Divergent properties and phylogeny of cyanobacterial 5-*enol*-pyruvyl-shikimate-3-phosphate synthases: evidence for horizontal gene transfer in the *Nostocales*.

Giuseppe Forlani[1], Michele Bertazzini[1], Donatella Barillaro[1] and Rosmarie Rippka[2]

[1] Department of Life Science and Biotechnology, University of Ferrara, Italy; [2] Institut Pasteur, Unité des Cyanobactéries, URA CNRS 2172, Paris, France

Author for correspondence: Giuseppe Forlani, Dept of Life Science and Biotechnology, via L. Borsari 46, I-44121 Ferrara, Italy. Phone # + 39-0532-455311, FAX # + 39-0532-249761, E-mail flg@unife.it.

Section: *Physiology and Development* or *Evolution*

Total word count for the main body of the text = 5,363 (Introduction = 847, Materials and Methods = 1,147, Results = 1,291, Discussion = 2,009, and Acknowledgements = 69)

Six figures in colour and 2 tables, and 5 figures and 2 tables in supporting information

Summary

- Since it represents the target of the successful herbicide glyphosate, great attention has been paid to the shikimate pathway enzyme 5-*enol*-pyruvyl-shikimate-3-phosphate (EPSP) synthase. However, inconsistent results have been described concerning the sensitivity of the enzyme from cyanobacteria, and consequent inhibitory effects on cyanobacterial growth.

- The properties of EPSP synthase were investigated in a set of 42 strains representative of the large morphological diversity of these prokaryotes. Publicly available protein sequences were analyzed, and related to enzymatic features.

- In most cases, the native protein showed an unusual homodimeric composition and a general sensitivity to micromolar doses of glyphosate. In contrast, 8 out of 15 *Nostocales* strains were found to possess a monomeric EPSP synthase, whose activity was inhibited only at concentrations exceeding 1 mM. Sequence analysis showed that these two forms are only distantly related, the latter clustering separately in a clade composed of diverse bacterial phyla.

- The results are consistent with the occurrence of a horizontal gene transfer event involving an evolutionarily distant organism. Moreover, data suggest that the existence of class I (glyphosate-sensitive) and class II (glyphosate-tolerant) EPSP synthases representing two distinct phylogenetic clades is an oversimplification due to the limited number of analyzed samples.

Key words: cyanobacteria, EPSP synthase, glyphosate, herbicide tolerance, horizontal gene transfer, subunit composition.

Introduction

In plants and bacteria the penultimate step in the common pre-chorismate pathway leading to aromatic amino acid biosynthesis, the addition of the carboxyvinyl group of phosho-*enol*-pyruvate (PEP) to shikimate-3-phosphate (S3P), is catalysed by a monofunctional enzyme, 5-*enol*-pyruvyl-shikimate-3-phosphate (EPSP) synthase (EC 2.5.1.19) (Herrmann, 1995). Being identified as the main target of the most successful pesticide ever, the phosphonate herbicide glyphosate (*N*-[phosphonomethyl]glycine) (Grossbard & Atkinson, 1985; Duke & Powles, 2008), EPSP synthase has been subject of extensive studies in the past decades, and is currently one of the best characterized enzymes (Funke *et al.*, 2007; 2009). Rapidly and completely degraded by soilborne microorganisms to water, carbon dioxide and inorganic phosphate (Torstensson, 1985, Forlani *et al.*, 1999), and showing little or no acute or chronic toxicity, and no apparent carcinogenic and mutagenic activity (Duke & Powles, 2008, Astiz *et al.*, 2009), glyphosate is considered both environmentally-friendly and safe to mammals, which do not possess a functional shikimate pathway. Formerly, glyphosate was only of limited use, since the compound does not distinguish between weeds and crops. However, bacterial genes encoding glyphosate-resistant EPSP synthases were subsequently identified, cloned, and used to transform plants (Funke *et al.*, 2006). During recent years, herbicide-tolerant seeds have become available for an increasing number of species of greatest agronomical value (Gianessi, 2005).

The current massive utilization of glyphosate in agricultural systems worldwide, estimated in around 650,000 t in 2011 (Research and Markets, 2012), may exert negative side effects on the soil microflora. Following its leaching from the root zone into drainage water or groundwater (Borggaard & Gimsing, 2008, Aparicio *et al.*, 2013), aquatic ecosystems may also be affected, leading to potential shifts in microbial or algal community structure. This may be particularly true in the case of cyanobacteria, the only group of prokaryotes capable of oxygenic photosynthesis, which are widely distributed even in strikingly different habitats. Cyanobacteria play a major role in both carbon and nitrogen cycling, since many taxa, either symbiotic or free-living, can accomplish biological N_2 fixation. Conflicting data were reported as to their sensitivity to glyphosate. Some strains showed severe growth inhibition when the herbicide was applied at micromolar concentrations (Issa, 1999; Balakumar & Ravi, 2002). In contrast,

some other species were found to exhibit a remarkable natural tolerance to this phosphonate (Powell *et al.,* 1991, Lipok *et al.,* 2010). In the case of *Nostoc* sp. PCC 7937 (cited by the authors as "*Anabaena variabilis*" ATCC 29413), herbicide tolerance was shown to reflect an insensitive form of EPSP synthase (Powell *et al.,* 1992). A glyphosate-susceptible form of EPSP synthase was on the contrary purified from "*Spirulina platensis*" C1 (later renamed *Arthrospira* sp. PCC 9438), whose activity was completely inhibited by the herbicide at micromolar levels (Forlani & Campani, 2001). Interestingly, while the enzymes from Nostoc sp. PCC 7937 ("*Anabaena variabilis*" ATCC 29413) and all other bacterial and plant species characterized to date show a monomeric structure, with a subunit relative molecular mass ranging from 40 to 60 kDa (Powell et al., 1992, Forlani, 1997), the "*S. platensis*" C1 EPSP synthase appeared to be homodimeric (Forlani & Campani, 2001).

In previous studies we analysed some cyanobacterial strains for glyphosate sensitivity (Lipok *et al.,* 2007, Forlani *et al.,* 2008). All strains showed a remarkable tolerance to the herbicide up to millimolar levels. Two out of six were found to possess a glyphosate-insensitive form of EPSP synthase, and four were able to use the phosphonate as the sole phosphorus source for growth. Low uptake rates were measured only under P-deprivation, but experimental evidence for glyphosate metabolism was obtained also for strains apparently unable to use the phosphonate as a P source (Forlani *et al.,* 2008). These results suggested that various mechanisms may concur in providing cyanobacteria with herbicide tolerance. However, due to the relatively low number of strains analysed, general conclusions could not be drawn. The availability of more information concerning the susceptibility to glyphosate of cyanobacterial EPSP synthases would shed more light on potential negative side effects of the herbicide in terrestrial and aquatic ecosystems. In addition, knowledge about the distribution of mono- and dimeric EPSP synthases within the cyanobacterial phylum may also be exploited for the classification of these organisms. Moreover, since they have codon usage preferences more similar to plants than bacteria (Campbell & Gowri, 1990), the genes for glyphosate-tolerant EPSP synthases from cyanobacteria may be more suitable for the creation of transgenic plants than are wild-type or mutant genes of bacterial origin (Powell *et al.,* 1992). On this basis, we analysed the properties of the enzyme from a selection of 42 strains, representative of the large morphological diversity typical of this group of prokaryotes. Here we report that a dimeric, glyphosate-sensitive form of

EPSP synthase is the predominant type of enzyme, being found in representatives of all five subsections into which cyanobacteria have been classified (Castenholz, 2001). However, a cluster of filamentous heterocystous strains could be distinguished from their relatives in subsection IV (*Nostocales*) by possessing a monomeric and herbicide-tolerant enzyme. Sequence analysis suggests that the latter form may have been acquired by a horizontal gene transfer event involving an evolutionarily distant eubacterium.

Materials and Methods

Strains and growth conditions

Strains from the Pasteur Culture Collection of Cyanobacteria (PCC) and the Algal Culture Collection at the University of Durham (D) were grown at 24 ± 1°C under 14-h days and 10-h nights in 125 ml Erlenmayer flasks containing 25 ml of the appropriate minimal culture medium (Tables 1 and 2). Light was provided by E27 ES 1700 lumen daylight lamps (GE Lighting) at 300 µmol m^{-2} s^{-1} PAR in the case of *Arthrospira* (including "*Spirulina platensis*" C1), at 150 µmol m^{-2} s^{-1} for all the other genera. Subculturing was done every 6 weeks by transferring 5 ml aliquots to 20 ml of fresh medium. Growth was followed by harvest: 0.5 to 1.0-ml aliquots were withdrawn, and cells were sedimented by centrifugation for 3 min at 14,000 *g*. Pellets were resuspended with 1.0 ml methanol, and solubilization was allowed to proceed for 30 min in the dark, with occasional mixing. Samples were then centrifuged as above, and chlorophyll content in the supernatant was determined spectrophotometrically on the basis of the Arnon's formula (Lichtenthaler, 1987).

Enzyme extraction

Cells in the exponential phase of growth were harvested either by vacuum filtration on filter paper or by centrifugation for 10 min at 2,500 *g*, resuspended in ice-cold extraction buffer (50 mM Hepes-KOH buffer, pH 7.4, containing 5% [v/v] glycerol, 0.5 mM dithiothreitol, 0.5 mM EDTA, and 10 µM ammonium molybdate added to inhibit non-specific phosphatases), sedimented by centrifugation and washed again with the same buffer. Pelleted material was frozen at -20°C for 1 h, then transferred into a precooled

mortar and ground with alumina (2 g [g cells]$^{-1}$) until a fine paste was obtained. All subsequent operations were carried out at 0 to 4°C. The homogenate was resuspended with 10 ml g^{-1} of extraction buffer, and clarified for 10 min at 14,000 *g*. Solid ammonium sulfate was added to the supernatant to give 70% saturation. Precipitated proteins were collected by centrifugation, resuspended with extraction buffer and desalted by passage through a Bio-Gel P6DG column (Bio-Rad) equilibrated with the same buffer. Desalted extracts were immediately used for the determination of specific activity levels.

