Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
BIOLOGIA EVOLUZIONISTICA ED AMBIENTALE

CICLO XXIV

COORDINATORE Prof. Barbujani Guido

Characterization of the epigenetic diversity in cultivated and wild tomato species

Settore Scientifico Disciplinare BIO/04

Tutore
Dr. Bernacchia Giovanni

_____

Dottorando
Dott. Rainieri Massimo

_____
(firma)

Tutore
Dr. Gallusci Philippe

_____
(firma)

Anni 2009/2011

SUMMARY

Solanum Sect. *Lycopersicon* is a relatively small monophyletic clade that consists of 14 closely related species including the domesticated tomato, *Solanum lycopersicum* (formerly *L. esculentum*).

Tomato and its wild relatives are native of western South America along the coast and high Andes from central Ecuador, through Peru and northern Chile. One species is also found in Mexico and one is endemic to the Galapagos Islands (Nakazato et al 2010).

Although the classification and phylogeny of *Solanum* section *Lycopersicon* is a complex issue that has not yet reached a widely accepted consensus, different works using different approaches based on morphology, gene sequence analysis and metabolomic have been use to characterize wild tomato species (Peralta and Spooner 2007; Spooner, 2005; Zuriaga et al.,2009; Steinhauser M.C. et al., 2010).

In addition the tomato wild species are an important source of germplasm to increase the resistance of the cultivated tomato to biotic and abiotic stress.

Aim of our study is to characterize wild tomato species by a fine analysis of different aspects of fruit developmental process, including fruit size, cytological characterization of pericarp development, ploidy level analysis and gene expression measurement. In addition, since genomic DNA is subjected to tissue specific changes in DNA methylation levels and patterns during fruit development (Teyssier at al, 2008), we have investigated various epigenetic parameters in fruit of wild tomato species: these include DNA methylation analysis and polycomb gene characterization.

Our results clearly showed that despite the different colour and morphology of the fruits other differences can be seen during fruit development and ripening. For example, a red species such as *S.l. cerasiforme*, when compared to *WVa106* and *S. pimpinellifolium*, showed a change in colour of its locular tissue that occurred before the same change in the pericarp. The pericarp analysis then showed that in *WVa106* the increase in the pericarp thickness correlated with cell size increase via cell expansion (as showed from Cheniclet et al., 2005) while in the wild tomato species (except *S. neorikii*) the cell division plays an essential function during the whole process of fruit development. Furthermore, the expression analysis of the ripening phase of *RIN* and *PSYI* genes, acting upstream and downstream of the ethylene cycle, showed that it is possible to correlate the

breaker, orange and red ripe stage of the red-fruited species with the 40, 50 and 60dpa of green fruited species.

The timing of fruit development is strongly connected to the endopolyploidization, which is correlated to cell and fruit size (Nafati et al., 2010; Cheniclet et al., 2005 and Bourdon et al., 2010). The analysis of this aspect in cultivated and wild tomato species showed that red-fruited species had a maximum C value varying between 128C (*S. pimpinellifolium*) and 256C (*WVa106* and *S.l. cerasiforme*) while green-fruited species had a C value between 64C (*S. neorikii, S. arcanum* and *S. pennellii*) and 128C (*S. corneliomulleri* and *S. huayalasense)*. We also showed that the linear correlation found between fruit size and endoreduplication and endoreduplication and fruit size *(Nafati et al., 2010*) was not present in all the species analyzed. Species such as *S. pimpinellifolium, S. pennellii* and *S. huayalasense* showed a cell size increase without increase in the endoreduplication level.

In an effort to analyze the epigenetic variations between species, we analysed the DNA methylation in leaves and pericarp DNA. The global DNA methylation analysed on leaf DNA by HPLC did not show differences between species. In addition the analysis of the 5S locus specific DNA methylation in leaves by DNA digestion using *Hpa*II and *Msp*I enzymes showed a high methylation level at CG sequences while at the CNG sequences *S. neorikii, S. arcanum, S. corneliomulleri, S. chilense, S. huayalasense* and *S. habrochaites* were more methylated as compared to *WVa106, S.l. cerasiforme, S. pimpinellifolium* and *S. chmielewskii*.

Similar analysis on genomic DNA from pericarp showed a tissue specific DNA methylation variation during fruit development at 5S rDNA sequences. *WVa106*, like *S. arcanum*, *S. corneliomulleri* and *S. huayalasense* had a low level of methylation at 20dpa with an increase at breaker and orange stage for the red fruited species *WVa106* and at 40 and 60dpa for the green-fruited species. No variation in DNA methylation was observed in *S. pennellii* while in *S.l. cerasiforme* and *S. pimpinellifolium* the 5S locus specific methylation at CNG increased during fruit development. The differences do not overlap with the differences observed at ploidy level. A global methylation analysis performed on genomic DNA of fruits at 10dpa showed that species with a high C value (*WVa106* and *S.l cerasiforme*) had a low level of methylation compared to the other species that showed a lower C value and high methylation level.

The epigenetic variability between species has been also analyzed by the sequencing of two paralogues genes *SlEZ2* and *SlEZ3* of the Enhencer of zeste (E(z)) family and members of the Polycomb group proteins (PcG). The PcG were first discovered in the

fruit fly *Drosophila melanogaster* and their characterization has revealed that they work in a complex way. While in mammalian two polycomb repressive complex (PRC1 and PRC2) have been identified, in plants only the PRC2 has been identified. Although the mechanism of action of the PcG is not completely understood, methylation of histone H3 lysine 27 (H3K27) is important in establishing PcG-mediated transcriptional repression. The gene diversity has been analyzed by sequencing the whole genes in cultivated tomato (*WVa106*) and the wild relatives (*S. pimpinellifolium, S. neorikii, S. arcanum, S. pennellii*).

By sequencing *SlEZ2* and *SlEZ3* transcripts we showed for the first time that *SlEZ3* is subjected to alternative splicing, which produces three transcripts one of which encodes a *SlEZ3* protein 841 amino acid long bearing all the characteristics domains of EZ proteins, thus suggesting that it could be completely functional. The two other *SlEZ3* transcripts encode a truncated protein lacking the SET domain which has the catalytic activity.

In addition the analysis of *SlEZ3* and *SlEZ2* transcripts (which did not show the alternative splicing) showed that they are highly conserved between species suggesting that the proteins are functional and conserved.

Expression analysis of *SlEZ2* and *SlEZ3* during fruit development showed that *SlEZ2* was actively transcribed in almost all the species (except *S. pimpinellifolium* that showed an increase of expression during fruit development) at the first stages of fruit development and that its expression decreased during fruit development. On the other hand, *SlEZ3* was generally expressed at low level with the peak of expression at 40 and/or 50 dpa.

Differences for the two genes were also evident at the genomic level. Although is thought that *SlEZ2* is a duplication of *SlEZ3*, *SlEZ2* is a gene approximately 9.5Kb-long organized in 16 exons while *SlEZ3* is approximately 17.2Kb-long and it is organized in 21 exons. In addition *SlEZ3* has the intron 17 varying in size between 6000bp (*S. neorikii*) and 6835bp (*S. arcanum*) in the different wild species. In this intron we identified a LINE-like retrotransposon classified TERT003 and an unclassified repeat OTOT000.

Although the gene sequences did not show important differences between species, in the promoter region of *SlEZ2* we identified a *Ty3/gypsy*-like retrotransposon member of the *Galadriel* family. Interestingly, this element is present at this locus only in red/orange fruited species (*WVa06, S.l. cerasiforme, S. cheesmaniae, S. pimpinellifolium*) and absent in the green-fruited species. In addition we also found that this retrotransposon is more abundant in the genome of the red-fruited specie than in the green ones and that its

retrotranscriptase (RT) is actively transcribed in almost all the species (except for *S. corneliomulleri*).

In the genome the retrotransposon are usually silenced by epigenetic mechanisms (Rigal and Mathieu, 2011) such as DNA methylation and histone modifications. The comparison between the methylation status in the promoter region of *SlEZ2* of the species with the retrotransposon and the species without retrotransposon showed differences between species. Red-fruited species showed locus specific variations in DNA methylation in agreement with the *SlEZ2* gene expression profile while green-fruited species without retrotransposon did not show any locus specific change in DNA methylation in agreement with gene expression profile thus suggesting that *SlEZ2* regulation in red fruited species the SlEZ2 gene is regulated by DNA methylation while it is independent of the DNA methylation in the in green-fruited ones.

The results of this work show that despite the different colour and morphology other differences can be identified between species concerning the dynamics of fruit development and ripening. This study has also showed that the cytological relationships known in the cultivated tomato are only partially observed in the wild species.

In addition, epigenetic diversity has been found during fruit development among the different species in terms of DNA methylation and organization of the *SlEZ2* and *SlEZ3* genes.

All together these observations suggest that the fruit development and ripening are a complex process still not completely understood and that wild tomato species represent a useful tool for these studies.

ABBREVIATIONS

A	Adenite
ACC	Aminocyclopropane-1-Carboxylate
ACS	Aminociclopropane-1-Carboxylate Synthase
ACO	Aminocyclopropane-1-Carboxylate Oxidase
aDMA	Asymmetric N^G,N^G-di-methylarginine
AFLP	Amplification Fragment Length Polymorphism
AGO	Argonauta
ASH	Absent Small Homeotic disc
BAC	Bacterial Artificial Chromosome
Br	Breaker
C	Cytosine
CDK	Cyclin-Dependent Kinase
CYC	Cyclin
CLF	Curly Leaf
CNR	Colorless Non Ripening
CMT	Chromo-Methyltransferase
CP	Capsid-like Protein
CTAB	Hexadecyltrimethylammonium bromide
CTR	Constitutive Triple Response
D	Aspartic acid
DCL	Dicer Like
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic Acid
Dnmt	DNA methyltransferase
Dpa	Days post anthesis
DTT	Dithiothreitol
DRM	DNA methyltransferase Rearranged Domain
dsRNA	double strand RNA
E	Glutamic acid
EDTA	ethylenediaminetetraacetic acid
EI	Endoreduplication Index
EIN	Ethylene Insensitive
EMF	Embryonic Flower
EN	Endonuclease
ESC	Extra Sex Combs
ETR	Ethylene Receptor
E(z)	Enhance of zeste
FIE	Fertilization independent endosperm
FLC	Flowering locus C
FIS	Fertilization Independent Seed
HAT	Histone Acetyl Transferase
HDAC	Histone Deacetilase
HKMT	Histone Lysine Methyltransferase
HMT	Histone MethylTransferase
HPLC	High Pressure Liquid Chromatography
INT	Integrase
IR	Inverted Repeat
K	Lysine

LB	Luria Bertani
LHI	Like Heterochromatin protein
LTR	Long Terminal Repeat
M	Methyonine
MBD	Methyl CpG Binding Domain
MCV	Mean C Value
MEA	Medea
MET1	Methyltransferase 1
MMA	Mono-MethylArginine
MSI	Multicopy Suppressor of Ira
NAD	Nicotinamide Adenine Dinucleotide
nDNA	nuclear Deoxyribonucleic Acid
NLS	Nuclear Localization Signal
NOR	Non Ripening
Nr	Never ripe
Or	Orange
ORF	Open Reading Frame
PBS	Primer Binding Site
PcG	Polycomb Group Protein
PCR	Polymerase Chain Reaction
Pepc	phosphoenolpiruvate carboxylase
PhoRC	Pleiohomeotic Repressive Complex
PPT	Polypurine Tracts
PSC	Posterior Sex Comb
PR	protease
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PRE	Polycomb Repressive Element
PRMT	Protein Arginine Methyltransferase
PSY	Phytoene Synthase
QTL	Quantity Trait Loci
R	Arginine
RdDM	RNA directed DNA methylation
RIN	Ripening Inhibitor
RNA	Ribonucleic Acid
ROS	Repressor of Silencing
RR	Red Ripe
RT	Reverse Transcriptase
SAM	S-Adenosilmethyonine
sDMA	Symmetric N^G,N^G-di-methylarginine
SDS	Sodium Dodecyl Sulphate
SINE	Short Interspersed Nuclear Element
siRNA	small interference RNA
SSC	Sodium chloride Sodium Citrate
SU(VAR)3-9	Suppressor of variegation 3-9
SWN	Swinger
T	Thymine
TE	Transposable Element
TGS	Transcriptional Gene Silencing
TGRC	Tomato Genetic Resource Center

TIR	Terminal Inverted Repeat
TRE	Trithorax Repressive Element
TRX	Trithorax
TrxG	Trithorax group Protein
TSD	Target Site Duplication
Tur	Turning
UTR	Untranslated Region
V	Valine
VIGS	Virus Induced Gene Silencing
VRN	Vernalization

CONTENTS

CHAPTER 1

TOMATO (*Solanum L.*) AND ITS WILD RELATIVES

1.1 - TOMATO AND ITS WILD RELATIVES (SOLANACEAE): ORIGIN, HISTORY AND CLASSIFICATION.

1.1.1 – Origin and distribution.

Solanum Sect. *Lycopersicon* is a relatively small monophyletic clade that consists of 14 closely related species including the domesticated tomato, *Solanum lycopersicum* (formerly *L. esculentum*) (Fig.1A). Tomato and its wild relatives are native of western South America along the coast and high Andes from central Ecuador, through Peru, to northern Chile. One species is also found in Mexico, and one is endemic to the Galapagos Islands (Fig.1B).

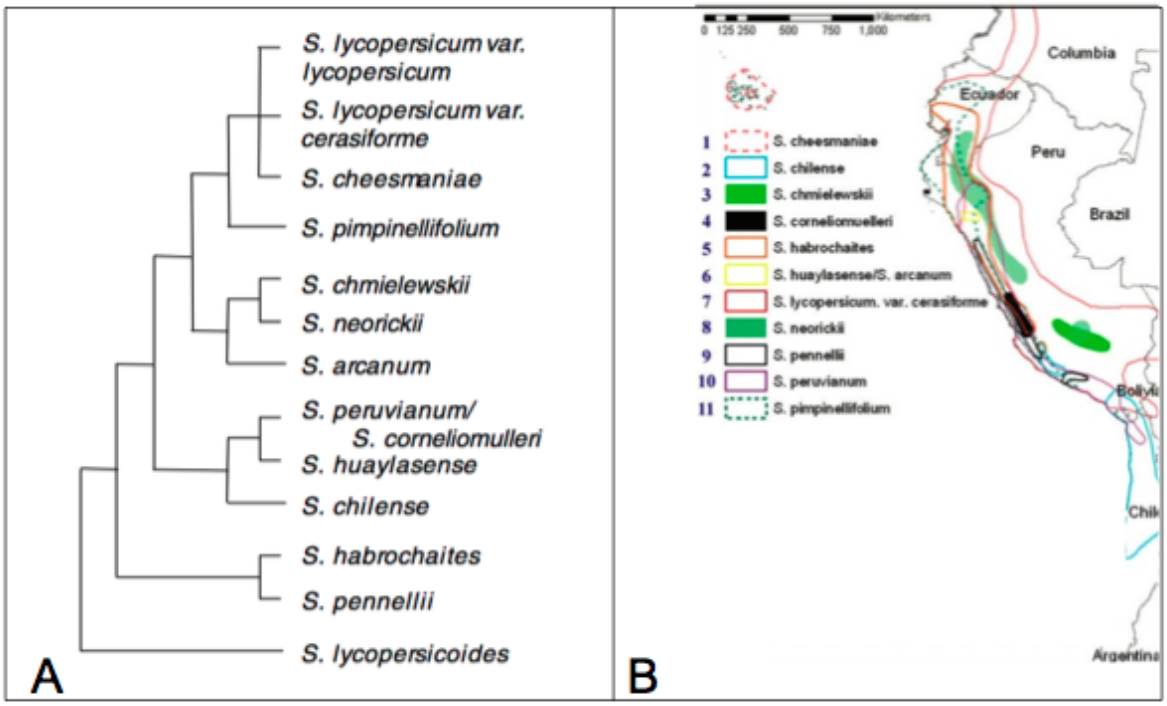

Fig.:1. Tomato phylogenetic tree and geographical distribution.

Although the cultivated tomato arrived in Europe in the early 1500s, it was mainly distributed only in Spain and Italy and it was not widely used until the late 1700. For two centuries the fragmented information about this plant (mainly used to make drugs) increased the enigma of the domestication place. To date, two competing hypotheses have been advanced to establish the place of the domesticated tomato. DeCandolle (1886) advanced the Peruvian hypothesis while Jenkins (1948) developed the Mexican hypothesis (Peralta and Spooner 2007). While the Peruvian hypothesis suggests that tomato was domesticated before the discovery of America but not much earlier, the Mexican hypothesis suggests an introduction and domestication in Mexico in pre-Columbian times (Peralta and Spooner 2007). The only putative archaeological evidence that can solve this enigma are flowers decorated on a ceramic produced by the Quimbaya culture (500-1000 AD) of Colombia (McMeekin, 1992). On the other hand, these flowers are not unequivocally tomato flowers but could be others *Solanum* flowers (possibly potato) (Peralta and Spooner 2007).

1.1.2 – Plant description

Tomato is a perennial herbaceous plant that grows in different habitats (from near sea level to over 3300m of altitude), and has had an impact on the evolution of the different tomato species and subspecies (to date, almost 1200 subspecies) (Rick 1973; TGRC).

Different tomato species and subspecies possess differences in terms of leaf (Fig.2), inflorescence, flower (Fig.3), fruit (Fig.4) and seed characters. The basic inflorescence is a cyma with different branching patterns and with or without axial bracts. Flowers are typically yellow with different sizes. Fruit size, colour and pubescence are also variable. Furthermore, the fruits are usually bilocular in the wild species and bilocular or multilocular in the cultivated varieties.

Moreover, the fertilization system can vary between cross-pollinating (allogamus) species to self-pollinating (autogamous) species (Chen et al., 2007). These two fertilization systems have been associated by Rick in the 1982 (Rick 1982a) with the morphology of flowers. His hypothesis was that large flowers and great stigma, exserted beyond the anthers, are more likely to receive pollen from other plants.

This idea has been recently confirmed at molecular level (Chen et al., 2007). The discovery of *Style2.1* gene as the major quantitative trait locus (QTL), responsible for the key floral attribute of style length, is associated with the evolution of self-pollination in cultivated species. This gene encode a putative transcription factor that regulates cell elongation in developing styles, furthermore it has been shown that the transition from cross-pollination to self-pollination is accompanied with a mutation in the *style2.1* promoter (Chen et al., 2007) (fig.5).

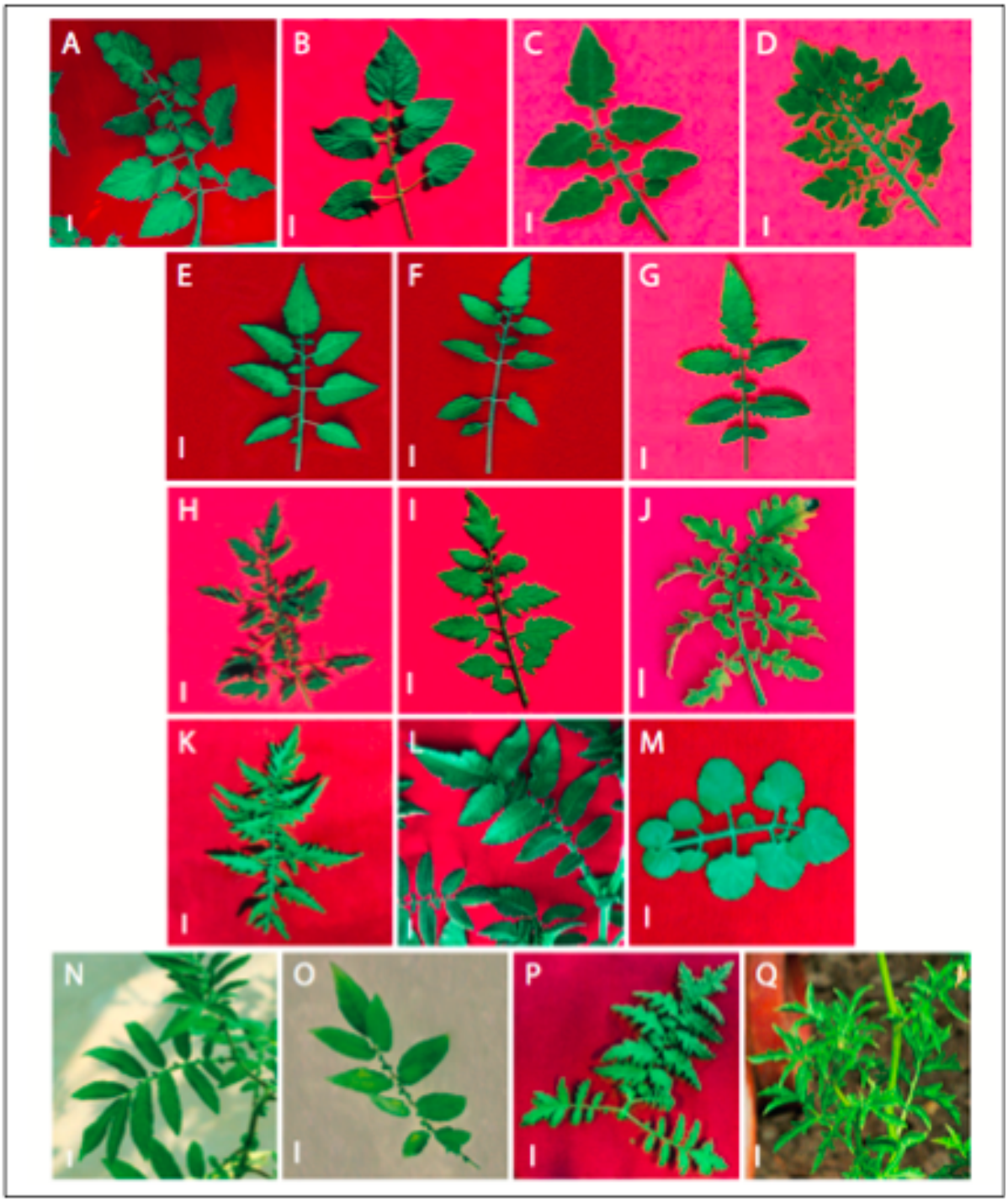

Fig.:2. Leaves of tomato and outgroup species. A, *S. lycopersicum*; B, *S. pimpinellifolium*; C, *S. cheesmaniae*; D, *S. galapagense*; E, *S. neorickii*; F, *S. chmielewskii*; G, *S. peruvianum* northern population; H, *S. peruvianum* northern population; I, *S. peruvianum* southern population; J, *S. peruvianum* southern population; K, *S. chilense*; L, *S. habrochaites*; M, *S. pennellii*; N, *S. ochranthum*; O, *S. juglandifolium*; P, *S. lycopersicoides*; Q, *S. sitiens*. Scale bars = 1 cm in C and D, 2 cm in A, B, E, F, G, H, I, J, K, M, P and Q, and 3 cm in L, N and O. (Spooer et al., 2005)

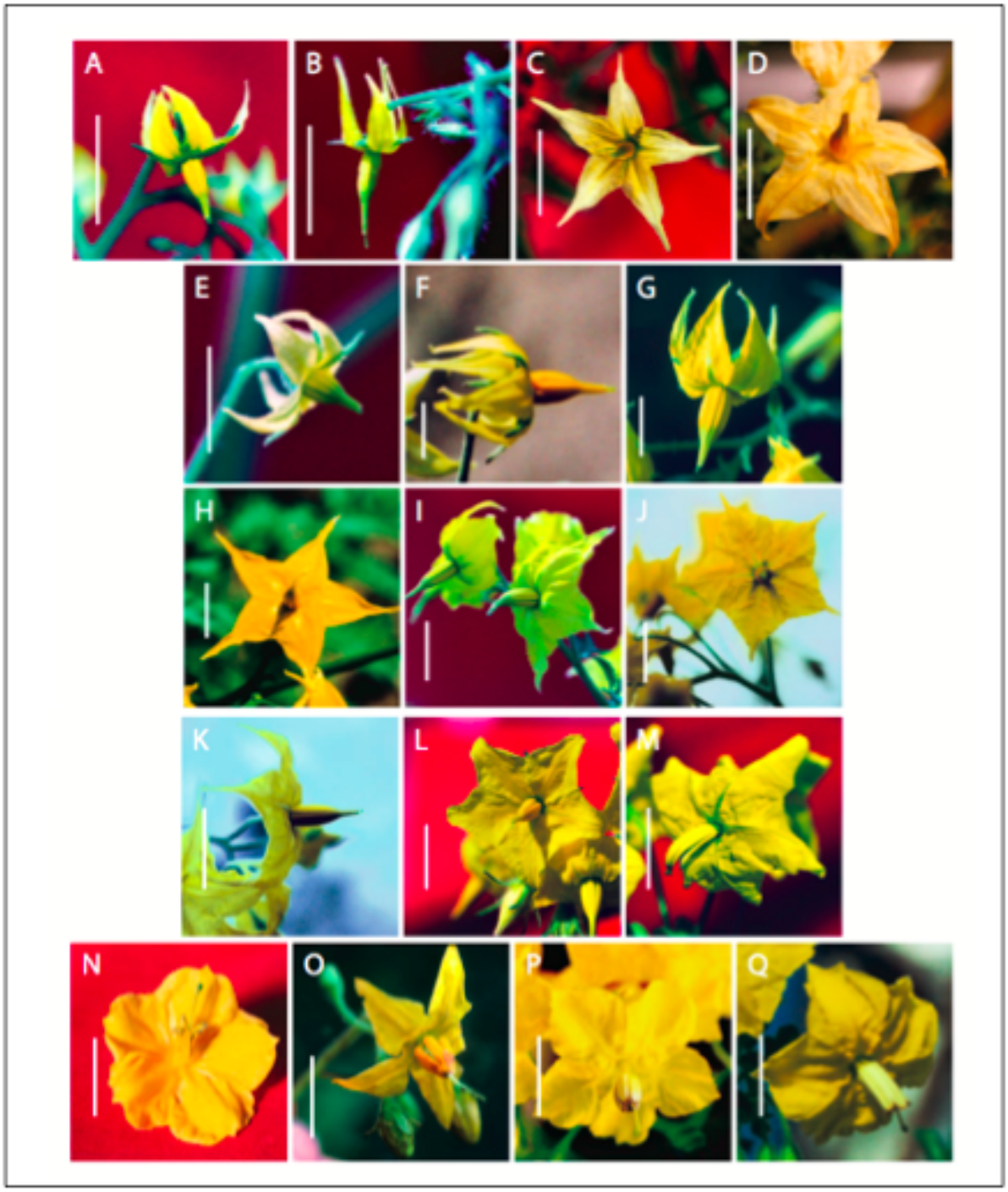

Fig.:3. Flowers of tomato and outgroup species. A, *S. lycopersicum*; B, *S. pimpinellifolium*; C, *S. cheesmaniae*; D, *S. galapagense*; E, *S. neorickii*; F, *S. chmielewskii*; G, *S. peruvianum* northern population; H, *S. peruvianum* northern population; I, *S. peruvianum* southern population; J, *S. peruvianum* southern population; K, *S. chilense*; L, *S. habrochaites*; M, *S. pennellii*; N, *S. ochranthum*; O, *S. juglandifolium*; P, *S. lycopersicoides*; Q, *S. sitiens*. Scale bars = 0.5 cm in E–G; 1 cm in A–D, H–Q. (Spooer et al., 2005).

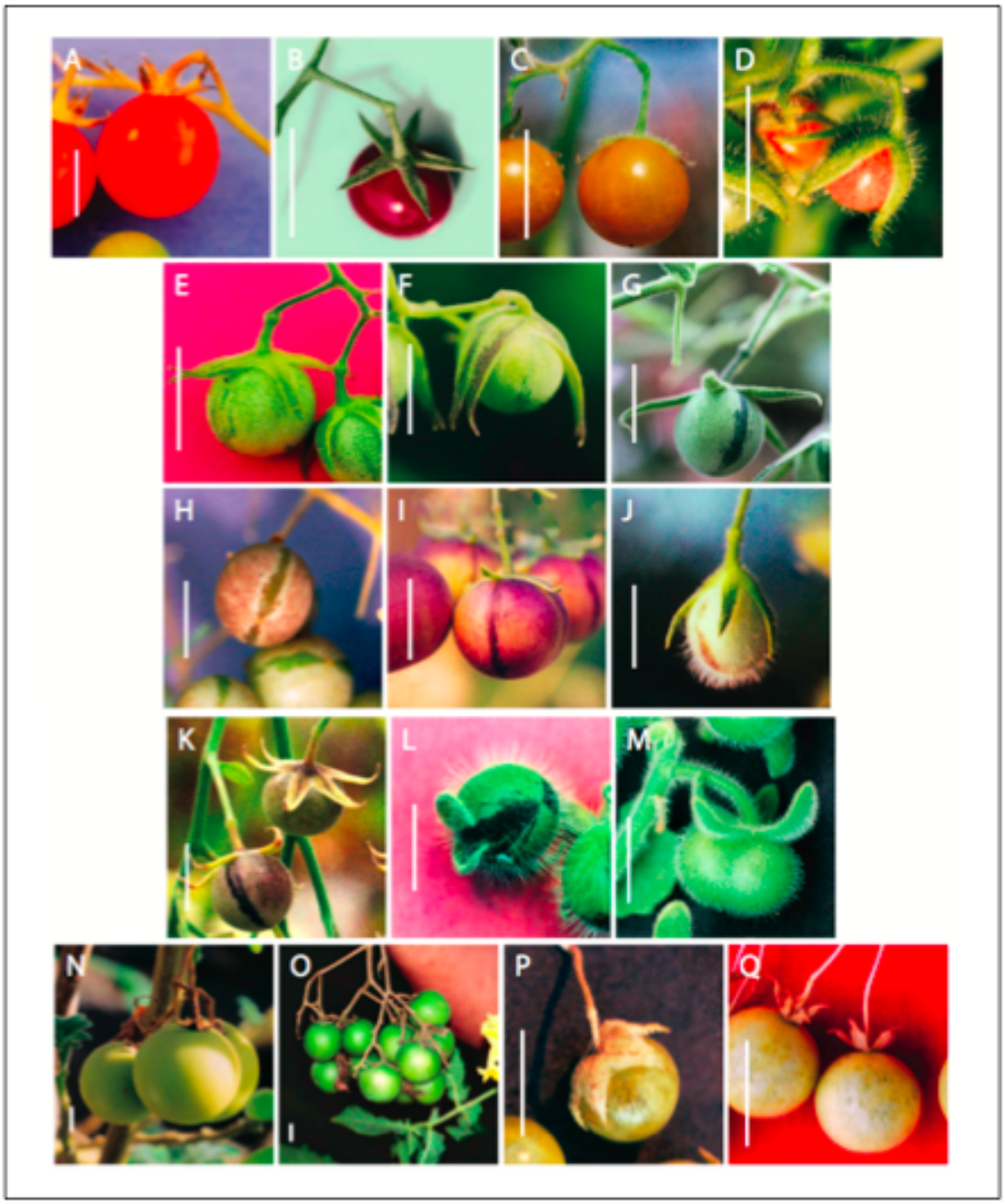

Fig.:4. Fruits of tomato and outgroup species. A, *S. lycopersicum*; B, *S. pimpinellifolium*; C, *S. cheesmaniae*; D, *S. galapagense*; E, *S. neorickii*; F, *S. chmielewskii*; G, *S. peruvianum* northern population; H, *S. peruvianum* northern population; I, *S. peruvianum* southern population; J, *S. peruvianum* southern population; K, *S. chilense*; L, *S. habrochaites*; M, *S. pennellii* (*LA716*); N, *S. ochranthum*; O, *S. juglandifolium*; P, *S. lyco- persicoides*; Q, *S. sitiens*. Scale bars = 1 cm throughout. (Spooer et al., 2005)

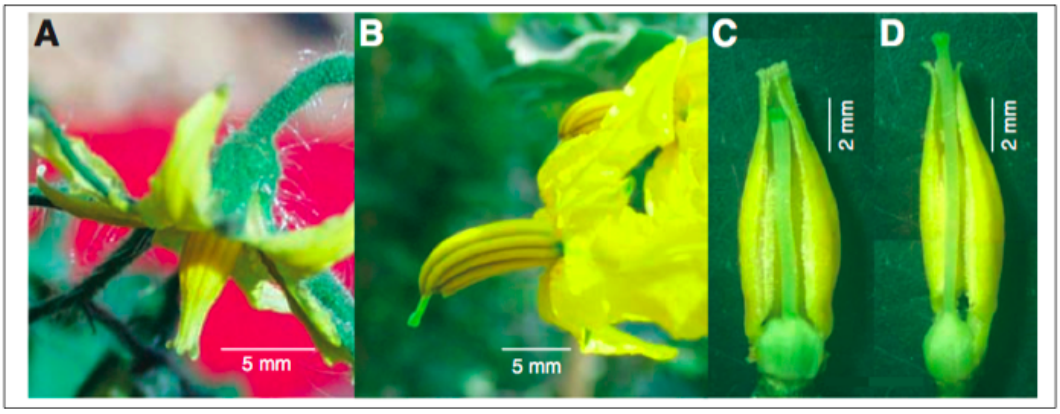

Fig.:5. (A) Flower from self-pollinating cultivated tomato (*S. lycopersicum*). (B) Flower from out-crossing, wild species *S. pennellii.* The stigma surface is recessed relative to the anther cone in *S. lycopersicum* but exserted in *S. pennellii.* (C) Cross section of a flower from *S. lycopersicum* with short-style allele (recessed stigma). (D) Cross section of a flower from *S. lycopersicum* with long- style allele (exserted stigma).

