹³C atoms, thus suggesting a clear separation between Ala and Glu assimilation pathways. The primary assimilation pathway of Asp seems to be through the TCA cycle, although part of its carbon skeleton is diverted to PEP formation (and possibly to gluconeogenesis) at T6. Quite interestingly, ¹³C atoms derived from Asp degradation are observed on purine metabolism product AXP and GXP from the first sampled time point to the last, despite the gluconeogenic metabolism precursor PEP does not display the same signal. We thus argue that the origin of Asp-derived labelling on AXP and GXP might be due to TCA intermediates playing a role in purine metabolism, e.g., 5-phosphoribosylamine. Finally, we confirmed the late entrance of Met into PhTAC125 metabolic network, being initially converted to homocysteine and then also rerouted towards the production

of trimethylamine (at T24), with a possible but still undisclosed pathway involving the formation of betaine or carnitine^{34–36}.

To further characterize the metabolic response of PhTAC125 to nutrients consumption, the whole intracellular metabolome was evaluated through untargeted ¹H NMR across four time points, i.e., early and late exponential growth (3 and 6 h) and early and late stationary phase (8 h and 30 min, 24 h). In the spectra, the signals of 26 metabolites could be unambiguously assigned and quantified (Supplementary Fig. 19 and 20). These metabolites were representative of six major metabolic modules, i.e., amino acids metabolism, purine and pyrimidine metabolism, sugars, amino sugars metabolism, nucleotide precursors and TCA. The trend of metabolites' concentration in time revealed that the relative concentration of purine and pyrimidines intermediates (namely uridine and inosine) decreases in time. A similar trend is observed for ribose and for the detected TCA intermediate (fumarate). This scenario is compatible with the decrease in growth rate and overall biomass production of PhTAC125 along the growth curve (Fig. 4c) and with a consequent decrease of DNA

synthesis activity (purine and pyrimidine precursors) and energy demand (TCA cycle). Conversely, sugars (e.g., glucose) and amino sugars (e.g., UDP-NAG) increase their cellular concentration in the latest growth stages, reaching values that are up to three times the initial ones. This latter finding can be interpreted in the light of two considerations: first, we have already shown that carbohydrates metabolic genes increase their expression at the final stages of PhTAC125 growth on peptones, suggesting the activation of sugar metabolism-related pathways at later growth stages. Second, the increase of sugar/amino sugars intracellular concentrations might have a role in PhTAC125 cell aggregation when nutrients concentration starts to deplete given that: (i) such metabolites are known to be involved in cell–cell contacts³⁷, (ii) PhTAC125 was shown to produce a biofilm that incorporates amino sugars and (iii) this is supposed to be a strategy to survive in poor nutrient conditions³⁷. Our findings are in line with a scenario in which the presence of a lower availability of nutrients induces a greater production of biofilm since the biofilm matrix can improve the capture of nutrients³⁸.

Modelling simultaneous and sequential amino acids uptake. At least two different explanations may account for the mixed sequential/diauxic nutrient uptake. Either this phenotype is “simply” determined by different uptake kinetics of the different compounds or the assimilation pattern is actively regulated by the cells. To discern between these two scenarios, we implemented two mathematical models accounting for cell growth and nutrients uptake during the 19 AA experiment. To reduce the complexity of the problem, the 19 amino acids were lumped into the four corresponding clusters shown in Fig. 4b, d. In this way we modelled the growth of PhTAC125 in a hypothetical growth medium embedding four different groups of carbon sources ideally representing the 19 AA medium. The first model is based on the Michaelis–Menten–Monod kinetics (MMM model) and is formulated as follows:

$$P + S_i \xrightarrow{r} 2P, \quad (1)$$

where P represents bacterial cells, S_i (with $i = 1, 2, 3, 4$) represents each of the four groups of pooled C sources and r the rate at which the reaction occurs. Specifically, r was modelled according to a canonical (Michaelis–Menten-derived) Monod kinetics with:

$$r = \frac{\beta_i \cdot \phi S_i}{k_i + \phi S_i}, \quad (2)$$

where β_i , ϕS_i and k_i represent the maximum rate constant for cell production, the concentration and the Michaelis–Menten constant for the i th group of amino acids, respectively. According to these formulas, the state variables model can be written as:

$$\frac{d\phi P}{dt} = \sum_{i=1}^4 \frac{\beta_i \cdot \phi S_i^2}{k_i + \phi S_i} \cdot \phi P - d \cdot \phi P, \quad (3)$$

$$\frac{d\phi S_i}{dt} = -\frac{\beta_i \cdot \phi S_i^2}{k_i + \phi S_i} \cdot \phi P, \quad (4)$$

where S_i (with $i = 1, 2, 3, 4$) represents each of the four lumped substrates and d the bacterial cells death rate.

The second model implemented here accounts for the effect of the regulatory processes of catabolite inhibition and activation that can be observed during microbial growth on multiple substrates. The model is a modified version of the cybernetic model proposed by ref. ³⁹, overall resembling the one proposed in ref. ⁴⁰.

The cybernetic modelling framework takes into account the (yet) unknown regulatory processes regulating the microorganisms' uptake kinetics. It assumes that microorganisms have evolved under the selective pressure to become optimal with respect to certain cellular objectives (in our case, maximization of

biomass production) and achieve this task by actively modulating the induction/repression and activation/inhibition of the key enzymes of substrates available in their external environment. Cybernetic variables (see below) are introduced in the model to account for the induction/repression and activation/inhibition of the key, bottleneck enzymes regulating cell growth, substrate consumption, and key enzyme production^{41,42}.

According to this model, the assimilation of substrate S_i by cells P (Eq. (1)) is assumed to be catalyzed by the set of enzymes E_i (with $i = 1, 2, 3, 4$). The assumption here is that enzymes responsible for the assimilatory pathway of each pool of nutrients are induced by the presence of S_i (and repressed by the presence of the other nutrients). This alternative model can be written as:

$$P \xrightarrow{r_i v_i} 2P, \quad (5)$$

$$\emptyset \xrightarrow{r_{E_i} u_i} E_i, \quad (6)$$

where E_i represents the key assimilatory enzyme. The rate equations for biomass production (Eq. (5)) and for enzyme synthesis (Eq. (6)) can be written as a modified form of Monod's equation and are respectively expressed as follows:

$$r_i = V_{max,i} \cdot \phi E_i \cdot \frac{\phi S_i}{K_{S_i} + \phi S_i}, \quad (7)$$

$$r_{E_i} = V_i \cdot \frac{\phi S_i}{K_{E_i} + \phi S_i}, \quad (8)$$

where ϕE_i represents enzyme concentration, V_{E_i} is the maximum rate constant for enzyme's biosynthesis and $V_{max,i}$ is the maximum rate constant for bacterial production P on the i th substrate. K_{S_i} and K_{E_i} are the Michaelis–Menten constant for the i th substrate and the synthesis of the i th enzyme, respectively.

The inhibition/activation effect due to the concentration of the different substrates is accounted for by two (control) variables, u_i and v_i , representing the fractional allocation of resource for the synthesis of E_i and the mechanism of controlling enzymes E_i activity, respectively. u_i is expressed as:

$$u_i = \frac{r_i}{\sum_{j=1}^4 r_j}, \quad (9)$$

with $0 \leq u_i \leq 1$ and $\sum_{j=1}^4 u_j = 1$. The other control parameter, v_i is expressed as:

$$v_i = \frac{r_i}{\max\{r_1, r_2, r_3, r_4\}}, \quad (10)$$

where the denominator accounts for the observation that priority is given to the consumption of the substrate(s) that guarantee the highest growth rate⁴³. The model further considers constitutive enzyme production rate (β_i), the effect of dilution of the specific enzyme level due to cell growth (α), constant protein decay in the cells and bacterial death rate (d) and can be written as follows:

$$\frac{d\phi S_i}{dt} = -r_i \cdot v_i \cdot \phi P, \quad (11)$$

$$\frac{d\phi P}{dt} = \sum_{j=1}^4 v_j \cdot r_j \cdot \phi P - d \cdot \phi P, \quad (12)$$

$$\frac{d\phi E_i}{dt} = u_i \cdot r_{E_i} - \sum_{j=1}^4 v_j \cdot r_j \cdot \phi P - \alpha \cdot \phi E_i + \beta_i. \quad (13)$$

The model parameters were determined by fitting the experimental data (shown in Fig. 4a–d) with model simulations and their values are reported in Supplementary Tables 7 and 8. As

shown in Fig. 6, the cybernetic model accurately reproduces the dynamics of all the species considered. This is even clearer in the case of nutrient concentration dynamics where the model implementing the MMM model, is not capable of producing a satisfactory approximation of the real data. Indeed, R^2 calculation indicates that the cybernetic modelling framework is performing better on four out of five of the species included in the model (Table 2). Likewise, when using the AIC metric, the comparison between cybernetic model and MMM was shown to follow the same trend.