Enzyme assays

EPSP synthase activity was measured in the forward direction at 35°C by determining the release of inorganic phosphate using the malachite green dye assay method (Forlani *et al.*, 1994). The reaction mixture contained 50 mM Hepes-KOH, pH 7.4, 1 mM S3P, 1 mM PEP and a limiting amount of enzyme (5 to 25 pkat) in a final volume of 0.1 ml. After incubation for up to 60 min, the reaction was stopped by the addition of 1 ml of the malachite green-molybdate-acid colorimetric solution followed, after 1 min, by 0.1 ml of 34% (w/v) Na citrate. After 10 min at room temperature, absorption at 660 nm was measured against exact blanks in which S3P had been omitted. Activity was calculated from the initial linear rate on the basis of an extinction coefficient for phosphate ranging from 45,000 to 60,000 M^{-1} cm^{-1}, evaluated experimentally for each batch of colorimetric solution. The ammonium salt of S3P was purified from the culture broth of *Klebsiella pneumoniae* strain ATCC 25597 and quantified as described previously (Forlani *et al.*, 1992).

Shikimate dehydrogenase was assayed in the forward direction at 35°C by determining NADP$^+$ reduction. The assay mixture contained 100 mM glycine-NaOH buffer, pH 10.6, 1 mM NADP$^+$ and 1 mM shikimic acid in a final volume of 1 ml. A limiting amount of enzyme (from 0.1 to 0.4 nkat) was added to the pre-warmed mixture, and the increase in absorbance at 340 nm was determined for up to 15 min by continuous monitoring of the sample against blanks from which shikimate had been omitted. Activity was determined from the initial linear rate, with the assumption of an extinction coefficient for NADPH of 6,220 M^{-1} cm^{-1}. Protein concentration was determined by the method of Bradford (1976), using bovine serum albumin as the standard.

Molecular size evaluation

Following salting out with 70% ammonium sulphate, pellets were resuspended with column buffer (50 mM Tris-HCl, pH 7.5, containing 0.5 mM DTT and 250 mM NaCl) so as to obtain a final concentration of about 5 mg protein ml^{-1}. Aliquots (3 ml) were loaded onto a Sephacryl S200 SF (Pharmacia) column (1.6 x 90 cm, 180 ml bed volume) equilibrated with column buffer. Elution proceeded at a flow rate of 12 ml h^{-1}, while collecting 2-ml fractions. Alternatively, 1-ml aliquots were loaded onto a Sephadex G100 (Pharmacia) column (1.5 x 28.5 cm, 50 ml bed volume) equilibrated with column buffer. Elution proceeded at a flow rate of 30 ml h^{-1}, while collecting 1-ml fractions. Retention patterns were used to correlate elution volumes with the logarithm of molecular weight; unknown native molecular masses were estimated by comparison with elution volumes obtained with protein markers (Pharmacia, Product No. 17-0441-01 and 17-0442-01), as shown in Figs. 1 and 2.

Determination of EPSP synthase sensitivity to glyphosate

Active fractions from gel permeation chromatography were pooled, and used to assess the effect of increasing levels of glyphosate on EPSP synthase activity. Proper dilutions of a 1 M solution (brought to pH 7.5 with KOH) of an analytical standard of the herbicide (Riedel-de Haën) were added to the reaction mixture to a final concentration ranging from 1 μM to 10 mM. Results were expressed as percentage of untreated controls, allowing the calculation of the concentrations causing 50% inhibition of enzyme activity (IC_{50}) and their confidence limits.

Database searches, sequence alignments and analysis

Protein sequences of cyanobacterial EPSP synthases were retrieved either from annotated sequences available in the Cyanobase databank (http://genome.microbedb.jp/CyanoBase#resources; 38 genes), or by similarity search in the NCBI databank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the sequences of the enzyme from *Synechocystis* sp. PCC 6803 and *Nostoc* sp. PCC 7120 as queries and limiting the results to cyanobacteria (taxid: 1117; 88 genes). Accession numbers are listed in supporting material S5. To extend the similarity search to phyla outside of cyanobacteria, sequences homologous to the *Gloeobacter violaceus* PCC 7421 EPSP synthase were retrieved from the UniProt KB/Swiss-Prot database using the BLAST-p

program with its default parameters, but excluding uncultured organisms and environmental samples, and setting the maximum target to 500 sequences. Accession numbers of the obtained sequences are reported in supporting material S6.

Multiple amino acid sequence alignment by the Clustal W method, divergence and percent identities were obtained by the MegAlign software (version 7.1.0), which also allowed the generation of a maximum likelihood phylogenetic tree of cyanobacterial EPSP synthases. A simplified phylogenetic tree by fast minimum evolution of sequences homologous to the *G. violaceus* PCC 7421 EPSP synthase was obtained by using the NCBI BlastP algorithm, with the following parameters: distance tree of results, radial view, 0.85 maximum sequence difference and Grishin [protein] distance.

Statistical analysis

Linear (enzyme activity assay) and non-linear (glyphosate IC_{50}, molecular mass evaluation) regression analyses were computed by using Prism 6 (version 6.03, GraphPad Software, Inc., USA).

Results

All *Arthrospira* strains possess a dimeric and glyphosate-sensitive form of EPSP synthase

The unusual quaternary structure of the EPSP synthase isolated from "*Spirulina platensis*" C1 (Forlani & Campani, 2001) could represent a peculiarity of this strain, currently assigned to the genus *Arthrospira* as strain *Arthrospira* sp. PCC 9438 (Scheldeman *et al.*, 1999; Baurain *et al.*, 2002), or be shared by other large-celled spirally coiled and gas vesicle containing members of this taxon of the *Oscillatoriales*. To investigate this aspect, five additional strains of *Arthrospira* were chosen for testing the properties of the enzyme. Results showed very similar patterns (Table 1), in agreement with the high degree of relatedness inferred for these strains based on 16S rRNA gene and ITS sequence analyses (Scheldeman *et al.*, 1999; Baurain *et al.*, 2002). In all cases the retention of enzyme activity during native gel permeation chromatography was lower than that of shikimate dehydrogenase, and the comparison of

elution volumes with those obtained under the same conditions for protein molecular markers allowed the estimation of a native molecular mass around 90 kDa (Fig. 1). Concerning the sensitivity to glyphosate, the concentrations causing 50% inhibition of enzyme activity (IC_{50}) ranged from 19 to 30 μM. Although only a limited number of strains have been analysed, such results suggest that a dimeric form of the enzyme represents another common feature for the members of this genus.

A dimeric EPSP synthase is prevalent among cyanobacteria, but a monomeric enzyme is present instead in a few strains of Subsection IV (*Nostocales*)

In order to obtain further information about the occurrence and distribution of this homodimeric and glyphosate-sensitive form of the enzyme, a larger number of cyanobacterial strains, representative of all five subsections, were analysed. Results are summarized in Table 2. As for *Arthrospira* strains, EPSP synthase from other members of the *Oscillatoriales* (subsection III) showed retention patterns upon gel filtration that are compatible with a native molecular mass ranging from 80 to 100 kDa. Similar results were obtained for all strains in subsections I (*Chroococcales*), II (*Pleurocapsales*) and V (*Stigonematales*), as well as for seven out of 15 strains of subsection IV (*Nostocales*), leading to the conclusion that a dimer is the typical form of the enzyme among oxygenic photosynthetic bacteria. In contrast, eight filamentous heterocystous strains differed from their relatives in subsection IV by possessing a monomeric EPSP synthase, which showed a relative molecular mass in the range 32 to 47 kDa (Fig. 2), very nearly the deduced molecular weight of the polypeptide subunit (Table 2).

The monomeric EPSP synthases among the *Nostocales* show a noteworthy resistance to glyphosate

When the sensitivity to glyphosate was compared, remarkable differences were found among the strains. In most cases, a high degree of sensitivity was again evident (Fig. 3), with IC_{50} values in the range from 10^{-5} to 10^{-4} M, similar to those determined for *Arthrospira* strains. For a few strains (*Synechococcus* PCC 7335, *Stanieria cyanosphaera* PCC 7301, and *Rivularia* sp. PCC 7116, respectively assigned to subclusters I, II and IV) an intermediate level of tolerance was shown, with IC_{50} values ranging from 10^{-4} to 10^{-3} M. Moreover, nine strains were found to possess an almost

insensitive enzyme, with IC_{50} values higher than 10^{-3} M. Interestingly, with only a single exception (*Gloeobacter violaceus* PCC 7421, subsection I), all the latter strains are members of the *Nostocales* (subsection IV). If both the quaternary structure and the sensitivity to the herbicide were plotted in the same graph (Fig. 4), a straightforward relationship became evident: all monomeric EPSP synthases are little affected by glyphosate, whereas the activity of all dimeric enzymes (with the above only exception) is progressively inhibited by the phosphonate in the micromolar range.

Sequence analysis shows a low degree of similarity between the monomeric and the dimeric forms of the cyanobacterial EPSP synthase

To obtain further information, the deduced amino acid sequences of EPSP synthases, whose nucleotide sequences are available in databanks for the strains herein analysed, were aligned and compared. Results (Supplemental material S1) provided an unexpected picture. Low amino acid identities and similarity scores were observed between the monomeric and the dimeric enzymes, and several gaps were evident as well. The monomeric forms exhibit less than 28 % amino acid sequence identities compared to the dimeric enzymes. In contrast, if the two forms were analysed separately, a high degree of conservation (> 50 % amino acid identities) was found within either group (Supplemental material S2). A maximum likelihood tree generated from the aligned sequences showed clearly that the proteins form two distinct clusters, at a noteworthy genetic distance (Supplemental material S3). Since the protein sequence data were available for only a few of the strains studied, and thus provided only a partial picture of the distribution of the two forms of enzymes, all sequences found in public databases for putative cyanobacterial EPSP synthases were downloaded and analysed. The resulting phylogenetic tree (Fig. 5), including two bacterial enzymes (*Chlorobium tepidum* ATCC 49652, and *Rhodopseudomonas palustris* CGA009) for comparison, once again grouped the cyanobacterial proteins into two separate clades, one of which contains the great majority of sequences. Within this major clade, the sequences of two unicellular strains (*Thermosynechococcus elongatus* BP1 and *Synechococcus* sp. PCC 6312) form *loner* branches, whereas the other protein sequences, including all the dimeric forms of EPSP synthases, are contained in a subclade that positions the evolutionarily ancient *Gloeobacter violaceus* strains at its base. This subclade is composed of two sister groups, one being formed uniquely by the sequences of

unicellular strains (*Chroococcales*), namely the two *Gloeobacter* strains, and the marine and freshwater strains collectively known as members of the *Prochlorococcus*/ *Synechococcus* clade (Shih *et al.*, 2013, and references therein). In contrast, the second sister subclade, comprising several clusters, encompasses representatives of all subsections (I-V), and shows a strain topology consistent with cyanobacterial phylogenies based on 16S rRNA gene sequences or genomic data (Shih *et al.*, 2013; Dagan *et al.*, 2013), particularly for the majority (20) of the filamentous heterocystous strains of subsections IV and V that all cluster together. A smaller set of protein sequences (7) of the *Nostocales* strains (subsection IV), however, differ strikingly from their relatives, and form a separate cluster in the second major cyanobacterial EPSP synthase clade.