1.2.3 – The classification and phylogeny of tomato and its wild relatives: a complicated puzzle.

The classification and phylogeny of *Solanum* section *Lycopersicon* is a complex issue that has not yet reached a widely accepted consensus.

Tomato plants were introduced in Europe in the sixteenth century and the early botanists recognized their relationship with the genus *Solanum.* Tournefort in the 1694 used the multilocular character of the fruit of cultivated tomatoes to differentiate them from *Solanum*. He was the first to consider cultivated tomatoes within distinct genera under the early name *Lycopersicon*.

Almost 60 years later, in the 1753 Linnaeus grouped all the cultivated forms of Tournefort under the specific name of *Solanum lycopersicum*. One year later, Miller reconsidered Tournefort's classification and formally described the genus as *Lycopersicon*. This classification has been maintained until the 1993 when the phylogenetic relationships within the Solanaceae were analyzed with a molecular approach using chloroplast DNA. The results supported tomato to be in the genus *Solanum* closed to the potato group (Spooner et al., 1993).

In 2005, a new phylogenetic classification was described, again assigning tomato to the genus *Solanum* (Peralta et al., 2005). In addition, this classification matches with the original conclusions of Linnaeus (1753) (Fig.6).

Tomato species can be divided into two big groups: the ingroup that includes the cultivated tomato and its wild relatives and an outgroup which includes *Solanum lycopersicoides.*

Early taxonomic studies subdivided the genus (ingroup) into two groups respectively called *Eulycopersicon* and *Eriolycopersicon*. The former group includes colour-fruited species while the latter includes the green-fruited species (Muller 1940). In the 1976, Rick separated the species into two different groups: the *esculentum* and the *peruvianum* complexes, based on their reproductive compatibility with the cultivated tomato (Rick 1976).

While the *esculentum* complex contains *S. l .pimpinnelifolium*, *l. cheesmanii*, *l. parviflorum* (now called *l. neorikii*), *l. chmieliwskii*, *l. hirsutum* (now called *l. habrochaites*) and *l. pennellii*, in the *peruvianum* complex there are: *l. chilense*, *l. peruvianum* and *l. peruvianum glandulosum* (Rick 1976).

Hypotheses on ingroup relationships within tomato also varied greatly depending on the criteria used. Tomato species classified on morphological features are quite different from a classification based on biological (interbreeding) criteria (Luckwill, 1943a; Rick, 1963). Moreover, Peralta and Spooner in 2001 produced a phylogeny of tomatoes based on DNA analysis of the single copy *GBSSI* (waxy) gene (Peralta and Spooner, 2001) and in 2005 Spooner did a new phylogeny based on Amplification Fragment Length Polymorphisms (AFLP) (Spooner, 2005).

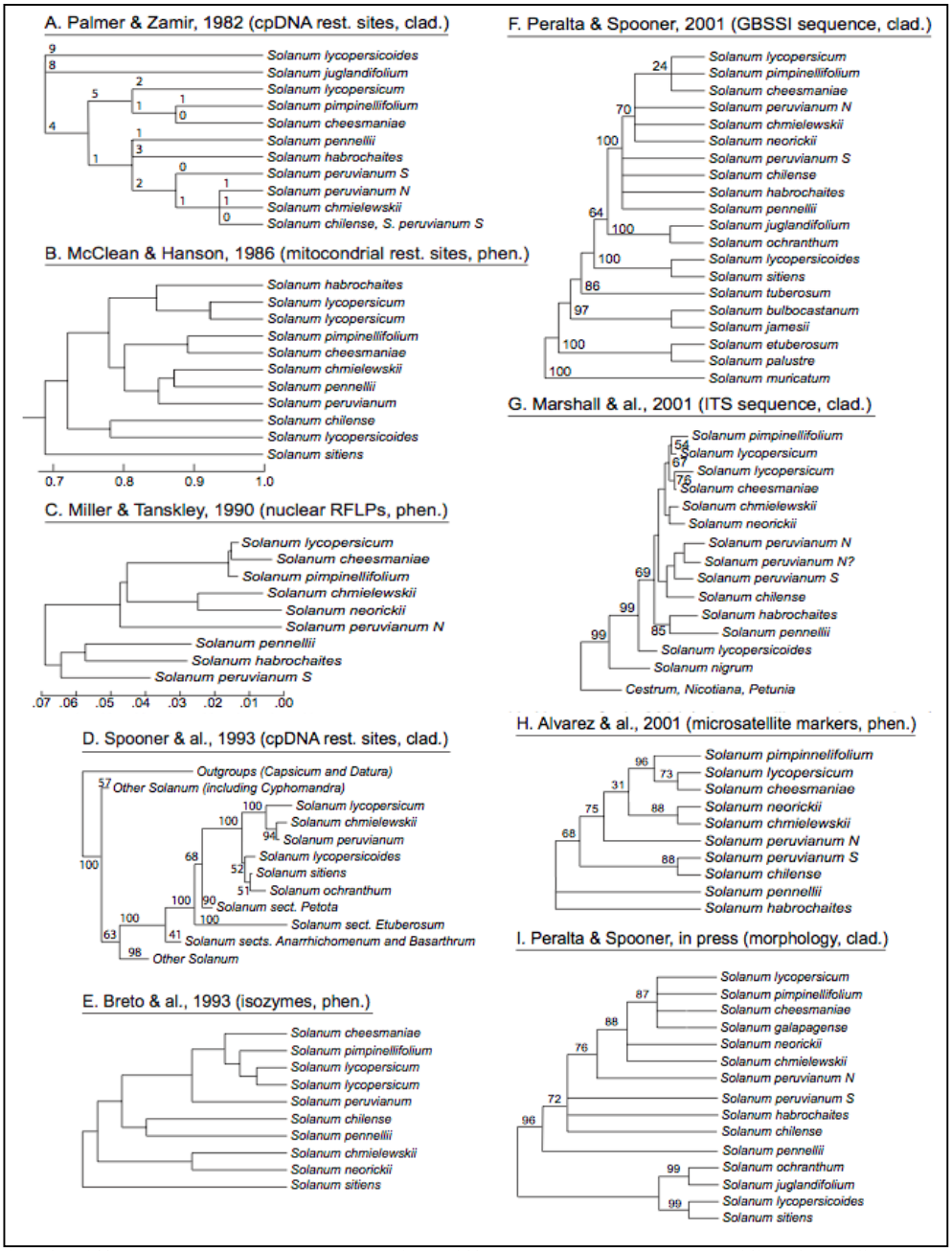

Fig.:6. Different phylogenetic trees using different parameters. (Modified from Spooer et al., 2005)

Recently, other phylogeny of tomato species were done using ecological and geographic parameters (Nakazato et al 2010; Moyle et al.,2008) but they did not result in a new phylogenetic tree but they only included new species within the tree produced from Spooner in 2005.

The complexity of the ingroup relationships depends on the high number of species involved (Fig.6). The species within the Solanaceae family are, so far, more than 3000 in about 90 genera and tomato and its wild relatives are 1160 (Knapp et al., 2004; Tomato Genetic Resource Center – TGRC).

In an effort to resolve this issue, an interesting work was published in 2009 (Zuriaga et al.,2009). Zuriga and colleagues tried to characterize by AFLP and two nuclear genes (CT179 and CT66) 210 accessions of tomato including all recognized species of *Solanum* section *Lycopersicon* and 3 accessions of *Solanum lycopersicoides*. Their results suggest a classification similar to those previously proposed although with some significant differences. Zuriga and colleagues suggested that the recently proposed species *S. corneliomulleri*, is indistinguishable from *S. peruvianum*. Furthermore, *S. arcanum* could represent a group of populations composed by two cryptic species probably due to their different geographic distribution. The CT179 data also showed a close relationship between *S. arcanum* and *S. huayalasense*. In terms of phylogenetic relationships the following groups were established: the Lycopersicon group (*S. pimpinellifolium, S. lycopersicum, S. cheesmaiae and S. galapagense*), the Arcanum group (*S .chmielewskii, S. neorikii, S. arcanum* and *S. huayalasense*) and the Eriopersicon group (*S. peruvianum* and *S. chilense*). *S. pennellii* and *S. habrochates* are not included in any group, but they are the closest to *S. lycopersicoides* (outgroup).

All these observations suggest that within the genus *Lycopersicon* (ingroup) closely related interspecies and interspecies heterogeneity have made very difficult the resolution of precise interspecific relationships. However, based on all the data available at the time, Nesbitt (Nesbitt and Tanksley, 2002) resolved the enigma of the speciation within the genus. Estimated divergence times based on pooled silent sites and a rate of $6{,}03 \times 10^{-9}$ silent substitutions per year suggest that the genus started its initial radiation almost 7 million years ago. Furthermore, *L. esculentum* and its nearest relatives *L. cheesmanii* and *L. pimpinellifolium* shared a recent common ancestor almost 1 million years before present (BP) (Nesbitt and Tanksley, 2002).

1.2 – ECONOMIC AND SCIENTIFIC IMPORTANCE OF TOMATO

1.2.1 – Economic importance

Although tomato was not used until the 1700, in the last 200 years the production of tomatoes in term of ton/ha^{-1} has continuously increased making tomato the major vegetable crop cultivated in the world. Tomato is used both fresh (salad) and in processed foods.

At the moment, the top five leading tomato producing countries are China, United States, Turkey, India and Egypt. In the past ten years, tomato cultivated areas have increased by 38% and production has increased by 45% and most of this worldwide increase in production has come from China.

In the United States the fresh market tomato has an estimated value of 1,17 billion dollars while the value of processed tomato production is 683 million dollars (USDA-NASS, 2003). In addition, there has been an expansion of greenhouse grown tomato. Greenhouse production of fresh market tomato is significant in Europe, especially in the Netherlands (Snyder, 1996).

The increasing economic importance of tomato is due to the high nutritional and low energetic value (≅20 kilocalories for 100g of product) of the tomato cherries. This is due to a high content in water (≅ 95%) while the rest includes: sucrose and fructose (≅3%), proteins (≅1%), fats (≅0,2%) and fibres (≅1,8%). Tomato cherries merit attention, even in terms of valuable micronutrients present at low concentration. It also contains carotenoids that are a considerable source of vitamin A, C and E (Abushita et al., 1997). Epidemiological studies indicated that carotenoids and vitamins play an important role in the prevention of cancer (Garewal, 1995) and heart diseases (Pandey et al.,1995). In addition, β-carotene is particularly important because is the precursor of vitamin A, and lutein. Both of them seem to reduce the risk of cancer (Abuscita et al., 2000). For all these reasons tomato has became an important agricultural commodity worldwide.

1.2.2 – Scientific importance

Tomato plant is not only important as an economic resource but it is also an excellent model system for both basic and applied plant research (Palma et al.,2011). This is due to many reasons, including ease of culture under a wide range of environments, short life cycle, photoperiod insensitivity, high self fertility and homozygosity, great reproductive potential and ease control of pollination (Fooland, 2007). In addition, several genetic/genomic tools are available and include so far: tomato wild species and mutant collections, F_2 synteny mapping population and permanent recombinant inbred (RI) mapping populations, BAC libraries and an advanced physical map, TILLING populations, tomato microarray, gene-silenced tomato lines and virus induced gene silencing (VIGS) libraries (Barone et al., 2008). For all these reasons tomato plant has been chosen as a model to study the fruit biology in climacteric fruits.

1.3 – WILD TOMATOES SPECIES AS A GENETIC RESOURCE.

In 1865 the first tomato field cultivation started in the United States using a cultigen called Tilden then, in 1870, it was substituted by Trophy and around 1910 public breeders started introducing the disease-resistant cultigen, Tennesee Red.
In 1940 closely related wild species within the genus *Lycopersicon* began to be screened for additional disease resistance and wild sources provided much of the breeding germplasm (Stevens and Rick 1986). To date, wild germplasm continues to play a major role in tomato breeding.

The major germplasm collections of tomato are maintained in the United States at the Plant Genetic Resource Unit, United States Department of Agriculture (USDA) and at the Tomato Genetic Resources Center (TGRC) located at the University of Davis, California. The collection at the TGRC has an emphasis on wild species where a number of accessions are available with tolerance to drought, flooding, high temperatures, aluminium toxicity, chilling injury, salinity-alkalinity and arthropod damage (http://tgrc.ucdavis.edu/). These stress tolerant wild species have been extensively used in tomato crop improvement (Rick and Chetelat 1995). This is because cultivated tomato, being a self-pollinated crop, has a reduced genetic variation. This can partially explain the slow rate of tomato improvement that was achieved until 1940, when the first use of wild species as a source of desired traits was reported (Bai and Lindhout, 2007).
L. pennellii has been found to be a promising source for drought tolerance and salt tolerance, furthermore also *L. cheesmanii*, *L. chmielewskii* and *L.hirsutum* seem to be interesting for fruit quality improvement. In these species important QTLs have been found that can improve the traits of the cultivated tomatoes (Eshed et al., 1996; Bernacchi et al., 1998a; Frary et al., 2003).

1.4 – STRUCTURAL AND FUNCTIONAL ORGANIZATION OF TOMATO GENOME.

The tomato (*Solanum lycopersicum*) is a diploid species with a genome composed of 12 chromosomes (2n = 2x = 24) totalling 950Mbp (Arumuganatha et al., 1991) encoding almost 35000 genes, the majority of them being located at distal euchromatic regions of the chromosomes with an approximate gene density of 6.7Kb/gene, similarly to *Arabidopsis* and rice (Fooland, 2007; Wang et al., 2006). Very little is known about the composition and organization of heterochromatin regions in the tomato genome. However, it has been show that the genomic DNA contains 59% of non coding sequences, almost 28% of coding sequences, 11% of transposons and 2% of organellar sequences (Barone et al., 2008). Furthermore most of the tomato genome (73%) is composed of single-copy sequences (Peterson et al., 1996) while the repeated sequences are around the 12% (Budiman et al., 2000).

A work published in the 2006 (Wang et al., 2006) focused its attention on the organization of euchromatin and heterochromatin in the tomato genome. The results of this research showed that 90% of the genes were present in contiguous stretches of euchromatin comprising only 25% of the total DNA in the tomato genome, moreover these regions appear largely devoid of repetitive sequences like retrotansposons. In contrast, the pericentromeric heterochromatin has a gene density 10-100 times lower than that of euchromatin and is largely occupied by retrotransposon of the *Jinling* family (Wang et al.,2006).

More information will be available when the sequencing of the tomato euchromatic genome (≅ 25%) will be concluded. Currently the sequencing of the 12 chromosomes has been split between 10 countries: Korea (chromosome 2), China (3), UK (4), India (5), Netherland (6), France (7), Japan (8), Spain (9), Italy (12) and United States (1, 10, 11) (Lee et al, 2007). In addition, chloroplast genome sequence is available and the mitochondrial genome will be sequenced by Argentina (Lee et al, 2007). Moreover, in the 2009 a complementary whole-genome shotgun approach was initiated, which in conjunction with other data yielded high quality assemblies (http://solgenomics.net/organism/Solanum_lycopersicum/genome).

1.5 – TOMATO FRUIT DEVELOPMENT AND RIPENING: ENDOREDUPLICATION, ETHYLENE PATHWAY, MOLECULAR AND EPIGENETIC MECHANISMS.

Different parts of the flower can contribute to the final structure of dry and fleshy fruits; thus the final form of the fruit depends on the number and type of floral organ components and how their different tissues grow and differentiate (Palma et al., 2011).

During fruit development and after ovary fertilization the ovary wall becomes the fruit pericarp (Fig.7A), which consists of three distinct layers: the endocarp, mesocarp and exocarp (Fig.7B and 7C). The septa of the carpels divide the ovary and fruit into two or more locules (Fig.7C) (Gillaspy et al 1993). An elongated axial placenta, to which the seeds are attached, is highly parenchimous and later gives rise to the tissue that fills the locular cavity (Fig.7C) (Gillaspy et al 1993).

The fusion of two or more carpels in fruits such as tomato results in a complex morphological structure in which it is difficult to discern the ontogenetical relationships of cells in the fusion zones (Fig.7D).

A distinct concentric vascular system is also present in the pericarp (Fig.7C). In addition this structure is covered on the outside by a thin cuticle, a skin that further consists of an epidermal layer (Gillaspy et al 1993).

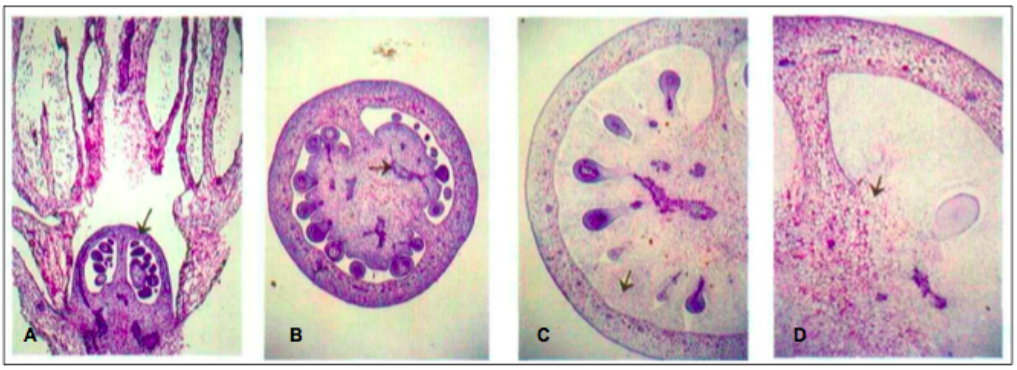

Fig.:7. Tomato fruit development.

The development of the fruits is classically described in four distinct phases: Fruit set (I), a phase of intense cell division (II), a phase of cells expansion (III) and finally ripening (IV) (Fig.8).

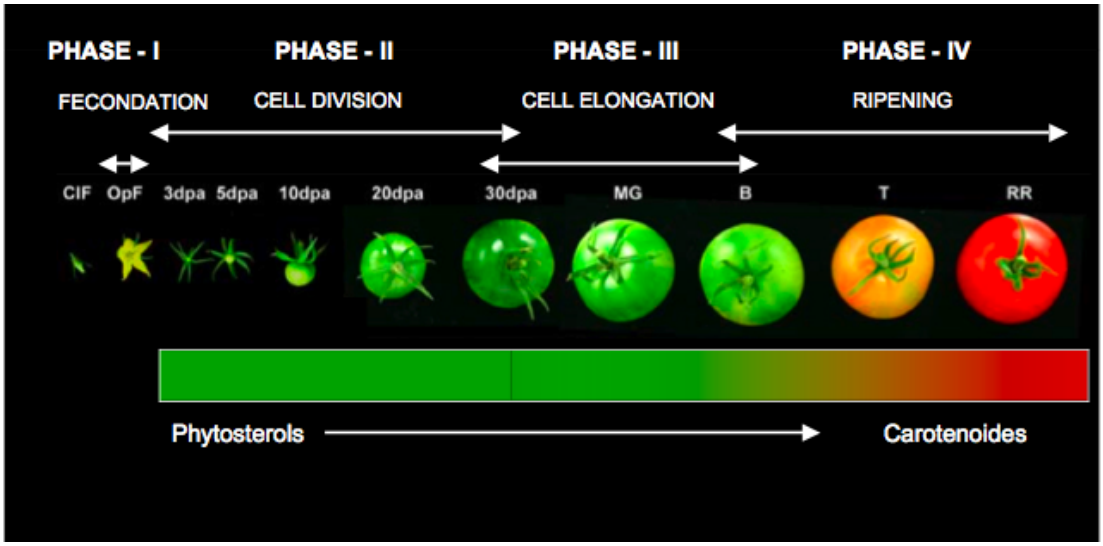

Fig.:8. Phases of fruit development.

The earliest phase (PHASE-I) involves the development of the ovary and the decision to abort or to proceed with further cell division and fruit development (Fig.8). This decision depends on the successful completion of pollination and fertilization. Then, the presence of fertilized ovules generally triggers the development of the ovary into a fruit. After fertilization, cell division is activated (PHASE-II) in the ovary and proceeds at high rate for 10 days and at a reduced pace for 10 more days. While the cell division phase ends, individual cells enlarge (PHASE-III), as does the entire fruit, for the following weeks. Finally, there is the ripening (PHASE-IV) that is an unique aspect of development starting after seed maturation has completed(Fig.8).

1.5.1 – The endoreduplication as a factor for fruit development.

One of the aspect of tomato fruit organogenesis involves fundamentals cellular processes such as cell division, cell expansion and cell differentiation, which have an impact on the final size, weight and shape of fruit. These developmental phenomena are under the control of complex interactions between internal signals (hormones) and external factors (environmental clues).

In particular, an important role is played by cell expansion, which requires a specific cell cycle where mitosis is bypassed. This modified cell cycle, called the endoreduplication

cycle or endocycle, consists of several round of DNA synthesis in the absence of mitosis (Chevalier et al., 2011).

In addition, it has been demonstrated that the endopolyploidization in tomato fruit does not lead to a doubling of the chromosomes number in the nucleus but produces chromosomes with 2n chromatids without any change in chromosome number (Bourdon et al., 2010).

A normal cell usually starts its cycle at the post-mitotic interphase (G1), replicates its DNA during the synthesis phase (S), grows further during the post-synthetic phase (G2) and then divides in mitosis (M). On the other hand, the endoreduplication cycle or endocycle is a truncated version of the canonical cell cycle (Bertin et al.,2007). The endocycle consists, in fact, in the reiteration of only two major stages: the post-mitotic interphase (G1) and the S phase. This situation is maintained for several rounds resulting in an exponential increase in the amount of nuclear DNA (Gutierrez, 2009).

The progression through the distinct phases of the plant canonical cell cycle and the endocycle requires the cyclin-dependent kinase (CDK) and a regulatory cyclin (CYC) subunit. At the boundary between the G1 and S phase the canonical A-type CDK-A are active, whereas a CDKB1 bound to a CYCA2, 3 is required to prevent a premature entry into the endocycle (Joubes et al 2000a; Boudolf et al.,2004) (Fig.9). Another gene involved in the control of the endocycle is WEE1. In tomato this gene acts in the G phase to allow the sufficient cell growth in response to nuclear DNA amplification (Chevalier et al., 2011) (Fig.9).

The endoreduplication process in tomato has considerable significance, not only because high levels of endoreduplication occur during fruit development (Bertin et al., 2007) but most importantly because it contributes to fruit growth in a developmentally and/or genetically regulated manner (Cheniclet et al., 2005; Chevalier, 2007).

Although the ability to form large cells is not fully restricted to endoreduplicating cells, in tomato a clear correlation has been demonstrated between cell size and ploidy levels (Cheniclet et al.,2005). In addition, Cheniclet and colleagues have been able to found a tight correlation between endoreduplication and fruit size (Cheniclet et al.,2005). It has been also reported that endoreduplication occurs always in fleshy fruits which develop rapidly comprising three to eight round of endocycle.

Many studies described the endoreduplication dynamic during the development of organs or tissues in various species. In maize (*Zea mays* L.) the endosperm nuclei reach a DNA content up to 690C (Larkins et al.,2001). Differently, in *Arabidopsis thaliana* moderate

endoreduplication has been reported in different tissues (up to 32C) (Galbraith et al.,1991). In tomato, large endoreduplicated cells are located in the mesocarp with DNA contents up to 256C or even 512C in cherry tomatoes as well as in large-fruit-size cultivars (Bertin et al.,2007). In addition in tomato cherries an interesting study has reported different endoreduplication levels for different tissues of tomato fruits. In this work the epidermis, pericarp and gel were analyzed (Joubes et al.,2000). At the red ripe stage, the results showed clearly that epidermis and gel have low endoreduplication levels (respectively up to 8C and 64C) while the pericarp reaches an endoreduplication level up to 256C (Joubes et al.,2000). Although many studies describe an endoreduplication dynamic during development, the real functional role of it remains controversial (Bertin et al., 2007).

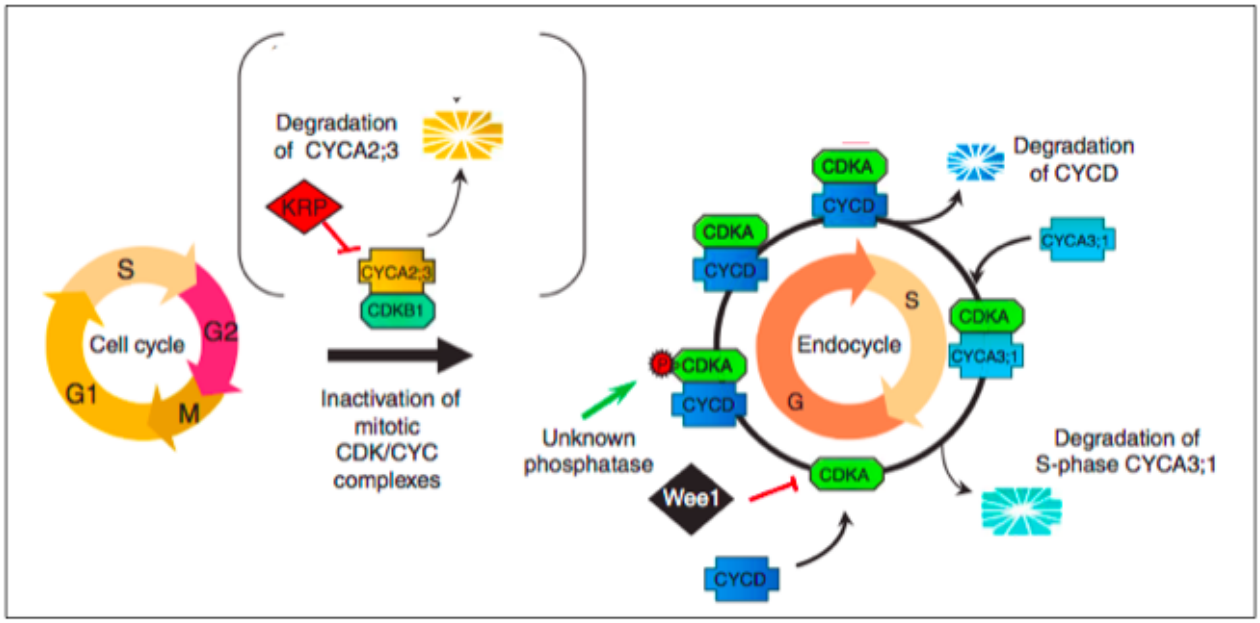

Fig.:9. Endoreduplication mechanism in tomato.

1.5.2 – *Fruit development and ripening: the ethylene pathway and molecular mechanisms.*

A second aspect of fruit development is the ripening which is the final phase of fruit development and involves deep metabolic changes in the biochemistry, physiology and gene expression (Palma et al., 2011). Furthermore, ripening involves softening of fruits tissues with an increased accumulation of sugar, acids and volatile compounds that

increase the palatability to animal. In this way the plant facilitates its own seed dispersion (Klee and Giovannoni, 2011).

Fleshy fruits are physiologically classified in climacteric (tomato, apple, banana and avocado) and nonclimacteric (citrus, strawberry and grape). Climacteric fruits are characterized by an increase in respiration and by a simultaneous increase in synthesis of the phytohormone ethylene upon initiation of ripening whereas non-climacteric fruits do not exhibit an increasing in respiration during ripening (Giovannoni, 2004).

1.5.2.1 – The synthesis of ethylene during climacteric ripening

Although the specific role of climacteric respiration in tomato fruit ripening remains unclear, the recruitment of ethylene as a coordinator of ripening serves to facilitate rapid and coordinated ripening (Giovannoni, 2004).

The biosynthetic pathway of ethylene is simple and consists of only two enzymes linked to the methionine metabolism. In this pathway, the S-adenosylmethionine is converted to 1-aminocyclopropane-1-carboxylate (ACC) by ACC-synthase (ACS). ACC is subsequently converted to ethylene by ACC oxidase (ACO) (Giovannoni, 2004; Klee and Giovannoni, 2011) (Fig.10).

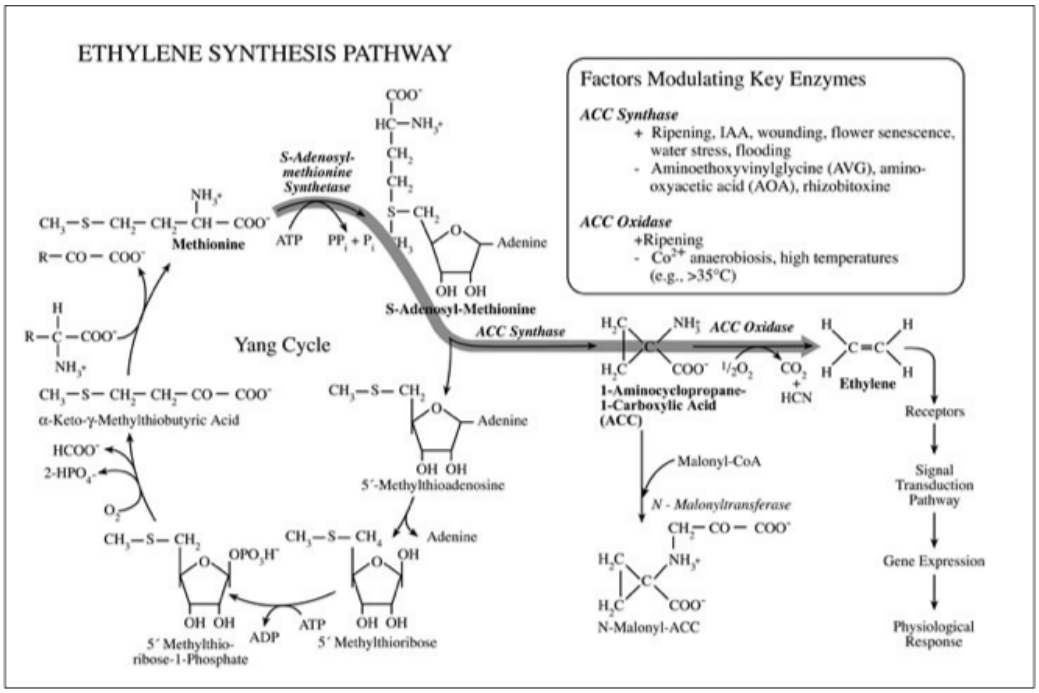

Fig.:10.Ethylene synthesis pathway.