These results suggest that, in the conditions tested, the uptake of nutrients is tightly regulated, leading to the simultaneous presence of diauxic and co-utilization strategies within the same growth curve. Hints on the main catabolic players involved in such assimilation patterns were obtained combining transcriptomic data from the complex-medium experiment and RT-PCR on specific targets (see Supplementary Note 10, Supplementary Fig. 21 and Supplementary Table 9).

Discussion

Our knowledge on the possible bacterial strategies for nutrients assimilation when multiple sources are available is biased by the fact that it has been mainly studied in a few model organisms, providing them with a reduced number of possible inputs (compared with those available in their source environment). Here, we have characterized a non-model response to nutrients switching and studied the process of bacterial nutrients uptake in experimental conditions that more closely resemble a natural setting, in terms of the availability of many different substrates simultaneously. Using a marine heterotrophic bacterium (*P. haloplanktis* TAC125) as a case study strain, we have shown that its response during growth lags do not resemble the one currently known for *E. coli*. Only 10% of the *E. coli* metabolic regulators and two (out of eight) main generic controllers (*rpoS* and *rpoD*) displayed an altered expression level in our experiments. Also, we showed that when the two microorganisms were independently cultivated in the same defined medium embedding 19 different amino acids, differences arose in the choice of the amino acids to utilize, in the timing of uptake and in the presence/absence of overall growth lags.

The poor overlap between PhTAC125 and *E. coli* transcriptional response suggests that, in marine bacteria, the response to nutritional switches and/or multi-axenic growth patterns may involve still untapped genetic circuits. As a matter of fact, PhTAC125 is known to lack the CCR system⁴⁴, which is currently referred to as the main driver in metabolic switches and diauxic phenotypes. Using time-resolved transcriptomics we have shown that growth lags in a nutritionally complex environment are probably due to the exhaustion of specific carbon sources and that such event has a deep impact also on other important gene categories including, for example, motility-related genes. In the second part of the curve, i.e., when nutrient concentration decreases, cell motility genes increase their expression, probably reflecting the need to explore the surrounding environment for other potential sources. This is in line with the observation that

many bacteria become motile when nutrients are scarce⁴⁵. Moreover, among the genes peaking their expression in correspondence of the first growth lag, those involved in post-translational modification, protein turnover and chaperon are

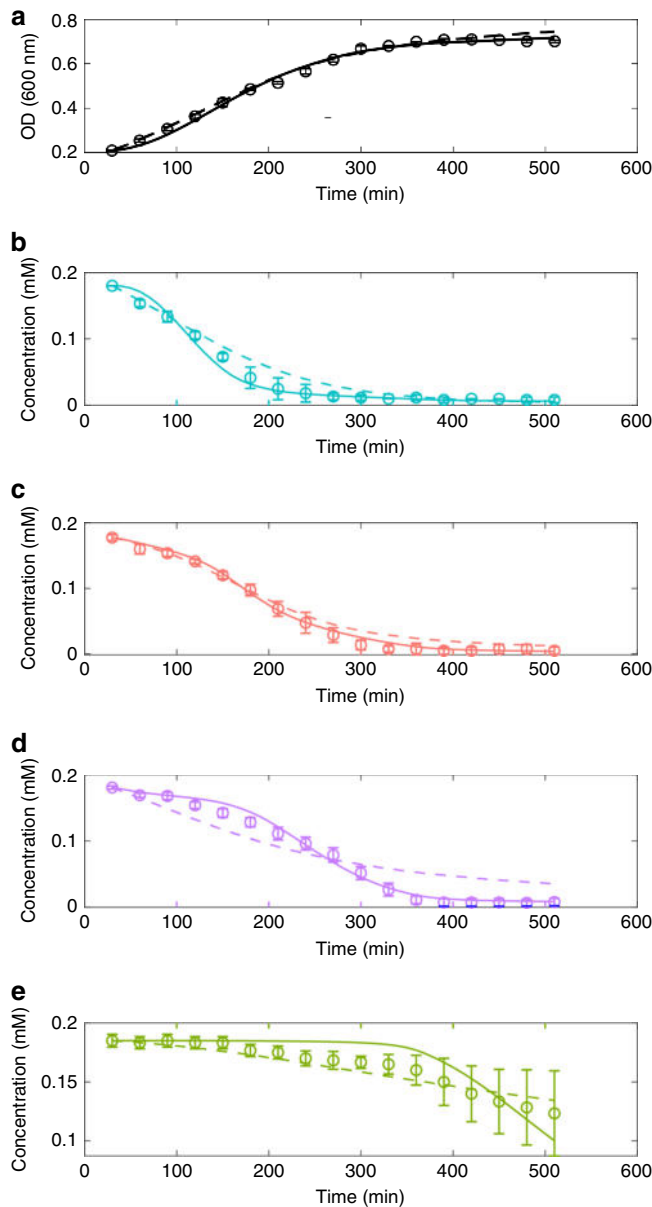


Fig. 6 Modelling the multi-axenic growth. Simulation outcomes for the two models implemented here for biomass (a) and lumped nutrients (b–e). Dashed lines represent the prediction for the MMM model. Solid lines represent the prediction for the cybernetic model. Open circles represent experimental data (same data shown in Fig. 4d). Error bars represent SD of two different cell cultures in two independent experiments.

Table 2 R^2 and AIC calculation for the different simulations using the two different models.

Model	Cluster A	Cluster B	Cluster C	Cluster D	Biomass
Pearson—MMM model	0.9844	0.9922	0.9584	0.9584	0.9959
Pearson—Cybernetic	0.9920	0.9973	0.9953	0.9193	0.9967
AIC—MMM model	−408.53	−418.46	−338.17	−476.82	−378.46
AIC—Cybernetic	−439.48	−475.38	−438.75	−395.97	−372.37

significantly over-represented. Generally speaking, proteins belonging to this functional category can be easily associated to the stress encountered by an exponentially growing batch culture that exhaust (part of) the readily available and preferred nutrients in the growth medium. At this moment cells undergo a regulated transition into stasis by activating a stereotypic stress response. Post-translational modifications have been shown to play a role in the starvation-induced growth arrest, for example in *S. coelicolor*^{28,46}. It is to be noticed that these proteins do not increase their expression in the later stages of the growth and, in particular, in the second growth arrest experienced by PhTAC125. This, together with the observation that both the overall number of DEGs and of those involved in many other functional processes (amino acids degradation, TFs and TCRS) is much higher in correspondence of the first time contrast, suggested that the same phenotype (the two growth lags) were mirrored by profoundly different cellular reprogramming patterns. This prompted us to investigate more in depth the regulation of amino acid assimilation pathways. This analysis highlighted the tight regulation required to efficiently exploit complex and amino acid enriched, nutritional conditions. A paradigmatic example of this capability is the (dis)regulation of several genes belonging to the same regulon (ArgR, TyrR1 and MetR), responsible for the activity of distinct and conflicting functional metabolic modules inside the same metabolic pathway. In principle, all the genes belonging to the same regulon are under the control the same TF. It is known that some TFs may function as either activators or repressors, often according to the positioning of the TF binding site in the target promoter, although this feature had not been described, to date, for amino acid assimilation pathways⁴⁷. Apparently, such a mechanism is at work in some of the amino acid catabolic pathways of PhTAC125, probably ensuring an efficient and correct exploitation of the amino acid mixture available in the surrounding environment.