The monomeric and glyphosate-resistant EPSP synthases from cyanobacteria are similar to enzymes from other, more evolutionarily distant lineages

To obtain further information, the consensus sequence found for either the dimeric or the monomeric enzyme (Fig. S1) were used as queries to search in databanks for homologous proteins. If the search was limited to the 1000 sequences showing the highest similarity scores, the monomeric enzyme, quite surprisingly, did not detect any dimeric cyanobacterial EPSP synthase as relatives, and *vice-versa* (data not shown). Only if 5000 nonredundant sequences showing the highest similarity scores were searched for, the results comprised both types (not shown). To obtain a visual record, the amino acid sequence of the phylogenetically intermediate enzyme of *G. violaceus* PCC 7421 was used to search within the manually annotated and reviewed sequences present in the UniProt Knowledgebase (Swiss-Prot), again excluding uncultured organisms and environmental samples. An automatic fast minimum evolution tree generated from the data (Fig. 6) showed that the dimeric forms clustered together as two sister clades, and were more closely related to various bacterial groups than to the monomeric cyanobacterial enzymes. The latter were found to be located at a considerable genetic distance, seemingly grouping closer to evolutionarily unrelated organisms, such as euryarcheotes, ascomycetes and dicotyledons. Although precise relationships can not be inferred from this preliminary approach, on the whole data suggest that the presence of a monomeric and glyphosate-tolerant form of EPSP synthase among cyanobacteria may have resulted from an extra-phylum horizontal gene

transfer event that occurred during the evolutionary history of the *Nostocales*.

Discussion

Despite the extensive studies on the influence of glyphosate in the environment, to date the sensitivity of cyanobacteria to this herbicide has been evaluated only sporadically. Some species have been included in wider screening of microorganisms for glyphosate sensitivity (Peterson *et al.*, 1994), or single strains have been characterized with respect to herbicide susceptibility (Powell *et al.*, 1991; Issa, 1999). As a consequence, inconsistent results have been reported on either the inhibitory effects on growth, or the ability of cyanobacteria to metabolize the phosphonic moiety (e.g. Powell *et al.*, 1991 *vs* Ravi & Balakumar, 1998). Contrasting evidence has been described as well concerning the sensitivity of the glyphosate target, EPSP synthase (Powell *et al.*, 1992; Forlani & Campani, 2001; Forlani *et al.*, 2008). To our knowledge, this is the first report in which the properties of the cyanobacterial enzyme have been investigated in a wide set of strains. Results clearly pointed out that the great majority of species possess an unusual and distinctive form of EPSP synthase, which upon gel permeation chromatography under native conditions showed elution patterns that are consistent with a homodimeric composition of the holoenzyme. With the only exception of ascomycete fungi, in which EPSP synthase is part of a 175 kDa pentafunctional "AroM" protein catalysing the direct conversion of 3-deoxy-D-arabino-heptulosonate-7-phosphate into EPSP (Richards *et al.*, 2006), the enzyme from all prokaryotes and other eukaryotes has proven to be a 40 to 50 kDa monofunctional polypeptide. The possibility that the 90 kDa protein of cyanobacteria characterized in this study may result from the formation *in vivo* of a complex between EPSP synthase and another enzyme of the shikimate pathway is ruled out by the appearance of only a single band upon SDS-PAGE (Forlani & Campani, 2001), as well as by a different elution pattern for shikimate dehydrogenase activity (Fig. 1), and the lack of both shikimate kinase and chorismate synthase activities in partially purified preparations (data not shown). Moreover, chromatography in the presence of high ionic strength (up to 1 M NaCl) did not change the elution properties, thus making the occurrence of protein aggregation during the run unlikely. Concerning the sensitivity to the inhibition brought about by glyphosate, a certain degree of

variability was found among these homodimeric forms of the enzyme, ranging from highly susceptible to mildly resistant (Tables 1 and 2). However, with only a few exceptions, IC_{50} values were lower than 100 μM, placing this cyanobacterial enzyme into the most sensitive of the three categories characterized earlier among eubacterial EPSP synthases (Schulz *et al.,* 1985).

This general picture was confirmed for all strains of subsections I, II, III and V, but not for those classified in subsection IV. Seven among the 15 *Nostocales* analysed in this work, comprising *Nostoc* sp. PCC 7937, previously shown (under the name "*Anabaena variabilis*" ATCC 29413) to be highly tolerant to glyphosate (Powell *et al.*, 1991; 1992), were found to possess a (more canonical) monomeric EPSP synthase. Interestingly, all these enzyme forms were almost insensitive to the herbicide, retaining appreciable catalytic rates in the presence of inhibitor concentrations higher than 10 mM. These results could imply that mutations may have occurred during the evolution of the *Nostocales* affecting both subunit dimerization and glyphosate binding. However, taking into account our own EPSP sequence analyses, as well as recent cyanobacterial phylogeny based on genome sequencing data (Shih *et al.,* 2013), this explanation can be excluded. In the latter study, *Nostoc* sp. PCC 7937 (cited as "*Anabaena variabilis*" ATCC 29413) and *Nostoc* sp. PCC 7120, both possessing monomeric and glyphosate-tolerant EPSP synthases (Table 2, Fig. 5), were shown to be closely related to *Nostoc* sp. PCC 7524 (see Fig. 1A in Shih *et al.,* 2013), which has a dimeric and glyphosate-sensitive enzyme (Table 2, Fig. 5), and are much more genetically distant to other strains with monomeric and tolerant enzymes, such as *Calothrix* PCC 7507 and *Nostoc punctiforme* PCC 73102. This would imply that similar mutations have arisen twice during the evolution of this group of cyanobacteria, which seems unlikely. More importantly, the monomeric and dimeric EPSP synthases did not show any relevant and consistent differences in the putative glyphosate binding site (Supplemental material S7), and only shared very low degrees of overall amino acid identities/similarities (Supplemental materials S1 and S2). Furthermore, the analysis of all available cyanobacterial EPSP synthase sequences clearly showed that the dimeric forms group together in a clade containing the large majority of enzymes, whereas a small number of sequences comprising all the monomeric forms cluster separately at a considerable genetic distance (Fig. 5). On the whole, the data therefore suggest that the presence of a few monomeric and glyphosate-tolerant forms of the enzyme may have resulted from a

horizontal gene transfer event that occurred during the phylogenetic radiation of the *Nostocales*. This would have led to a common ancestor in a representative of subclade 7, as defined previously (Shih *et al.,* 2013), in which both enzyme forms were present. Thereafter, either the monomeric or the dimeric enzyme was lost, leading to descendants with only one or the other of the two proteins. This hypothesis is further strengthened by the results depicted in Fig. 6, showing that the monomeric EPSP synthases from cyanobacteria are more similar to the enzymes from genetically distant organisms, such as euryarcheotes, ascomycetes and dicotyledons, than to the dimeric forms that seem to be more typical for these photosynthetic eubacteria. Increasing evidence has been reported concerning the occurrence of horizontal gene transfer events during the evolution of the shikimate pathway. For instance, a comprehensive analysis of the corresponding genes in diverse organisms (prokaryotes, oomycetes, ciliates, diatoms, basidiomycetes, zygomycetes, green algae, red algae and higher plants) indicated that plants initially inherited all the shikimate pathway genes from the cyanobacterial plastid progenitor genome, but subsequently five genes were obtained from a minimum of two other eubacterial genomes (Richards *et al.*, 2006). The same authors also demonstrated a high frequency of loss and replacement events, and estimated that at least 50 gene/domain losses occurred during the eukaryotic genome evolution. More recently, new insights into the early evolution of the shikimate pathway in prokaryotes have been obtained, showing the existence in *Archaea* of non-homologous isofunctional enzymes, and - once again - the occurrence of many bidirectional horizontal gene transfer events between the two prokaryotic domains (Zhi *et al.,* 2014). Interestingly, the sequences coding for EPSP synthases analysed in the latter study clustered into two phylogenetic clades separated by a considerable genetic distance, and reliable divergence. Because only subfamily 1 proteins were identified in *Archaea*, whereas both subfamily 1 and 2 enzymes were distributed among bacteria, the latter form was probably generated by gene duplication just after the divergence of the two prokaryotic domains. Since the enzyme from *Synechococcus* sp. JA-2-3B'a grouped in subfamily 2 (Zhi *et al.*, 2014), and is positioned in the main clade of cyanobacterial EPSP synthases that comprise the dimeric proteins (Fig. 6), the monomeric enzyme nowadays present in some of the *Nostocales* may have been derived from a donor in which the ancestral archaeal gene had been retained.

Strain clustering in the major clade of the EPSP synthase tree (Fig. 5), seemingly

representative of the dimeric form of the enzyme, is overall consistent with phylogenies inferred from 16S rDNA gene sequences (Fuller *et al.,* 2003; Tomitani *et al.,* 2006; Schirrmeister *et al.,* 2011) and genome sequencing data (Dagan *et al.,* 2013; Shi *et al.,* 2013), showing that the gene encoding this type of protein was vertically transmitted among cyanobacteria. Therefore, this gene could also serve as useful additional molecular marker for taxonomic identification. For instance, the EPSP sequences readily permit to distinguish between the spiral strains of *Arthrospira* from those of the genus *Spirulina*, represented by *Spirulina subsalsa* PCC 9445 (Fig. 5), in agreement with previous studies (Scheldeman et al., 1999; Shih et al., 2013). Some of the rather surprising positioning of the strains, such as the grouping of the two *Acaryochloris* EPSP synthases together with those of *Arthrospira* and *Trichodesmium* strains, or the positioning of *Synechococcus elongatus* PCC 6301 and PCC 7942, distant from strains of the *Synechococcus*/*Prochlorococcus* clade, may be due to different algorithms used for tree construction, and/or insufficient sampling of sequences, but may possibly also indicate that additional intra-phylum gene transfer events have occurred during cyanobacterial evolution.