In tomato both genes are encoded by multigenic family, eight genes encode the ACS and seven the ACC four of which have been characterized (Giovannoni, 2004). Furthermore, it has been shown that both LeACS1A and LeACS4 are responsible for the initiation of ripening ethylene and both are under developmental control.

Due to the presence of a high number of genes LeACOs and LeACSs two systems of ethylene action have been proposed in higher plants (Lelievre et al.,1998). System I is functional during normal vegetative growth, is ethylene auto inhibitory, and is responsible for the synthesis of the basal levels of ethylene detectable in all tissues including non-ripening fruits (Fig.11). System II operates during the ripening of climacteric fruits and during petal senescence. In this system the production of ethylene is autostimulated and requires different LeACS and LeACOs (Burry et al., 2000) (Fig.11).

All these genes act at the beginning of the ripening process, downstream, through the action of ethylene receptors, different players are recruited and activated controlling all the processes associated with ripening. On the other hand, LeACOs and LeACSs are themselves regulated by genes that work upstream the ethylene pathway.

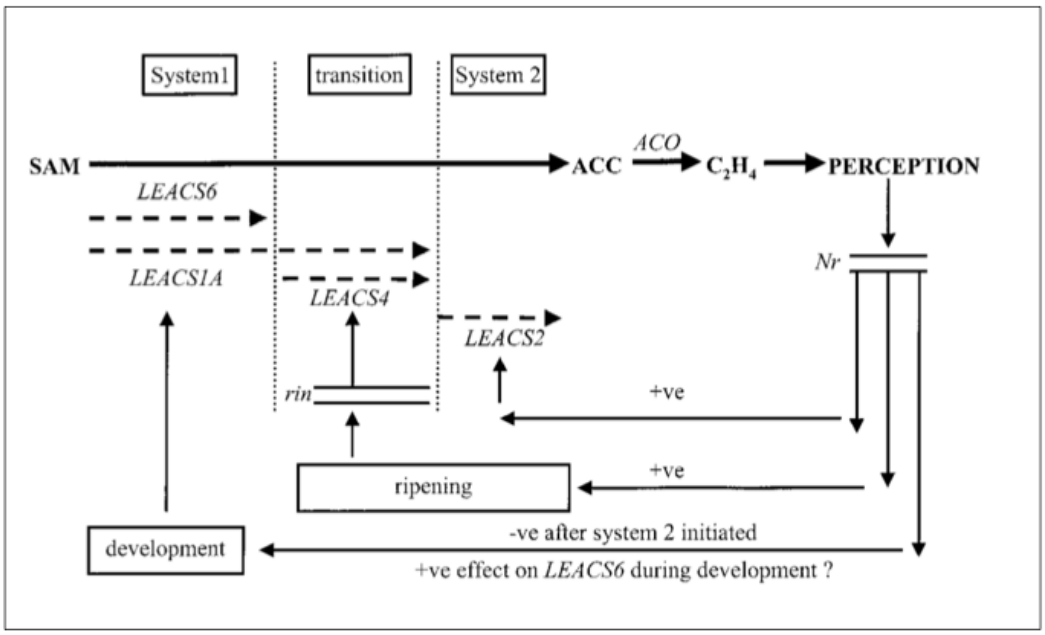

Fig.:11.Model proposing the regulation of ACS gene expression during the transition from system-I to system-II ethylene synthesis in tomato. (from Barry et al., 2000).

Upstream of the ethylene pathway, three genes seem to play a major role in developmental control of fruit ripening: RIPENING INHIBITOR (RIN) and COLORLESS NON-RIPENING (CNR) and NON RIPENING (NOR) (Giovannoni,

2007). These genes code for putative transcription factors and mutation of them blocks the ripening process (Vrebalov et al., 2002; Manning et al, 2006; Giovannoni, 2004)

In particular, the importance of Le-RIN gene is clear (Vrebalov et al., 2002) and mutation of RIN blocked the ripening process and resulted in a mutant fruit that failed to produce elevate levels of ethylene (Giovannoni, 2007), this because the RIN proteins play an integral role within the transition period from system I ethylene to system II ethylene in tomato fruits (Barry et al., 2000) (Fig.12). This data suggest that RIN is one of the earliest acting-ripening regulators required for both ethylene-dependent and ethylene-independent pathway (Fujisawa et al.,2011). This hypothesis has been recently confirmed from Li and colleagues; they have shown that RIN transcription factor has an impact on ethylene biosynthesis through the transcriptional regulation of Le-ACS2, Le-ACS4, Le-ACO1 and Le-ACS6 (Li et al., 2011) (Fig.12). In addition, Le-RIN plays a role in the regulation of other genes involved in cell wall modification, aroma and flavour development, pathogen defence, thus suggesting that Le-RIN may control multiple ripening processes (Fujisawa et al.,2011). Moreover, Le-RIN activity has an effect on the regulation of NOR and CNR (Fig.12).

The COLORLESS NON-RIPENING (*LeSPL-CNR*) gene has also been well characterized and it appeared to regulate fruit ripening upstream of ethylene biosynthesis (Manning et al., 2006; Giovannoni, 2004). It has been shown that its mutation in tomato inhibits normal ripening and produces a severe phenotype with reduced ethylene production, an inhibition of softening, a yellow skin, and a nonpigmented pericarp, suggesting that carotenoid biosynthesis is absent (Giovannoni, 2004). In addition *Colorless-non-ripening* mutants also showed a reduction in cell-to-cell adhesion and a study conducted in 2004 showed that a group of cell-wall degrading enzymes are linked to this loss of cell adhesion (Eriksson et al., 2004). In particular, enzymes such as pectinesterases, polygalacturonases, chitinases are involved whose genes have been found less expressed in the mutant as compared to wild-type plants (Eriksson et al., 2004).
Cnr locus has been mapped within the euchromatin region in the middle of long arm of chromosome 2 close to the CT277 marker (Manning et al., 2006).
Interestingly the CNR mutation is an epimutation that occurs at the promoter level: the genomic sequence of mutant and wild type are the same whereas the methylation is higher in the mutant (Manning et al.,2006). Notably, in the mutant most of the methylated

cytosines are in a symmetrical sequence context (CpG and CpNpG), which is believed to be maintained by Methyltransferase 1 (MET1) and chromo-methyltransferase 3 (CMT3), respectively (Manning et al., 2006). In addition, differences in DNA methylation levels have also been found between wild-type of different cultivars. In Liberto, the DNA in the LeSPL-CNR region appeared more methylated than in Ailsa Craig in both leaves and fruits.

1.5.2.2 – Ethylene signalling in tomato.

Downstream the ethylene biosynthesis, the regulation of ripening is modulated by a series of ethylene receptors able to activate different pathways to complete the ripening process. The first ethylene receptor was identified in tomato through the study of the Never-ripe (Nr) fruit-ripening locus (Wilkinson et al.,1995). This observation was confirmed by sequence analysis of the gene which resulted structurally similar to the *Arabidopsis* ETHYLENE RESPONSE SENSOR receptor. Finally, in tomato seven ethylene receptor genes have been identified (Le-ETR1, Le-ETR2, NR, Le-ETR4, Le-ETR5, Le-ETR6, Le-ETR7) (Wilkinson et al.,1995; Tieman et al., 1999; Wilkinson et al., 1995; Zhou et al., 1996) (Fig.12).

However, to date, their signal transduction pathway(s) has not yet been fully elucidated (Klee and Giovannoni, 2011). The best model proposed to explain their role suggests that the receptors are in a functionally "on" state in absence of ethylene. In the presence of ethylene these receptors turn in an "off" state permitting the ethylene signalling to proceed (Klee and Giovannoni, 2011).

Ethylene receptors in *Arabidopsis* have been shown to interact with the CONSTITUTIVE TRIPLE RESPONSE1 (CTR1) (Clark et al., 1998) and then, at the end of the signalling pathway, sets of transcription factors are found called ETHYLENE INSENSITIVE3 (*EIN3*). In tomato, three CTR genes, showing high DNA homology (Giovannoni, 2004), and EIN transcription factors (*Le-EIL*) were also identified, the latter being able to activate ethylene-responsive genes (Tieman et al., 2001) (Fig.12). One of the genes under strong positive ethylene control during ripening is the primary enzymatic regulator of flux into the carotenoid pathway: the PHYTOENE SYNTHASE (PSY) which is involved in the conversion of chloroplasts to chromoplasts by carotenoid accumulation. The accumulation of these carotenoids provides a visual indication that the fruit is mature and

suitable for consumption (Klee and Giovannoni, 2011; Giovannoni, 2004; Giorio et al., 2008) (Fig.12).

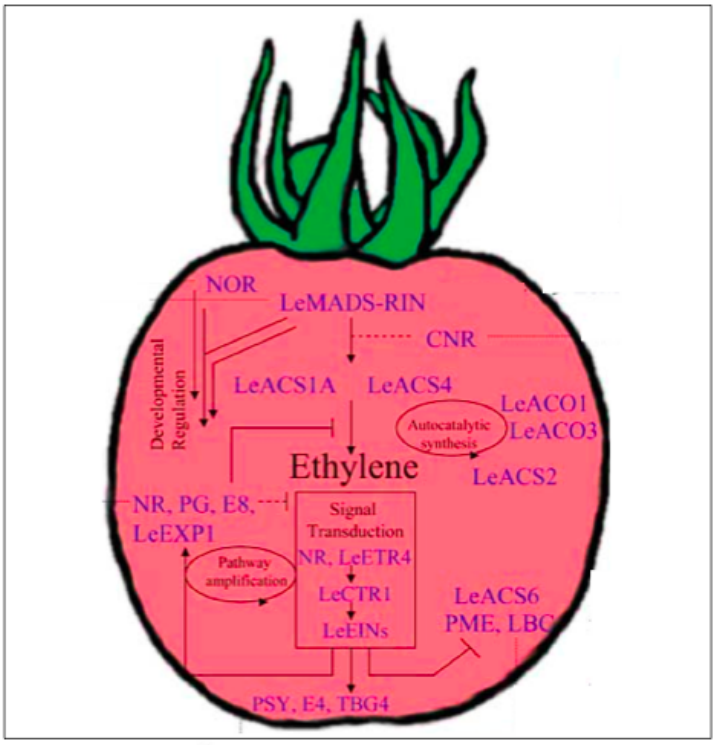

Fig.:12. Gene regulation during fruit development.

1.5.3 – Epigenetic regulation of fruit development.

The main objective of breeding is to improve the potential yield by introducing new genetic resources to increase stress tolerance of crops. In agriculture and breeding different type of yield are considered subdivided in three groups: potential, attainable and actual. The potential yield is the maximum yield a crop variety can reach under optimal growth and harvest condition and is determined by the genetic and epigenetic features of the crop. In addition, the plant productivity and stress tolerance are in relationship with energy metabolism (Nunes-Nesi et al., 2005a,b) and, although the photosynthesis is the main driver of plant productivity, it is the cellular respiration that controls it via the conversion of the fixed carbon into energy, which is then used for growth and maintenance. However, plant performance is determined not only by the energy content but by a whole range of pathways that are regulated by the concentration and ratios of energy metabolites: nicotinamide adenine dinucleotide (NAD^+), NADH and their derivates (Koch-Nolte et al.,2009). Thus, the ratio between energy content and respiration

can defines the energy efficiency that determines plant performance and productivity (Hauben et al., 2009).

On the other hand, the availability of metabolites, such as the methyl donor, S-adenosylmethionine (SAM), acetyl-CoA, and NAD^+, is one of the factors that determine the epigenetic state and epigenetic flexibility at specific loci (Mark De-Block et al.,2011). Furthermore, it has been demonstrated that energy efficiency can have an epigenetic component (Hauben et al., 2009). A study conducted using *Brassica napus* in the 2009 analyzed the possibility that an epigenetic mechanism can be used to improve the potential yield crop. In this study, two populations with high and low respiration performance were selected using isogenic line. Based on AFLP results these two lines were indistinguishable suggesting that the distinct physiological characteristic of the lines could have an epigenetic basis. This was then confirmed, the two lines showed different epigenetic information in DNA methylation and at the histone level (Hauben et al., 2009). In addition, it has been shown that this epigenetic information was transgenerationally stabilized for over eight generations. These results showed that the energy use efficiency is a distinct feature of plant vigour and yield and that it has an epigenetic component that can be used for artificial selection (Hauben et al., 2009)

These observations show as the epigenetic is increasingly recognised as a normal and essential mechanism for co-ordinating genome activity to regulate many aspects of development or response to the environment. A particular interest for the epigenetic breeding are the epialleles, which offer adaptive benefit to stress response and can be stably inherited (De-Block et al.,2011; Finnegan 1998; Tsaftaris and Polidoros, 2000). To date, in tomato, *Cnr* is the only known and well-characterized natural and stably epiallele and its mutants show a high methylation level within the promoter that suppresses gene expression (Manning et al., 2006; Giovannoni, 2004). High methylation levels have also been found in almost the 5% of gene in *Arabidopsis*. Promoter-methylated genes have a higher degree of tissue-specific expression suggesting that these are preferential sites for fine *cis*-regulation during fruit development (Zhang et al., 2006; Zilberman et al.,2007).

Another aspect of epigenetic information involves the microRNAs and it has been recently shown that the ortologue of *LeSPL-CNR* in *Arabidopsis* (*AtSPL3*) is regulated by microRNA suggesting that the transcription factors as CNR or RIN can be regulated during fruit development by small class of RNAs (Seymour et al., 2008).

Finally it is worth noting that adaptation of species to their environment may involve novel methylation that is subsequently inherited (Kalisz et al., 2004). This is an important

consideration that could have a significant impact on strategies for crop breeding (Seymour et al., 2008).

1.6 – EPIGENETICS MECHANISMS

The epigenome regulation of chromatin structure and genome stability is essential for the interpretation of genetic information and ultimately for the determination of phenotype. Furthermore, the transcriptomic activity of a plant at a certain stage of development is controlled by genome-wide combinatorial interactions of epigenetic modifications.

In the cells of all eukaryotes the genomic DNA is associated with a set of histone proteins in a highly compacted complex called chromatin. The five major types of histone proteins, termed H1, H2A, H2B, H3 and H4, are rich in positively charged basic amino acids which interact with the negatively charged phosphate groups in the DNA (Kornberg, 1974; Carter, 1978).

The basic unit of the chromatin complex is the nucleosome which is an octamer containing two copies of each histone H2A, H2B, H3 and H4 that form the histone core which binds and wraps 146bp of DNA (1,7 turns of DNA) (Fig.13). Finally, a fifth histone, H1, is located at the position where the DNA enters and exits the nucleosome core, thus sealing 20 more bp thus wrapping two full turns of DNA (Kornberg, 1974).

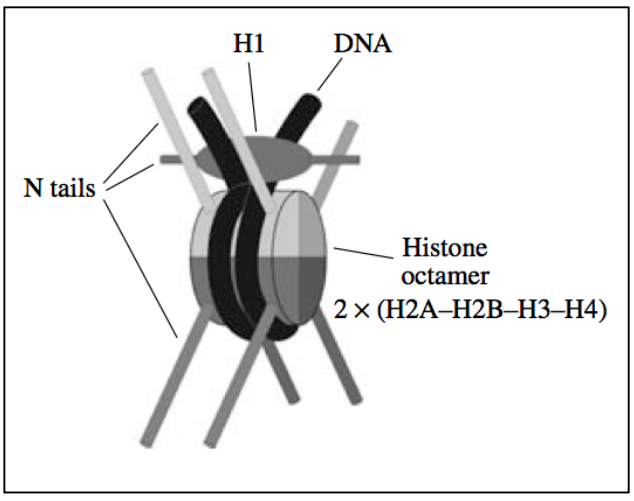

Fig.:13. nucleosome (from Koryakov 2006)

To obtain a fully compacted, methaphase chromosome different levels of chromatin condensation are necessary (Fig.14).

The first level of condensation is the association of the DNA with the histone proteins to form the “beads-on-a-string” form of chromatin, which has a diameter of 11nm. This structure develops in a 30nm chromatin fibre of packed nucleosomes through the interactions between neighbouring nucleosomes (Fig.14). The 30nm fibre binds to the chromosome scaffold and it coils into a helix whose highly condensed structure will be characteristic of the methaphasic chromosome (Fig.14).

The flexibility of the 30 nm chromatin fibre agrees with the modern notion of the role of chromatin in the regulation of genome operation. In this phase the chromatin can be present in two different states: heterochromatin and euchromatin (Fig.14).

The heterochromatin is a region of condensed chromatin, which can correspond to a constitutively condensed region (centromeric, pericentromeric and telomeric regions) and/or to a facultative region such as a gene-rich region that can be found in a less condensed state as euchromatin.

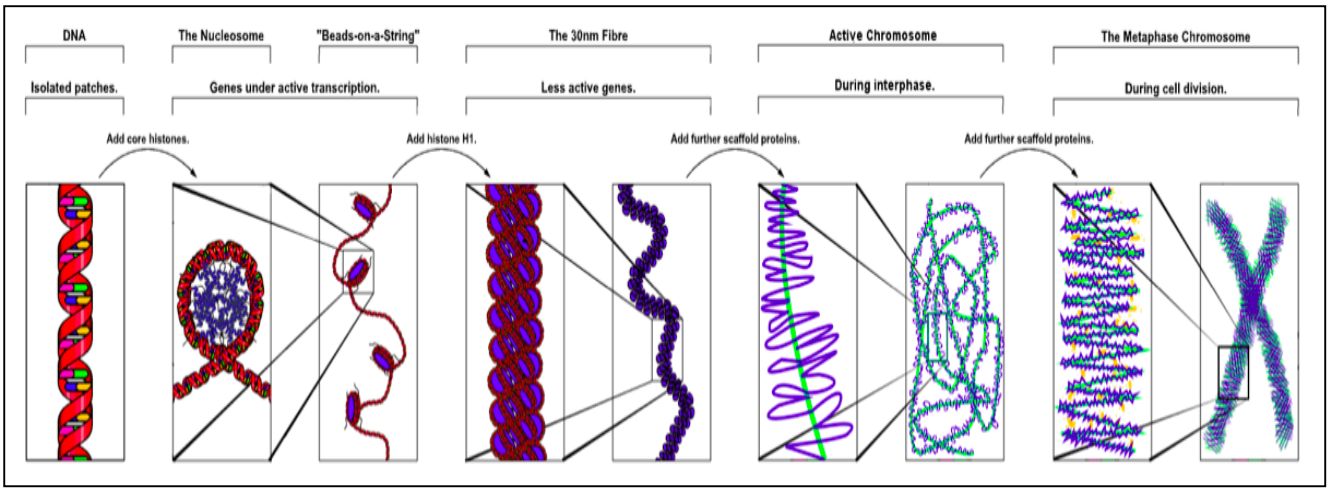

Fig.:14. Chromatin condensation

Active and/or inactive chromatin is greatly stabilized by interactions between the N-terminal tails of histones of neighbouring nucleosomes. The fibre structure can be substantially changed by various modifications of these domains. The first and main covalent histone modification that was found is the acetylation of lysine residues (Morales et al.,2001). Subsequently other covalent modifications were found such as Lys and Arg methylation, Ser phosphorylation, Glu poly(ADP)-ribosylation, ubiquitination, sumoylation and variants of the histones (Kauzarides T, 2002). It was also shown that the

specific position of these modifications has a significance (Kauzarides T, 2002). The combination of these signals exposed on the nucleosome surface constitutes a specific epigenetic code, also known as the histone code (Kauzarides T, 2002).). This code can be read by various proteins which are involved in DNA replication, transcription, repair and other genetic processes (Kauzarides T, 2002).

1.6.1 - The epigenetic marks have an effect on the chromatin structure.

The term "post-genomic era", which is often used to classify the present scientific period, does not only stress the fact that the scientific community has finally reached beyond the mere deciphering of genomes, it also indicates that there is another level of genomic information: the epigenome which involves the epigenetic information (Brero et al., 2006).

The epigenetic term derives from greek *epi (επί* - over, above, outer) and *genetic* and involves a set of modifications that does not affect the original nucleotide sequence but has an impact on gene expression (Brero et al., 2006). These changes may remain throughout the mitotic cell divisions for several generations but while the DNA sequence is identical for all cell types in a multicellular organism, the epigenome is potentially dynamic and cell type specific.

To date, the epigenetic marks involve: DNA sequence, histone tails and histone variants.

1.6.2 - DNA methylation in plants.

Plant DNA methylation share many similarities with the animal counterpart, but it has also specific features (Vanyushin, 2006).

A specific feature of plant genomes is a high degree of nuclear (nDNA) methylation in terms of 5-methylcytosine (m^5C) and rarely N^6methyladenine (m^6A). These chemical groups are added to nucleotides by specific enzymes called DNA-methyltransferases (Dnmt) able to transfer methyl groups from the universal donor, S-adenosyl-methionine (SAM or AdoMet), onto cytosine and adenine residues located in specific DNA regions. Cytosine DNA methylation controls plant growth and development and, similarly to

animals, it controls practically all genetic processes including transcription, replication, DNA repair, cell differentiation and it is predominantly involved in gene silencing and transposition (Vanyushin, 2006).

Methylation patterns are transmitted through cycles of DNA replication by maintenance methyltransferases thus allowing their stability throughout the generations (McClintock, 1967).

Using methylation-sensitive restriction enzymes it has been shown that the distribution of DNA methylation in plants occurs predominantly in cytosines within symmetrical sequences such as CpG. Furthermore, the amount of m^5C located in CpNpG sequences in plant DNA may correspond to up to about 30% of the total m^5C content in the genome (Finnegan et al, 1998). It was also shown that in plants the m^5C can also be found, in a lower quantity, in asymmetrical sequences CpNpN (where N is A, T or C) (Oakeley et al. 1996).

The first comprehensive analysis of the entire genome of *Arabidopsis thaliana* showed that pericentromeric heterochromatic repetitive sequences are heavily methylated (Zilberman et al., 2007).

However *Arabidopsis thaliana* with reduced levels of DNA methylation shows a range of abnormalities including loss of apical dominance, reduced stature, altered leaf size and shape, reduced root length, homeotic transformation of floral organs and reduced fertility (Kakutani et al.,1998; Ronemus et al., 1996).

1.6.2.1 - Plant DNA methyltransferases

Plants have genes coding for at least three classes of cytosine methyltransferase.

The first class of DNA methyltransferase, METHYLASE 1 (MET1) (Fig.15), was discovered in *Arabidopsis thaliana* and is similar in structure to the mouse methyltransferase Dnmt1 (Finnegan and Dennis, 1993) with a 50% of amino acid identity within the methyltransferase domain. The high homology between the amino terminus of MET1 and Dnmt1 suggests that this region may have similar functions, further it has been shown that this domain is able to direct the enzyme to the nucleus (Bestor and Verdine 1994). This class of enzyme is involved in the maintenance of the methylation during the DNA replication thanks to its ability to discriminate hemimethylated and unmethylated

DNA with a strong preference for the hemimethylated strand (Bestor, 1992; Li et al. 1992).

The Arabidopsis MET1 gene is member of a small multigene family that include MET2a, MET3 and MET2b (Fig.15). MET1 is the most transcribed gene (Genger et al. 1999) and it is expressed in vegetative and floral tissues (Ronemus et al.1996). *MET1* homologues have now been identified in carrot, pea, tomato and maize (Finnegan and Kovac, 2000). Two genes encoding proteins of the MET1 class have been identified both in carrot and maize. The two carrot genes are over 85% similar, with the major difference being the presence of a repeated sequence of 171 bp, which is represented five times in one gene, but only once in the other (Bernacchia *et al.*, 1998).

In *Arabidopsis thaliana* a second class of methyltransferase has been identified coding for enzymes called chromomethylases (CMT) (Henikoff and Comai, 1998) due to the presence of a chromodomain, they are coded by a small gene family with at least 3 members, CMT1, CMT2 and CMT3 (Henikoff and Comai, 1998) (Fig.15). While the conserved motifs in CMT are relatively homologous to that of MET1, the amino terminal domain is very variable (Genger et al., 1999).

In *Arabidopsis thaliana* CMT3 takes part in the methylation of the SUPERMAN gene and is responsible for maintaining epigenetic gene silencing, furthermore cmt3 mutants display a wild type morphology but exhibit decreased CpNpG methylation at the SUPERMAN gene and at other sequences in the genome (Lindroth et al., 2001). Another work from Tompa and colleagues (2002) showed that the methylation by CMT3 is important to maintain the silencing and the inactivation of the retrotransposons (Tompa et al., 2002).

A cytosine DNA methyltransferase containing a chromodomain has been isolated from maize (ZMET2), it is similar to CMT1 and CMT3 of *Arabidopsis* and it has been shown that be required for *in vivo* methylation of CpNpG sequences (Papa et al., 2001).

To date, CMT seem to be unique to plants because no methyltransferase of this class have been identified in other species (Genger et al., 1999).

A third class of methyltransferase include the enzymes called DRM1 and DRM2 (respectively DNA methyltransferase Rearranged Domain 1 and 2) which have a catalytic domain homologous to the mammalian Dnmt3 (Finnegan and Kovac, 2000). This class of enzymes similarly to the mammalian Dnmt3 seem to be involved in *de novo* methylation of non/CpG DNA sequences. Moreover, it has been shown that the DRM are involved in a methylation process via RNA called RNA directed DNA methylation (Cao et al.,2003).

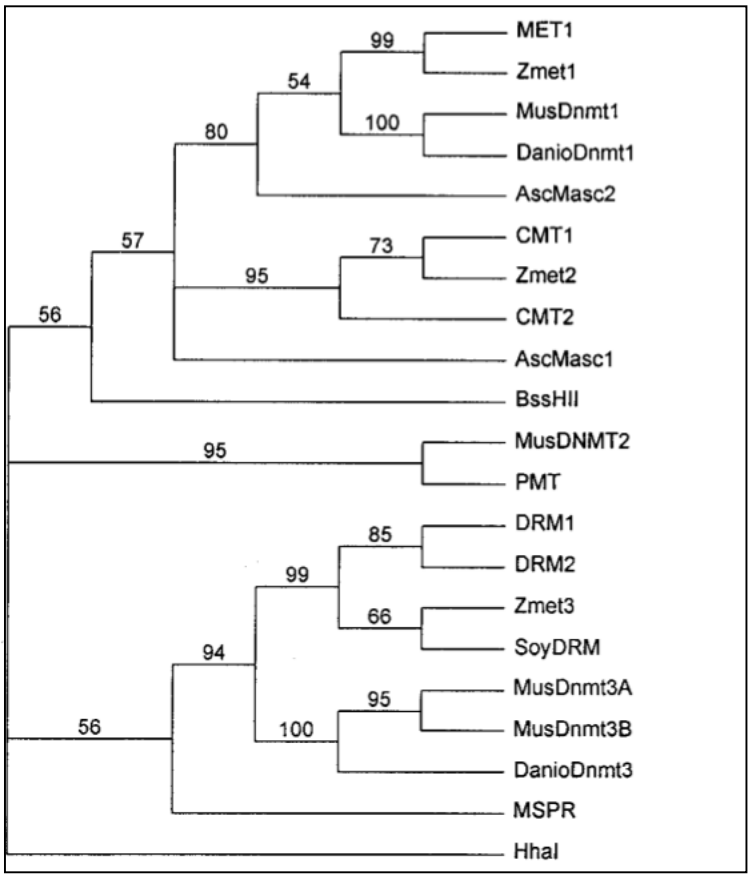

Fig.:15.Phylogenetic three of mammalian and plant DNA methyltransferases.

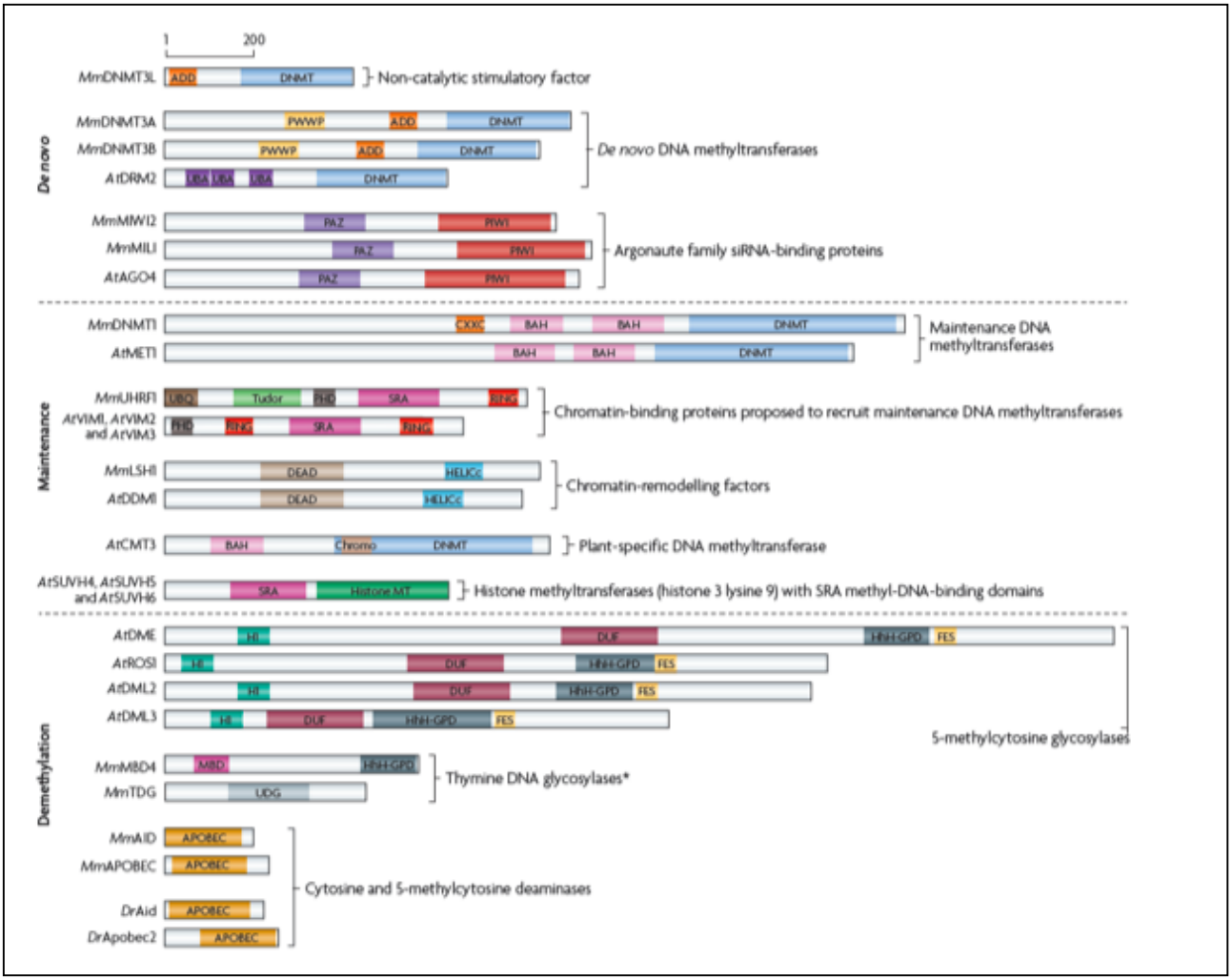

Fig.:16. Conserved domains in mammalian and plant DNA methyltransferases

1.6.2.2 – Plant RNA-directed DNA methylation

RNA-directed DNA methylation (RdDM) in plants refers to a specific process in which small interfering RNA molecules (siRNAs) guide *de novo* methylation of cytosine in all sequence contexts CG, CpNpG, CpNpN at homologous DNA regions.

This mechanism has been discovered in plants by studying the relationship between RNA silencing and DNA methylation in transcriptional gene silencing (TGS) events characterized by promoter methylation (Jones et al.,1999; Chan et al.,2004).

Recent works showed that the RNA-mediated silencing is an evolutionarily conserved mechanism of defense against viruses and transposable elements (Finnegan et al., 2003; Eamens et al., 2008).