Despite being closer to the environmental natural setting, growth in complex medium does not allow a precise understanding of the usage of all the available carbon sources. For this reason, we have assembled a defined but nutritionally rich medium containing 19 amino acids as the sole carbon sources for the bacterial cells. Tracking their concentration in time, we showed that the two main feeding strategies commonly thought to be exclusive to each other (i.e., sequential and simultaneous uptake of nutrients) can coexist in the same growth curve. Clustering metabolized amino acids into four major groups revealed that amino acids belonging to the same group are co-utilized, whereas the switch among the different clusters is tightly modulated. A “canonical” diauxic shift is finally observed at the end of the growth, when the consumption of a set of previously untapped nutrients begins. Thus, similar to the complex-medium growth experiment (Fig. 1), the two growth lags apparently underlay different cellular states. Indeed, despite the exhaustion of nutrients is common to both growth lags, in one case (the first lag) cells are already metabolizing alternative compounds when the exhaustion of the preferred sources occurs.

The order in which nutrients are utilized can be partially explained by the biomass yield and growth rates obtained when each single amino acid is provided as single carbon sources. Those amino acids allowing the highest growth rate and biomass production are those that are consumed first in the 19 AA experiment.

Using ¹³C-labelled amino acids and time-resolved NMR spectra, we were able to follow the degradation pathways of four selected amino acids (Glu, Ala, Asp and Met) and derive their fate inside the cell. We showed that Ala is readily converted into pyruvate and then probably used to fuel sugar metabolism (i.e.,

gluconeogenesis and pentose phosphate pathway), leading to the production of ribonucleotides. Asp and Glu are instead promptly used to fuel the TCA cycle, with two notable exceptions, i.e., the entrance of Glu-derived carbons into the biosynthesis of NAD precursors and the conversion of (part of) the initial amount of Asp into PEP and its usage for ribonucleotide biosynthesis. The fate of Met inside the metabolic network remained hard to decipher, and the actual contribution to the growth of PhTAC125 will require further investigation. Further, results coming from the untargeted evaluation of the overall intracellular metabolome is in line with an increased importance of sugar metabolism/intermediates upon exhaustion of the available amino acids in the medium, probably reflected in the characterized production of biofilm in poor nutrients conditions.

Finally, we have shown that the dynamics of nutrients degradation can be explained using a theoretical model that accounts for gene regulation and, in general, for the proper resource allocation for the synthesis of the main assimilatory pathways. This modelling framework can accurately interpret the pattern of nutrients degradation in a nutritionally rich environment.

In conclusion, we would like to stress the importance of cultivating and studying microorganisms in nutritional conditions that more closely resemble the ones most found in nature, for example for what concerns the contemporary availability of many distinct possible carbon sources as done here. By doing so and using a combination of computational and experimental (transcriptomics and NMR-based metabolomics) approaches, we have shown that, despite diauxic and co-utilization strategies have been usually thought as conflicting phenotypes, they can coexist in the same growth curve and give rise to a diversified ensemble of feeding strategies.

The use of different sources depending on the phase of cell growth and, most of all, a distinct metabolic fate inside the cell for each of the metabolized compounds, is a common feature of intracellular bacteria (e.g., *Legionella pneumophila*, *Listeria monocytogenes* or *Coxiella burnetii*^{17,19,48–50}) suggesting that plastic strategies for carbon assimilation might be evolved in response to nutritionally poor and highly variable conditions.

It will be interesting to investigate which are the molecular mechanisms allowing the implementation of this mixed and apparently unconventional feeding strategy and, in particular, the fine-tuned regulatory circuits that are probably responsible for the efficient switching among all the available carbon sources. Future efforts will be also devoted to understanding the effect of fluctuations (in the number of cells and/or in nutrients concentration) and of possible population heterogeneity^{10,11} on the resulting growth dynamics of heterotrophic marine bacteria.

Methods

Bacterial strain, media and growth condition. *P. haloplanktis* TAC125⁵¹ cells were routinely grown in Marine Agar (MA) or Broth (MB) (Condalab, Spain) under aerobic condition at 21 °C. The stock suspension of the strain was stored in 20% [v/v] glycerol solution at –80 °C. For growth curves experiment, Schatz salts⁵² (1 g/l KH₂PO₄, 1 g/l NH₄NO₃, 10 g/l NaCl, 0.2 g/l MgSO₄ × 7H₂O, 0.01 g/l FeSO₄ × 7H₂O, 0.01 g/l CaCl₂ × 2H₂O) were supplemented with 5 g/l Peptone N-Z-Soy BL 7 (Sigma-Aldrich S.r.l) (complex medium) or with 19 amino acids (19 AA medium) each one at a final concentration of 0.2 mM (cysteine was not included due to its rapid oxidation to cystin²⁹). To confirm the order of AA we used Schatz salts supplemented with (i) 19 AA at a final concentration of 1 mM each, (ii) 19 AA at a final concentration of 2 mM each and (iii) 16 AA at a final concentration of 0.2 mM and histidine, methionine and tryptophan at a final concentration of 2 mM. The experiments with ¹³C AA were performed using Schatz salts supplemented with 18 standard AA and one of the four marked amino acid (L-Alanine-¹³C₃, L-Aspartic acid-¹³C₄, L-Glutamic acid-¹³C₅ and L-Methionine-¹³C₅) all at a final concentration of 0.2 mM. In all cases the pH was adjusted to 7.0. All the amino acids were purchased from Sigma-Aldrich S.r.l. All the experiments were performed under aerobic condition at 21 °C.

E. coli Dh5α (laboratory stock suspension stored in 20% [v/v] glycerol solution at –80 °C) were grown in Luria Bertani (LB)⁵³, agar or broth, and in M9 media⁵³

supplemented with 19 AA at a final concentration of 0.2 mM (pH 7.0) (19 AA M9 medium). The experiments were performed under aerobic condition at 37 °C.

Growth curve experiments. The growth curve experiments with PhTAC125 were performed after two pre-cultures, as in Wilmes et al.²⁵ with some adaptations. For the complex medium growth curves, a first preculture was grown for 20 h, in 20 ml MB medium in a 100 ml flask. Then this preculture was diluted 1:1000 in a final volume of 100 ml of the complex medium in a 1 l flask. After 20 h of growth, the optical density (OD₆₀₀) of this second preculture was measured to be used to inoculate the final flask (1 l) in a final volume of 150 ml of Schatz salts and Peptone with a starting OD₆₀₀ ~ 0.1. For the growth curves in the 19 AA medium, for all the concentrations used, the first preculture in MB was diluted 1:100 in a final volume of 100 ml of the 19 AA medium (for each experiment the AA concentration used for the final growth curves was used) in a 1 l flask. After 22 h of growth, the second preculture was washed, resuspend and used to inoculate the final flask (1 l) in a final volume of 200 ml of the 19 AA medium with a starting OD₆₀₀ ~ 0.1. In all experiments, the pre-cultures and the final growth cultures were incubated at 21 °C with shaking. Each experiment was performed in duplicate. Cell growth was monitored measuring the OD₆₀₀ every hour in the experiments in complex medium, and every half an hour or an hour, depending on the concentration of AA used, in the experiments with the 19 AA medium. Three different measures were performed at each time point for each biological replicate.

The growth curve experiments with *E. coli* were performed after a first preculture of 20 h in 20 ml of LB broth in a 100 ml flask and a second preculture in 100 ml of 19 AA M9 medium.

After 20 h of growth, the optical density (OD₆₀₀) was measured to be used to inoculate the final flask (1 l) in a final volume of 200 ml of 19 AA M9 medium. The cells were grown at 37 °C with shaking in duplicate and the growth was monitored measuring the OD₆₀₀ every hour.

Sampling. Two biological replicates of the growth curves performed in complex medium were used for RNA-seq experiment. Every hour, in correspondence of the OD₆₀₀ measurements, two replicates of 500 µl each for each curve, were treated with the RNA protect bacteria reagent (Qiagen S.r.l.) and conserved at -80 °C.

During the experiment in the 0.2 mM 19 AA medium, at each time point, two replicates of 500 µl each for each curve, were treated with the RNA protect bacteria reagent like above. During all the experiment in the 19 AA medium, regardless of the concentration of AA used, and in the growth curve with *E. coli*, two replicates of 1 ml each for each curve were filtered at each time point (Filtropur 0.2 µm, SARSTED AG & Co. KG) to remove bacterial cells and conserved at -20 °C for NMR metabolomic.