Due to the biotechnological interest for obtaining herbicide-resistant crop plants, an increasing number of glyphosate-tolerant EPSP synthases have been identified and described to date (Funke *et al.,* 2006, and references therein). As a function of their susceptibility to the herbicide, two classes of enzymes have been defined. Class I comprises EPSP synthases from plants and bacteria that are naturally sensitive to the phosphonate, whereas class II enzymes are of bacterial origin, and show a higher tolerance to glyphosate. The amino acid sequences of representatives of both classes of proteins formed two well separated clusters by phylogenetic inference (Cao *et al.,* 2012). Quite surprisingly, if the consensus sequences of the cyanobacterial forms of the enzyme were included in an analysis carried out with these same bacterial proteins, the monomeric and glyphosate tolerant EPSP synthase clustered with class I enzymes, whereas the dimeric and glyphosate sensitive form grouped with class II proteins (supplemental material S4). Type II enzymes have most often been found in soilborne bacterial strains selected for their ability to grow in the presence of millimolar concentrations of the herbicide, or the genes encoding these proteins were cloned and introduced into susceptible recipients strains by transformation after extracting DNA directly from soil contaminated with high concentrations of glyphosate (Cao *et al.,*

2012, and references therein). The results presented in this study, and the recent data on the occurrence among prokaryotes of two phylogenetic clades with considerable genetic distance (Zhi *et al.,* 2014), strongly suggest that in both phylogenetic lineages, previously considered as representative of class I and class II enzymes, glyphosate-susceptible as well as glyphosate-tolerant proteins may occur, and that the presence of glyphosate-resistant forms in one set could have been overestimated by the above-cited selection protocols. Furthermore, the grouping of the cyanobacterial glyphosate-sensitive dimeric EPSP synthases with monomeric glyphosate-tolerant proteins opens the question of whether other bacterial representatives may exist that also possess a dimeric form of the protein.

During recent years increasing information has become available concerning the molecular bases for glyphosate resistance in several plant and bacterial species. A few point mutations have been identified that resulted in the conversion of glyphosate-sensitive EPSP synthases into herbicide-tolerant enzymes (Pollegioni *et al.*, 2011). Moreover, amino acid residues that are essential for both PEP and glyphosate binding to the active site have been identified (Schönbrunn *et al*., 2001). However, a complex picture has emerged, since since several different sites in the protein are important for substrate/inhibitor binding (Supplemental material S7), and also mutations in other regions of the enzyme have been reported to influence the susceptibility to glyphosate (Pollegioni *et al.,* 2011). Furthermore, a remarkable dissimilarity has been found over the entire amino acid sequences between class I and class II EPSP synthases (Cao *et al.*, 2010, this work), showing that highly conserved residues in the former are not present in the latter, and *vice-versa*. On the other hand, no information is available at all concerning amino acid residues involved in protein dimerization, a feature found to date only for cyanobacterial EPSP synthases. It is thus not possible to ascertain whether dimerization influences the three-dimensional structure of the protein cleft important for herbicide binding. As a consequence, at this point of the research any discussion about the basis of the differential sensitivity to glyphosate shown by the cyanobacterial enzymes, as well as its relationship with the dimeric/monomeric structure of the protein, would be pure speculation. Similarly, it is difficult to explain the noteworthy tolerance to the herbicide found for the enzyme extracted from *G. violaceus* PCC 7421, which shows high sequence similarity to the sensitive forms of the enzyme (Fig. S7). Work is currently in progress in our laboratory to shed more light on these aspects, as well as to

investigate the kinetic properties of glyphosate-tolerant EPSP synthases from the *Nostocales*.

Acknowledgements

The technical assistance of Drs Alberto Campani and Anna Trzęsiok is acknowledged. The authors thank Prof. A.M. Sanangelantoni (University of Parma, Italy) for providing *Arthrospira* strains from the Durham collection. This work was supported in part by grants from the University of Ferrara (*Fondo di Ateneo per la Ricerca*). R. Rippka (retired) acknowledges the Institut Pasteur and the CNRS URA-2172 for funding during the time this work was initiated.

References

Aparicio VC, De Gerónimo E, Marino D, Primost J, Carriquiriborde P, Costa JL. 2013. Environmental fate of glyphosate and aminomethylphosphonic acid in surface waters and soil of agricultural basins. *Chemosphere* **93**: 1866–1873.

Astiz M, Zirulnik F, Giménez MS, De Alaniz MJT, Marra CA. 2009. Overview of glyphosate toxicity and its commercial formulations evaluated in laboratory animal tests. *Current Topics in Toxicology* **6**: 1–15.

Balakumar T, Ravi V. 2002. Catalytic degradation of the herbicide glyphosate by the paddy field isolates of cyanobacteria. *In* Algae and their Biotechnological Potential. Edited by Chen F, Jiang Y. pp. 195–206. Heidelberg: Springer.

Baurain D, Renquin L, Grubisic S, Scheldeman P, Belay A, Wilmotte A. 2002. Remarkable conservation of internally transcribed spacer sequences of *Arthrospira* ("*Spirulina*") (Cyanophyceae, cyanobacteria) strains from four continents and of recent and 30-year-old dried samples from Africa. *Journal of Phycoogy* **38**: 384–393.

Borggaard OK, Gimsing AL. 2008. Fate of glyphosate in soil and the possibility of leaching to ground and surface waters: a review. *Pest Management Science* **64**: 441–456.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry* **72**: 248–254.

Campbell WH, Gowri G. 1990. Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiology* **92**: 1–11.

Cao G, Liu Y, Zhang S, Yang X, Chen R, Zhang Y, Lu W, Liu Y, Wang J, Lin M, Wang G. 2012. A novel 5-enolpyruvylshikimate-3-phosphate synthase shows high glyphosate tolerance in *Escherichia coli* and tobacco plants. *PLoS One* **7**:e38718.

Castenholz RW. 2001. Phylum BX. Cyanobacteria - Oxygenic Photosynthetic Bacteria. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn., Vol. 1. Edited by Boone DR, Castenholz RW, pp. 473–487. New York, Springer Verlag.

Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, Tandeau de Marsac N, Gugger M, Lockhart PJ, Allen JF, Brune I, Maus I, Pühler A, Martin WF. 2013. Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. *Genome Biology and Evolution* **5**: 31–44.

Dev A, Tapas S, Pratap S, Kumar P. 2012. Structure and function of enzymes of shikimate pathway. *Current Bioinformatics* **7**: 374–391.

Duke SO, Powles SB. 2008. Glyphosate: a once-in-a-century herbicide. *Pest Management Science* **64**: 319–325.

Forlani G .1997. Properties of the 5-*enol*-pyruvyl-shikimate-3-phosphate synthase isoforms isolated from maize cultured cells. *Journal of Plant Physiology* **150**: 369–375.

Forlani G, Campani A. 2001. A dimeric 5-*enol*-pyruvyl-shikimate-3-phosphate synthase from the cyanobacterium *Spirulina platensis*. *New Phytologist* **151**: 443–450.

Forlani G, Mangiagalli A, Nielsen E, Suardi MC. 1999. Degradation of the phosphonate herbicide glyphosate in soil: evidence for a possible involvement of unculturable microorganisms. *Soil Biology and Biochemistry* **31**: 991–997.

Forlani G, Nielsen E, Racchi ML. 1992. A glyphosate-resistant 5-*enol*-pyruvyl-shikimate-3-phosphate synthase confers tolerance to a maize cell line. *Plant Science* **85**: 9–15.

Forlani G, Parisi B, Nielsen E. 1994. 5-*Enol*-pyruvyl-shikimate-3-phosphate synthase from *Zea mays* cultured cells: purification and properties. *Plant Physiology* **105**: 1107–1114.

Forlani G, Pavan M, Gramek M, Kafarski P, Lipok J. 2008. Biochemical bases of a widespread tolerance of cyanobacteria to the phosphonate herbicide glyphosate. *Plant and Cell Physiology* **49**: 443–456.

Forlani G, Riccardi G, De Rossi E, De Felice M. 1991. Biochemical evidence for multiple forms of acetohydroxy acid synthase in *Spirulina platensis*. *Archives of Microbiology* **155**: 298–302.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. 2003. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Applied and Environmental Microbiology* **69**: 2430-2443.

Funke T, Han H, Healy-Fried ML, Fischer M, Schonbrunn E. 2006. Molecular basis for the herbicide resistance of Roundup Ready crops. *Proceedings of the National Academy of the Sciences of the U.S.A.* **103**: 13010–13015.

Funke T, Healy-Fried ML, Han H, Alberg DG, Bartlett PA, Schonbrunn E. 2007. Differential inhibition of class I and class II 5-enolpyruvylshikimate-3-phosphate synthases by tetrahedral reaction intermediate analogues. *Biochemistry* **46**: 13344–13351.

Funke T, Yang Y, Han H, Healy-Fried ML, Olesen S, Becker A, Schonbrunn E. 2009. Structural basis of glyphosate resistance resulting from the double mutation Thr^{97}→Ile and Pro^{101}→Ser in 5-enolpyruvylshikimate-3-phosphate synthase from *Escherichia coli*. *Journal of Biological Chemistry* **284**: 9854–9860.

Gianessi LP. 2005. Economic and herbicide use impacts of glyphosate-resistant crops. *Pest Management Science* **61**: 241–245.

Grossbard E, Atkinson D. 1985. The Herbicide Glyphosate. Pp. 490. London: Butterworths.

Herrmann KM. 1995. The shikimate pathway: early steps in the biosynthesis of aromatic compounds. *The Plant Cell* **7**: 907–919.

Issa AA. 1999. Interference of glyphosate with the shikimate pathway by Cyanobacteria in chemostat culture. *Microbios* **100**: 47–55.

Lichtenthaler HK. 1987. Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods in Enzymology* **148**: 350–382.

Lipok J, Owsiak T, Młynarz P, Forlani G, Kafarski P. 2007. Phosphorus NMR as a tool to study mineralization of organophosphonates – The ability of *Spirulina* spp. to degrade glyphosate. *Enzyme and Microbial Technology* **41**: 286–291.