The factors involved in the RdDM machinery have been identified by genetic studies and they are the RNA-dependent RNA polymerase 2, DCL-3 and the RNA polymerase IVa

(PolIVa). These factors are necessary to process the dsRNA into siRNAs (Eamens et al., 2008; Kanno et al., 2005; Herr et al.,2005) (Fig.17).

The RNA polymerase IV does not work in overlap with the RNA polymerases I, II and III but is a novel enzyme formed by a large subunit encoded by NRPD2a together with two alternative subunits encoded by NRPD1a and NRPD1b.

Mutants for RNA polymerase IVa show a decreased level of methylation on cytosine in the CpG, CpNpG and CpNpNp contexts thus suggesting a relationship between PolIV and methyltransferase proteins responsible for the RNA-dependent *de novo* methylation such as DRM2, DRM1 and CMT3 (Onodera et al.,2005; Cao et al., 2003). It has been suggested that methylated regions of genomic DNA provide the template, either directly or indirectly, for the PolIVa and that the resultant PolIVa transcripts are copied by RNA-dependent RNA polymerase 2 (RdRP2) to generate dsRNA (Pontes et al.,2006) (Fig.17).

Subsequently DCL3 processes the dsRNA into siRNA that are incorporated into an AGO4 effector complex that directs the *de novo* DNA methylation of homologous loci in association with PolIVb, a domains rearranged methyltransferase 2 (DRM2) and a protein involved in chromatin remodeling DRD1 (Pontes et al.,2006; Kanno et al. 2005; Kanno et al.,2004) (Fig.17).

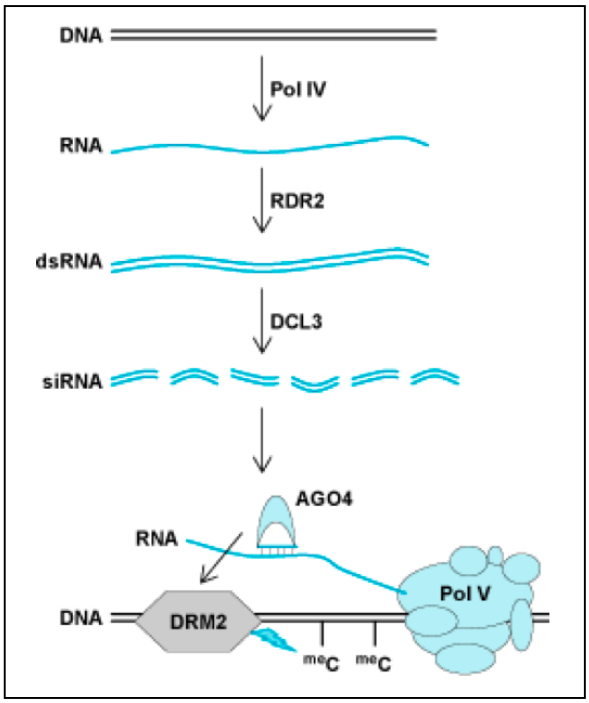

Fig.:17. RNA-directed DNA methylation

1.6.2.3 – DNA demethylation in plants

Although the DNA methylation is an important process that controls practically all genetic phenomena in the cell, in some cases, during development or in response to environmental alterations, a reset of the epigenetic state can be necessary. This kind of modifications involves a processed called DNA demethylation.

Demethylation of DNA can be passive and/or active. Passive DNA demethylation occurs when maintenance methyltransferases are inactive during the cell cycle following DNA replication, while active DNA demethylation involves one or more enzymes called demethylase and can occur independently of DNA replication (Bhutani et al., 2011).

One proposed mechanism of active DNA demethylation involves base excision repair, which is initiated by DNA glycosylases that cleave the glycosidic bond between the m^5C base and the deoxyribose and then the gap is filled by DNA polymerase and DNA ligase.

In *Arabidopsis*, DNA demethylation is mediated by the DNA glycosylase of the DEMETER family, which requires three proteins: repressor of silencing (ROS1) demeter-like 2 and 3 (DML2 and DML3, respectively) (Bhutani et al., 2011).

It has been found that regions demethylated by the DML enzyme are enriched for small interfering RNAs and generally contain sequence repeats and transposons. These results (Penterman et al., 2007; Zhu, 2009) suggest a relationship between demethylation and RNA-mediated DNA methylation. Moreover, down-regulation of ROS1 results in the accumulation of CG methylation (Penterman et al., 2007) suggesting that DNA demethylation by DML enzyme could play a role in the protection of the genome from repetitive sequences, as a defence pathway.

1.6.2.4 - DNA methylation in the tomato *(Solanum lycopersicum)* genome.

The DNA methylation of the tomato genome has been analysed in a limited number of studies. Messeguer and colleagues (1991) performed for the first time an epigenetic analysis on genomic DNA extracted from *Solanum l. esculentum*.

Their results showed that the percentage of G+C content could be estimated around 37.4%, which is the lowest, reported for any plant species. Non-coding regions have a

G+C content even lower (32%) and coding regions are considerably richer in G+C (46%) (Messeguer et al.,1991) (Tab.1).

The 5-methyl cytosine was the only modified base observed by HPLC analysis and it has been estimated to be approximately 23%, which is high compared to animal species, but well within the range reported for other plants (0-37%) (Tab.1). However considerable variation was observed in the levels of methylation across different stages/tissues. Immature tissues showed a lower level of m^5C (20%) than the mature tissues (25%). Seeds showed the higher value (27%) suggesting that de novo methylation might occur after pollination and during seed development (Messeguer et al.,1991).

Species	% G + C	% 5mC/(5mC + C)
Man	41	4.4
Drosophila	41	≤0.2
Yeast	37	<0.03
Dicotyledonous plants		
Arabidopsis thaliana	41.4	4.6
Lycopersicon esculentum (mature leaves)	37.4	25.0
Solanum tuberosum (mature leaves)	38.0	24.6
Nicotiana tabacum (mature leaves)	40.7	27.8
Pisum sativum	41.9	23.2
Sinapsis alba	39.3	12.2
Vicia faba	39.1	30.5
Gossypium hirsutum	34.2	26.6
Helianthus	37.6	37.2
$\bar{x}$ *(dicots)* =	38.9	
Monocotyledonous plants		
Zea mays	46.0	26.7
Triticum aestivum	48.0	22.4
Oryza sativa	42.1	18.6
$\bar{x}$ *(monocots)* =	45.4	

Tab.1. G+C and m^5C content of nuclear DNA from various plant and animal species

It has been also estimated that 55% of the CpG and 85% of the CpNpG sites are methylated.

The work reports also the average m^5C in mature pollen (22%) from *S. peruvianum* (PI 128657) (Messeguer et al.,1991).

A similar, but more precise analysis was conducted on pericarp and locular tissue from S.l. cv Ailsa Craig (Teyssier et al., 2008). Genomic DNA was extracted and studied from fruits at different developmental stages (20 days post anthesis, dpa, 30dpa, Mature Green MG, Breaker B, Turning T and Red Ripe RR) and leaves.

The results of this study, in agreement with those from Messeguer et al. (1991), showed that the level of m^5C analyzed by HPLC in mature leaves was around 22.3% (25% for Messeguer et al.) while differences were detected in fruits. No change was observed in m^5C content in the locular tissues during development (around 20%) (Teyssier et al.,2008).

Even if Messeguer did not observe changes in m^5C levels during fruit development (Messeguer et al.,1991). Teyssier and colleagues, on the other hand, revealed interesting variations in pericarp tissues. The early stages of development showed a stable level of m^5C of 30%, this level decreased in the pericarp during the ripening stages of turning and Red Ripe where the m^5C content of pericarp DNA decreased until 20% (Teyssier et al.,2008).

These results clearly indicate tissue-specific variation of the global DNA methylation level during fruit development in tomato.

The methylation profile was also analysed at repetitive DNA sequences by Southern blotting using the methylation-sensitive enzymes *Hpa*II/*Msp*I. Three types of repetitive elements were analyzed: the 5S rDNA, the 18s rDNA and a dispersed repetitive element, the Ty3-gypsy like retrotransposon.

The results showed an increase in the methylation level of CCGG sites in pericarp during fruit growth for all the loci also including the CNG sites. Moreover, the variations in the methylation level of the 5S locus appeared more significant during fruit growth as compared to the later stages of fruit ripening (Teyssier et al., 2008).

During this study the expression levels of eight putative tomato methyltransferase (DMT) were also analysed. It has been shown that *SlMET1* homologue of the *Arabidopsis* DMT1 was highly expressed in young plant organs. The genes coding putative chromometilase (CMT) showed a different expression profile: the *SlCMT3* was expressed in all plant tissues while *SlCMT2* was preferentially expressed in stems and *SlCMT4* in flowers and young leaves.

Finally the domain rearranged methyltransferase (DRM) *SlDRM5* and 8 were expressed in all the tissues while *SlDRM6* and 7 were detected predominantly in flowers.

In another recent article (Hobolth et al.,2006), a codon-based model was developed to analyse the effects of CpG and CpNpG methylation in coding regions. The model has been tested using a data set of 369 tomato genes and it showed that there is a very little effect of CpNpG methylation but a strong effect of CpG methylation on almost all genes (Hobolth et al.,2006).

These results have suggested different roles for CpG and CpNpG methylation, with the second one playing a specialized role in the defence against transposons and RNA viruses (Hobolth et al.,2006).

1.6.2.5 - The Methyl-CpG-Binding Proteins: a link between DNA methylation and histone modification.

The epigenetic picture is further complicated by the interaction of different pathways inducing heterochromatin formation. Recent data suggest that cytosine methylation co-operates in a network of interactions with histone modifications to modulate epigenetic control on gene expression.

In fact, histone and DNA modifications are not independent processes, for example the histone deacetylases are often recruited by DNA methylation to induce transcriptional repression and gene silencing (Loidl, 2004).

In animal, DNA methylation can lead to the recruitment of specific m^5C-binding proteins (MBDs) able to bind methylated CpGs. The first mammalian MBD proteins included MeCP2, MBD1, MBD2, MBD3 and MBD4, which have high sequence homology within the MBD domain (Ballestar and Wolfe, 2001). It has been shown that all mammalian MBDs (which include MeCP2) are able to complex with different factors (HMT or SWI/SNF) taking part in the formation of a unique gene silencing complex leading to histone modification, chromatin condensation and therefore gene repression (Bird and Wolf, 1999). This suggests a strong link between DNA methylation and chromatin structure (Tariq and Paszkowski, 2004). MBDs proteins are also able to work in association with PRMT.

In *Arabidopsis thaliana* 13 putative MBDs genes (AtMBD1-13) showed homology to the mammalian MBD domain of MeCP2 but only few of them appeared able to bind methylated DNA (Berg et al.,2003, Scebba et al., 2003). Furthermore, only few of these AtMBDs bind the canonical methyl CpG site (AtMBD6) while AtMBD5 recognizes CpNpN context and show a nuclear localization (Scebba et al.,2003).

AtMBD7 has three MBD domains and it is able to interact with AtPRMT11, homologous to the mammalian PRMT1 (Scebba et al.,2007) and shares common subcellular locations with AtMBD7 (Scebba et al.,2007).

Therefore, also in plants MBDs could act as a link between two levels of the epigenetic information with DNA methylation acting together with histone modification to perpetuate and maintain a repressed chromatin state.

1.7 - TRANSPOSABLE ELEMENTS AND THEIR EPIGENETIC CONTROL.

Transposable elements (TEs) are fragment of DNA that can insert themselves into new chromosomal locations by duplication or by excision and insertion in a different position. Almost seventy years ago, Barbara McClintock discovered this class of genomic elements in maize since then, the intensive sequencing efforts have revealed that most eukaryotic genome do not simply contain genes but are filled with transposable elements. Furthermore, they can be the major constituents of a genome (Rigal et al., 2011).
TEs have been classified in two classes based on their transposition strategy: the elements of class I and II (Fig.18).

The class I TEs require a reverse transcription of a RNA intermediate for their duplication. This class is further subdivided into two subclasses with or without long terminal repeats.
Retrotransposons that have long terminal repeats (LTRs) are autonomous elements containing at least two genes called *gag* and *pol*. The *gag* gene encodes a protein called capsid-like and the *pol* gene encodes a polyprotein that is responsible for transposition.
Within the LTRs are U3, R, and U5 regions that contain signals for initiation and termination of transcription that starts at the 5'-end of R within the 5' LTR and terminates at the 3' end of R within 3' LTR. The genes within the retrotransposons encode capsid-like proteins (CP), endonucleases (EN), integrases (INT), proteases (PR), reverse transcriptases (RT), and RNAse-H. Other features are primer binding sites (PBS), polypurine tracts (PPT), nucleic acid binding moiety (NA), inverted terminal repeats (IR), target site duplication (TSD), 5' untranslated region (5' UTR), 3' untranslated region (3' UTR), and Pol III A- and B-promoter recognition sites for RNA polymerase III. Furthermore, the LTR retrotransposons can be divided into two groups: *Ty1-copia* and *Ty3-gypsy* which differentiate each other for the order of the genes only.

Among the non-LTR retrotransposons we can find the long interspersed nuclear elements (LINE) and short interspersed nuclear elements (SINE). The coding region of the LINE includes an ORF1, a gag-like protein; EN, endonuclease and a RT, reverse transcriptase. These retrotransposons increase their copy number after each mobilization.

The class II transposons transpose by excision and insertion into a new chromosomal location by means of a specific enzyme, called transposase, encoded by transposon itself. They usually contain short terminal inverted repeats called TIR.

This class contains also the more recently identified helitrons which do not contain inverted repeats and appear to duplicate through a rolling-circle mechanism which requires helicase and replicase proteins.

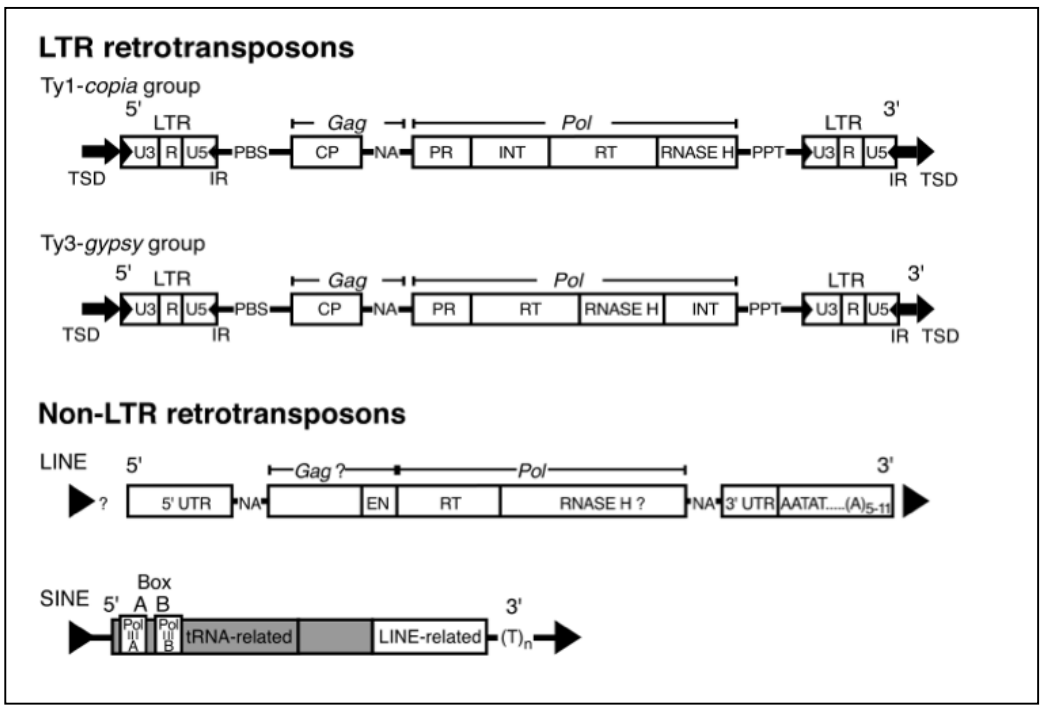

Fig.:18. General structure of the *Ty1-copia*, *Ty3-gypsy*, LINE and SINE retrotransposons (image adapted from Amar Kumar,2004).

Although TEs play an important role in the genome evolution, their incorrect mobilization represents a threat for genome integrity. To limit TEs harmful potential, host genomes have developed sophisticated mechanisms that counteract TE activation and maintain TEs in a silent state.

These mechanisms are epigenetic in nature because they do not result from genetic mutation but generate a repressive chromatin environment.

An important role in the repression of transposable elements is played by DNA methylation via the DNA methyltransferases and the RNA-directed DNA methylation (RdDM).
In addition, the TEs are rich not only in CG methylation but also in CpNpG and CpNpN methylation, which is consistent with histone modifications such as H3K9me2 and RdDM pathway.

The first evidence for a role of DNA methylation in controlling TEs activity resulted from the studies of Barbara McClintock on the class II TEs in maize. The molecular analysis of the Activator (Ac), Suppressor-mutator (Spm) and Mutator (Mu) elements revealed that the inactivation of these elements was correlated with the methylation of their DNA sequences. Further, high-throughput profiling studies of DNA methylation confirmed that TEs sequence represent the most highly methylated sequences of the genome.

Also RNA silencing is widely used by eukaryotes to control TEs activity at the transcriptional and post transcriptional levels. In plants there are evidence that small interference RNAs (siRNAs) are involved in RdDM and gene silencing. The mechanism involves the RNA-dependent RNA polymerase2 (RDR2) which generate dsRNA from single-strand RNA matrix, the endonuclease DICER-LIKE 3 cleaves then the dsRNA into fragments 24nt-long (siRNA) which are bound by AGO4 and AGO6. These factors recruit the DNA methyltransferase 2 (DRM2) on the DNA at sites homologous to the siRNA. Paradoxically, transcription from silent genomic targets is necessary for RdDM.
This transcription seems to be made by two plant specific RNA polymerases: the RNA pol IV and V and polV transcripts interact with the AGO4/siRNA to induce DNA methylation and H3K9me2 (Fig.19).

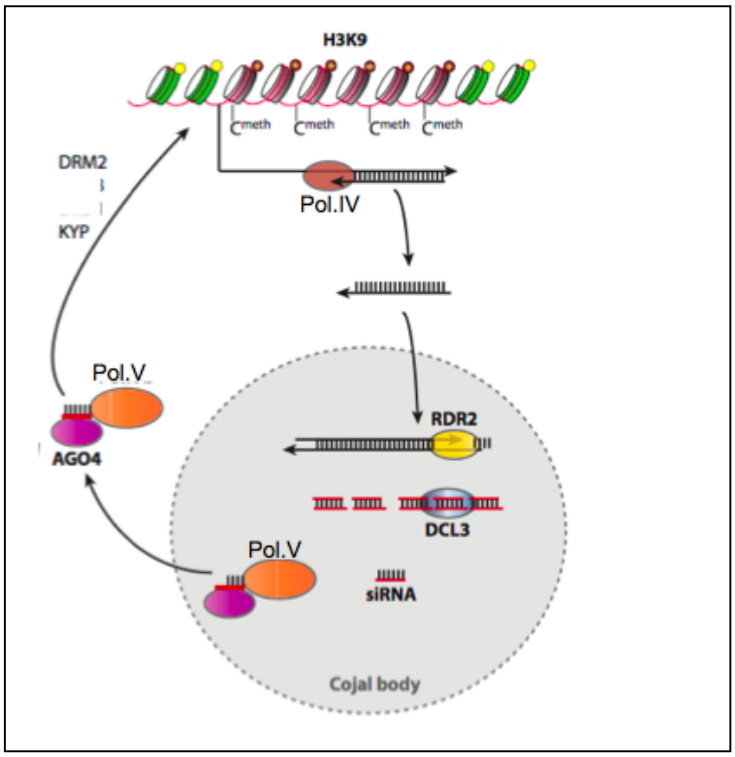

Fig.:19 Retrotransposon silencing by RdDM pathway (adapted from Damon Lisch, 2009)

In addition to DNA methylation, TEs are associated with various post-translational histone modifications, which are characteristic of a repressive chromatin state. In particular, two epigenetic marks seem to be involved in this silencing: the H3K9me2 and H3K27me1.

Some studies suggest that the H3K9me2 plays a role in TEs silencing and requires the activity of HKMTs such as KYP SUVH5 and 6 and other histone modifications and but is difficult to delineate the exact contribution of each mark, given the strong link between DNA methylation and histone modifications.

Although the TEs have to be silenced, their silencing has been shown in *Arabidopsis* to be developmentally controlled and reversed in specific cell types. These regulations involve DNA methylation from DNA methyltransferase and RdDM.

TEs can be also activated by stress and recent reports have demonstrated that heat stress can overcome TE silencing in *Arabidopsis* at least at the transcriptional level.

This release from silencing is mainly transient and the stress that induce the destabilization of silencing does not alter the common epigenetic marks such as DNA methylation and H3K9me2 indicating that this marks are not sufficient for efficient transcriptional silencing.

1.8 - HISTONE MODIFICATIONS.

In contrast to DNA, where methylation is the only covalent mark identified to date, the N terminal tails of histones are subjected to a variety of post translation modifications including acetylation, phosphorylation, methylation, ubiquitination, ADP ribosylation and sumoylation (Fig.20).

All these marks are involved in the regulation of the histone code: these covalent modifications might in fact recruit specific effector proteins therefore translating the code into functional outcomes (Kouzarides T.2007).

The acetylation and methylation of selected lysine (K) residues in the N terminal tails of histone H3 and H4 seem to have a crucial role in heterochromatin formation.

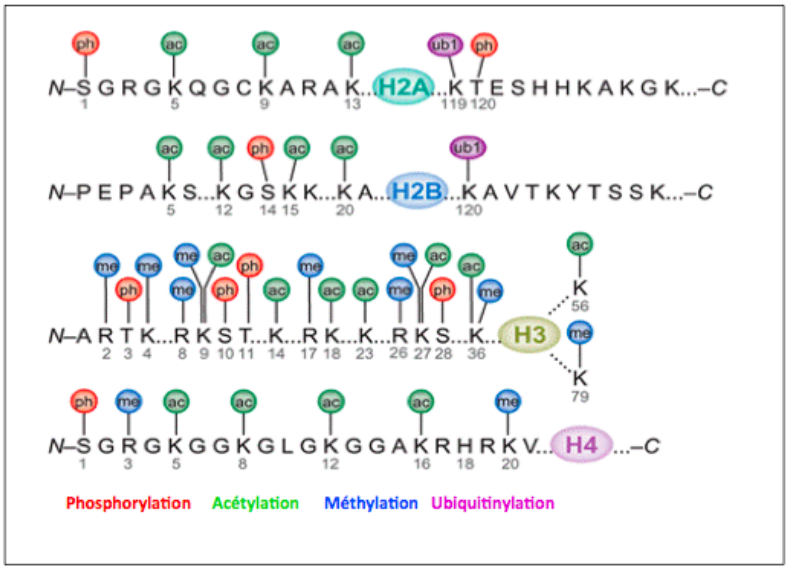

Fig.:20. Plant histone modifications.

1.8.1 – Histone acetylation and deacetylation by HISTONE ACETYL TRANSFERASE (HAT) and HISTONE DEACETYLASE (HDAC).

The acetylation of the lysine (K) obtained by the addition of an acetyl group (CH_3CO-) is one of the best known histone modification. This mark neutralizes the positive charge of

the lysine which is not able to interact with the negative charge of the phosphate group on the backbone of the DNA therefore resulting in a decreased chromatin condensation.
Acetylation and deacetylation of lysine residues in the N-terminal core of the histone are catalysed by two classes of enzymes: the histone acetyltransferase (HAT) and the histone deacetylase (HDACs, HDAs and HDs).
In *Arabidopsis thaliana* 18 members of a putative histone deacetylase family have been identified (AtHD) (Pandey et al., 2002). The analysis of mutants for AtHD1 showed various developmental abnormalities suggesting that this histone deacetylase act as a putative global transcriptional regulator (Thian and Chen, 2001). In a further work (Lawrence et al., 2004) it has been shown that AtHD1 catalyzes histone deacetylation in stress response and flower development suggesting that the reversible modification of histones (acetylation and deacetylation) is a dynamic mechanism of gene regulation in response to changes in environmental cues and developmental programs.
Other works on AtHD6 showed that AtHD6 is a key component in the epigenetic switch mechanism that silences rRNA genes by means of changes in histones and cytosine methylation (Murfett et al., 2001).
It has been shown (Zhou et al., 2010) that the H3K9 acetylation targets non transposable elements, transposable elements and genes. Furthermore in *Arabidopsis* athd19 mutants showed high levels of H3K9 acetylation, which indicates that AtHD19 plays an important role in the regulation of the level of H3K9 (Zhou et al., 2010).

1.8.2 – Histone arginine methylation by PROTEIN ARGININ METHYLTRANSFERASE (PRMT).

Differently from the lysine the arginine could be only mono- or di- methylated in a symmetrical or asymmetrical manner by a class of enzymes called Protein Arginine Methyltransferase (PRMTs) which are able to transfer methyl groups from S-AdoMet to the arginine with the formation of methyl-arginine and S-adenosylhomocysteine (Kouzarides et al., 2002; Bedford et al.2007) (Fig.21)

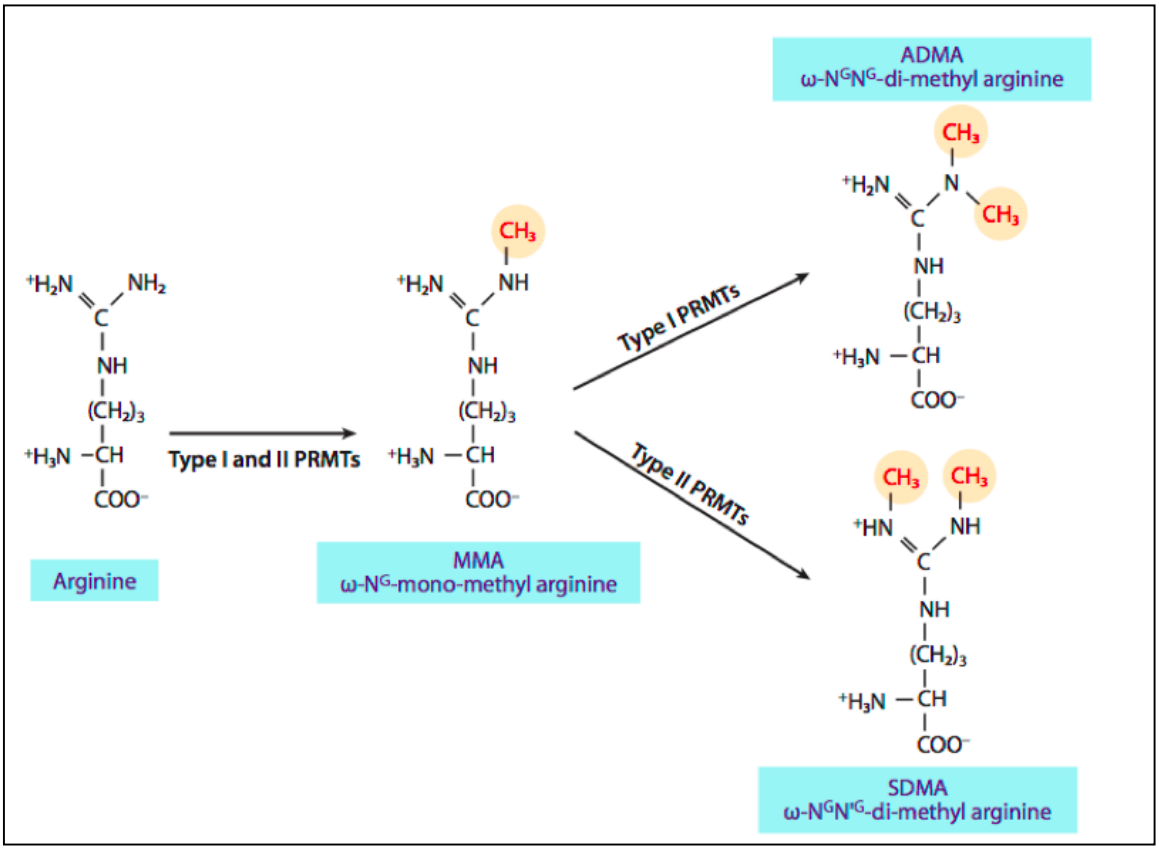

Fig.:21. Arginine methylation by the activity of the PRMT type I and type II

Arginine methylation is a histone modification correlated with gene transcription without alterations of the charge of the nucleosomes (Tariq and Paszarkowski, 2004; Sims et al., 2003; Pal and Sif, 2007).

This family of proteins show a high amino acid homology within the domain able to bind the S-AdoMet and the catalytic domain with similarities with proteins from *Schizosaccharomyces pombe*, *Arabidopsis*, *Caenorhabditis elegans* and also in *Drosophila*, *Xenopus*, zebrafish, rice and tomato suggesting that the PRMTs are highly conserved in eukaryotes. On the other hand the C-terminal domain does not show similarity between species. (Zhang et al 2000) (Fig.21).

Two types of PRMTs have been identified to date: the type I catalyses the formation of N^G-mono-methylarginine (MMA) and asymmetric N^G,N^G-di-methylarginine (aDMA) while the type II enzyme catalyses the formation of N^G- mono-methylarginine and symmetrical N^G,N'^G-di-methylarginine (sDMA) (Gary and Clarke, 1998; Smith et al., 1999; McBride and Silver., 2001).

While 11 PRMTs have been identified in mammals, only a few genes have been identified by sequence homology in plants: in *Arabidopsis thaliana* 8 genes coding for putative PRMTs have been found (Bedford, 2007) (Fig.21).
The first studies on the plant PRMTs were published in 2007 with the characterization of the PRMT10 (Niu et al.,2007), PRMT11 also called AtPRMT1b (Scebba et al., 2007; Yan et al., 2007) and the PRMT12 or AtPRMT1a (Yan et al., 2007).
The study of AtPRMT10 showed that *in vitro* it acts as a type I methyltransferase on the arginine 3 of the histone H4 and also on the H2, furthermore prmt10 mutants showed longer flowering time suggesting that this enzyme is necessary to flowering (Niu et al.,2007).
Also AtPRMT12 (or AtPRMT1a) is a type I PRMT that in vitro shows a methylation activity on the H4R3, this result has been confirmed also in vivo. Furthermore, this enzymes can use as substrates not only the R on the histones tails but also arginines in proteins located in the cellular membranes thus suggesting a nuclear as well as a cytoplasmic activity (Yan et al., 2007).
Another type I PRMT enzyme is AtPRMT11 (or AtPRMT1b), which has its active on arginine 3 of the histone H4 and also into the cytoplasm. This activity has been shown in vitro and in vivo (Yan et al., 2007). In another work, it has been confirmed that AtPRMT11 possesses the characteristic features of a type I protein arginine methyltransferase (Scebba et al.,2007) and it can methylate in an asymmetrical fashion histones as well as cellular proteins. Through two-hybrid screening it was also shown that AtPRMT11 interacts with the COOH-terminal portion of AtMBD7 containing the third MBD domain. Furthermore it has been observed that AtMBD7 is the substrate of the activity of AtPRMT11. The methyltransferase is in fact able to methylate the region containing the second MBD domain of AtMBD7 which is particularly rich in RG and RXR contexts (Scebba et al.,2007). These data suggested the existence of an interplay of different epigenetic mechanisms in plant cells (Scebba et al.,2007).

The characterization of PRMT5, a type II enzyme, showed a methylation activity on the arginine 3 of the histone H4 which was necessary to induce vernalization through the repression of the FLOWERING LOCUS C (FLC) (Schmitz et al.,2008).
AtPRMT4a and 4b have also been identified in plants by homology with the mammalian PRMT14 and 13 (Niu et al., 2008). These proteins showed in vitro the capacity to di-methylate the arginine 2, 17, 26 on histone H3 (Niu et al., 2008).