Growth curves with ¹³C amino acids. Four different experiments were performed using one uniformly ¹³C-labelled amino acid each time (¹³C-Glu, ¹³C-Ala, ¹³C-Asp and ¹³C-Met). For each experiment, two pre-cultures as described above were used, while the final growth experiments were performed in quadruplicate. Cell growth were monitored measuring the OD₆₀₀ every hour. At four time points, early and late exponential growth (3 and 6 h) and early and late stationary phase (8 h and 30 min, 24 h), one of the four replicates was analyzed. Overall 1 ml of the medium were filtered (Filtropur 0.2 µm, SARSTED AG & Co. KG) to remove bacterial cells and conserved at -20 °C. The remaining culture (199 ml), was pelleted by centrifuging for 10 min at 11,000 rpm at 4 °C and resuspended in 500 µl of PBS⁵⁴. Then, cells were sonicated for 20 min, with cycle of 1 s of activity and 9 s of rest (292.5 W, 13 mm tip), with contemporary cooling on ice. After lysis, the samples were centrifugated for 25 min at 4 °C, at 8000 g, as described in ref. ⁵⁴.

RNA extraction and sequencing. For RNA-seq, a preliminary sequencing (data not shown) was performed on an Illumina HiSeq 50 platform (Genomix4Life S.r.l., Italy). Total RNA was extracted with a RNeasy Tissue Mini Kit (Qiagen S.r.l.) following manufacturer's instructions. For improving the lysis step proteinase k and lysozyme were added to the lysis solution and the samples were homogenized using Tissue lyser II (Qiagen S.r.l.). The concentration and purity of RNA were analyzed using a NanoDrop ND-1000 (Thermo Fisher Scientific) and a Bioanalyzer (Agilent Technologies, Inc.). rRNA was removed from the sample using the Ribo-Zero Magnetic Kit (Bacteria) (Illumina, Inc.). The quality of the RNA depletion was then checked using Bioanalyzer (Agilent RNA 6000 PICO Assay, Agilent Technologies, Inc.). The ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina, Inc.) is then used to make the RNA-Seq library from the Ribo-Zero treated RNA. For each library 1 µg of RNA (rRNA depleted) was used following manufacturer's instructions. The quality of the libraries was evaluated using Bioanalyzer (Agilent Technologies, Inc.).

For the final experiment total RNA was extracted from a total of 20 samples of the growth curve in complex medium (five time points, two technical and two biological replicates) and library sequencing has been carried out at Genomix4Life S.r.l. (Italy) on an Illumina NextSeq500 (single-end sequencing strategy, 1 × 75 bp, ~25 reads/sample).

For real-time PCR (qRT-PCR), total RNA was extracted from the samples of five time points (T4, T6, T8, T10 and T12) for each biological replicate of the growth curve performed in the 19 AA medium, using a RNeasy Mini Kit (Qiagen S.

r.l.), following the manufacturer's instruction. DNA was then removed from the samples using a RNase-free DNase (Qiagen S.r.l.). Overall 10 µl of the extracted RNA was reverse-transcribed using a Superscript II Reverse Transcriptase (Invitrogen) with Random primers (Invitrogen) following the manufacturer's instruction.

Quantitative real-time PCR (qRT-PCR). qRT-PCR reactions were performed in a final volume of 10 µl containing 1 µl of a 1:10 dilution of each cDNA, 5 µl of Powrup Sybr Master Mix (Life Technologies) and 1 µM of each primer (Supplementary Table 9). Primers were designed using the Primer3 software⁵⁵. Each sample was spotted in triplicate.

A first experiment using known amounts of DNA of the PhTAC125 strain (1-0.1-0.01-0.001 ng) were performed to obtain a standard curve and calculate the amplification efficiency for each primer pairs (data not shown). *rplM* and *dnaA* genes were used as internal references to normalize mRNA content.

All the reactions were performed on a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems by Life Technologies). Cycling conditions were: hold stage [50 °C for 2' and 95 °C for 10'], PCR stage [40 cycles of: 95 °C for 30'', 59 °C for 1', 72 °C for 15''], melt curve stage [95 °C for 15'' 60 °C for 1', 95 °C for 15'].

RNA-seq data analysis. Bowtie 2 (v2.2.3)⁵⁶ was used to align raw reads to *P. haloplanktis* TAC125 reference genome (GCA_000026085.1_ASM2608v1). rRNA depletion, strand specificity and gene coverage were evaluated using BEDTools (v2.20.1)⁵⁷ and SAMtools (v0.1.19)⁵⁸ to verify the library preparation and sequencing performances. Raw read counts were then used to calculate TPM values for each PhTAC125 gene. Clusters of co-regulated genes were identified using the Clust tool⁵⁹ using the following parameters: k-means clustering method, tightness weight equal to 0.3 and Q3s outliers threshold equal to 2.0.

Differentially expressed genes between the various contrasts were identified using the R (R Development Core Team, 2012, <https://www.r-project.org/>) package DeSeq2⁶⁰ using default parameters and the following thresholds: adjusted *p* value < 0.01 and log₂FC > 0.75 or log₂FC < -0.75. The clustering of genes based on their FC was performed using the Pheatmap R package. Visualization of Pearson correlation was performed using "corplot" R package.

Functional enrichment analysis and regulon identification. To conduct functional enrichment, each gene whose upstream intergenic region was clustered in one of the three clusters was assigned to a specific functional category using a BLAST⁶¹ search against the COG database⁶², with default parameters and considering a hit as significant if *E*-value < 1e-20. The exact binomial test implemented in the R package was used to assess over- and under-represented functional categories against the corresponding genomic background. Available information on PhTAC125 amino acid metabolism regulons were retrieved using RegPrecise database^{63,64}. The RegPrecise includes information for 7 PhTAC125 regulons (Supplementary Table 5).

Motif finding. Shared, conserved upstream motifs were searched up to 200 bp upstream of the genes belonging to the same regulon. These sequences were retrieved were retrieved from the *P. haloplanktis* TAC125 reference genome and fed into the MEME suite⁶⁵ v. 5.1.1. MEME was used in combination with MAST (version 5.0.5)⁶⁶ for identifying the most plausible shared motifs upstream of the selected genes. MEME was used setting the following parameters: -nmotifs 5, -minw 6, -maxw 30, -objfun classic, -revcomp, -markov_order 0, -minsites 1, -maxsites 3. All the other parameters were set as default. MAST was used using default parameters. In all cases, only the best scoring motif was considered for further analyses, provided that the search produced a significant result (*e*-value < 0.05). The conservation of identified shared motifs was represented using WebLogo⁶⁷.

Modelling. The deterministic system was simulated by numerically integrating differential equations using the Matlab built-in function ode45 v. 2019a. To estimate the unknown parameters of the model from experimental data we used a stochastic curve-fitting in-house Matlab software. The algorithm is based on the paper by Cardoso et al.⁶⁸ and consists in the combination of the non-linear simplex and the simulated annealing approach to minimize the squared deviation function. To increase the points available for curve-fitting, we used the spline interpolation function implemented in MATLAB on the measured values of nutrients and biomass concentration. The same function was adopted to increase the points available for growth rate estimation during the experiment reported in Fig. 1a.

The codes used to perform the simulations reported in this work and the details about the options of the curve-fitting environment, are available at <https://multiauxic.sourceforge.io>.

To assess the quality of the fit of the MMM vs. the cybernetic model *R*² and AIC were computed. *R*² values between experimental data and model predictions were

computed using the built-in *cor.test* function in R⁶⁹. AIC values were computed as:

$$AIC = N \times \log\left(\frac{SSE}{N}\right) + 2 \times p.$$

With p being the number of parameters in the models (19 and 9 for cybernetic and MMM models, respectively), N being the number of training cases and SSE being the sum of squared errors for each training set. The interpolated datasets used to compute the fit with the models were also used when computing AIC (i.e., $N = 49$). The AIC values were computed using MATLAB 2019a.

Growth rates estimation on interpolated growth data. To increase the points available for curve-fitting, we used the spline interpolation function implemented in MATLAB on the measured values of biomass concentration (OD) obtained from growing PhTAC125 cells on peptone.

Metabolomic assay and data analysis. NMR spectra were acquired on (i) cell media to monitor the uptake of the various amino acids by measuring their levels in samples collected at different time points of cell growth; (ii) cell lysates to characterize the intracellular metabolome and its variation over time.

Medium samples were prepared in 5.00 mm NMR tubes by mixing 60 μ L of a potassium phosphate buffer (1.5 M K₂HPO₄, 100% (v/v) ²H₂O, 10 mM sodium trimethylsilyl [2,2,3,3-²H₄] propionate (TMSp), pH 7.4) and 540 μ L of sample. Cell lysate samples were prepared in 5.00 mm NMR tubes by mixing 60 μ L of ²H₂O and 540 μ L of samples.