Lipok J, Studnik H, Gruyaert S. 2010. The toxicity of Roundup® 360 SL formulation and its main constituents: glyphosate and isopropylamine towards non-target water photoautotrophs. *Ecotoxicology and Environmental Safety* **73**: 1681–1688

Peterson HG, Boutin C, Martin PA, Freemark KE, Ruecker NJ, Moody MJ. 1994. Aquatic phyto-toxicity of 23 pesticides applied at expected environmental concentrations. *Aquatic Toxicology* **28**: 275–292.

Pollegioni L, Schonbrunn E, Siehl D. 2011. Molecular basis of glyphosate resistance-different approaches through protein engineering. *FEBS Journal* **278**:2753–2766.

Powell HA, Kerby NW, Rowell P, Mousdale DM, Coggins JR. 1992. Purification and properties of a glyphosate-tolerant 5-enolpyruvylshikimate 3-phosphate synthase from the cyanobacterium *Anabaena variabilis*. *Planta* **188**: 484–490.

Powell HA, Kerby NW, Rowell P. 1991. Natural tolerance of cyanobacteria to the herbicide glyphosate. *New Phytologist* **119**: 421–426.

Ravi V, Balakumar H. 1998. Biodegradation of the C–P bond in glyphosate by the cyanobacterium *Anabaena variabilis* L. *Journal of Scientific and Industrial Research* **57**: 790–794.

Research and Markets. 2012. Outlook for China glyphosate industry 2012-2016. Report 2101356. www.researchandmarkets.com/reports/2101356/outlook_for_china_glyphosate_industry_20122016.

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryotic Cell* **5**: 1517–1531.

Scheldeman P, Baurain D, Bouhy R, Scott M, Mühling M, Whitton BA, Belay A, Wilmotte A. 1999. *Arthrospira* (*"Spirulina"*) strains from four continents are resolved into only two clusters, based on amplified ribosomal DNA restriction analysis of the internally transcribed spacer. *FEMS Microbiology Letters* **172**: 213–222.

Schirrmeister BE, Antonelli A, Bagheri HC. 2011. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology* **11**:45.

Schönbrunn E, Eschenburg S, Shuttleworth WA, Schloss JV, Amrhein N, Evans JN, Kabsch W. 2001. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proceedings of the National Academy of the Sciences of the U.S.A.* **98**:1376–1380.

Schulz A, Krüper A, Amrhein N. 1985. Differential sensitivity of bacterial 5-enolpyruvyl-shikimate-3-phosphate synthases to the herbicide glyphosate. *FEMS Microbiology Letters* **28**: 297–301.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of the Sciences of the U.S.A.* **110**: 1053–1058.

Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proceedings of the National Academy of the Sciences of the U.S.A.* **103**: 5442-5447.

Torstensson L. 1985. Behaviour of glyphosate in soils and its degradation. *In* The Herbicide Glyphosate. Edited by Grossbard E., Atkinson D. pp. 137–150. London: Butterworths.

Zhi XY, Yao JC, Li HW, Huang Y, Li WJ. 2014. Genome-wide identification, domain architectures and phylogenetic analysis provide new insights into the early evolution of shikimate pathway in prokaryotes. *Molecular Phylogenetics and Evolution* **75**: 154–164.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 MegAlign (version 7.1.0) multiple amino acid sequence alignment for EPSP synthases from cyanobacterial strains analysed in this work, whose sequences are available in public databases.

Fig. S2 MegAlign (version 7.1.0) pair distances obtained following multiple amino acid sequence alignment of cyanobacterial EPSP synthases analysed in this work.

Fig. S3 Phylogenetic tree of cyanobacterial EPSP synthases compared with that of 16S ribosomal DNAs.

Fig. S4 Maximum likelihood phylogenetic tree of a selection of enzymes that are representative of class I and class II EPSP synthases.

Table S5 Accession numbers of sequences used to generate the phylogenetic tree shown in Fig. 5.

Table S6 Accession numbers of sequences used to generate the phylogenetic tree shown in Fig. 6.

Fig. S7 MegAlign (version 7.1.0) multiple amino acid sequence alignment for EPSP synthases that are representative of class I and class II enzymes.

Table 1 Properties of EPSP synthases in strains of the genus *Arthrospira*

Strain		16S rDNA cluster	---shikimate dehydrogenase--- specific activity (nkat mg^{-1})	native mass (kDa)	----------EPSP synthase---------- specific activity (pkat mg^{-1})	native mass (kDa)	glyphosate IC_{50} (M)
Arthrospira sp. PCC 9438 (*S. platensis* C1)	-----	Ia	0.77 ± 0.11	58.3 ± 3.1	35 ± 5	91.4 ± 2.2	28 ± 2 x 10^{-6}
Arthrospira sp. PCC 9223	D933	Ia	0.82 ± 0.02	51.8 ± 2.6	43 ± 5	81.4 ± 6.9	32 ± 3 x 10^{-6}
Arthrospira sp. var. "Lonar"	D920	Ia	0.54 ± 0.14	61.2 ± 3.8	50 ± 11	105.5 ± 11.4	22 ± 4 x 10^{-6}
Arthrospira sp. PCC 8005	D914/H	Ib	0.63 ± 0.03	53.5 ± 2.8	35 ± 2	84.3 ± 7.4	19 ± 7 x 10^{-6}
Arthrospira fusiformis Hegewald 1976/83	D910/H	II	0.97 ± 0.15	55.7 ± 3.0	50 ± 7	111.2 ± 12.7	27 ± 4 x 10^{-6}
Arthrospira sp. (*platensis*) Compere 86.79	D905	II	1.22 ± 0.16	60.1 ± 3.6	41 ± 3	100.5 ± 10.3	32 ± 8 x 10^{-6}
Arthrospira sp. var. "Orovilca"	D921	II	0.93 ± 0.27	52.1 ± 2.6	43 ± 14	82.1 ± 7.0	22 ± 3 x 10^{-6}

With the exception of *Arthrospira* sp. PCC 9438, previously reported as "*Spirulina platensis*" C1 (Forlani *et al.*, 1991), strains were obtained from the culture collection at Durham University (UK) and are cited under the respective Durham strain designations (second column); 16S rDNA clusters are defined according to Scheldeman *et al.* (1999). Specific activity levels are mean ± SE over 3 independent determinations. Native molecular mass was estimated from retention patterns on Sephacryl S200 column; data from two runs carried out with independent enzyme preparations were subjected to non-linear regression analysis (Sigmoidal, 4PL, X is log[concentration]). Glyphosate susceptibility was assessed in two independent experiments in which active fractions from gel permeation were assayed in triplicates in the presence of increasing concentrations of the phosphonate ranging from 1 to 100 μM. Data were combined and analysed (log[inhibitor] vs. normalized response - variable slope). Values are reported with their 95% confidence intervals.

Table 2 Properties of EPSP synthases in strains of different genera representing the five subsections into which cyanobacteria are classified

Strain	subsection	PCC culture medium	EPSP synthase				
			specific activity (pkat mg^{-1})	subunit mass (kDa)	native mass (kDa)	holomer structure	glyphosate IC_{50} (M)
Gloeobacter violaceus PCC 7421	I	1539	29 ± 6	45.3	83.0 ± 6.9	dimer	$2.3 \pm 1.1 \times 10^{-3}$
Gloeothece sp. PCC 6909	I	1539	54 ± 11	n.a.	98.0 ± 6.2	dimer	$48 \pm 15 \times 10^{-6}$
Microcystis aeruginosa PCC 7941	I	1539	101 ± 21	47.6	79.8 ± 5.0	dimer	$55 \pm 9 \times 10^{-6}$
Synechococcus elongatus PCC 6301	I	1539	72 ± 14	47.5	95.4 ± 10.3	dimer	$26 \pm 2 \times 10^{-6}$
Synechococcus sp. PCC 6715	I	1539	42 ± 9	n.a.	80.1 ± 3.4	dimer	$78 \pm 11 \times 10^{-6}$
Synechococcus sp. PCC 7002	I	1540	172 ± 19	47.3	85.2 ± 1.2	dimer	$14 \pm 1 \times 10^{-6}$
Synechococcus sp. PCC 7335	I	1534	112 ± 5	42.4	110.7 ± 5.9	dimer	$120 \pm 37 \times 10^{-6}$
Synechococcus elongatus PCC 7942	I	1539	117 ± 19	47.5	100.2 ± 6.9	dimer	$28 \pm 3 \times 10^{-6}$
Synechocystis sp. PCC 6701	I	1539	36 ± 10	n.a.	92.2 ± 7.4	dimer	$7 \pm 3 \times 10^{-6}$
Synechocystis sp. PCC 6803	I	1539	68 ± 12	47.0	89.0 ± 6.3	dimer	$7 \pm 2 \times 10^{-6}$
Chroococcidiopsis thermalis PCC 7203	II	1539	51 ± 14	47.5	93.8 ± 11.7	dimer	$35 \pm 4 \times 10^{-6}$
Stanieria cyanospheaera PCC 7301	II	1534	33 ± 5	n.a.	91.5 ± 2.0	dimer	$168 \pm 89 \times 10^{-6}$
Lyngbya sp. PCC 7419	III	1539	122 ± 12	n.a.	82.8 ± 1.0	dimer	$80 \pm 21 \times 10^{-6}$
Coleofasciculus chthonoplastes PCC 7420*	III	1539	116 ± 9	n.a.	99.3 ± 1.8	dimer	$24 \pm 4 \times 10^{-6}$
Oscillatoria sp. PCC 6304	III	1539	26 ± 5	47.9	83.0 ± 1.8	dimer	$29 \pm 10 \times 10^{-6}$
Oscillatoria sp. PCC 7112	III	1539	51 ± 11	47.5	92.5 ± 5.4	dimer	$50 \pm 12 \times 10^{-6}$
Oscillatoria sp. PCC 7515	III	1539	23 ± 3	n.a.	84.7 ± 3.6	dimer	$21 \pm 7 \times 10^{-6}$
Leptolyngbya boryana PCC 6306	III	1539	84 ± 14	n.a.	87.2 ± 2.3	dimer	$17 \pm 2 \times 10^{-6}$