1.8.3 – Histone lysine methylation by PROTEIN LYSINE METHYLTRANSFERASE (HKMT)

Histone methylation is one of the most elaborate modifications; not only it occurs on different residues (lysine (K) and arginine (R)) and distinct sites but also differs in the number of methyl groups added. In *Arabidopsis*, histone lysine methylation occurs mainly at lysine 4, 9, 27 and 36 of histone H3. It has been observed that in *Arabidopsis*, differently from mammals, the lysine 20 in histone H4 can acetylated or monomethylated (H4K20me1) (Naumann K. et al., 2005) (Fig.22).

This epigenetic mark does not change the net charge of the nucleosomes but increases the hydrophobicity and may alter intra or/and intermolecular interactions or/and create new binding surface for proteins that bind preferentially to the methylated domain.
The enzymes involved in this epigenetic modification are called histone lysine methyltransferase (HKMTs) (Fig.22) and they encode for SET domain proteins.

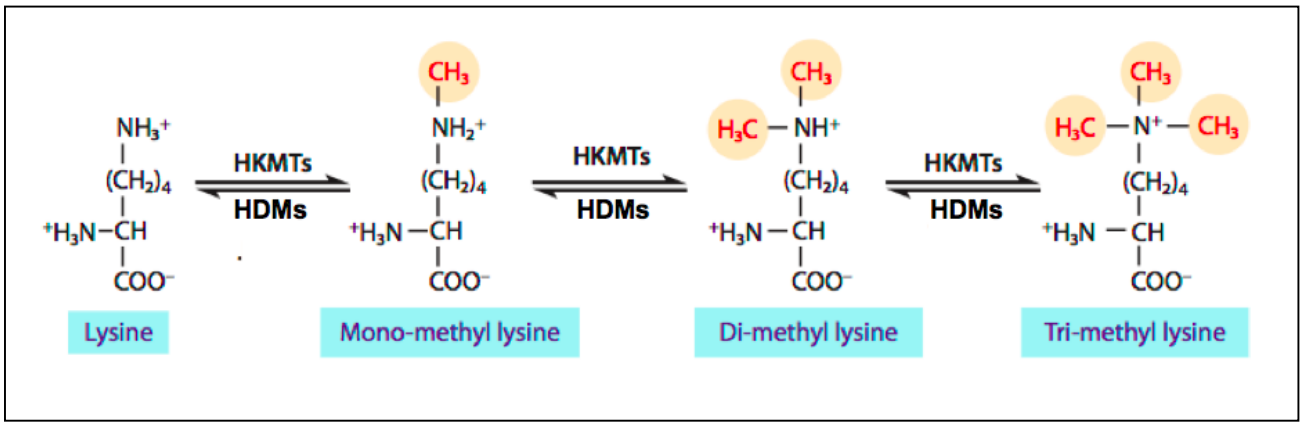

Fig.:22. Lysine methylation by the activity of HKMTs

Based on sequence homology with the animal counterparts, 41 and 37 SET domain proteins have been identified in *Arabidopsis* and rice, respectively. This SET domain contains approximately 130 amino acids and it is rich in cysteine. SET domain proteins in plant are classified in four group: SU(VAR)3-9, Enhancer of zeste (E(z), Trithorax (TRX) and Absent small homeotic disc 1 (ASH1) (Baumbusch et al.,2001; Zhao et al., 2004).
It has been shown that the methylation of the histone H3K9 and H3K27 is associated with silenced regions while H3K4 and H3K36 methylation is associated with active chromatin (Berger SL. 2007).

In *Arabidopsis* the H3K9 is mainly mono- or dimethylated and it has been detected in the chromocenters, repeated sequences and retrotransposons, while H3K9 tri- and also dimethylation can be detected in euchromatin (Bernatavichute et al., 2008; Mathieu et al., 2005).
The first plant H3K9 methyltransferase identified was KRYPTONITE (KYP) (Jackson et al.2002) also known as SU(VAR)3-9 homologous 4 (SUVH4).

Similarly to H3K9, H3K27 can be mono-, di- or tri-methylated. In plants there are only two different protein complexes known to methylate H3K27, they are called: TRITHORAX GROUP PROTEIN (TrxG) and POLYCOMB GROUP PROTEIN (PcG), respectively.
Some TrxG and PcG proteins have intrinsic histone methyltransferase activity, which is mediated by the evolutionary conserved 130-residues SET domain (for SU(VAR)3-9, Enhancer of zeste E(Z), TRX), suggesting that the maintenance of cellular memory involves methylation of histones.

Histone methylation is mediated by Trithorax group proteins (TrxG), which acts antagonistically to Polycomb group proteins (PcG). While PcG proteins are generally required for maintaining a repressive state, the TrxG proteins are responsible for the maintenance of an active state.
The *Arabidopsis thaliana* genome contains five TRX homologs and seven TRX related proteins.
ARABIDOPSIS TRITHORAX_RELATED PROTEIN 5 (ATXR5) and ATXR6 are the only enzymes which have been proved by biochemical assay to cause the mono-methylation of H3K27 (Jacob et al.,2009).

1.8.4 – Polycomb group protein (PcG).

Polycomb group proteins (PcG) were first discovered in the fruit fly *Drosophila melanogaster* and they characterization has revealed that these proteins work in a complex way. Distinct complexes called PRC1 (Polycomb Repressive Complex 1), PRC2

(Polycomb Repressive Complex 2) and PhoRC (Pleiohomeotic Repressive Complex) have been identified (Schwartz and Pirrotta, 2008*)*.

Analysis based on sequence homology showed that the presence of PcG proteins is conserved in *Drosophila* as well as in other taxa. This group of protein is necessary in *Drosophila* for the control of body segmentation by preventing inappropriate expression of homeotic genes (Hox).

1.8.4.1 - Drosophila PcG: Recruiting and gene silence mechanisms.

In mammals and *Drosophila* the action of PcG is based on two principal types of multiprotein complexes, PRC1 and PRC2.

PRC1 contains a core of four proteins, the chromodomain protein polycomb (PC), Posterior Sex Comb (PSC), Polyhomeotic (PH) and dRING (Schwartz and Pirrotta, 2008).

In addition, other proteins able to interact with the PRC1 were characterized: these non-PcG proteins include the protein zeste and several TAFIIs that are associated to TBPs (TATA-Binding-Protein). The TAFIIs are component of the transcriptional factor TFIID.

Zeste protein is able to bind the DNA at specific consensus sequences called PREs (Polycomb Repressive Elements) (Saurin et al., 2001). The PREs are cis-regulatory sequences placed upstream of genes and seem to be necessary to recruit PcG complexes (Sengupta et al.,2004). The PcGs are not able to bind the DNA, thus they depends on other protein factors that can bind the DNA at the PREs level and then recruit the PcGs.

In addition to Zeste, GAGA (GAF) and PLEIOHOMEOTIC (PHO) have been identified in *Drosophila* as factors able to bind the PREs (Brown et al.,1998; Horard et al., 2000).

Different works showed the possibility for PHO to bind the PRC1 and also the PRC2 while GAGA not only binds the PRE but also the TRE (Trithorax Response Element) (Poux et al 2001a; Wang et al, 2004).

PRC1 is not only able to interact with the H3K27me3 through the PC chromodomain but it also has the histone modification activity dRING able to ubiquitylate H2A K119 (Wang et al., 2004) (Fig.23).

The *Drosophila* PRC2 complex, which is responsible for the characteristic chromatin mark histone H3K27me3, includes E(Z), Su(Z)12 and ESC (Schwartz and Pirrotta, 2008; Pirrotta, 2003) and also other non PcG proteins such as RPD3 that is a

histone deacetylase, the histone binding protein p55 (Tie et al, 2001). In *Drosophila* a larger PRC2 complex containing PCL (Polycomb-like) has been also identified (Schwartz and Pirrotta, 2008).

In addiction to the PRC1 and PRC2 a third PcG complex has been identified in *Drosophila* called PhoRC, that has affinity to H4K20me1,2 (Klymenko et al.,2006) (Fig.23).

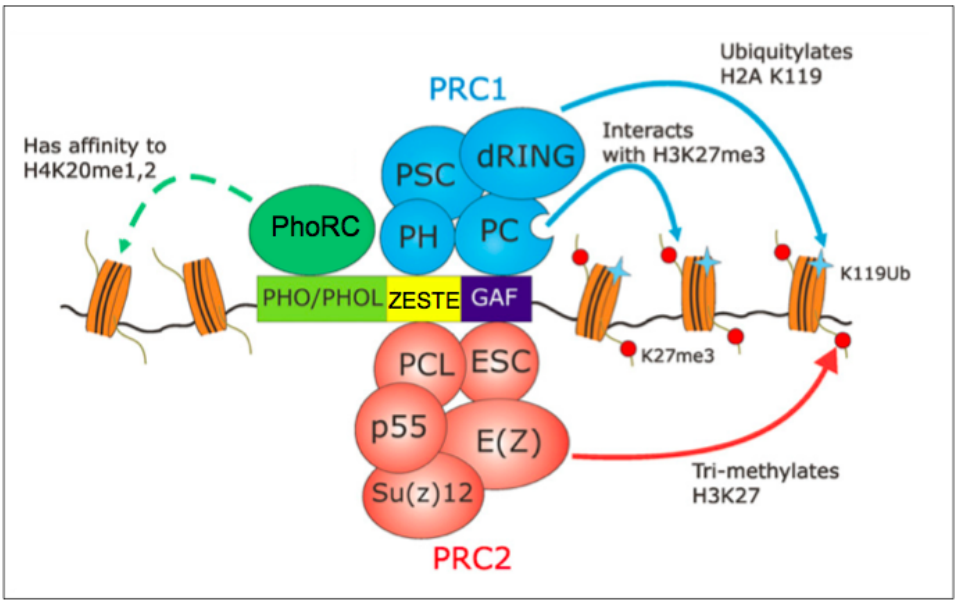

Fig.:23. Biochemical activity of the PcG complexes in *Drosophila melanogaster* (modified from Schwartz and Pirrotta, 2008).

In *Drosophila* the mechanism of gene silencing is not completely understood. The first step is played by the PRC2 complex. It is recruited at the PRE level by PHO, ZESTE and GAGA proteins and by other unknown factors. The PRC2 binds the nucleosome by Su(Z)12, p55 and ESC (Necrasov et al., 2005) and then other factors are recruited. One of these is a HDAC which deacetylates the H3, hence causing its methylation at K9 and K27 residues (Muller et al.,2002).

Afterwards, the PRC1 is recruited by the PC proteins at the level of the PRE (Paro et Hogness, 1991; Cao et al.,2002) via the factors GAGA, ZESTE and PHO (Saurin et al 2001; Poux et al., 2001b).It is not yet clear how the PRC1 modify the chromatin organization. However, it has been shown that a second histone modification is associated with PcG complexes, the ubiquitination of K119 of the histone H2A, which seems to be performed by the dRING component (Schwartz and Pirrotta, 2008). In addition, it has

been shown that the PC component of PRC1 is able to recognize and bind the trimethylation (Schwartz and Pirrotta, 2008).
Furthermore little is known how the silenced state is maintained during cell cycles (Francis and Kingston, 2001).

1.8.4.2 - The plant Polycomb Group Proteins.

Goodrich and co-workers in 1997 (Goodrich et al., 1997) characterize a Polycomb Group gene in *Arabidopsis*. Then, a complete screening of the whole Arabidopsis genome has been made by sequence homology using the well-known Polycomb genes from *Drosophila*.

While *Drosophila* has three different PcG complexes, at the moment only the PCR2 has been clearly identified in plants (Makarevich et al., 2006). In addition, contrarily from animals, plant PcG proteins are encoded by a small gene family whose members encode for proteins involves in the regulation of different developmental pathways (Schwartz and Pirrotta, 2007)

While in the *Drosophila* PRC2 contains Enhancer of zeste (E(z)), suppressor of zeste 12 (Su(z)12), extra sex comb (ESC), in *Arabidopsis* there are three E(Z) homologues MEDEA (MEA), CURLY LEAF (CLF) and SWINGER (SWN); three SUZ12 homologues: FERTILIZATION INDEPENDENT SEED2 (FIS2), VERNALIZATION 2 (VRN2) and EMBRYONIC FLOWER 2 (EMF2); the remaining PRC2 proteins include two WD40 motif proteins, FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) and MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) homologs of the *Drosophila* ESC and p55 proteins (fig:) (Reyes and Grossniklaus et al., 2003) .

Even if no protein homologous to PRC1 members have been found in plants, in the last years five putative PRC1 RING-finger homologs have been identified in *Arabidopsis*: two RING1A/1B and three BMI1A/1B/1C (Sanchez-Pulido et al.,2008).

In *Arabidopsis* AtBMI1A and AtBMI1B were identified as necessary for the monoubiquitination of H2AK121 and are implicated in repression of embryonic and stem cell regulators (Bratzel et al.,2010) while AtMBI1C physically interacts with AtRING1A71B and may be involved in flowering regulation (Li et al., 2011) (Fig.24).

AtRING1A and AtRING1B, homologous to PSC of *Drosophila*, also associate with LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) (Henning and Derkacheva, 2009) to form a complex similar to the animal PRC1(Fig.24).

Two other proteins have been proposed to be involved in PRC1-like functions. These proteins are VERNALIZATION1 (VRN1) and EMBRYONIC FLOWER1 (EMF1), which are able to bind and act together with LHP1 and AtBMI1A/1B (Bratzel et al., 2010) (Fig.24).

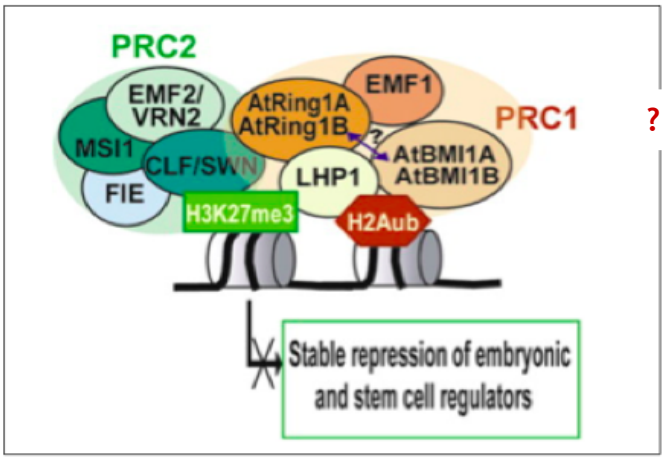

Fig.:24. Organization of thePRC2 and the hypothetical PRC1 in *Arabidopsis thaliana* (image from Bratzel et al., 2010)

1.8.4.3 - The proteins involved in the plant PRC2 complex.

PRC2 proteins coded by the different plant gene families show a high level of homology in the domains involved in the interaction with DNA and with other proteins.

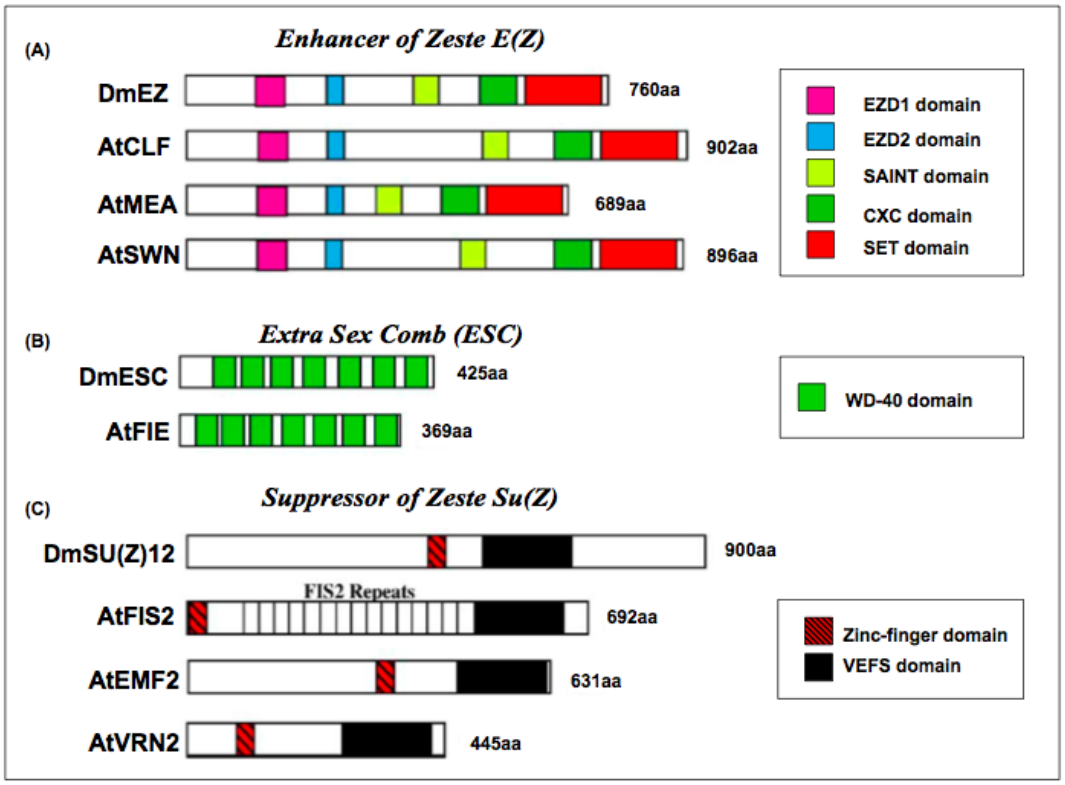

Fig.:25. Schematic domain organization of *Arabidopsis* polycomb proteins.

The proteins of the class Enhancer of Zeste contain five functional domains. The E(Z) domain EZD1 and D2 that contain 70 and 44 amino acids respectively, with the EZD2 domain having a stretch of 5 cysteines that are necessary to bind the protein of the class Suppressor of Zeste (Fig.25A.) and the SANT domain (SWI3, ADA2, N-Cor and T-FIIB) for DNA binding followed by a cysteines-rich domain (CXC) which plays an important role in the activity of the SET domain. This latter domain (130aa-long) is located in the C-terminal region of the protein and bears the methyltransferase activity (Fig.25A.).

Proteins of the Extra Sex Comb class are characterized by repetitions of WD-40 domains bearing a tryptophan-aspartic acid (W-D) dipeptide (Ohad et al.1999). Several of these repeats are combined to form a protein with domains specific for protein-protein interactions (Komachi et al.,1994) (Fig.25B.).

Finally, the Suppressor of zeste proteins are characterized by the presence of two conserved domains. The first is a zinc-finger domain necessary for DNA binding and a VEFS (VRN2-EMF2-FIS2-Su(z)12) domain which mediates the interaction with the EZD2 domain of the Enhancer of zeste class (Yamamoto et al.,2004) (Fig.25C).

1.8.4.4 - Plant Polycomb Repressive Complexes: the functions.

Based on molecular and biochemical evidence, at least three PRC2 complexes co-exist in *Arabidopsis*, harbouring different paralogs of E(z) and Su(z)12 proteins families, each complex controlling a particular developmental program.

PcG proteins dictate the transcriptional status of target genes and therefore control the choice between alternative development programs (Fig.26).

The reproductive FIS2-PRC2 complex, which also contains MEA/SWN, FIE and MSI1, is involved in the regulation of the female gametophyte and seed development through the silencing of target genes (Fig.26).

A second complex called EMF2-PRC2 involved CLF/SWN FIE and MSI1 and, by silencing target genes, suppresses a premature transition from the vegetative to the reproductive stage, furthermore, it takes part with the VRN2-PRC2 in the regulation of floral organs development after vernalization (Fig.26).

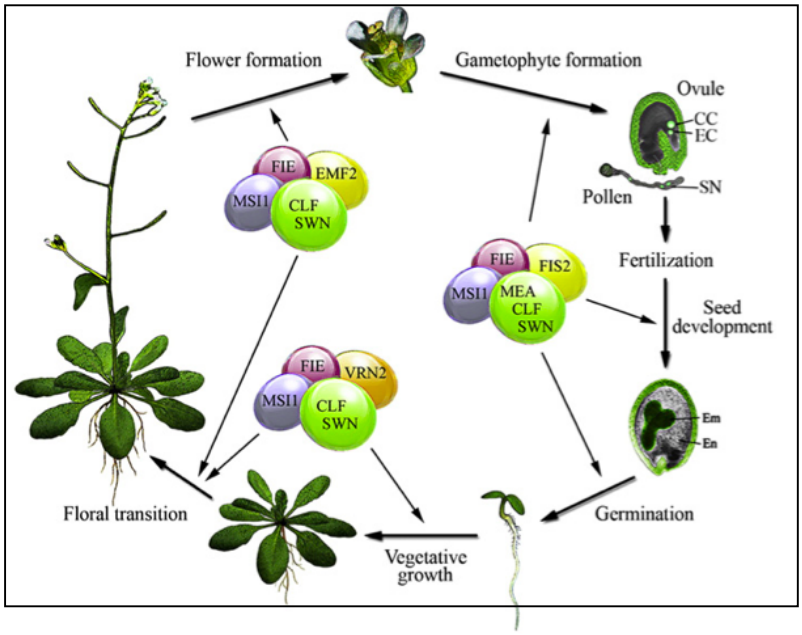

Fig.:26. Regulation of the *Arabidopsis* life cycle by PRC2 complexes.

1.8.4.5 – Role of PcG proteins during floral induction.

The switch to reproductive development (flowering) largely determines the reproductive success of plants and is controlled by endogenous factors such as plant hormones and environmental factors such as photoperiod and temperature (Kohler and Grossniklaus, 2002).

One of the well-known mechanisms for this switch is played by PcG proteins. In *Arabidopsis* two PRC2 complexes are involved in this signalling network. Firstly, the EMF2 complex suppresses precocious flowering by repressing the transcription of flowering activators such as *FLOWERING LOCUS T (FT)*, that is the main flowering time regulator, and *AGAMOUS LIKE 19 (AGL19)* (Kohler and Grossniklaus, 2002; Henning and Derkacheva, 2009; Yoshida et al.,2001).

Upon vernalization that usually occurs at the seedling stage there is a promotion of flowering. This promotion is regulated by the VRN2-PRC2 complex that promotes flowering by downregulating the expression of the MADS-box gene *FLOWERING LOCUS C (FLC),* which acts as a strong floral repressor of the expression of several flowering promoting genes (Kohler and Grossniklaus, 2002; Henning and Derkacheva, 2009).

1.8.4.6 – Role of PcG proteins during seed development.

Both *MEA* and *FIS2* are imprinted in the endosperm. MEA is homologous to *Drosophila* E(z) whose SET domain has methyltransferase activity on lysine 27 of histone 3 (H3K27). FIS2 is a zinc-finger transcription factor homologous to *Drosophila* Suppressor of zeste12 [Su(z)12]. The *FIS* class gene products, MEA, FIS2, and FIE appear to function in a large PcG complex along with additional components. This PcG complex is thought to repress gene transcription via histone modification and chromatin remodeling, and the established patterns are stably propagated through mitotic cell cycles. In addition, direct interactions between AtFIE and AtMSI1, AtFIE and AtMEA, and AtFIE and AtFIS2 have been demonstrated. Furthermore, in fertilized seeds, alterations in the FIS2-PRC2 complex lead to abnormal embryo development and overproliferation of the endosperm. (Kohler and Grossniklaus, 2002;Pien and Grossniklaus, 2007; Grossniklaus et al., 2001).

1.8.4.7 - The PRC2 homologs in *Solanum lycopersicum*.

In a recent work, (How Kit et al.,2010) three genes coding proteins of the Enhancer of zeste class called *SlEZ1*, *SlEZ2* and *SlEZ3* were identified by sequence homology in tomato.

These genes contain two CLF-homologous sequences *SlEZ2* (*SlCLF1*) and *SlEZ3 (SlCLF2)* and one SWN-homologous *SlEZ1* (*SlSWN*) (Fig.27).

It has also been shown that these genes bear the typical signature domains of E(z) proteins EZD1, EZD2, the SANT domain, the CXC domain and the SET domain which has the methyltransferase activity (Fig.27).Further, SlEZ3 encodes a truncated peptide lacking the SET domain.

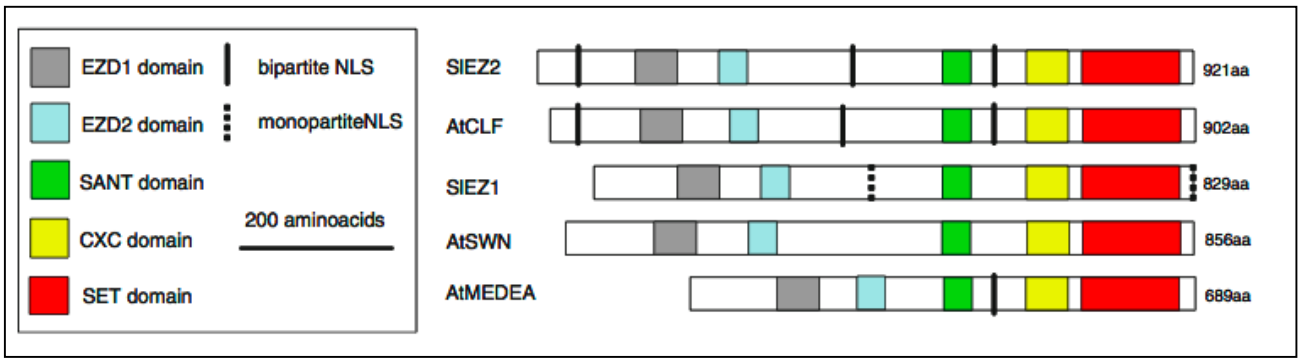

Fig.:27. Conserved domains in SlEZ1 and SlEZ2

By phylogenetic analysis using full-length cDNAs of E(z) proteins from *Drosophila*, *Arabidopsis*, maize, rice and petunia it has also been shown that the similarity between SlCLF2 and PhCLF1, or SlSWN and PhSWN is higher than between CLF and SWN homologues within each species.

In addition, using chimeric constructs GFP-SlEZ1 and SlEZ2, How Kit and colleagues showed that both proteins are targeted into nucleus (How Kit et al.,2010).

It has been shown that SlEZ1 is ubiquitously expressed and its transcripts are abundant in flowers, in the pericarp of fruits and in the jelly mixed with seeds at 30dpa. On the other hand, SlEZ2 transcripts are strongly expressed in open flowers and young fruits (10dpa). Both SlEZ2 and SlEZ1 were detected in young seeds and developing fruit tissues but only SlEZ1 retained its expression level during fruit ripening suggesting different functions for the two proteins (How Kit et al.,2010).

RNAi plants analyzed the activity of the SlEZ1 and the mutants clearly showed an abnormal development of flower organs, with modified petals and lacking the typical anther cone.

1.9 - AIM OF THE STUDY.

The fruit development is a complex set of mechanisms that work together to allow the seed formation and dispersion.
Tomato has been chosen as a model plant to study the fleshy fruits and the mechanisms involved in the different aspect of fruit development are not yet fully understood. On the other hand the tomato *Solanum* section *Lycopersicon* is a relatively small monophyletic clade that consists of 14 closely related species including the domesticated tomato, *Solanum lycopersicum* (formerly *L. esculentum*).

Although the wild tomato species are an important source of germplasm to improve the stress resistance and quality of the cultivated tomato, they have been mainly studied at morphological levels in an effort to understand the phylogenetic relationship inside the clade. In the last years new enzymatic and metabolomic approaches have been developed to analyze the characteristic of the wild species but these studied are limited to few species.
Our work, complementary to the approaches already employed, will try to characterize the whole clade of wild species with a particular focus on fruit development and ripening. This will be achieved by analysing the kinetic of fruit growth development, the expression of genes involved in the process of development and ripening and the cytological structure of the pericarp. The relationships between these different aspects will be analysed in the wild tomato species and compared to what is already known in cultivated tomato.

In addition we will analyse the epigenetic diversity of the wild tomato species by two different approaches. The first consists in the analysis of the DNA methylation at global level and at repetitive sequences in leaves and fruit at different developmental stages while the second approach required the sequencing and analysis of the gene family Enhancer of zeste E(z), which are the catalytic subunits of the Polycomb repressive

complex 2 previously identified in *Solanum lycopersicum* var. *WVa106*. In this case, our attention will be focused on *SlEZ2* and *SlEZ3* genes which are thought to be result of a recent duplication event (How Kit et al., 2010).

CHAPTER 2

CHARACTERIZATION OF WILD TOMATO SPECIES.

2.1 – INTRODUCTION

Although much attention has been devoted to characterizing the wild tomato species (Rick 1973; Spooner et al., 1993; Peralta and Spooner 2007; Spooner et al., 2005) until now most studies have focused on a global morphological characterization of the vegetative part of the plants and on the flower and fruit morphology.

One recent work performed by Peralta and Spooner make use of a comparative morphological trait analysis between wild tomato species to improve the previous classification of the tomato clade which was based on the sequence analysis of single-copy nuclear GBSSI or *waxy* gene (Peralta and Spooner 2001). They analyzed a total of 66 accessions belonging to 10 wild tomato different tomato species, namely: *S. chilense, S. chmielewskii, S. cheesmaniae, S. galapagense, S. habrochaites, S. lycopersicoide Dunal, S. neorikii, S. pennellii, S. peruvianum, S. pimpinellifolium and S. lycopersicum L.*. For each species they considered stem architecture, leaf, inflorescence and flower morphology, fruit structure , and the presence or absence or plant trichomes. Not less than 66 traits were been used in a phenetic analysis, which is widely to classify organisms on the base of taximetrics and/or morphological character (Peralta and Spooner 2007).

Their results supported the existence of at least 10 species as different and the existence of a northern and southern population of *S. peruvianum* as separate taxa. In addition, qualitative traits such as flowers and leaves can be used for species description and cladistic analysis (Peralta and Spooner 2007).

Similar studies characterized the genetic diversity between wild tomato species using AFLP approaches and/or sequence divergence between nuclear genes. Zuriga and colleagues (Zuriaga et al., 2009) analysed 210 different accessions of tomato wild species using AFLP and sequence comparison of two unliked nuclear loci: *CT179* and *CT66* which encode respectively a putative tonoplast intrinsic protein Δ-type and an Arginine decarboxilase. Their results led to a classification similar to those previously proposed

although with some significant differences. Hence the recently proposed species *S. corneliomulleri* was indistinguishable from *S. peruvianum*. Furthermore, *S. arcanum* could be seperated in two cryptic species probably due to their different geographic distribution (Zuriaga et al., 2009).

Recently, new approaches were developed aimed at characterizing the enzymatic activity and/or the metabolomic conposition of fruit and leaves in wild tomato species including *S. pimpinellifolium, S. neorikii, S. chmielewskii, S. habrochaites* and *S. pennellii* (Schauer et al., 2005; Steinhauser M.C. et al., 2010). Typical results showed a high variance in metabolite content in both leaves and fruits of the wild species (Schauer et al., 2005). Furthermore, analysis of leaf metabolite content, show *S. pimpinellifolium* was the closest wild species to *S. lycopersicum* followed by *S. pennellii*, *S. chmielewskii* and *S. neorikii* while *S. habrochaites* was the most distinct. These observations are quite different from those obtained using genetic and morphological approaches. Furthermore, all this information can be used for metabolomic engineering in wide breeding strategies (Schuauer et al 2005).

Although a lot of works have been done on wild tomato species the literature lacks of complete work on the dynamic of fruit development concerning kinetic of fruit development and weight, pericarp thickness, cell size, endoreduplication as well as gene expression analysis of gene regulatory the fruit development and ripening. A fine characterization of all this aspects and their relationship could be complementary to what is already known and could show new aspect of the fruit development.

In this study we want to analyse the relationship between wild and cultivated tomato species considering fruit growth and development characteristics that it has already been well described in cultivated tomato species.

In the following chapter we aim at providing a fine characterization of the development of fruits of a selected set of wild and cultivated tomato species. Various aspects of fruit development including morphological, physiological, cytological have been considered. We focused our attention on the pericarp structure (number of cell layer, cell surface) and its evolution during fruit development. Although these data are available for cultivated tomato (Cheniclet et al.,2005; Faurobert et al., 2007) there is no comprehensive analysis of these parameters in fruits of wild tomato species. In addition, the expression of a set of genes known to characterize the different phases of development of cultivated tomato fruits has been analysed in the species under study.