Spectral acquisition and processing were performed according to procedures developed at CERM^{54,70–74}. All the spectra were recorded using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm PATXI ¹H-¹³C-¹⁵N and ²H-decoupling probe including a z axis gradient coil, an automatic tuning matching and an automatic and refrigerate sample changer (SampleJet). A BTO 2000 thermocouple served for temperature stabilization at the level of \sim 0.1 K at the sample. Before measurement, samples were kept for 5 min inside the NMR probe head, for temperature equilibration at 300 K.

For media, NMR spectra were acquired with water peak suppression and (i) one-dimensional (1D) ¹H standard NOESY pulse sequence⁷⁵ using 128 scans, 65,536 data points, a spectral width of 12,019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s and a mixing time of 0.01 s; (ii) two-dimensional (2D) ¹H-¹³C heteronuclear single quantum coherence spectroscopy (HSQC) pulse sequence (hsqcetgpsisp2, Bruker). A total of 80 scans were collected using a spectral width of 12,019 for f_2 and of 30,178 for f_1 , f_2 and f_1 acquisition time of 0.085 and 0.002 s, respectively, and a relaxation delay of 2 s.

For lysates, NMR spectra were acquired with water peak suppression and (i) 1D ¹H standard NOESY pulse sequence using 64 scans, 98,304 data points, a spectral width of 18,028 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s and a mixing time of 0.01; (ii) 1D ¹H Carr–Purcell–Meiboom–Gill sequence using 64 scans⁷⁶, 73,728 data points, a spectral width of 12,019 Hz, an acquisition time of 3.07 s and a relaxation delay of 4 s; (iii) 2D ¹H-¹³C HSQC pulse sequence (hsqcetgpsisp2, Bruker)⁷⁷. A total of 80 scans were collected using a spectral width of 12,019 for f_2 and of 30,178 for f_1 , f_2 and f_1 acquisition time of 0.085 and 0.002 s, respectively, and a relaxation delay of 2 s.

The raw data were multiplied by a 0.3 Hz exponential line broadening before applying Fourier transformation. Transformed spectra were automatically corrected for phase and baseline distortions. All the spectra were then calibrated to the reference signal of TMSp at δ 0.00 ¹H chemical shift (ppm) using TopSpin 3.5 (Bruker BioSpin srl). ¹H-¹³C HSQC spectra were also calibrated to the methyl signal of alanine at δ 19.03 ¹³C chemical shift (ppm).

The signals deriving from each metabolite were assigned using an internal NMR spectral library of pure organic compounds, spiking NMR experiments and literature data. Matching between the present NMR spectra and the NMR spectral library was performed using the AMIX and Assure software (Bruker BioSpin srl). The relative concentrations of the various metabolites were calculated by integrating the corresponding signals in defined spectral range, using a home-made R 3.0.2 script.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

RNA-seq data that support the findings of this study have been deposited in NCBI SRA archive with the accession codes SAMN12207305 to SAMN12207324. Metabolomics data have been deposited at MetaboLights (<https://www.ebi.ac.uk/metabolights/>) under code unique identifier MTBLS1699 (www.ebi.ac.uk/metabolights/MTBLS1699).

Databases used in this work: COG (<https://www.ncbi.nlm.nih.gov/COG/>), KEGG (<https://www.genome.jp/kegg/>), RegPrecise (<http://regprecise.snpdiscovery.org:8080/WebRegPrecise/>).

The authors declare that the other data supporting the findings of this study are available within the paper. No restrictions apply to data availability. The source data underlying Figs. 1a–c, 2a–d, 2b–e, 4a, c, d, 5a, b, S6a, S8a–d, S9a, b, 10a, b, S20 and S21 are provided as a Source Data files.

Code availability

Code and scripts used to implement the model are available at <https://multiauxic.sourceforge.io>.

Received: 11 September 2019; Accepted: 26 May 2020;

Published online: 19 June 2020

References

- Lambert, G. & Kussell, E. Memory and fitness optimization of bacteria under fluctuating environments. *PLoS Genet.* **10**, e1004556 (2014).
- Monod, J. The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
- Joshua, C. J., Dahl, R., Benke, P. I. & Keasling, J. D. Absence of diauxie during simultaneous utilization of glucose and Xylose by *Sulfolobus acidocaldarius*. *J. Bacteriol.* **193**, 1293–1301 (2011).
- Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA* **43**, 553 (1957).
- Kim, H. U. et al. Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol. Syst. Biol.* **7**, 460 (2011).
- Xie, L. & Wu, X. L. Bacterial motility patterns reveal importance of exploitation over exploration in marine microhabitats. *Part I Theory Biophys. J.* **107**, 1712–1720 (2014).
- Son, K., Menolascina, F. & Stocker, R. Speed-dependent chemotactic precision in marine bacteria. *Proc. Natl Acad. Sci. USA* **113**, 8624–8629 (2016).
- Rochex, A. & Lebeault, J. M. Effects of nutrients on biofilm formation and detachment of a *Pseudomonas putida* strain isolated from a paper machine. *Water Res.* **41**, 2885–2892 (2007).
- Petrova, O. E. & Sauer, K. Sticky situations: key components that control bacterial surface attachment. *J. Bacteriol.* **194**, 2413–2425 (2012).
- Solopova, A. et al. Bet-hedging during bacterial diauxic shift. *Proc. Natl Acad. Sci. USA* **111**, 7427–7432 (2014).
- Succurro, A., Segre, D. & Ebenhoh, O. Emergent subpopulation behavior uncovered with a community dynamic metabolic model of *Escherichia coli* diauxic growth. *mSystems* **4**, e00230-18 (2019).
- Narang, A. & Pilyugin, S. S. Bacterial gene regulation in diauxic and non-diauxic growth. *J. Theor. Biol.* **244**, 326–348 (2007).
- Wang, X., Xia, K., Yang, X. & Tang, C. Growth strategy of microbes on mixed carbon sources. *Nat. Commun.* **10**, 1279 (2019).
- Hermesen, R., Okano, H., You, C., Werner, N. & Hwa, T. A growth-rate composition formula for the growth of E.coli on co-utilized carbon substrates. *Mol. Syst. Biol.* **11**, 801 (2015).
- Chen, F. et al. Differential substrate usage and metabolic fluxes in *Francisella tularensis* subspecies holarctica and *Francisella novicida*. *Front Cell Infect. Microbiol.* **7**, 275 (2017).
- Steeb, B. et al. Parallel exploitation of diverse host nutrients enhances *Salmonella* virulence. *PLoS Pathog.* **9**, e1003301 (2013).
- Abu Kwaik, Y. & Bumann, D. Host delivery of favorite meals for intracellular pathogens. *PLoS Pathog.* **11**, e1004866 (2015).
- Abu Kwaik, Y. & Bumann, D. Microbial quest for food in vivo: ‘nutritional virulence’ as an emerging paradigm. *Cell Microbiol.* **15**, 882–890 (2013).
- Eisenreich, W., Rudel, T., Heesemann, J. & Goebel, W. To eat and to be eaten: mutual metabolic adaptations of immune cells and intracellular bacterial pathogens upon infection. *Front Cell Infect. Microbiol.* **7**, 316 (2017).
- Enke, T. N., Leventhal, G. E., Metzger, M., Saavedra, J. T. & Cordero, O. X. Microscale ecology regulates particulate organic matter turnover in model marine microbial communities. *Nat. Commun.* **9**, 2743 (2018).
- Seymour, J. R., Marcos & Stocker, R. Chemotactic response of marine microorganisms to micro-scale nutrient layers. *J. Vis. Exp.* **4**, e203 (2007).
- Qi, W. et al. New insights on *Pseudoalteromonas haloplanktis* TAC125 genome organization and benchmarks of genome assembly applications using next and third generation sequencing technologies. *Sci. Rep.* **9**, 16444 (2019).
- Casillo, A. et al. Anti-biofilm activity of a long-chain fatty aldehyde from Antarctic *Pseudoalteromonas haloplanktis* TAC125 against *Staphylococcus epidermidis* Biofilm. *Front. Cell Infect. Microbiol.* **7**, 46 (2017).
- Giuliani, M. et al. Process optimization for recombinant protein production in the psychrophilic bacterium *Pseudoalteromonas haloplanktis*. *Process Biochem.* **46**, 953–959 (2011).
- Wilmes, B. et al. Fed-batch process for the psychrotolerant marine bacterium *Pseudoalteromonas haloplanktis*. *Micro. Cell Fact.* **9**, 72 (2010).
- Fondi, M., Bosi, E., Presta, L., Natoli, D. & Fani, R. Modelling microbial metabolic rewiring during growth in a complex medium. *BMC Genomics* **17**, 970 (2016).