Calothrix sp. PCC 7102	IV	1539	46 ± 10	n.a.	87.4 ± 2.0	dimer	38 ± 9 x 10^{-6}
Calothrix sp. PCC 7507	IV	1539	57 ± 5	46.6	47.1 ± 1.7	monomer	9.9 ± 2.5 x 10^{-3}
Cylindrospermum licheniforme ATCC 29412	IV	1539	83 ± 17	n.a.	84.8 ± 4.6	dimer	13 ± 1 x 10^{-6}
Cylindrospermum sp. PCC 7604	IV	1539	403 ± 40	n.a.	105.5 ± 8.9	dimer	53 ± 6 x 10^{-6}
Nostoc muscorum PCC 7906	IV	1539	31 ± 7	n.a.	38.0 ± 0.3	monomer	8.7 ± 5.1 x 10^{-3}
Nostoc sp. PCC 6719	IV	1539	55 ± 13	n.a.	36.7 ± 1.1	monomer	9.8 ± 1.9 x 10^{-3}
Nostoc sp. PCC 7119	IV	1539	41 ± 11	n.a.	36.6 ± 1.0	monomer	5.6 ± 1.9 x 10^{-3}
Nostoc sp. PCC 7120	IV	1539	62 ± 10	46.5	31.6 ± 0.5	monomer	5.8 ± 0.9 x 10^{-3}
Nostoc sp. PCC 7937**	IV	1539	88 ± 11	46.7	33.2 ± 0.1	monomer	6.8 ± 0.7 x 10^{-3}
Nostoc punctiforme PCC 73102	IV	1539	44 ± 6	46.7	34.6 ± 0.5	monomer	>10 x 10^{-3}
Nostoc sp. PCC 7413	IV	1539	44 ± 7	n.a.	41.5 ± 0.5	monomer	>10 x 10^{-3}
Nostoc sp. PCC 7524	IV	1539	47 ± 11	47.5	83.7 ± 2.5	dimer	13 ± 4 x 10^{-6}
Rivularia sp. PCC 7116	IV	1540	54 ± 11	47.5	86.6 ± 1.8	dimer	253 ± 115 x 10^{-6}
Scytonema hofmanni PCC 7110	IV	1539	67 ± 15	n.a.	86.7 ± 4.0	dimer	49 ± 7 x 10^{-6}
Tolypothrix sp. PCC 7601 (*Fremyella diplosiphon*)	IV	1539	60 ± 8	n.a.	88.3 ± 2.2	dimer	14 ± 1 x 10^{-6}
Fischerella muscicola PCC 73103	V	1539	66 ± 9	n.a.	99.1 ± 2.1	dimer	46 ± 17 x 10^{-6}
Fischerella thermalis PCC 7521	V	1539	69 ± 13	n.a.	85.6 ± 9.4	dimer	34 ± 11 x 10^{-6}

* Previously named "*Microcoleus chthonoplastes*" PCC 7420; ** corresponds to "*Anabaena variabilis*" ATCC 29413. Specific activity levels are mean ± SE over 6 independent determinations. Subunit mass was calculated on the basis of available deduced protein sequences (supporting material S1); n.a., not available. Native molecular mass was estimated from retention patterns on Sephadex G100 column; data are mean ± SE over three runs carried out with independent enzyme preparations. Glyphosate susceptibility was assessed in three independent experiments in which active fractions from gel permeation were assayed in duplication in the presence of increasing concentrations of the phosphonate ranging from 3 μM to 10 mM. Data were combined and analysed (log[inhibitor] vs. normalized response - variable slope).

Legends to figures

Fig. 1. Evaluation of native molecular mass of EPSP synthases from *Arthrospira* strains (subsection III, *Oscillatoriales*). Extracts from cells harvested in the exponential phase of growth were fractionated by gel permeation on a Sephacryl S200 column, and retention patterns were compared to those obtained with protein molecular markers. Non-linear regression analysis led to an estimated mass of the holomer ranging from 81 to 105 kDa, and suggested for the enzyme a homodimeric structure. As an additional term of comparison, the elution profiles of shikimate dehydrogenases were also determined. Since in both bacteria and higher plants the latter are homodimers with relative molecular masses of about 60 kDa (Dev *et al.*, 2012), higher elution volumes for EPSP synthase than for shikimate dehydrogenase are consistent with this conclusion.

Fig. 2. Evaluation of native molecular mass of EPSP synthases from 15 cyanobacterial strains of the *Nostocales* (subsection IV). Extracts from cells harvested in the exponential phase of growth were fractionated by gel permeation on a Sephacryl S200 column, and retention patterns were compared to those obtained with protein molecular markers. Non-linear regression analysis yielded heterogeneous results, with two datasets of estimated masses, one ranging from 32 to 47 kDa, the other from 84 to 106 kDa. The deduced molecular mass of the subunit being around 46 kDa (Powell *et al.*, 1992, Table 2 in this work), the data suggest that either monomeric or homodimeric enzymes may occur among the representatives of subsection IV.

Fig. 3. Sensitivity to glyphosate of EPSP synthases from three cyanobacterial strains of the *Nostocales* (subsection IV). Following gel permeation chromatography, the enzyme was assayed in the presence of increasing concentrations of the herbicide. Results were expressed as percent of activity in untreated controls, and are mean ± SE over six replicates obtained with three different enzyme preparations. Non-linear regression of data allowed the calculation of IC_{50} values, which were $13 \pm 1 \times 10^{-6}$ M, $5.3 \pm 0.6 \times 10^{-5}$ M and $5.8 \pm 0.9 \times 10^{-3}$ M for *Cylindrospermum licheniforme* ATCC 29412, *Cylindrospermum* sp. PCC 7604 and *Nostoc* sp. PCC 7120, respectively.

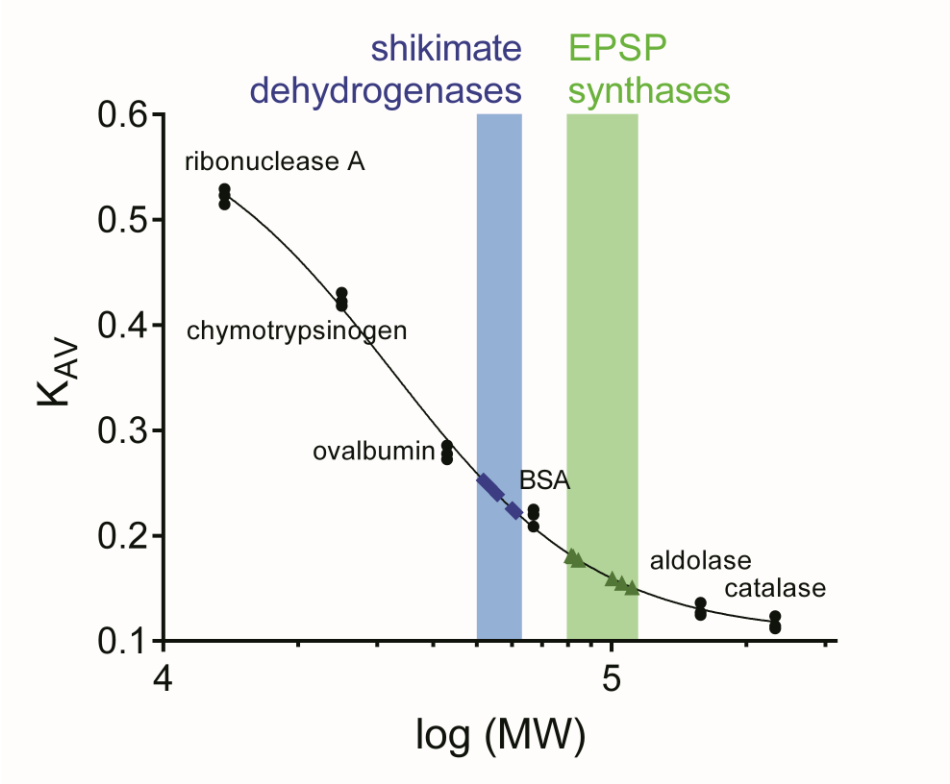
Fig. 4. Relationship between quaternary structure and sensitivity to glyphosate of

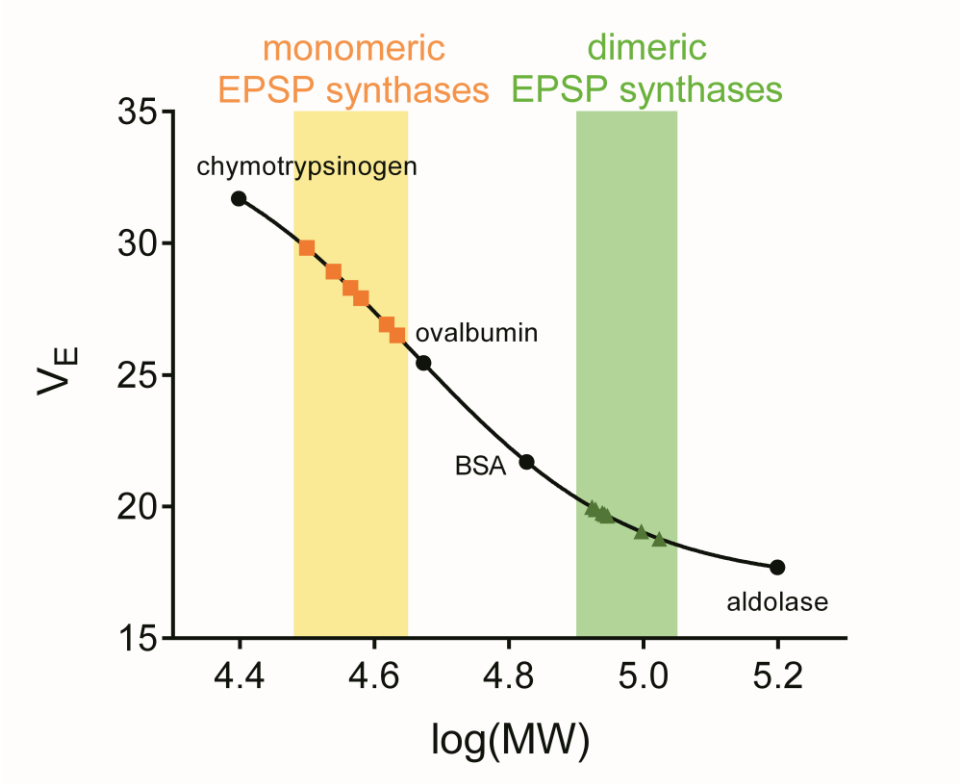
cyanobacterial EPSP synthases. IC_{50} values were plotted against the estimated relative molecular mass under native conditions. With the only exception of the enzyme from the evolutionarily ancient *Gloeobacter violaceus* PCC 7421, which displayed a significantly higher tolerance to the herbicide, all dimeric EPSP synthases showed IC_{50} values in the 10^{-5} to 5 x 10^{-4} M range. Conversely, all monomeric enzymes showed a remarkable tolerance to glyphosate, with IC_{50} values higher than 5 x 10^{-3} M. Strains are labelled according to their subdivision: I, *Chroococcales*; II, *Pleurocapsales*; III, *Oscillatoriales*; IV, *Nostocales*; V, *Stigonematales*.