2.2 – MORPHOLOGICAL AND PHYSIOLOGICAL CHARACTERIZATION OF FRUITS FROM WILD TOMATO SPECIES (*Solanum* l. sect. Lycopersicon)

Wild tomatoes species are native from western South America and grow in a variety of habitats, from near the sea level to over 3300m in numerous valleys of the Andes. It has been shown that the major influences on species distribution appear to be precipitation; temperature and vegetation cover suggesting that relatively few environmental conditions determine the different species habitats (Moyle, 2008). These various environmental conditions have been the driving forces leading to the actual diversity of the wild tomato species. For the time being, the monophyletic clade of *Solanum* section *Lycopersicum* includes 14 related species with almost 1500 accessions. (see general introduction, Zuriaga et al., 2009)

In this study one representative accession for each species was provided by C.M. Rick Tomato Genetics Resource Center (TGRC, http://tgrc.ucdavis.edu/) of the University of Davis, California. These include: *WVa106* (cherry tomato), *S.l. var. cerasiforme* (LA1226), *S. cheesmaniae* (LA0930), *S. pimpinellifolium* (LA0722), *S. chmielewskii* (LA1330), *S. neorikii* (LA1326), *S. arcanum* (LA2152), *S. corneliomullieri* (LA0103), *S. huayalasense* (LA1982), *S. chilense* (LA1930), *S. habrochaites* (LA1353), *S. pennellii* (LA1926). Four to five plants for each species were grown in greenhouse under controlled conditions. Among these only eight species produced flowers and fruits (Fig.28). Depending on the species self-pollination was performed by vibrating the flowers (*WVa106*, *S.l. cerasiforme*, *S. pimpinellifolium* and *S. neorikii*) while for the few species allogamous and self-incompatible (*S. arcanum, S. corneliomullieri, S. huayalasense* and *S. pennellii*) pollen was harvested at flower anthesis and used in manual pollination.

Although two independent cultures with a minimum of 4 plants were performed in 2009 and 2010 using similar accessions, a few species (*S. chesmaniae, S. chmielewskii, S. chilense* and *S. habrochaites*) did not produce fruits in the green house conditions used in this study (Tab.2).

		Flowering conditions		Greenhouse conditions	
Specie:taxonomy	**Accession**	**Long Day**	**Short Day**	**Flower**	**Fruit**
WVa106				✔	✔
S.l.cerasiforme	LA1226		✔	✔	✔
S.chesmaniae	LA0930	✔		✖	✖
S.pimpinellifolium	LA0722		✔	✔	✔
S.chmielewskii	LA1330		✔	✔	✖
S.neorikii	LA1326		✔	✔	✔
S.arcanum	LA2152		✔	✔	✔
S.corneliomulleri	LA0103		✔	✔	✔
S.huayalasense	LA1982		✔	✔	✔
S.chilense	LA1930	✔		✖	✖
S.habrochaites	LA1353	✔		✔	✖
S.pennellii	LA1926		✔	✔	✔

Tab.2. Table of the species cultivated with their accession number and flowering conditions. In the greenhouse columns are indicated the species that produce flowers and fruits. Signs: positive (✔) and negative (✖).

For all other accessions, fruits were harvested at 10, 20, 30 days post anthesis (dpa) and at Breaker (Br.), Turning (Tu.), Orange (Or.) and Red Ripe (RR.) stages of ripening for the red-fruited species. Since most wild species do not accumulate carotenoids and remain green during ripening, their development lacks clear criteria that characterize the ripening process. Fruit from green-fruited species were therefore harvested every ten days from 10 to 60 dpa (Fig.28).

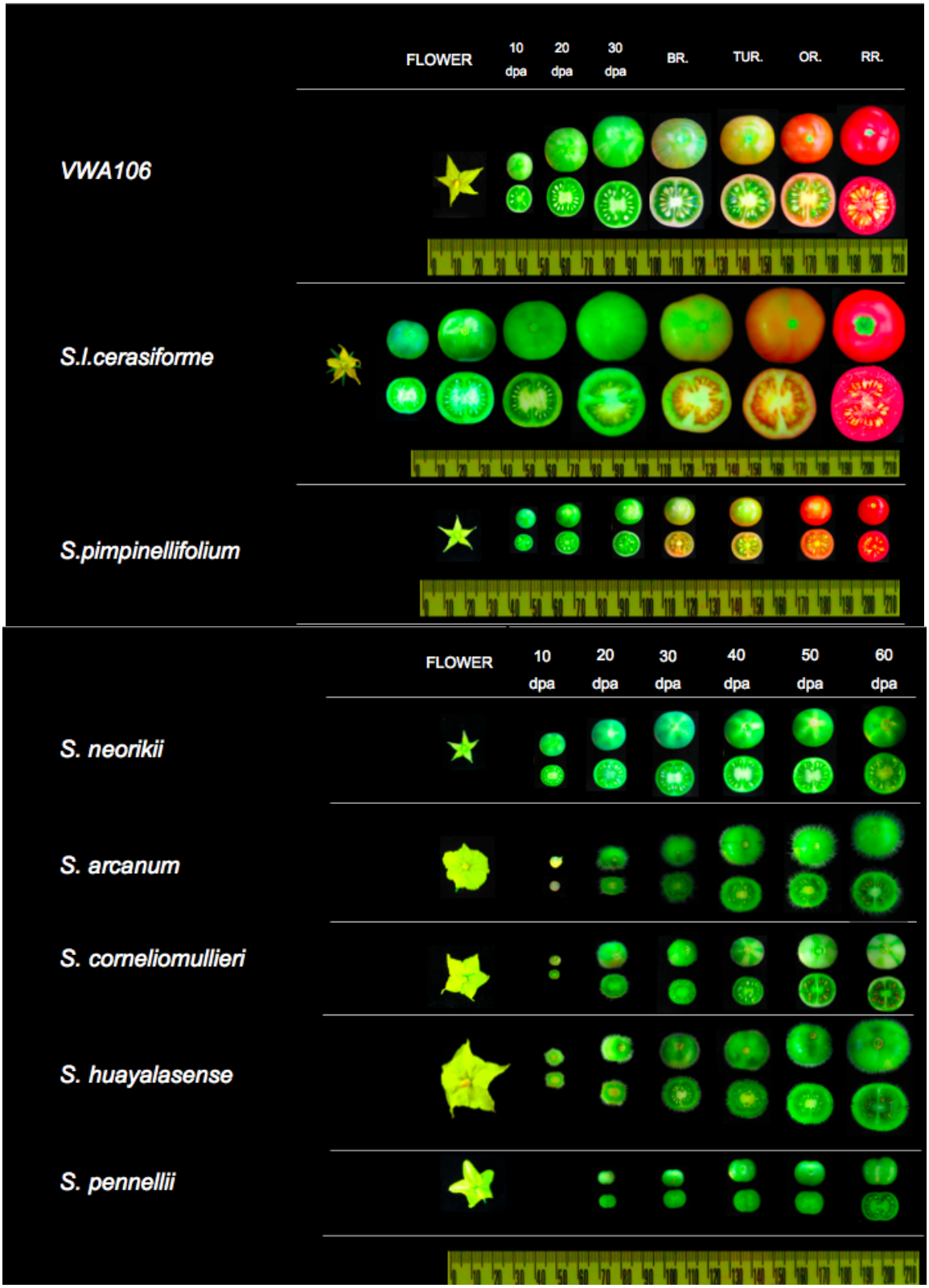

Fig.:28. Flowers and fruits from cultivated tomato and wild relatives during development. On the left the species: *WVa106, S.l.cerasiforme* (LA1226), *S.pimpinellifolium* (LA0722), *S.neorikii* (LA1326), *S.arcanum* (LA2152), *S.corneliomulleri* (LA0103), *S.huayalasense* (LA1982), *S.pennellii* (LA1926). On the right, the fruits harvest at different developmental stage. The fruits were harvested at 10, 20 30 days post anthesis (dpa) and a breaker (BR.), turning (TU.), orange (OR.) and red ripe (RR.) stages for red fruited species. Green fruits were harvested at 10, 20, 30 40, 50, 60dpa. Flower of the different species are also showed.

Red-fruited species developed fruits characterized by a significant difference in size (Fig.28). At 10dpa *S. pimpinellifolium* fruits had a diameter of 7.7±0.67 mm and *WVa106* has diameter of 11.88±1.7 mm. The largest fruits with a diameter of 17±1.14 mm at 10 dpa developed on *S.l. cerasiforme* (Fig.29A). Despite these initial size differences, the kinetic of fruit size increase was globally similar with an initial rapid growth followed by a plateau at the breaker stage. Fully ripen fruits were in all cases approximately twice larger than 10 dpa fruits with a diameter of 14.08±1.13 mm for *S. pimpinellifolium*, 25.3±1.07 mm for *WVa106* and 32.58±2.59 mm for *S.l. cerasiforme* (Fig.29A).

It is worth noting that, although all red-fruited fruits reached the fully ripen stage at approximately the same age (42 to 45 dpa), the developmental kinetic showed some differences. In the case of *S.l. cerasiforme* the breaker stage was reached in average 8 days (41dpa) late than in the *WVa106* variety (33dpa) and in *S. pimpinellifolium* (34dpa). This is consistent with the fact that the growth phase of *S.l. cerasiformae* is longer than for the two other species (Fig.29). Inversely the ripening phase was quicker and lasted only 4 days in *S.l. cerasiformae* contrary to *S lycopersicum lycopersicum* variety *WVa106* and to the wild species *S. pimpinellifolium* which have a ripening period of 10 days. It should be noted that colour changes did not occur similarly in all species developing red fruits. In *S.l. cerasiforme* the locular tissue was coloured before the pericarp changed colour whereas an opposite situation was observed in *WVa106*. In fruits of *S. pimpinellifolium* the change in colour occurred simultaneously in the locular and the pericarp tissues.

The observation noticed between WVa106 and *S.l.cerasiforme* could suggest a different dynamic of fruit ripening. For all the species there is a linear increasing in weight during fruit development and ripening with a weight at 10dpa of 0.09±0.04g for *S. pimpinellifolium*, 0.82±0.34g *WVa106* and 2.13±0.52g for *S.l. cerasiforme.*

Fruits from green-fruited species are very diverse (Fig.28). Some such as fruits from *S. arcanum* and *S. huayalasense* present a hairy aspect while *S.neorikii, S.corneliomullieri* and *S.pennellii* have not.

The colour of the fruit change between different species: *S.huayalasense* and *S.arcanum* showed a change in colour from light green to dark green. *S.neorikii* showed to be dark green with five line light yellow or white in the upper part. Opposite situation was

presented on *S.corneliomulleri* while *S.pennellii* did not showed change in colour during fruit development.

S. neoriki, *S. arcanum* and *S. corneliomullieri* (Fig.29B) similarly to the cultivated species, are characterized by a rapid increase of fruit size from 10 to 30-40dpa before reaching a plateau after 40 dpa. Other species, such as *S. huayalasense* and *S. pennellii*, show a bimodal kinetic of size increase with a first increase occuring between 10 and 30dpa and a second one starting at 50 dpa (Fig.29 B and C).

When considering the increase in fruit weight (Fig.29E and F), species could also be separated in two groups. The first group, *S. neorikii*, *S. corneliomullieri* and *S. Arcanum,* (Fig.29E) behave similarly to the red-fruited species and is characterized by a linear increase of fruit weight. . In contrast fruits from *S. huayalasense* and *S. pennellii* show a bimodal kinetic of fruit weight, which overlap the fruit size, increase (Fig.29F).

Based on all this parameters, species could be separated in two groups: group 1 which include all the red-fruited species and *S.neorikii, S.corneliomulleri* and *S.arcanum* that are characterized by a linear increase in fruit weigh and a fruit size that increase until 30dpa and/or 40dpa for *S.neorikii, S.corneliomulleri and S.arcanum*; group 2 involve *S.huayalasense* and *S.pennellii* which showed an increase in fruit size and weigh in a bimodal fashion.

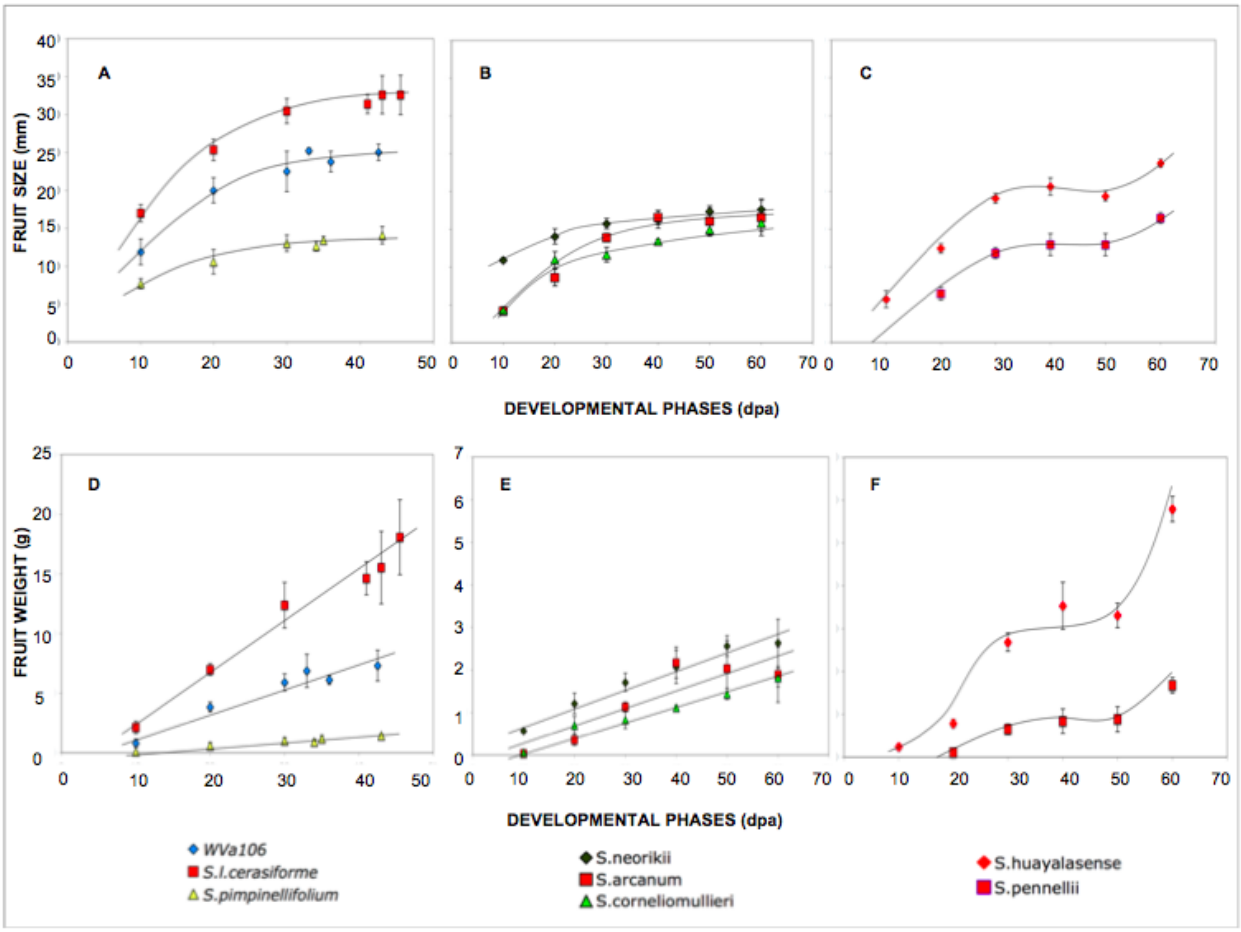

Fig.:29. Kinetic of fruit development. Panels A, B and C describe the increase in fruit size in the different species while panels D, E and F describe the increase in fruit weight. In addition the species are organized according to the phylogenetic tree and subdivided on the basis of the colour of the cherries. Panels A and D contain the red-fruited species: *WVa106*, *S.l. cerasiforme* (LA1226), *S. pimpinellifolium* (LA0722); panels B, C, E, and Fshow the green-fruited species. In B and E: *S. neorikii* (LA1326), *S. arcanum* (LA2152) and *S. corneliomullieri* (LA0103) while C and F *S. huayalasense* (LA1982) and *S. pennellii* (LA1926). The black line (—) shows the trend increasing.

2.3 – FRUIT PERICARP ANALYSIS IN CULTIVATED TOMATO AND WILD RELATIVES.

2.3.1 – Pericarp description.

Although the fruit is complex organ that involve deep biochemical and physiological changing of their different structures (epidermal cells, pericarp, locular tissues and seeds) we focused our attention on the pericarp structure. Pericarp is subjected to different

modification mechanisms that involve its own cells during all fruit development and ripening. These change involve the endoreduplication, cell size as well as gene involved in the processes of fruit growth, ripening and softening. In addition the pericarp can be easily studied compared to the other parts of the fruit.

In this part of the study we analyzed the pericarp thickness during fruit development and ripening to analyze if it evolves like in the well know cultivated species.

During the development of *WVa106* fruits, cell division are mainly located within the outer and inner subepidermal cell layers which are the source of most of the new cell layers Inversely, the mesocarp cells are less subjected to cell division and more to cell expansion. In this variety, the cell layer number is already fixed at 5 dpa and maintained during all fruit development and ripening (Cheniclet et al., 2005). Similar results have been obtained on the cultivar Cervil although in this case, the cell layer number will increase during the first 14 dpa (Faurobert et al.,2007).

In the present study, three to five fruits were selected at each developmental stage listed above and use to measure cell size, pericarp thickness, number of cell layer, and endoreduplication levels (see Chapter 3).

A first analysis of the pericarp structure showed that both in red-fruited and in green-fruited species the classical cellular organization of the pericarp is clearly visible. The external epidermis can be easily observed, under which the exocarp or outer subepidermal cell layers, a mesocarp and the endocarp or inner subepidermal cell layer are found (Fig30A - B). In addition, as already observed in the *WVa106 variety* (Cheniclet et al.,2005), at 10dpa outer and inner subepidermal cell layer are composed of small cells presenting the characteristic of actively dividing cells. Inversely the mesocarp contains a set of already enlarged cells that are unlikely to divide actively (Fig30A - B). However, all species are not equivalent at this stage as fruit mesocarp of red-fruited species contains already very large cells that are not found in species developing green fruits at this stage. Thus, red-fruited species such as *WVa106* showed a general increase in the size of cells of the inner and outer subepidermal layers, while green species showed more complex pattern of cell size increase. The aspect of the cell size was also analyzed and will be discus later in this chapter.

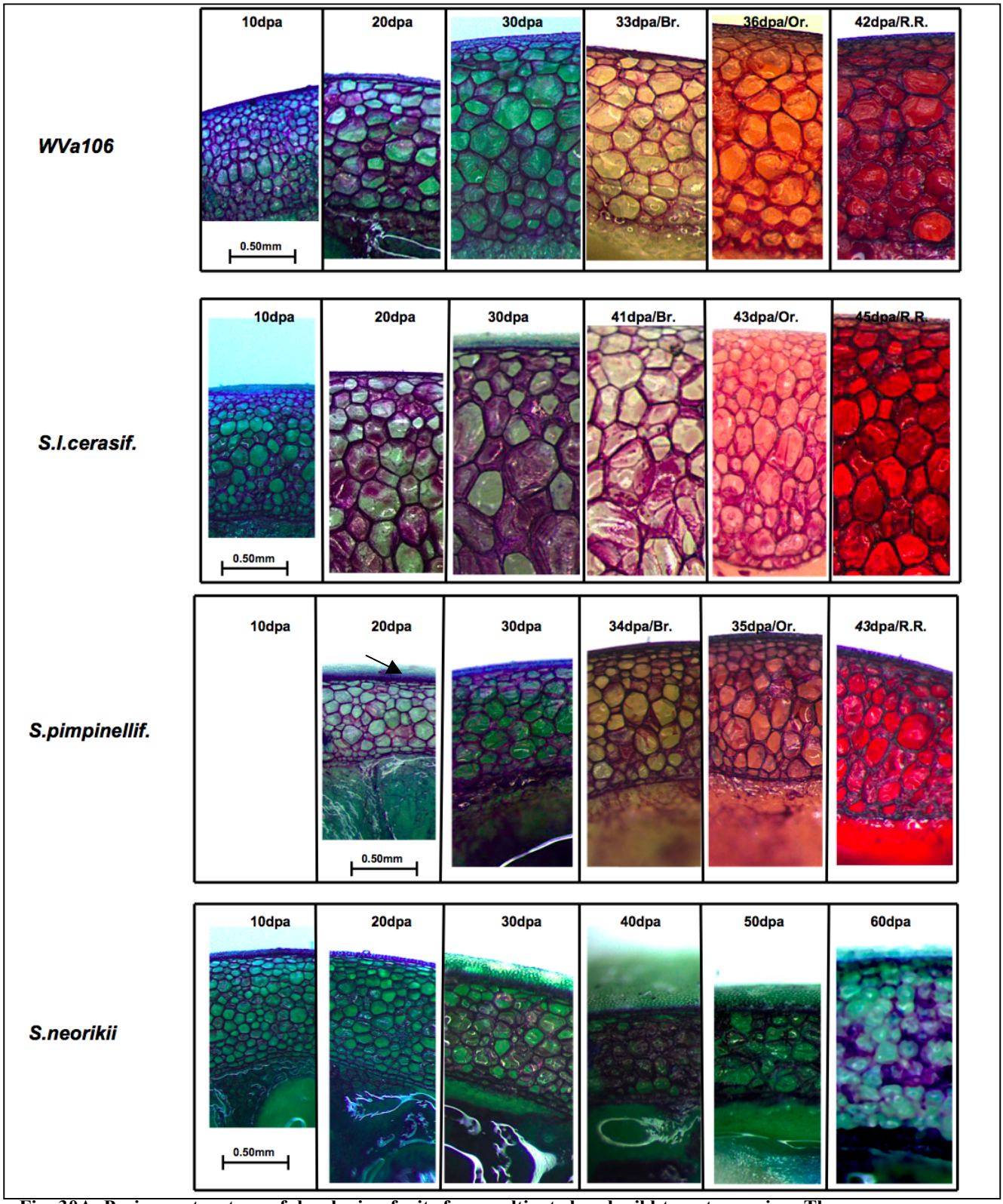

Fig.:30A. Pericarp structure of developing fruits from cultivated and wild tomato species. The name of the species is written on the left and developmental stages are indicated within each panel. *WVa106, S.l.cerasiforme*, *S.pimpinellifolium, S.neorikii*. Pictures were done after hand cutting of the fruits, 0.05% touloidine blue staining and observation under a LEICA MZFLIII stereomicroscopy. All the picture are representative of the pericarp stricture at each developmental stage. Pictures are representative images from 802 photos. *S.pimpinellifolium* pictures at 10dpa are not showed because not informative. Bars =0,50mm. Black arrows show the outer subepidermal cell layer.

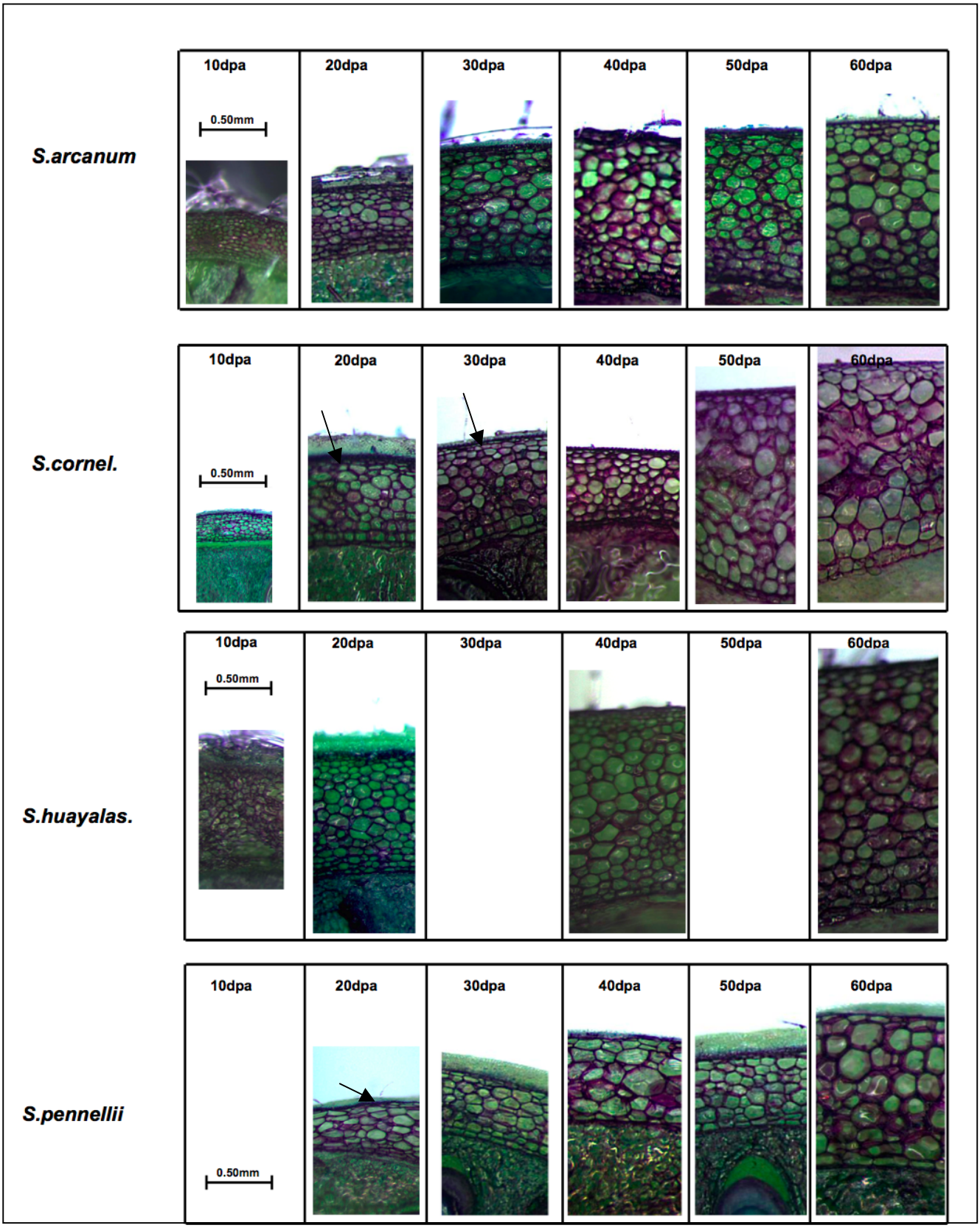

Fig.:30B. Pericarp structure in cultivated and wild tomato species during fruit development and ripening. The name of the species is indicated on the left. *S.arcanum*, *S.corneliomulleri*, *S.huayalasense*, *S.pennellii*. Pictures were done after hand cutting of the fruits, 0,05% touloidine blue staining and observation under a LEICA MZFLIII stereomicroscopy. All the picture are representative of the pericarp stricture at each developmental stage. Pictures are representative images from 802 photos. *S.huayalasense* pictures at 30 and 50dpa get lost. Fruits of S.pennellii at 10dpa are not present. Bars =0,50mm. Black arrows show the outer subepidermal cell layer.

Pericarp thickness was analysed at all stages of fruit development. At least 3 fruits for each stage and specie were analysed which represent a total of 144 fruit and 288 measurements. It is noteworthy that in *WVa106, S. pimpinellifolium, S. corneliomulleri, S. arcanum,S. huayalasense* and *S.pennellii* fruit pericarp thickness increased in a linear way, while in *S. neorikii* seemed to be stable until 50dpa and than it increase its size up to 60dpa. *S.l. cerasiforme* showed a bimodal increase of the pericarp thickness. Obviously thickness of fruit pericarp differed between species, although in all species pericarp thickness was positively correlated with fruit size. Hence, small fruits, such as those *S. pimpinellifolium* or of *S. corneliomulleri* were in general characterized by a very narrow pericarp (0.74±0.11 mm and 0.21±0.07mm large at 10 dpa respectively), whereas the pericarp of the large fruits from *S.l. cerasiforme,* was already 1.15±0.14 mm large at this stage.

On the base of pericarp thickness analysis (Fig.31), the species can be separate in three groups. The first one includes: *WVa106, S. pimpinellifolium, S. corneliomulleri, S. arcanum, S.huayalasense* and *S.pennellii* that showed a linear increase of the pericarp thickness during almost all the fruit development. A second group includes *S. neorikii* which showed a thick of its pericarp constant during the first 50dpa. Finally, *S.l. cerasiforme* had a pericarp thickness that increase in a bimodal fashion.
As pericarp thickness is determined by the number of cell layers and by the average cell size, we analysed at all developmental stages these two parameters, to determine in each species their relative contribution.

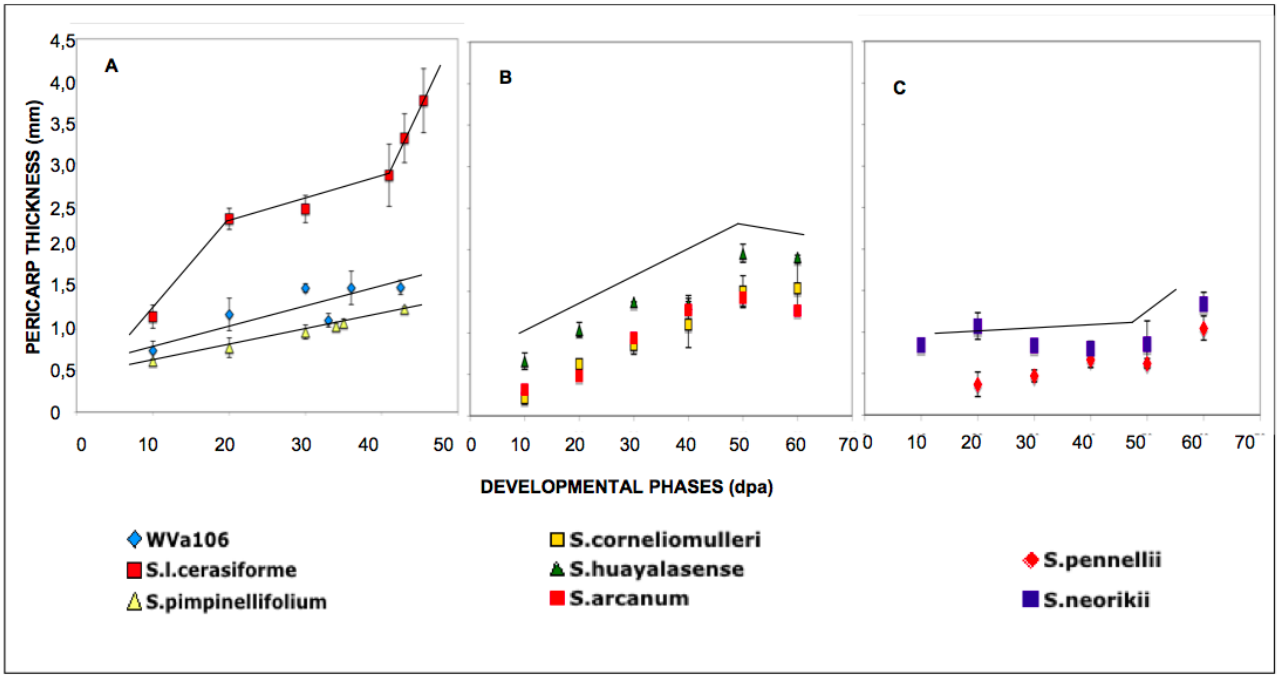

Fig.:31. Pericarp thickness increase during fruit development. (A) red-fruited species (B) green-fruited species *S. corneliomulleri, S. huayalasense and S. arcanum* while in (C) *S. pennellii* and *S. neorikii*. The increasing rate calculated by slope was approximately the same for *WVa106, S. pimpinellifolium, S. arcanum, S. huayalasense*, *S. corneliomulleri* and *S. pennellii* (average 0,022±0,004). In *S.l. cerasiforme* the pericarp thickness increases in a bimodal fashion, a first increasing between 10 and 20 dpa and the second during the ripening phase). *S. neorikii,* showed an increase in pericarp thickness only between 50 to 60dpa (slope =0.005). The black line (—) shows the trend of the pericarp thickness increase.