27. Vigentini, I., Merico, A., Tutino, M. L., Compagno, C. & Marino, G. Optimization of recombinant human nerve growth factor production in the psychrophilic *Pseudoalteromonas haloplanktis*. *J. Biotechnol.* **127**, 141–150 (2006).
28. Haverkorn van Rijsewijk, B. R., Nanchen, A., Nallet, S., Kleijn, R. J. & Sauer, U. Large-scale ¹³C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol. Syst. Biol.* **7**, 477 (2011).
29. Yang, Y. et al. Relation between chemotaxis and consumption of amino acids in bacteria. *Mol. Microbiol.* **96**, 1272–1282 (2015).
30. Mocali, S. et al. Ecology of cold environments: new insights of bacterial metabolic adaptation through an integrated genomic-phenomic approach. *Sci. Rep.* **7**, 839 (2017).
31. Hosie, A. H., Allaway, D., Galloway, C. S., Dunsby, H. A. & Poole, P. S. *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184**, 4071–4080 (2002).
32. Hosie, A. H. & Poole, P. S. Bacterial ABC transporters of amino acids. *Res. Microbiol.* **152**, 259–270 (2001).
33. Sannino, F. et al. *Pseudoalteromonas haloplanktis* produces methylamine, a volatile compound active against Burkholderia cepacia complex strains. *N. Biotechnol.* **35**, 13–18 (2017).
34. Bazire, P. et al. Characterization of l-carnitine metabolism in *Sinorhizobium meliloti*. *J. Bacteriol.* **201**, e00772-18 (2019).
35. Marciani, P. et al. L-carnitine and carnitine ester transport in the rat small intestine. *Pharmacol. Res.* **23**, 157–162 (1991).
36. Unemoto, T., Hayashi, M., Miyaki, K. & Hayashi, M. Formation of trimethylamine from DL-carnitine by *Serratia marcescens*. *Biochim. Biophys. Acta* **121**, 220–222 (1966).
37. Plumbridge, J. Regulation of the utilization of amino sugars by *Escherichia coli* and *Bacillus subtilis*: same genes, different control. *J. Mol. Microbiol. Biotechnol.* **25**, 154–167 (2015).
38. Ricciardelli, A. et al. Environmental conditions shape the biofilm of the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Microbiol Res* **218**, 66–75 (2019).
39. Kompala, D. S., Ramkrishna, D. & Tsao, G. T. Cybernetic modeling of microbial growth on multiple substrates. *Biotechnol. Bioeng.* **26**, 1272–1281 (1984).
40. Boianelli, A. et al. A non-linear deterministic model for regulation of diauxic lag on cellobiose by the pneumococcal multidomain transcriptional regulator CelR. *PLoS One* **7**, e47393 (2012).
41. Mandli, A. R. & Modak, J. M. Cybernetic modeling revisited: a method for inferring the cybernetic variables u_i from experimental data. *Ind. Eng. Chem. Res.* **54**, 10190–10196 (2015).
42. Ramkrishna, D. & Song, H. S. Dynamic models of metabolism: review of the cybernetic approach. *AIChE J.* **58**, 986–997 (2012).
43. Kompala, D. S., Ramkrishna, D., Jansen, N. B. & Tsao, G. T. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. *Biotechnol. Bioeng.* **28**, 1044–1055 (1986).
44. Medigue, C. et al. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* **15**, 1325–1335 (2005).
45. Koirala, S. et al. A nutrient-tunable bistable switch controls motility in *Salmonella enterica* serovar Typhimurium. *mBio* **5**, e01611–e01614 (2014).
46. Novotna, J. et al. Proteomic studies of diauxic lag in the differentiating prokaryote *Streptomyces coelicolor* reveal a regulatory network of stress-induced proteins and central metabolic enzymes. *Mol. Microbiol.* **48**, 1289–1303 (2003).
47. van Hijum, S. A., Medema, M. H. & Kuipers, O. P. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol Mol. Biol. Rev.* **73**, 481–509 (2009). Table of Contents.
48. Fuchs, T. M., Eisenreich, W., Kern, T. & Dandekar, T. Toward a systemic understanding of *Listeria monocytogenes* metabolism during infection. *Front. Microbiol.* **3**, 23 (2012).
49. Häuslein, I. et al. Multiple substrate usage of *Coxiella burnetii* to feed a bipartite-type metabolic network. *Front. Cell. Infect. Microbiol.* **7**, 285 (2017).
50. Omsland, A., Hackstadt, T. & Heinzen, R. A. Bringing culture to the uncultured: *Coxiella burnetii* and lessons for obligate intracellular bacterial pathogens. *PLoS Pathog.* **9**, e1003540 (2013).
51. Birolo, L. et al. Aspartate aminotransferase from the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC 125: cloning, expression, properties, and molecular modelling. *Eur. J. Biochem.* **267**, 2790–2802 (2000).
52. Papa, R., Rippa, V., Sannia, G., Marino, G. & Duilio, A. An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125. *J. Biotechnol.* **127**, 199–210 (2007).
53. Miller, J. et al. A short course. *Bacterial Genetics* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY, 1992).
54. Checucci, A. et al. Creation and characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *ACS Synth. Biol.* **7**, 2365–2378 (2018).
55. Koressaar, T. et al. Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics* **34**, 1937–1938 (2018).
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
57. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
58. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Abu-Jamous, B. & Kelly, S. Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. *Genome Biol.* **19**, 172 (2018).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
61. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
62. Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
63. Leyn, S. A., Maezato, Y., Romine, M. F. & Rodionov, D. A. Genomic reconstruction of carbohydrate utilization capacities in microbial-mat derived consortia. *Front. Microbiol.* **8**, 1304 (2017).
64. Leyn, S. A. et al. Comparative genomics and evolution of transcriptional regulons in Proteobacteria. *Micro. Genom.* **2**, e000061 (2016).
65. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
66. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).
67. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
68. Cardoso, M. F., Salcedo, R. L. & De Azevedo, S. F. The simplex-simulated annealing approach to continuous non-linear optimization. *Comput. Chem. Eng.* **20**, 1065–1080 (1996).
69. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2018). Available online at <https://www.R-project.org/>.
70. Vignoli, A. et al. High-throughput metabolomics by 1D NMR. *Angew. Chem. Int. Ed. Engl.* **58**, 968–994 (2019).
71. Takis, P. G., Ghini, V., Tenori, L., Turano, P. & Luchinat, C. Uniqueness of the NMR approach to metabolomics. *TrAC Trends Anal. Chem.* **120**, 115300 (2019).
72. Bernacchioni, C. et al. NMR metabolomics highlights sphingosine kinase-1 as a new molecular switch in the orchestration of aberrant metabolic phenotype in cancer cells. *Mol. Oncol.* **11**, 517–533 (2017).
73. Ghini, V. et al. Evidence of a DHA signature in the lipidome and metabolome of human hepatocytes. *Int. J. Mol. Sci.* **18**, 359 (2017).
74. D'Alessandro, G. et al. (1)H-NMR metabolomics reveals the Glabrescione B exacerbation of glycolytic metabolism beside the cell growth inhibitory effect in glioma. *Cell Commun. Signal.* **17**, 108 (2019).
75. McKay, R. T. How the 1D-NOESY suppresses solvent signal in metabolomics NMR spectroscopy: an examination of the pulse sequence components and evolution. *Concepts Magn. Reson. Part A* **38**, 197–220 (2011).
76. Carr, H. Y. & Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **94**, 630 (1954).
77. Castañar, L., Parella, T. Recent advances in small molecule NMR: improved HSQC and HMQC experiments. In: *Annual Reports on NMR Spectroscopy*. (Elsevier, 2015).