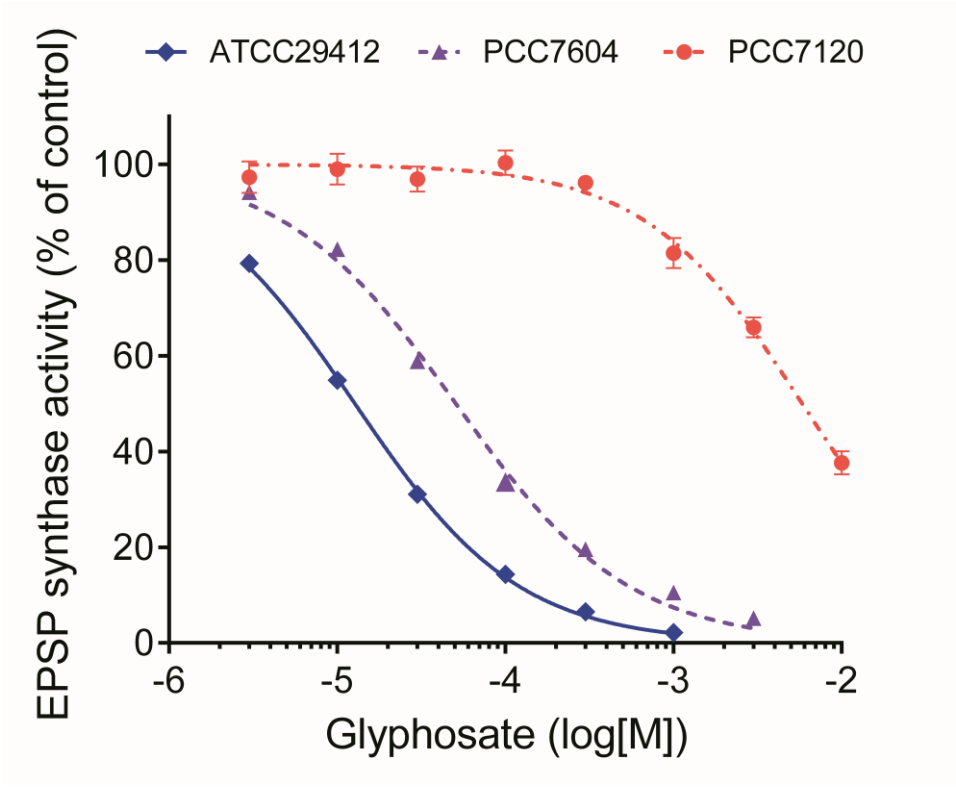
Fig. 5. Maximum likelihood phylogenetic tree of cyanobacterial EPSP synthases. Available sequences were retrieved from Cyanobase (http://genome.microbedb.jp/CyanoBase#resources) and NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) databanks. Accession numbers are reported in supporting material S5. The tree was generated using MegAlign (version 7.1.0) after multiple amino acid sequence alignment by the Clustal W method. Branch length is proportional to the phylogenetic distance in amino acid substitutions *per* site. All strains are labelled with the respective identifiers, followed by coloured symbols for their appropriate subsection assignments, which correspond to the following botanical orders: I, *Chroococcales*; II, *Pleurocapsales*; III, *Oscillatoriales*; IV, *Nostocales*; V, *Stigonematales*. Two bacterial sequences (*Chlorobium tepidum* ATCC 49652, and *Rhodopseudomonas* CGA009) were also included for comparison. The strains analysed in this work, whose sequences are reported in supporting material S1, are emphasized in yellow (monomeric EPSP synthases) or green (dimeric enzymes). Two separate and remarkably distant cyanobacterial clades are evident. The sequence shown for *Nostoc* sp. PCC 7937 corresponds to that published under the strain designation "*Anabaena variabilis*" ATCC 29413.

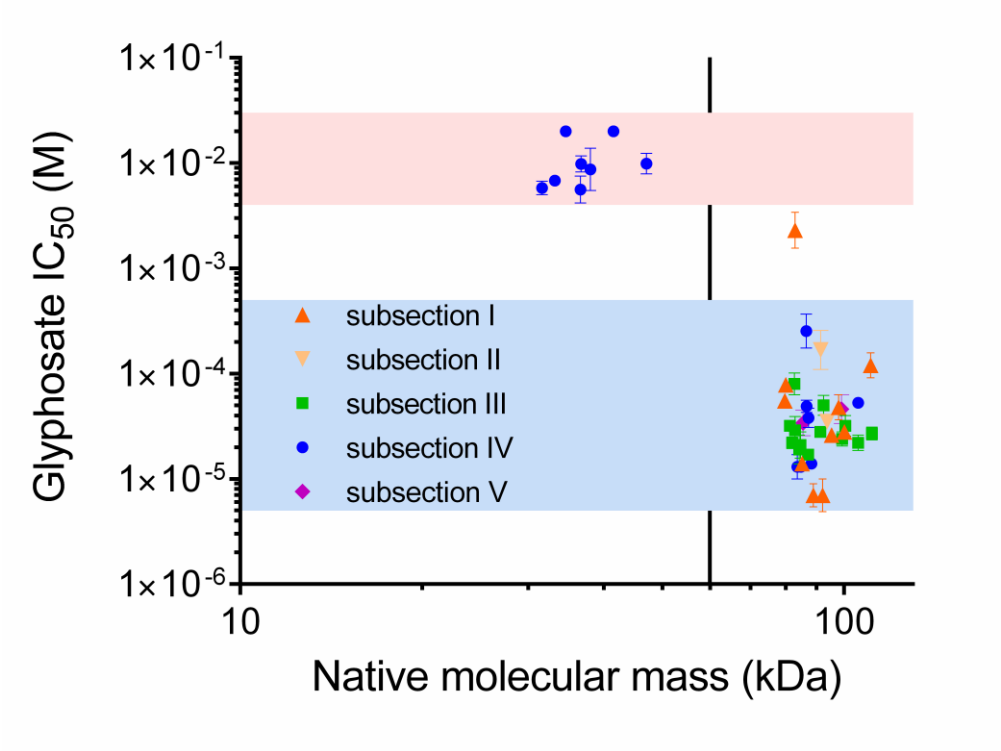
Fig. 6. Simplified fast minimum evolution phylogenetic tree (with a 0.85 maximum sequence difference and Grishin [protein] distance) of manually annotated and reviewed sequences coding for EPSP synthases. Sequences homologous to the *Gloeobacter violaceus* PCC 7421 EPSP synthase, as listed in supporting material S6, were retrieved from the UniProt KB/Swiss-Prot database using the BLAST-p program with its default parameters, but excluding uncultured organisms/environmental samples, and setting the maximum target to 500 sequences. The three dimeric, glyphosate-sensitive forms of the

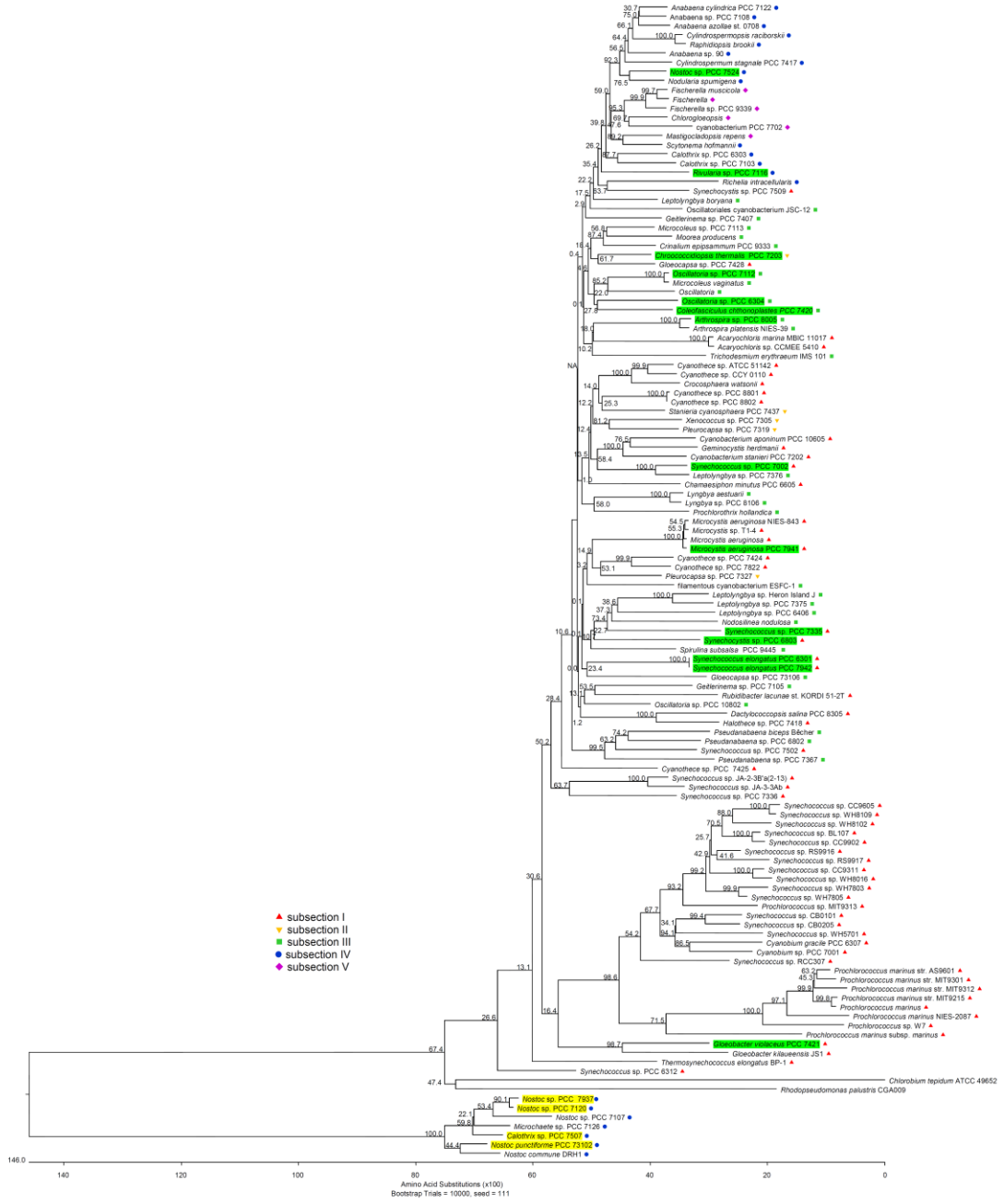
enzymes included in this tree clustered together with their cyanobacterial relatives (light-green-shaded area), near the proteins of phylogenetically close bacterial phyla. The distance of this cyanobacterial clade is extensive compared to the three monomeric, glyphosate-resistant cyanobacterial enzymes of the *Nostocales* (light-yellow-shaded area), which are positioned near evolutionarily unrelated phyla, such as euryarcheotes, ascomycetes and dicotyledons. The proteins of the strains characterized in this work are emphasized in deep-yellow (monomeric) or deep-green (dimeric enzymes). The plant enzymes are indicated with an arrow.