2.3.2 – The impact of the cell layer number on the pericarp structure and dynamic of development

The number of cell layers, measured at the red ripe stage for species developing red fruits and at 60dpa for species with green fruits varied significantly between species. Species were separated in three groups based on the Tukey's statistical test (Fig.32, see methods for detail). A first group which is composed of plants with fruits having a cell layer number ranging between 16 and 18 includes *S.l. cerasiforme, S. arcanum* and *S. huayalasense*. A second group which includes *WVa106, S. pimpinellifolium, S. corneliomulleri* and *S. neorikii* is characterized by fruits with an average cell layer number at maturity of 12,78±1,78. *S. pennellii* correspond to the third group with a number of cell layers of approximately 8, significantly lower than all other species. Although before date has not been found any correlation has been found before between pericarp thickness and number of cells layer during fruit development, the pericarp of the

species with a high number of cell layers (In *S.l. cerasiforme, S. arcanum* and *S. huayalasense)* is characterized by small cells located in the inner and outer subepidermal layer. These small cells are most probably in an active state of cell division and hence responsible for the final number of cell layers.

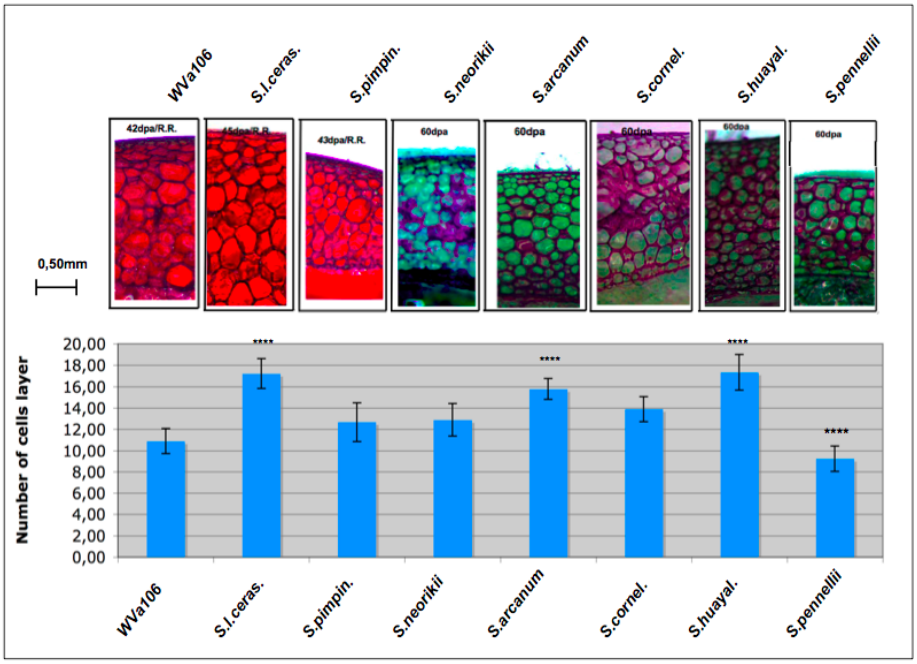

Fig.:32.Number of cell layers in cultivated tomato and wild relatives at red ripe or 60dpa stages. Data are means ± standard deviation. The p-value obtained by statistic Tukey's test lower than 5, 1, 0,1 and 0,001% is indicated respectively with: *, **, * and ****; (n=72).**

As we previously observed that both fruit growth kinetics and fruit pericarp thickness changes differed between species (Fig.29 and 31), cell layer number was counted at all developmental stages. It has already been described that in the cultivated species *S. Lycopersicum lycopersicum*, variety *WVAa106*, that the cell layer number is already fixed after 4 to 5 dpa (Cheniclet et al., 2005). Obviously a similar trend is observed in our study for this species and for *S neorikii* with no variation in cell layer number from 10 dpa to the red ripe stage and to 60 dpa respectively (Fig.33A). At the contrary, all other species were characterized by a progressive increase in the number of cell layers within the pericarp during fruit development, although with different kinetics. Hence *S.l. cerasiforme* fruit pericarp contained 12.33±1.12 cell layers at 10dpa. Cell layer number did not change until 40 dpa and increased during ripening to reach 17.22±1.39 at the RR stage. Late increase of cell layer number was also observed in *S. pimpenifolium* with a progressive increase after 20 dpa. In the remaining species, *S. arcanum, S. huyalasense,*

S. corneliomulleri and S. pennellii, the cell layer number increases during fruit development but the kinetic differed clearly between species (Fig 33C).

S.huayalasense and *S. pennellii* as already observed for fruit size and weight present two phases of increase in cell number layer. The first one takes place between 10 and 30dpa and the second one extends from 50 to 60dpa.

This observation agrees with the initial analysis of the pericarp structure (Fig.30A and B) as in *S. pimpinellifolium, S. corneliomulleri*, and *S. pennellii* the inner and outer subepidermal cell layers were visible during the first 20, 30 and 40dpa (Fig.30A and B) and suggest that the increase of the pericarp thickness involves cell division also in the central and later phases of fruit development.

In addition, as already observed in *WVa106*, the new cell layers seem also to be originated in the wild species predominantly from the outer subepidermis layer and not from the inner one.

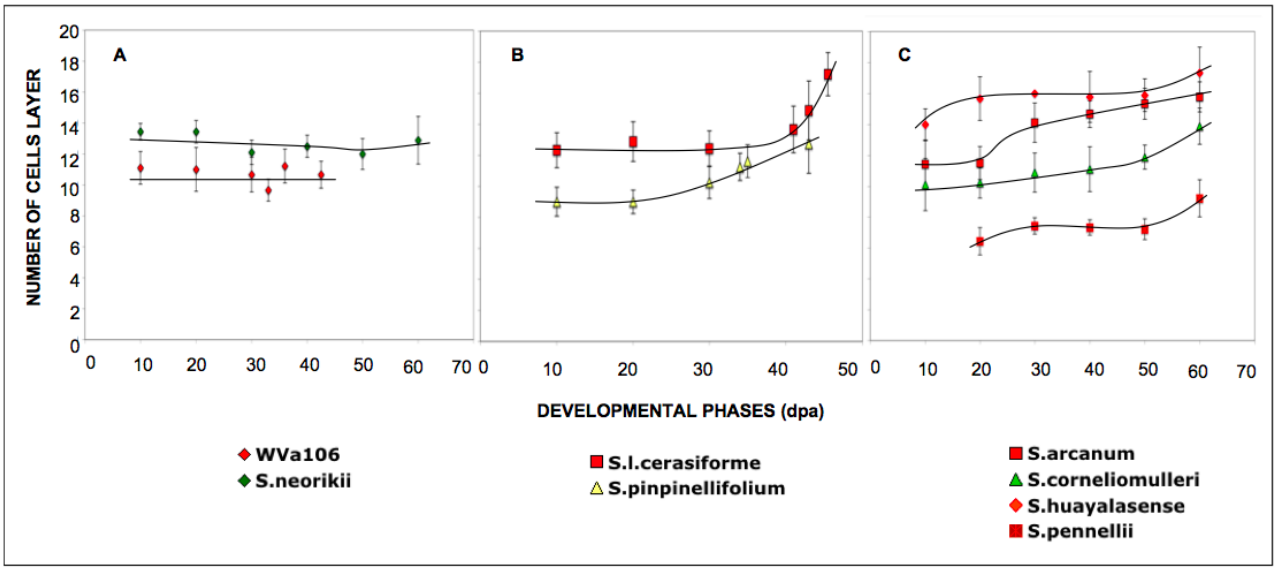

Fig.:33. Change in cell layer number in pericarp during fruit development. Although the number of cell layers is genetically programmed during the first 3-10 dpa and it is maintained during fruit development in the cultivated species (*WVa106*) (A), the wild species have a different behaviour (B and C). the red and green-fruited wild species show an increase in the number of cell layers. Interestingly *S. neorikii* shows the same behaviour as *WVa106*. The black line (—) shows the trend of the C value increase.

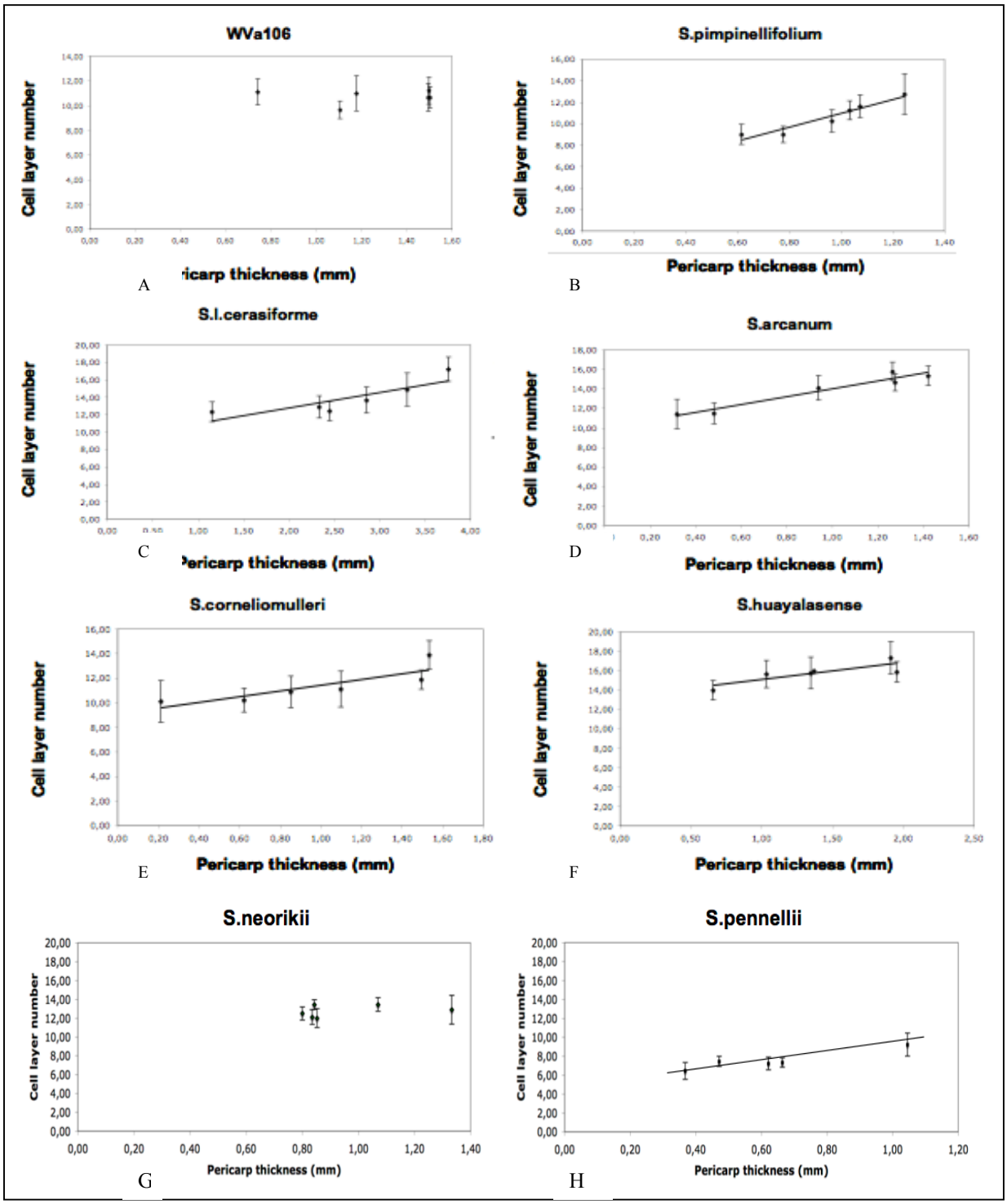

Fig.:34.Relationship between pericarp cell layer number and pericarp thickness. In *WVa106* (A) and *S.neorikii* LA1326(G) the number of cells do not change with the increasing of the pericarp thickness and fruit development. Wild tomato species (B,C,D,E,F and H) showed a relationship approximately linear between pericarp thickness and cell number layer.

Our observations suggest that the role of cell division and cell expansion have a different impact on fruit development in cultivated and wild relatives. Unlikely that in *WVa106* where the cell division is involved in fruit development mainly during the first days after

anthesis whereas the cell expansion is more important during fruit growth. Hence, *WVa106* fruit growth is not correlated with increase in cell layer number (Fig.34A) as attended. Opposite situation was found in wild relatives (*S.l. cerasiforme, S. pimpinellifolium, S. arcanum, S. corneliomulleri, S. huayalasense* and *S. pennellii* Fig.34B-F) where the increasing of the fruit thickness is linked with an increase in cell layer number. In S.neorikii has showed a different situation compared to the wild species, in fact it did not show an increase in cell layer number during all the fruit development like *WVa106.*

In contrast to *WVa106* and *S. neorikii*, other wild tomato species (*S.l. cerasiforme, S. pimpinellifolium, S. arcanum, S. corneliomulleri*, *S. huayalasense* and *S. pennellii* Fig.34B-F) showed a linear correlation between the pericarp thickness and the number of cell layer. Thus, cell division is maintained at all fruit development stages and participates to the increase of pericarp thickness. This contrast with the *WVa106* situation, as in this case, cell expansion has been shown to be the major contributor to the increase in pericarp size.

2.4 – CELL SIZE ANALYSING.

The switch from cell division to cell expansion plays a major role during fruit development in cultivated tomato species. In these species, pericarp cell expansion is involved in the increase of pericarp thickness (Cheniclet et al., 2005; Bertin 2005). In order to determine the contribution of this process in the wild species on pericarp thickness, cell size was analysed at various developmental stages using the fruits samples described above.

The mean size of pericarp cells was determined in cross sections of parenchymatous (not vascular) part of the mesocarp. The cell size was measured from cell of the mesocarp, which showed homogeneous cells easily identifiable. For these reason we excluded the most outer and inner layers of pericarp cells; we also choose to exclude the cells of the vascular bundles due to the heterogeneous cells.

At all stages, *S.l. cerasiforme* pericarp contains cells with the highest average surface. However, in all red fruited species the increase in cell surface followed a similar

kinetic. Cell increase is rather slow between 10 dpa and the breaker stage and again during ripening. Between these two phases a rapid increase in cell size occurred between the breaker stage and the orange stage. A first step of increase occurs from anthesis to the breaker stage where the mean cellular surfaces were 0.07±0.005 mm^2 for *S.l. cerasiforme*, 0.03±0.007 mm^2 for *WVa106* and 0.01±0.001 mm^2 for *S. pimpinellifolium*.

Between breaker and orange stage *WVa106* and *S. pimpinellifolium* showed the maximum increase in cell size. In one (*S. pimpinellifolium*) or 2 (*WVa106*) days the cell surface increases from 0.030±0.007 mm^2 up to 0.04±0.004 mm^2 in *WVa106* whereas in *S. pimpinellifolium* increases from 0.01±0.001 mm^2 up to 0.02±0.002 mm^2.

A second step of low increase in cell surface was observed during the last phase of the ripening phase (from orange to red ripe) where the cells had a surface of 0.044±0.007 mm^2 for *WVa106*, 0.076±0.007 mm^2 for *S.l. cerasiforme* and 0.032±0.004 mm^2 for *S. pimpinellifolium* (Fig.35).

Pericarp cells from green fruit showed a bimodal increase in cell size. These two steps proceeded in a species-specific manner (Fig.35). In *S. arcanum, S. corneliomulleri, S. huayalasense* the first step of increase occurs during the first 30dpa which is then followed, after a transition phase, by the second increase phase at 40 dpa. *S. pennellii* showed a transition phase longer than the other green fruits and the second phase of increase starts at 50dpa (Fig.35).

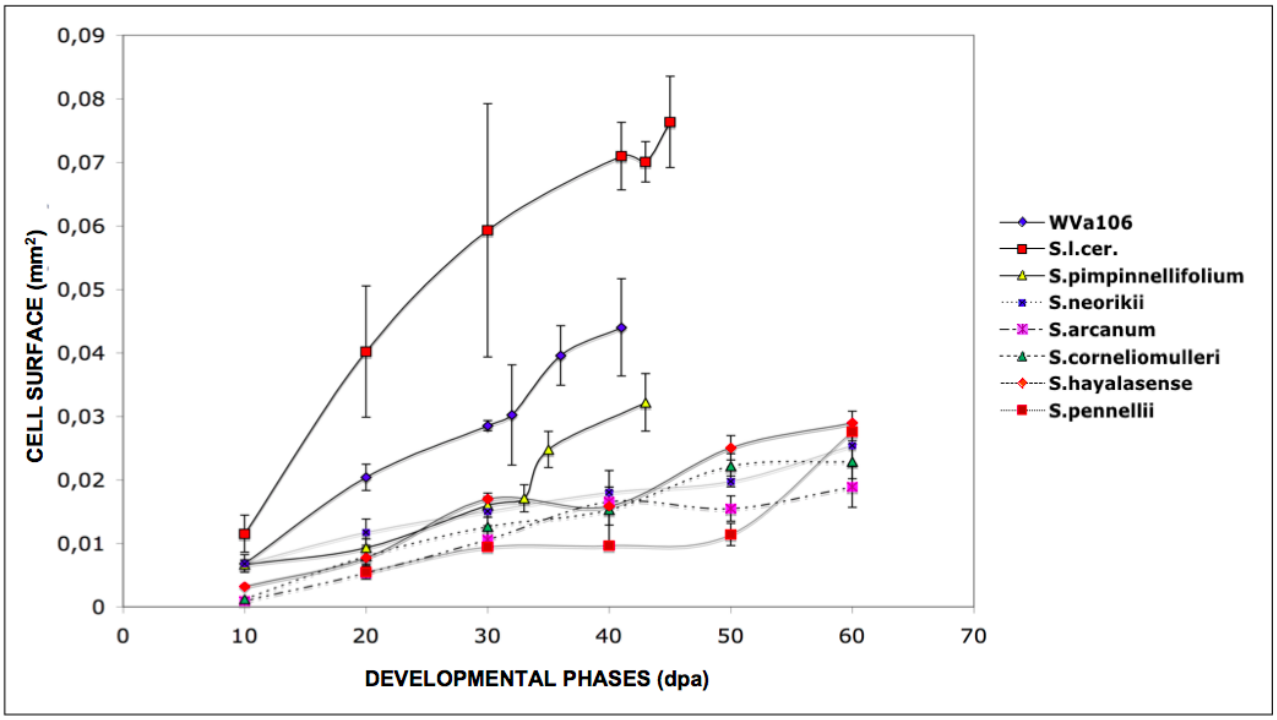

Fig.:35.Evolution of pericarp cell surface during fruit development. In red-fruited species (*WVa106, S.l. cerasiforme* LA1226, *S. pimpinellifolium* LA0722), the increase of cell surface is higher than that of the green-fruited species (*S. neorikii* LA1326, *S. arcanum* LA2152, *S. corneliomullieri* LA0103, *S. huayalasense* LA1982 and *S. pennellii* LA1926). In addition, while in red fruits the increase is linear during the first 30 dpa and higher during the ripening, the second increase in green fruits starts at 40 or 50dpa.

The correlation analysis of the cell surface compared to the increase of the pericarp thickness showed interesting results. Compared to *WVa106* where this correlation has never been showed, the wild tomato species showed an approximately linearity between both of parameters in all the red and green-fruited species (excluded *S. neorikii*) (Fig.36). This could be strongly linked to the variation in number of cell layers. As showed, all the wild tomato species (excluded *S. neorikii*, see Fig.34) showed an approximately linear increase of the cell layer number during fruit development. Finally, if the ration between cell layer number and cell width (a factor that have an influence on pericarp thickness) is maintained constant, the result could be a linear increase of cell surface and pericarp thickness.

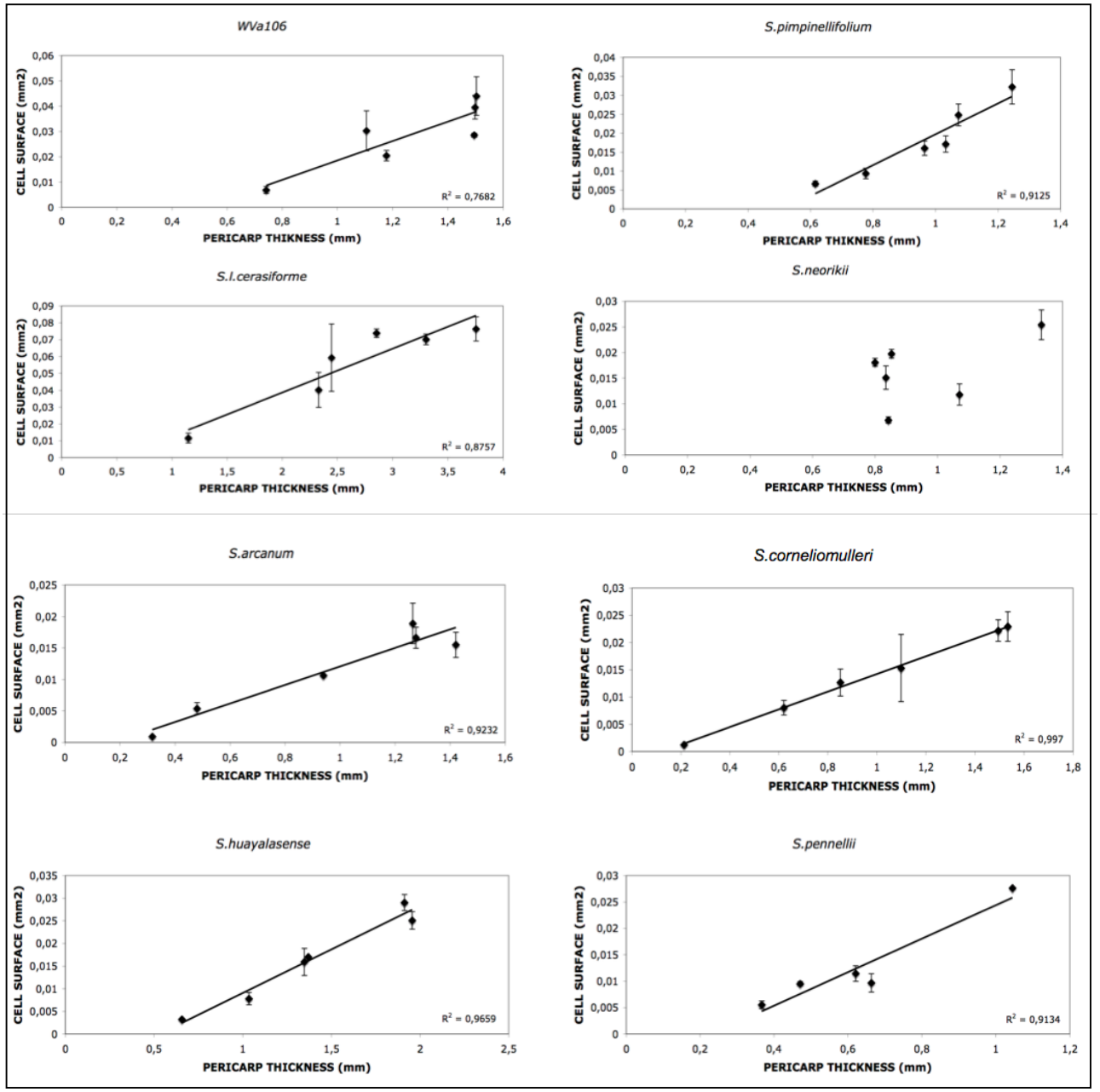

Fig.:36. Relationship between pericarp thickness and cell surface in wild and cultivated tomato species. Cultivated tomato (*WVa106*) and the wild species *S.neorikii* (LA1326) did not showed a linear correlation between the pericarp thickness and cell surface that were foud for all the other wild species.

2.5 – GENE EXPRESSION ANALYSIS OF TOMATO FRUIT DEVELOPMENT AND RIPENING.

2.5.1 – Introduction

The development of the fruits is classically described in four distinct phases: fruit set, a phase of intense cell division, a phase of cell expansion and ripening (Gillaspy et al.,1993).

The earliest phase involves the development of the ovary following pollination and fertilization. Then, the presence of fertilized ovules triggers the development of the ovary into a fruit. After fertilization, cell division is activated in the ovary and proceeds at high rate for some days that differer between species. While the cell division phase ends, individual cells enlarge, as does the entire fruit, for the following weeks. All these phases are regulated by a set of genes involved in cell duplication and cell expansion. Finally, ripening is a unique aspect of development starting after seed maturation has completed (Gillapsy et al., 1993). This phase involves deep metabolic changes in the biochemistry, physiology and gene expression such as (softening of fruits tissues, an increased accumulation of soluble sugar, acids and volatile compounds that increase the palatability to animal. In this way, plants in the wild facilitate their own seed dispersion (Palma et al., 2011, Klee and Giovannoni, 2011).

Fleshy fruits are physiologically classified in climacteric (tomato, apple, banana and avocado) and nonclimacteric (citrus, strawberry and grape). Climacteric fruits are characterized by an increase in respiration and by a simultaneous increase in synthesis of the phytohormone ethylene upon initiation of ripening whereas non-climacteric fruits do not exhibit an increasing in respiration during ripening (Giovannoni, 2004). In the tomato fruit, upstream to the ethylene pathway, three genes encoding putative transcription factors seem to play a major role in developmental control of fruit ripening: RIPENING INHIBITOR (RIN) and COLORLESS NON-RIPENING (CNR) and NON RIPENING (NOR) start to be encoded at the mature green stage (Giovannoni, 2007). Downstream to the ethylene biosynthesis, the regulation of ripening is modulated by a series of ethylene receptors able to activate different pathways to complete the ripening process. However, to date, their signal transduction pathway(s) has not yet been fully elucidated (Klee and

Giovannoni, 2011). One of the genes under positive ethylene control during ripening is the primary enzymatic control of carbon flux entering the carotenoid pathway: the PHYTOENE SYNTHASE (PSY), which catalyse the head to head condensation of geranyl geranyl pyrophosphate to form phytoene, the direct precursors of all carotenoids. Accumulation of carotenoid occurs concomitantly to the chloroplasts to chromoplasts conversion typical of the ripening process in the red tomato fruit. Accumulation of carotenoids, mainly lycopene, provides a visual indication that the fruit is fully ripen and suitable for consumption (Klee and Giovannoni, 2011; Giovannoni, 2004; Giorio et al., 2008).

When considering wild tomato species, only a limited number develop fruits that change colour during ripening. For this reason, we analysed the expression of genes characteristic of the different fruit developmental phases as defined in the cultivated species.

In the following part we will present results concerning the expression analysis of the Pepc2 gene (Phosphoenolpyruvate carboxylase, AJ313434) characteristic of the cell elongation phase and of two genes expressed during the ripening process, RIN and PSY1 (phytoene synthase, PI114490)

2.5.2 – Gene expression analysis

The analysis of gene expression were performed by real-time quantity PCR, the expression levels were normalized using *ACTIN* as reference gene and following the procedure described from Perikless Simon in the 2003 (Perikless, 2003). The results are shown as normalized copy number for one microgram of mRNA extracted.

The fruit specific phosphoenolpyruvate carboxylase (*PEPC2)* is involved in the initial fixation of the atmospheric CO_2, the maintenance of cytoplasmatic pH and ionic balance (Latzko and Kelly 1983), in the synthesis of the main organic acids, malic and citric acids, which accumulate during fruit development (Guillet et al.,2002). In tomato fruit organic acids accumulation began at the end of the cell division phase and PEPCase activity decline before the ripening. It has been shown that in the fruits of *WVa106*, *PEPC2* is strongly expressed during the cell expansion phase and could contribute to the synthesis of organic acids as counter-ions for potassium that accumulates in the vacuole

therefore permitting the cell enlargement, which occurs during the rapid growth phase of the fruit (Guillet et al., 2002).

In red-fruited species used in this study, the *PEPC2* gene is highly expressed at 10dpa and decrease until the breaker stage. Similar expression profiles were also observed for *S. pimpinellifolium* and *S.l. cerasiforme* (Fig.37A). In addition, *S.l. cerasiforme* showed a high increase of pericarp thickness between breaker stage and red-ripe but at this stage was not detected any pick of *PEPC2* expression. This suggest that the increasing in pericarp thickness could be linked to cell expansion not mediated by *PEPC2* expression. Green-fruited species showed two different situations: *S. neorikii, S. arcanum* and *S. corneliomulleri* had the pick of expression at 30 (*S. neorikii* and *S. arcanum*) or at 40 dpa (*S. corneliomulleri*) (Fig.37B). *S.huayalasense* and *S.pennellii* showed a double expression profile with picks at 20 and 40dpa (Fig.37C). In these later cases the two picks of PEPC2 gene expression are correlated with the of cell size increase (Fig.35).

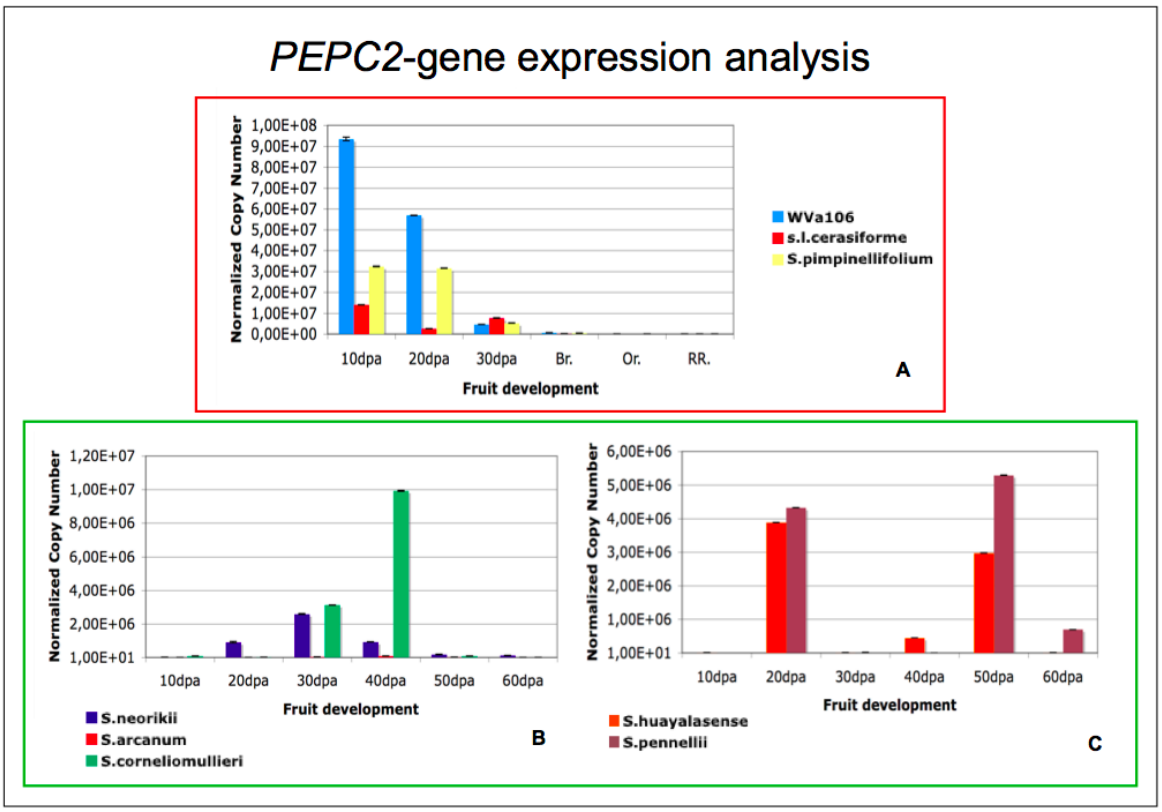

Fig.:37.Gene expression analysis of *PEPC2* during fruit development and ripening. The PEPC2 gene expression shows differences between species. While the red-fruited wild species (*S.l.cerasiforme* LA1226 and *S.pimpinellifolium* LA0722) show a high expression during the early stages of development like *WVa106*, the green-fruited species (*S.neorikii* LA1326, *S.arcanum* LA2152 and *S.corneliomullieri* LA0103), have a high expression between 30 to 40dpa. In the green cherries from *S.huayalasense* LA1982 and *S.pennellii* LA1926, show two pick of expression at 20 and 50dpa. The normalized copy number shows the number of copy of transcript in one microgram of RNA extracted.