Acknowledgements

P.T. and V.G. acknowledge the support and the use of resources of Instruct-ERIC, a Landmark ESFRI project, and specifically the CERM/CIRMMP Italy Centre. M.F. and E.P. would like to thank Prof. Alessio Mengoni for the help in designing the experiments and his critical evaluation of the work. This project was supported by a PNRA (Programma Nazionale di Ricerca in Antartide) grant (grant PNRA16_00246).

Author contributions

M.F., E.P. and R.F. initiated the project. M.F. and E.P. designed all the experiments. E.P. performed and/or supervised to all the experiments performed in this work. P.T. and V.G. performed the metabolomic experiments. B.C. and L.C. performed preliminary

transcriptomic experiments. MG contributed during growth and metabolomic experiments and RNA-seq data analysis under the supervision of E.P. and M.F. C.F. contributed to the realization of growth experiments. MF performed/supervised to all the computation reported in this work. M.F. and F.D.P. wrote the theoretical kinetic models and performed the simulations. M.F. wrote the paper. All the authors contributed to the editing of the paper. All the authors have read and approved the final version of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16872-8>.

Correspondence and requests for materials should be addressed to M.F.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

List of Publications

- Kuzmanovic N., Fagorzi C., Mengoni A., Lassalle F., diCenzo G. C., (in preparation) Taxonomy of Rhizobiaceae revisited: proposal of a new framework for genus delimitation
- Giacomazzo, G.; Conti, L.; Guerri, A.; Pagliai, M.; Fagorzi, C.; Sfragano, P.; Palchetti, I.; Pietraperzia, G.; Mengoni, A.; Valtancoli, B.; Giorgi, C. (in preparation) Nitroimidazole-based ruthenium(II) complexes: playing with structural parameters to design photostable and light-responsive antibacterial agents
- Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali C, diCenzo GC, Mengoni A. (2021) Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis. *mSystems*. 2021 Jan 12;6(1):e00974-20. doi: 10.1128/mSystems.00974-20. PMID: 33436514; PMCID: PMC7901481.
- Conti, L., Mengoni, A., Giacomazzo, G. E., Mari, L., Perfetti, M., Fagorzi, C., ... & Giorgi, C. (2021). Exploring the potential of highly charged Ru (II)-and heteronuclear Ru (II)/Cu (II)-polypyridyl complexes as antimicrobial agents. *Journal of Inorganic Biochemistry*, 111467.
- Bianco, C., Andreozzi, A., Romano, S., Fagorzi, C., Cangioli, L., Prieto, P., ... & Defez, R. (2021). Endophytes from African Rice (*Oryza glaberrima* L.) Efficiently Colonize Asian Rice (*Oryza sativa* L.) Stimulating the Activity of Its Antioxidant Enzymes and Increasing the Content of Nitrogen, Carbon, and Chlorophyll. *Microorganisms*, 9(8), 1714.
- Cangioli, L., Checcucci, A., Mengoni, A., & Fagorzi, C. (2021). Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations. *Biological Communications*, 66(1), 47-54.
- Niccolai, E., Russo, E., Baldi, S., Ricci, F., Nannini, G., Pedone, M., ... & Amedei, A. (2020). Significant and conflicting correlation of IL-9 with *Prevotella* and *Bacteroides* in human colorectal cancer. *Frontiers in immunology*, 11.
- Fagorzi, C., Ilie, A., Decorosi, F., Cangioli, L., Viti, C., Mengoni, A., & diCenzo, G. C. (2020). Symbiotic and Nonsymbiotic Members of the Genus *Ensifer* (syn. *Sinorhizobium*) Are Separated into Two Clades Based on Comparative Genomics and High-Throughput Phenotyping. *Genome biology and evolution*, 12(12), 2521-2534.
- Vassallo, A., Miceli, E., Fagorzi, C., Castronovo, L. M., Del Duca, S., Chioccioli, S., ... & Fani, R. (2020). Temporal Evolution of Bacterial Endophytes Associated to the Roots of *Phragmites australis* Exploited in Phytodepuration of Wastewater. *Frontiers in microbiology*, 11, 1652.
- Perrin, E., Ghini, V., Giovannini, M., Di Patti, F., Cardazzo, B., Carraro, L., ... & Fondi, M. (2020). Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nature communications*, 11(1), 1-16.
- Fagorzi, C., Del Duca, S., Venturi, S., Chiellini, C., Bacci, G., Fani, R., & Tassi, F. (2019). Bacterial Communities from Extreme Environments: Vulcano Island. *Diversity*, 11(8), 140.

- Chiellini, C., Pasqualetti, C., Lanzoni, O., Fagorzi, C., Bazzocchi, C., Fani, R., ... & Modeo, L. (2019). Harmful effect of *Rheinheimera* sp. EpRS3 (Gammaproteobacteria) against the protist *Euplotes aediculatus* (Ciliophora, Spirotrichea): insights into the ecological role of antimicrobial compounds from environmental bacterial strains. *Frontiers in microbiology*, 10, 510.
- Bellabarba, A., Fagorzi, C., diCenzo, G. C., Pini, F., Viti, C., & Checcucci, A. (2019). Deciphering the symbiotic plant microbiome: translating the most recent discoveries on rhizobia for the improvement of agricultural practices in metal-contaminated and high saline lands. *Agronomy*, 9(9), 529.
- Fagorzi, C., Checcucci, A., diCenzo, G., Debiec-Andrzejewska, K., Dziewit, L., Pini, F., & Mengoni, A. (2018). Harnessing Rhizobia to Improve Heavy-Metal Phytoremediation by Legumes. *Genes*, 9(11), 542.
- Checcucci, A., Ghini, V., Bazzicalupo, M., Beker, A., Decorosi, F., Dohlemann, J., Fagorzi, C. ... & Turano, P. (2018). Creation and multi-omics characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *bioRxiv*, 296483.
- Chiellini, C., Miceli, E., Bacci, G., Fagorzi, C., Coppini, E., Fibbi, D., Bianconi, G., Mengoni, A., Canganella, F., Fani, R. (2018) Spatial structuring of bacterial communities in epilithic biofilms in the Acquarossa river (Italy). *FEMS Microbiology Ecology*
- DiCenzo, G. C., Debiec, K., Krzysztoforski, J., Uhrynowski, W., Mengoni, A., Fagorzi, C., ... & Drewniak, L. (2018). Genomic and biotechnological characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14. *Genes*, 9(8), 379.
- Maggini, V., Miceli, E., Fagorzi, C., Maida, I., Fondi, M., Perrin, E., ... & Fani, R. (2018). Antagonism and antibiotic resistance drive a species-specific plant microbiota differentiation in *Echinacea* spp. *FEMS microbiology ecology*, 94(8), fiy118.
- Russo, E., Bacci, G., Chiellini, C., Fagorzi, C., Niccolai, E., Taddei, A., ... & Miloeva, M. (2018). Preliminary Comparison of Oral and Intestinal Human Microbiota in Patients with Colorectal Cancer: A Pilot Study. *Frontiers in microbiology*, 8, 2699.

Book chapters

- Camilla Fagorzi, Giovanni Bacci, Alessio Mengoni (2019) Il microbiota delle piante: la biodiversità delle comunità batteriche. *Accademia Nazionale dei Lincei, XLVI seminario sulla Evoluzione Biologica e i Grandi Problemi della Biologia*
- Camilla Fagorzi, Alice Checcucci (2021) A compendium of Bioinformatic Tools for Bacterial Pangenomics to be used by wet-lab scientists. *Bacterial Pangenomics*, Edited By Marco Fondi, Alessio Mengoni, Giovanni Bacci
- Camilla Fagorzi (2021) Challenges of Multi-Omics in Improving Microbial-Assisted Phytoremediation. *Rhizomicrobiome Dynamics in Bioremediation*, Edited by Vivek Kumar