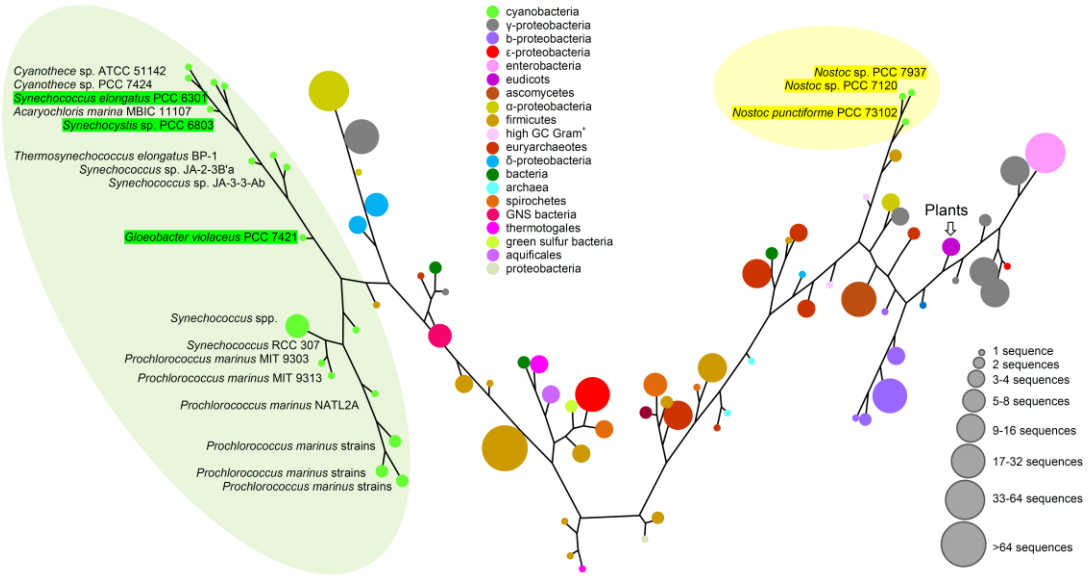

shikimate dehydrogenases
EPSP synthases
ribonuclease A
chymotrypsinogen
ovalbumin
BSA
aldolase
catalase
K_{AV}
0.6
0.5
0.4
0.3
0.2
0.1
4
5
log (MW)


monomeric EPSP synthases
dimeric EPSP synthases
chymotrypsinogen
ovalbumin
BSA
aldolase
V_E
35
30
25
20
15
4.4
4.6
4.8
5.0
5.2
log(MW)


ATCC29412
PCC7604
PCC7120
EPSP synthase activity (% of control)
100
80
60
40
20
0
-6
-5
-4
-3
-2
Glyphosate (log[M])


Glyphosate IC50 (M)
1×10-1
1×10-2
1×10-3
1×10-4
1×10-5
1×10-6
subsection I
subsection II
subsection III
subsection IV
subsection V
10
100
Native molecular mass (kDa)


subsection I
subsection II
subsection III
subsection IV
subsection V
Anabaena cylindrica PCC 7122
Anabaena sp. PCC 7108
Anabaena azollae st. 0708
Cylindrospermopsis raciborskii
Raphidiopsis brookii
Anabaena sp. 90
Cylindrospermum stagnale PCC 7417
Nostoc sp. PCC 7524
Nodularia spumigena
Fischerella muscicola
Fischerella
Fischerella sp. PCC 9339
Chlorogloeopsis
cyanobacterium PCC 7702
Mastigocladopsis repens
Scytonema hofmannii
Calothrix sp. PCC 6303
Calothrix sp. PCC 7103
Rivularia sp. PCC 7116
Richelia intracellularis
Synechocystis sp. PCC 7509
Leptolyngbya boryana
Oscillatoriales cyanobacterium JSC-12
Geitlerinema sp. PCC 7407
Microcoleus sp. PCC 7113
Moorea producens
Crinalium epipsammum PCC 9333
Chroococcidiopsis thermalis PCC 7203
Gloeocapsa sp. PCC 7428
Oscillatoria sp. PCC 7112
Microcoleus vaginatus
Oscillatoria
Oscillatoria sp. PCC 6304
Coleofasciculus chthonoplastes PCC 7420
Arthrospira sp. PCC 8005
Arthrospira platensis NIES-39
Acaryochloris marina MBIC 11017
Acaryochloris sp. CCMEE 5410
Trichodesmium erythraeum IMS 101
Cyanothece sp. ATCC 51142
Cyanothece sp. CCY 0110
Crocosphaera watsonii
Cyanothece sp. PCC 8801
Cyanothece sp. PCC 8802
Stanieria cyanosphaera PCC 7437
Xenococcus sp. PCC 7305
Pleurocapsa sp. PCC 7319
Cyanobacterium aponinum PCC 10605
Geminocystis herdmanii
Cyanobacterium stanieri PCC 7202
Synechococcus sp. PCC 7003
Leptolyngbya sp. PCC 7376
Chamaesiphon minutus PCC 6605
Lyngbya aestuarii
Lyngbya sp. PCC 8106
Prochlorothrix hollandica
Microcystis aeruginosa NIES-843
Microcystis sp. T1-4
Microcystis aeruginosa
Microcystis aeruginosa PCC 7941
Cyanothece sp. PCC 7424
Cyanothece sp. PCC 7822
Pleurocapsa sp. PCC 7327
filamentous cyanobacterium ESFC-1
Leptolyngbya sp. Heron Island J
Leptolyngbya sp. PCC 7375
Leptolyngbya sp. PCC 6408
Nodosilinea nodulosa
Synechococcus sp. PCC 7336
Synechocystis sp. PCC 6803
Spirulina subsalsa PCC 9445
Synechococcus elongatus PCC 6301
Synechococcus elongatus PCC 7942
Gloeocapsa sp. PCC 73106
Geitlerinema sp. PCC 7105
Rubidibacter lacunae st. KORDI 51-2T
Oscillatoria sp. PCC 10802
Dactylococcopsis salina PCC 8305
Halothece sp. PCC 7418
Pseudanabaena biceps Bêcher
Pseudanabaena sp. PCC 6802
Synechococcus sp. PCC 7502
Pseudanabaena sp. PCC 7367
Cyanothece sp. PCC 7425
Synechococcus sp. JA-2-3B'a(2-13)
Synechococcus sp. JA-3-3Ab
Synechococcus sp. PCC 7336
Synechococcus sp. CC9605
Synechococcus sp. WH8109
Synechococcus sp. WH8102
Synechococcus sp. BL107
Synechococcus sp. CC9902
Synechococcus sp. RS9916
Synechococcus sp. RS9917
Synechococcus sp. CC9311
Synechococcus sp. WH8016
Synechococcus sp. WH7803
Synechococcus sp. WH7805
Prochlorococcus sp. MIT9313
Synechococcus sp. CB0101
Synechococcus sp. CB0205
Synechococcus sp. WH5701
Cyanobium gracile PCC 6307
Cyanobium sp. PCC 7001
Synechococcus sp. RCC307
Prochlorococcus marinus str. AS9601
Prochlorococcus marinus str. MIT9301
Prochlorococcus marinus str. MIT9312
Prochlorococcus marinus str. MIT9215
Prochlorococcus marinus
Prochlorococcus marinus NIES-2087
Prochlorococcus sp. W7
Prochlorococcus marinus subsp. marinus
Gloeobacter violaceus PCC 7421
Gloeobacter kilaueensis JS1
Thermosynechococcus elongatus BP-1
Synechococcus sp. PCC 6312
Chlorobium tepidum ATCC 49652
Rhodopseudomonas palustris CGA009
Nostoc sp. PCC 7937
Nostoc sp. PCC 7120
Nostoc sp. PCC 7107
Microchaete sp. PCC 7126
Calothrix sp. PCC 7507
Nostoc punctiforme PCC 73102
Nostoc commune DRH1
146.0
140
120
100
80
60
40
20
0
Amino Acid Substitutions (x100)
Bootstrap Trials = 10000, seed = 111


cyanobacteria
γ-proteobacteria
b-proteobacteria
ε-proteobacteria
enterobacteria
eudicots
ascomycetes
α-proteobacteria
firmicutes
high GC Gram+
euryarchaeotes
δ-proteobacteria
bacteria
archaea
spirochetes
GNS bacteria
thermotogales
green sulfur bacteria
aquificales
proteobacteria
Cyanothece sp. ATCC 51142
Cyanothece sp. PCC 7424
Synechococcus elongatus PCC 6301
Acaryochloris marina MBIC 11107
Synechocystis sp. PCC 6803
Thermosynechococcus elongatus BP-1
Synechococcus sp. JA-2-3B'a
Synechococcus sp. JA-3-3-Ab
Gloeobacter violaceus PCC 7421
Synechococcus spp.
Synechococcus RCC 307
Prochlorococcus marinus MIT 9303
Prochlorococcus marinus MIT 9313
Prochlorococcus marinus NATL2A
Prochlorococcus marinus strains
Prochlorococcus marinus strains
Prochlorococcus marinus strains
Nostoc sp. PCC 7937
Nostoc sp. PCC 7120
Nostoc punctiforme PCC 73102
Plants
1 sequence
2 sequences
3-4 sequences
5-8 sequences
9-16 sequences
17-32 sequences
33-64 sequences
>64 sequences