Overall activities

First author publications

- Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali C, diCenzo GC, Mengoni A. (2021) Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis. *mSystems*. 2021 Jan 12;6(1):e00974-20. doi: 10.1128/mSystems.00974-20. PMID: 33436514; PMCID: PMC7901481.
- Fagorzi, C., Ilie, A., Decorosi, F., Cangioli, L., Viti, C., Mengoni, A., & diCenzo, G. C. (2020). Symbiotic and Nonsymbiotic Members of the Genus *Ensifer* (syn. *Sinorhizobium*) Are Separated into Two Clades Based on Comparative Genomics and High-Throughput Phenotyping. *Genome biology and evolution*, 12(12), 2521-2534.
- Fagorzi, C., Del Duca, S., Venturi, S., Chiellini, C., Bacci, G., Fani, R., & Tassi, F. (2019). Bacterial Communities from Extreme Environments: Vulcano Island. *Diversity*, 11(8), 140.
- Fagorzi, C., Checcucci, A., diCenzo, G., Debiec-Andrzejewska, K., Dziewit, L., Pini, F., & Mengoni, A. (2018). Harnessing Rhizobia to Improve Heavy-Metal Phytoremediation by Legumes. *Genes*, 9(11), 542.
- Camilla Fagorzi, Giovanni Bacci, Alessio Mengoni (2019) Il microbiota delle piante: la biodiversità delle comunità batteriche. *Accademia Nazionale dei Lincei, XLVI seminario sulla Evoluzione Biologica e i Grandi Problemi della Biologia*
- Camilla Fagorzi, Alice Checcucci (2021) A compendium of Bioinformatic Tools for Bacterial Pangenomics to be used by wet-lab scientists. *Bacterial Pangenomics*, Edited By Marco Fondi, Alessio Mengoni, Giovanni Bacci
- Camilla Fagorzi (2021) Challenges of Multi-Omics in Improving Microbial-Assisted Phytoremediation. *Rhizomicrobiome Dynamics in Bioremediation*, Edited by Vivek Kumar

Corresponding author publications

- Cangioli, L., Checcucci, A., Mengoni, A., & Fagorzi, C. (2021). Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations. *Biological Communications*, 66(1), 47-54.

Co-author publications

- Kuzmanovic N., Fagorzi C., Mengoni A., Lassalle F., diCenzo G. C., (in preparation) Taxonomy of Rhizobiaceae revisited: proposal of a new framework for genus delimitation
 - Giacomazzo, G.; Conti, L.; Guerri, A.; Pagliai, M.; Fagorzi, C.; Sfragano, P.; Palchetti, I.; Pietraperzia, G.; Mengoni, A.; Valtancoli, B.; Giorgi, C. (in preparation) Nitroimidazole-based ruthenium(II) complexes: playing with structural parameters to design photostable and light-responsive antibacterial agents
 - Conti, L., Mengoni, A., Giacomazzo, G. E., Mari, L., Perfetti, M., Fagorzi, C., ... & Giorgi, C. (2021). Exploring the potential of highly charged Ru (II)-and heteronuclear Ru (II)/Cu (II)-polypyridyl complexes as antimicrobial agents. *Journal of Inorganic Biochemistry*, 111467.
 - Cangioli, L., Checcucci, A., Mengoni, A., & Fagorzi, C. (2021). Legume tasters: symbiotic rhizobia host preference and smart inoculant formulations. *Biological Communications*, 66(1), 47-54.

- Niccolai, E., Russo, E., Baldi, S., Ricci, F., Nannini, G., Pedone, M., ... & Amedei, A. (2020). Significant and conflicting correlation of IL-9 with *Prevotella* and *Bacteroides* in human colorectal cancer. *Frontiers in immunology*, 11.
- Vassallo, A., Miceli, E., Fagorzi, C., Castronovo, L. M., Del Duca, S., Chioccioli, S., ... & Fani, R. (2020). Temporal Evolution of Bacterial Endophytes Associated to the Roots of *Phragmites australis* Exploited in Phytodepuration of Wastewater. *Frontiers in microbiology*, 11, 1652.
- Perrin, E., Ghini, V., Giovannini, M., Di Patti, F., Cardazzo, B., Carraro, L., ... & Fondi, M. (2020). Diauxie and co-utilization of carbon sources can coexist during bacterial growth in nutritionally complex environments. *Nature communications*, 11(1), 1-16.
- Chiellini, C., Pasqualetti, C., Lanzoni, O., Fagorzi, C., Bazzocchi, C., Fani, R., ... & Modeo, L. (2019). Harmful effect of *Rheinheimera* sp. EpRS3 (Gammaproteobacteria) against the protist *Euplotes aediculatus* (Ciliophora, Spirotrichea): insights into the ecological role of antimicrobial compounds from environmental bacterial strains. *Frontiers in microbiology*, 10, 510.
- Bellabarba, A., Fagorzi, C., diCenzo, G. C., Pini, F., Viti, C., & Checcucci, A. (2019). Deciphering the symbiotic plant microbiome: translating the most recent discoveries on rhizobia for the improvement of agricultural practices in metal-contaminated and high saline lands. *Agronomy*, 9(9), 529.
- Checcucci, A., Ghini, V., Bazzicalupo, M., Beker, A., Decorosi, F., Dohlemann, J., Fagorzi, C. ... & Turano, P. (2018). Creation and multi-omics characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *bioRxiv*, 296483.
- Chiellini, C., Miceli, E., Bacci, G., Fagorzi, C., Coppini, E., Fibbi, D., Bianconi, G., Mengoni, A., Canganella, F., Fani, R. (2018) Spatial structuring of bacterial communities in epilithic biofilms in the Acquarossa river (Italy). *FEMS Microbiology Ecology*
- diCenzo, G. C., Debiec, K., Krzysztoforski, J., Uhrynowski, W., Mengoni, A., Fagorzi, C., & Drewniak, L. (2018). Genomic and biotechnological characterization of the heavy-metal resistant, arsenic-oxidizing bacterium *Ensifer* sp. M14. *Genes*, 9(8), 379
- Maggini, V., Miceli, E., Fagorzi, C., Maida, I., Fondi, M., Perrin, E., ... & Fani, R. (2018). Antagonism and antibiotic resistance drive a species-specific plant microbiota differentiation in *Echinacea* spp. *FEMS microbiology ecology*, 94(8), fiy118.
- Russo, E., Bacci, G., Chiellini, C., Fagorzi, C., Niccolai, E., Taddei, A., ... & Miloeva, M. (2018). Preliminary Comparison of Oral and Intestinal Human Microbiota in Patients with Colorectal Cancer: A Pilot Study. *Frontiers in microbiology*, 8, 2699.

Congress presentations

- CSM-SCM 2021 - June 14-17, 2021 Online Conference
Poster presentation: "Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis"
- SMRT Leiden 2021 - May 26-27, 2021 Online Conference
Poster presentation: "Unraveling the bacterial epigenome and its influence on gene transfer and transcriptomic control"
- 1st Florence ISME DAY - November 21, 2020 Online Conference
Oral communication: "Toward and integrated and predictive view of symbiotic nitrogen fixation"
- FEMS 2019 –May 2019, Glasgow (Scotland)
Poster presentation: "Comparative genomics and deep phenotyping of the plant-associated genus *Ensifer*"
- 14th GIM, 2019 - September 8-11, 2019 Pisa (Italy)
Poster presentation: "Comparative genomics and deep phenotyping of the plant-associated genus *Ensifer*"

- SIBM 2019 - June 10-14, 2019 Livorno (Italy)
Poster Presentation: "Impatto Portuale su Aree Marine Protette"
- Symbioza8 - May 17-19, 2019, Warsaw (Poland)
Poster presentation and oral communication: "Comparative genomics and deep phenotyping of the plant-associated genus *Ensifer*"
- Scuola di Genetica in Cortona, AGI - June 14-15, 2019 Cortona (Italy)
Poster presentation: "Comparative genomics and deep phenotyping of the plant-associated genus *Ensifer*"
- XXXIII SIMGBM Congress - June 19-22, 2019 - Florence (Italy)
Oral communication: "Comparative genomics and deep phenotyping of the plant-associated genus *Ensifer*"
- Cortona Procarioti 2018 - May 17-19, 2018 - Cortona (Italy)
Oral communication: "Towards microbe-assisted therapy of colon cancer: exploring oral and intestinal human microbiota in patients with CRC"

Courses external to the PhD program attended

- ICBEB Seminar - November 26-30, 2018 - Trieste (Italy) "Translational aspects of plant microbiome research"

Visiting in external laboratories

- May 20-24, 2019 - Visiting student at the Department of Plant Protection and Biotechnology, University of Gdansk. Head of the laboratory: Prof. Ewa Łojkowska.

Research and development

- Admitted to the 19th pre-incubation mission of the Incubatore Universitario Fiorentino (IUF) with the project "Eco-Rehab" aimed at developing an innovative procedure to separate and recover lithium from ion-lithium exhausted batteries.