In order to define more precisely the ripening process in the green fruited species, we analysed the expression levels of two the genes involved upstream and downstream of the ethylene pathway: Sl*RIN* and Sl*PSYI* respectively. The RIN protein play an essential role that determine the transition from system I to system II for ethylene production in tomato fruits (Barry et al., 2000).

The results of quantity PCR showed that in red-fruited species the *RIN* is activated at the breaker stage. In *WVa106* and *S.l. cerasiforme* the picks of expression were detected at orange stage and later at the red ripe stage for *S .pimpinellifolium* . In green-fruited species including *S. neorikii* and *S. arcanum* the activation of *RIN* occured at 40 dpa followed with a maximum expression level at 50dpa. This maximum expression level was reached at 60 dpa in fruits of *S. corneliomulleri* and *S. huayalasense*. In fruits of *S. pennellii*, the highest RIN expression level occurred at 50dpa but remained lower compared to all other species.

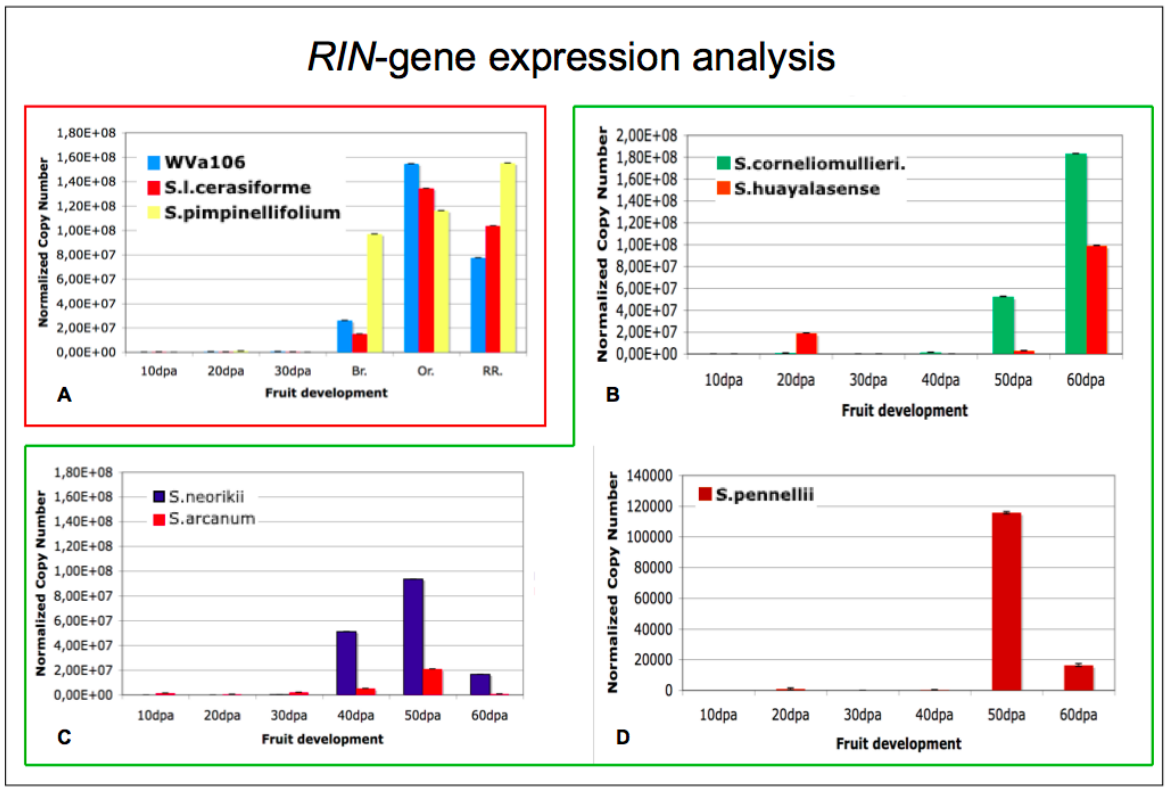

Fig.:38.Gene expression analysis of *RIN* during fruit development and ripening. While into the red fruits (A) *WVa106, S.l.cerasiforme* LA1226, *S.pimpinellifolium* LA0722, RIN is activated at Breaker (Br.) stage and increase at Orange (Or.), into the green fruits as *S.neorikii* LA1326, *S.arcanum* LA2152 and *S.pennelli* LA1926 the maximum expression levels has been detected at 50dpa (C and D) or at 60dpa for *S.corneliomullieri* LA0103 and *S.huayalasense* LA1982 (B). The normalized copy number shows the number of copy of transcript in one microgram of RNA extracted.

The analysis of *PSYI* gene expression (Fig.39) showed that the gene in red fruit is activated at breaker stage concomitantly to the induction of the carotenoid accumulation. Its maximum expression level was measured at the orange stage (*WVa106* and *S. pimpinellifolium*) and at the red ripe stage in *S.l. cerasiforme* fruits. Although in green-fruited species there is not conversion of chloroplast to chromoplast, a low expression level of *PSYI* were detected in all the green-fruited specie. *S. neorikii* and *S. corneliomulleri* showed a pick at 50dpa and *S.huayalasense* at 60dpa, while *S.pennellii* and *S. arcanum* showed only a basal expression of the gene suggesting that PSYI is not particularly involved in the ripening process.

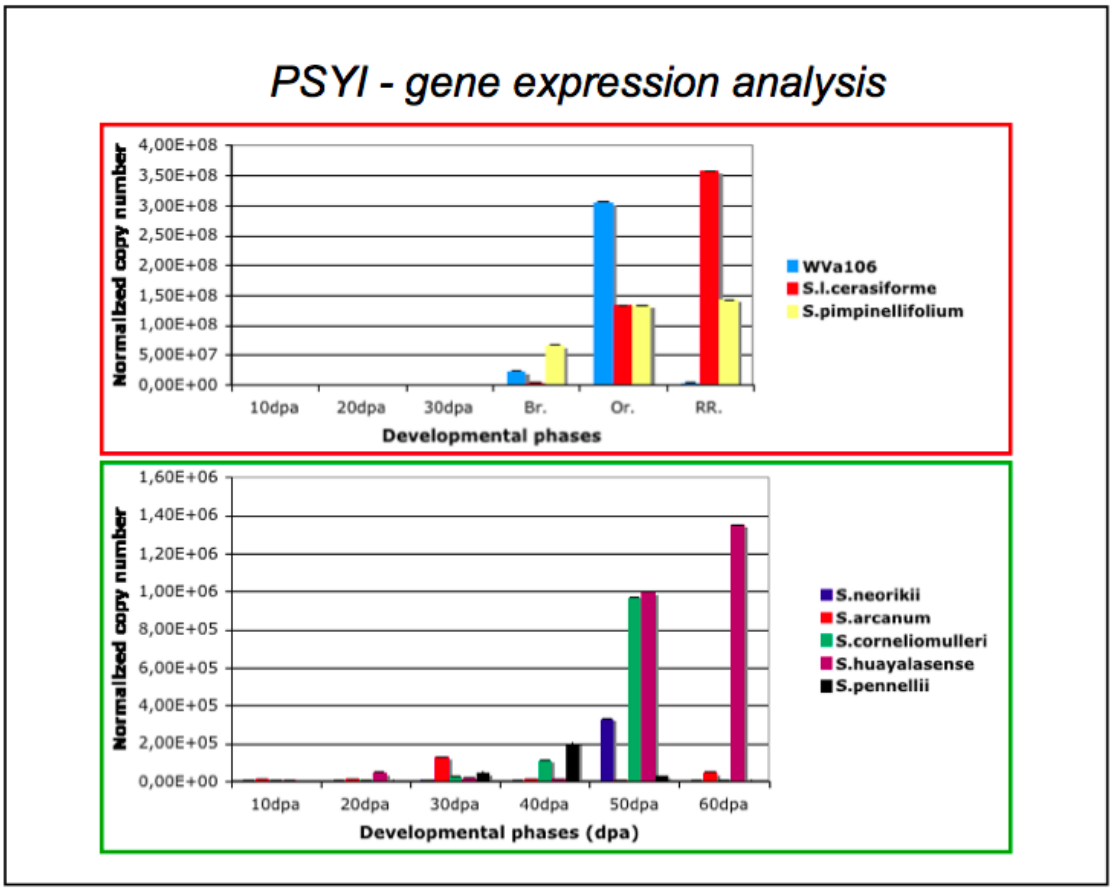

Fig.:39. Gene expression analysis of *PSYI* during fruit development and ripening. The *PSYI* gene expression shows differences between species. While the red-fruited wild species (*S.l.cerasiforme* LA1226 and *S.pimpinellifolium* LA0722) show a high expression during the ripening phase, the green-fruited species have a low expression. In the green cherries of *S.arcanum* LA2152 and *S.pennellii* LA1926 there were not expression during all the developmental phases. The normalized copy number shows the number of copy of transcript in one microgram of RNA extracted.

2.6 – DISCUSSION AND CONCLUSION

The analysis of fruits from red and green-fruited species revealed morphological and physiological variations concerning size and colour and morphology. For most of the species, despite differences, there was an approximately linear increase in weight during the first 30 days after anthesis, as well as in size till it reached a plateau at 30dpa. Only *S. huayalasense* and *S. pennellii* underwent a second increase in weight between 50 and 60dpa as well as in size.

Analysis of cross sections of tomato cherries also revealed differences in the dynamics of ripening. In *S.l. cerasiforme*, for example, the locular tissues changed colour from green to orange before the pericarp suggesting a different dynamics of development as compared to *WVa106* or *S. pimpinellifolium*. Similarly the pericarp thickness increased more during the ripening phase while the other species showed a linear increase during the earlier stages.

Differences between species were found also at the level of the pericarp structure which is important for ripening and softening. The study of the pericarp structure showed different dynamics of pericarp growth. Differently from *WVa106*, the wild relatives showed a pericarp still containing inner and outer subepidermal cell layers also at the last stages of development (*S. arcanum* and *S. huayalasense*) in a species-specific manner. The data suggest that in these species the number of cell layer is not genetically determined during the first days after anthesis. This observation was also confirmed by the analysis during fruit development of the number of cell layers. While our analysis of the cultivated species *WVa106* confirmed previous results from Cheniclet and colleagues (Cheniclet et al., 2005), which showed that the number of cell layers is stable during fruit development, wild species behaved differently. *S.l. cerasiforme*, for example, has a higher number of cell layers as compared to the other wild relatives as well as *S .neorikii* while surprisingly all the other wild species showed an increase in the number of cell layers according to the stage as well as the continuous presence of thin outer and inner subepidermal cell layers.

As already shown in cultivated species, the dynamics of the pericarp thickness and structure is linked to the size of the cells and to the ploidy levels. The measurement of the

cell size confirmed what observed before. *S.l. cerasiforme* had the largest cell size, moreover all the red-fruited species showed a similar behaviour with a first cell size increase between 10 and 30dpa and a second one that occurred during the ripening stage. This was also observed in green-fruited species though the second size-increasing phase was species specific and started at 40 or 50dpa up to 60dpa.

In addition, in contrast to *WVa106* and *S.neorikii* that did not showed an increase of cell layer number, the increase of this parameter in all the other wild relatives during fruit development showed to be involved in the increase of the pericarp thickness in a linear way with the cell surface increase.

This observation suggest that compared to *WVa106* where the increase of the pericarp thickness involve mainly the cell size by cell expansion, in wild tomato species the increase of the pericarp thickness involve cell expansion as well as cell division during all fruit development. This observation it has been clearly supported by the analysis of the relationships between cell surface and pericarp thickness and between the number of cell layer and pericarp thickness.

The development of the fruits is classically divided in four phases: fruit set, intense cell division, cell expansion and finally ripening, which involve deep metabolic changes in the biochemistry and physiology regulated by a set of genes involved in the different mechanisms.

Although in red-fruited species is easily recognize the beginning of ripening, in green-fruited species follow fruit development and ripening is more complicated. On a set of genes, we choose to analyze the expression levels of gene involved in cell expansion (*PEPC2*) to analyze the fruit developmental phases, while the analysis of the ripening process were done using a gene upstream (*RIN*) and downstream (*PSYI*) of the ethylene cycle. On the bases of these observations we can conclude that PEPC2 had an expression profile characteristic in the different species. Hence, *S.huayalasense* and *S.pennellii* showed a double pick at 20 and 50dpa that are strongly correlated with the two pahses of increasing in fruit size and weight (Fig.29C and F). This correlation could be related to the accumulation of products into the vacuole with an increase cell enlargement, in agreement with the proposed role of PEPCase suggested postulated from Guillet in the 2002 (Guillet et al., 2002)

On the other hand, the analysis of the genes involved in the ripening process suggested that is possible make a correlation between the different developmental stages of red and green fruits. On the base of RIPENING INHIBITOR gene expression analysis, we can

propose in an approximate way that the breaker, orange and red ripe stage in red fruit could be compared respectively to 40, 50 and 60dpa in green-fruited species. This observation can be only partially confirmed by the expression analysis of *PSYI* gene. Although *S.neorikii*, *S.corneliomulleri* and *S.huayalasense* showed a lower expression level of *PSYI* compared to the red-fruited species, they maintain a similar profile with an activation of the gene to 40dpa with a maximum expression level at 50dpa (S.neorikii and *S. corneliomulleri*) and/or 60dpa (*S. huayalasense*). On the other hand, *S. arcanum* and *S. pennellii* showed a basal expression level of PSYI suggesting that its expression varying in a specie specific manner.

CHAPTER 3

ENDOREDUPLICATION AND DNA MTHYLATION IN CULTIVATED AND WILD TOMATO SPECIES.

3.1 – INTRODUCTION.

The switch from cell division to cell expansion plays a major role during fruit development in cultivated tomato species (Cheniclet et al., 2005; Bertin 2005). In these species, pericarp cell expansion is involved in the increase of pericarp thickness and is accompanied by the endoreduplication process. In order to determine the contribution of this process in the wild species on pericarp thickness, cell size was analysed at various developmental stages using the fruits samples described above.

3.2 – PLOIDY ANALYSIS OF THE PERICARP TOMATO FRUITS DURING FRUIT DEVELOPMENT AND RIPENING

3.2.1.- Introduction

Tomato fruit development results from the interplay between cell division and cell expansion which determine the cell number and cell size both contributing to fruit size determination (Bohner and Bangerth, 1988). It has been well described that the transition from cell division to cell growth is accompanied by a dramatic increase in cell ploidy level, a phenomenon also called endopolyploidization. A recent work published in 2010 has demonstrated that endopolyploidization in tomato fruit tissues does not lead to a doubling of the chromosome number in the nucleus as expected for endomitosis but to the production of chromosomes with 2n chromatids without changes in chromosome number (Bourbon et al., 2010).

A positive correlation was also shown between cell size and ploidy level in developing tomato fruit (Cheniclet et al.,2005). Yet, the relationship between fruit size and ploidy

level is not clear and the endoreduplication process has been suggested to be involved in the control of the rate of organ growth rather than to its size. In a recent analysis Bourbon et al., (2010) was reported that endoreduplication occurs in fleshly fruits, (strewberry, melon, cucumber, pepper) that develop rapidly (in less than13 weeks) comprising three to eight round of endocycle, but is not detected in fruits with slower developmental processes.

To analyse to which extend endoreduplication is maintained in wild tomato species as compared to cultivated one, we analyzed the ploidy level of pericarp cells at all the stages of fruit development described in chapter 1.

3.2.2 – Pericarp ploidy level during development of wild and cultivated tomato fruits

The pericarp from three to five tomato fruits was analyzed at all developmental stages for each species by flow cytometry after Dapi staining (see methods). The ploidy histograms showed a clear difference in ploidy levels between red and green-fruited species (Fig.40A).. At maturity, fruits from red-fruited species contain nuclei with a maximum C value ranging from 128C (*S. pimpinellifolium*) up to 256C (*WVa106* and *S.l. cerasiforme*). At contrary, fruits from green-fruited species, are characterized by lower C value ranging from 64C (*S. neorikii, S. arcanum* and *S. pennellii*) to 128C (*S. corneliomulleri* and *S. huayalasense*) (Fig.40A and B). However, in all species analysed fruit development and growth is characterized by a significant increase in the maximum C value in the pericarp cells.

In addition to the maximum C value, the distribution of the nuclei in each ploidy class and the average C value provide a way to evaluate the endoreduplication process in cells of pericarp tissues. Indeed, species that showed a low maximum C value accumulate a higher level of nuclei in the 2C class (*S. arcanum* 43.05%±4.32% and *S. huayalasene* 27.57%±1.52%) as compared to *WVa106* (6.58%±1.40%) at 10dpa. In addition, 10dpa fruits in *WVa106* and of *S.l. cerasiforme* species already had nuclei in the 32C class while at this developmental stage highest ploidy level observed in fruits of green-fruited species was 8C (*S.arcanum* and *S.corneliomulleri*) or 16C (*S.neorikii* and *S.huayalasense*). It is noteworthy that for *WVa106* and *S. pimpinellifolium* and *S.l.cerasiforme* the ploidy distribution changes during ripening with an increase number of nuclei with high ploidy

levels, whereas no more evolution is observed after 30dpa in *S. arcanum, S. neorikii* and *S. pennellii*. Among the green-fruited species *S.huayalasense* presents a different behaviour as the proportion of nuclei of the 4C class decrease and the proportion of nuclei of higher ploidy level increases after 50 dpa (Fig.40A and B).

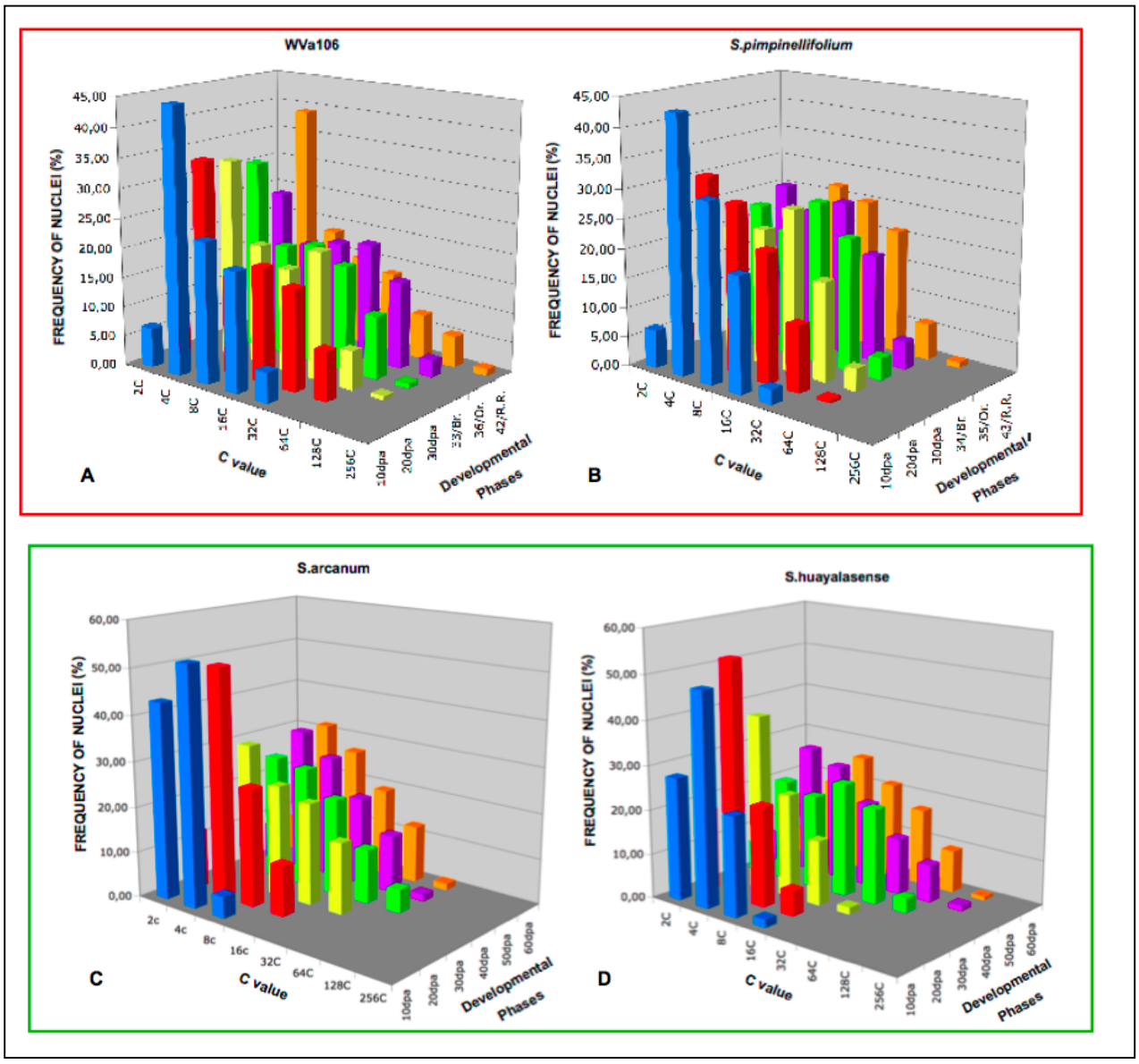

Fig.:40A. Evolution of ploidy levels during tomato fruit development and ripening. The red panel describes the red-fruited species (*WVa106* and *S. pimpinellifolium*) whereas the green panel shows the green-fruited species (*S. arcanum* and *S. huayalasense*). In the pericarp of the cultivated species *WVa106* (A) the maximum C value is up to 256C while in the wild species is generally lower. The wild red species *S. pimpinellifolium* (B) shows a C value lower than *WVa106* and similar to the green species *S. huayalasense* (D). *S. arcanum* shows a maximum C value of 64C. In addition differences can be seen at 10dpa when in the green-fruited species the 2C and 4C classes show a high percentage of nuclei compared to the red ones. Finally, the red-fruited species have an endoreduplication rate higher than the green ones. *WVa106* (A) and *S. pimpinellifolium* (B) show a C value up to 256C and 128C after 42-43dpa while the green-fruited species *S. arcanum* (C) and *S. huayalasense* (D) have a C value of 64C at 40dpa. These measurements were obtained also for the other species.

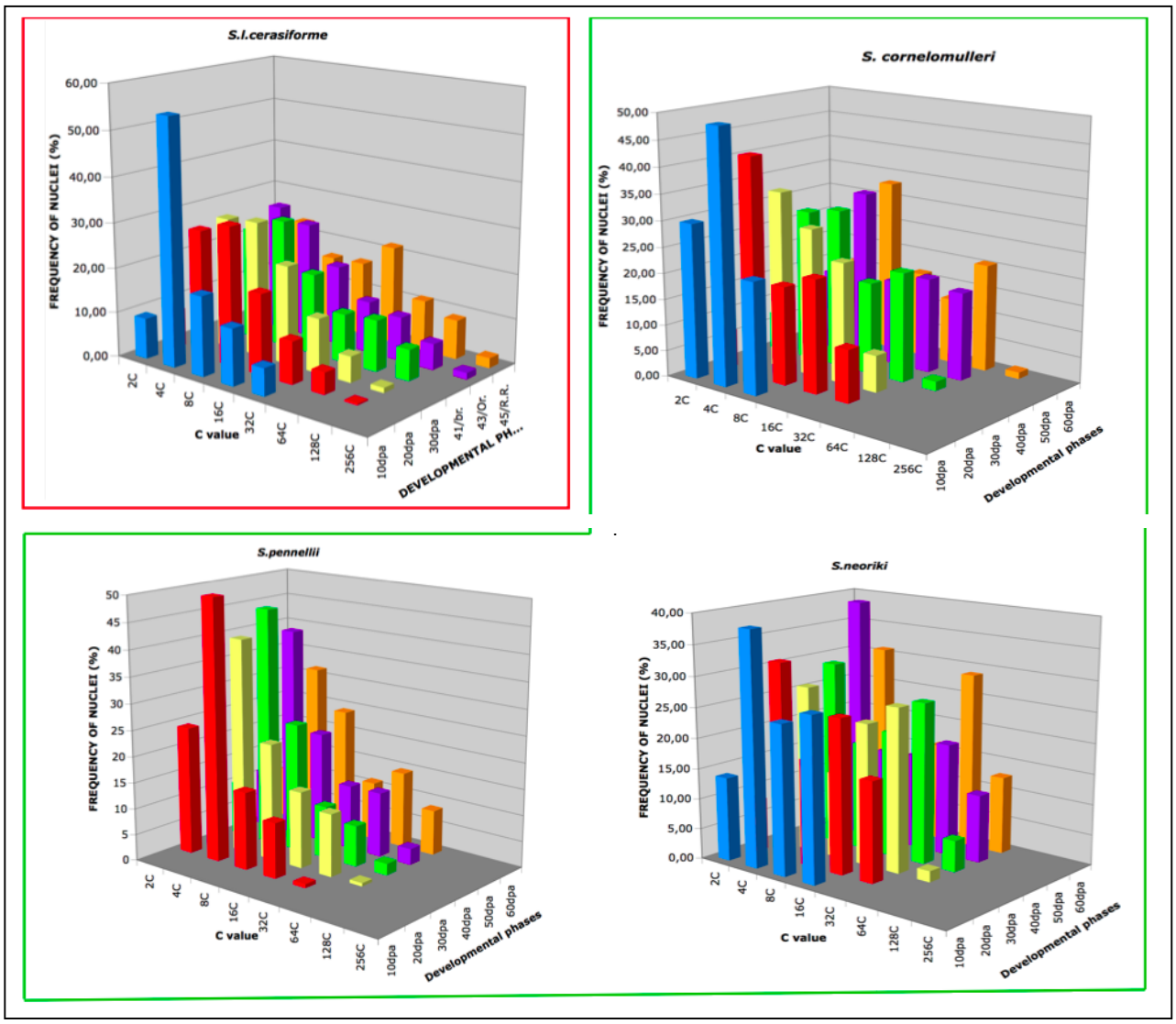

Fig.:40B. Evolution of ploidy levels during tomato fruit development and ripening. The red panel describes the red-fruited species *S.l.cerasiforme* whereas the green panel shows the green-fruited species (*S. neorikii*, *S. corneliomulleri and S.pennellii*). In the pericarp of the cultivated species *S.l.cerasiforme* the maximum C value is up to 256C while in the wild species is generally lower. The wild red species *S. neorikii* shows a C value lower and similar to *Spennellii* . *S. corneliomulleri* shows a maximum C value of 128C.

The increasing in C class of nuclei well support the increase in cell size observed red-fruited species: between breaker stage and orange and orange and red ripe stage. Hence the increase in the ploidy level observed in the red-fruited species during the ripening phase occurs concomitantly with the increase in cell size observed after breaker stage in red-fruited (Fig.35).

Similarly this observation can be done also for green-fruited species, where the long transition phase before the second increase in cell size could be due to the stop of the endoreduplication to 40 dpa (*S. arcanum, S. pennellii* and *S. neorikii*).

The analysis of the mean C value (MCV), which considers the frequency of nuclei present in each C class and represent the average cell ploidy level of pericarp cells at a given developmental_stage showed contrast situation between species. The MCV was calculated in according the formula:

$$Mean\ C\ Value = \Sigma_{ni=1} (C_I \times N_I)/N_{sample}$$

N: number ploidy class in the sample; C_i: C value of a given class of Nucleus n_i; N_i: number of nuclei in peak n_i; N_{sample}: total number of nuclei.

As expected from the C max analysis and the distribution of nuclei within ploidy classes during fruit development, red and green fruits show differences in fruit MCV. Red-fruited species (*WVa106, S.l. cerasiforme* and *S. pimpinellifolium*), are characterized by an exponential increase in the MCV (Fig.41A). The maximum MCV value was observed at the red ripe stage in all three species, with a value of 35.88±4.96 in *S.l. cerasiforme,* 24.27±7.46 *WVa106* and 25.72±2.11 *S. pimpinellifolium,*. Green-fruited species (*S. neorikii, S. arcanum, S. corneliomulleri* and *S. pennellii*) presented a progressive increase evolution of the MCV during fruit development between 10 and 30dpa and also between 50 and 60dpa, while a stationary pahse is present between 30 and 40dpa (Fig.41B). The behaviour of *S. huayalasense* was characterized by (Fig.41C) a low increase in the C value during the first 30dpa and then a second increase from 40 to 60dpa but the two increasing phases were separated by a rapid increase of the MCV between 30 and 40dpa showing an opposite situation compared to the other green species.

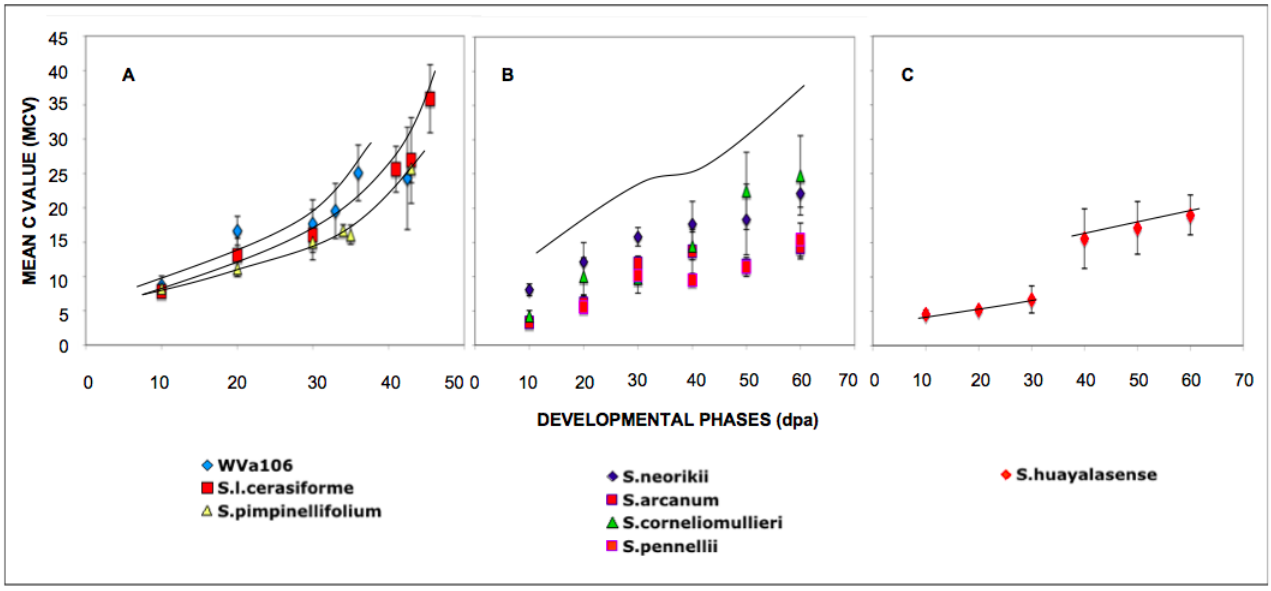

Fig.:41.Change of the mean C value in pericarp during fruit development. In (A) the red-fruited species (*WVa106*, *S.l. cerasiforme*, *S. pimpinellifolium)* show a linear increasing of the C value that becomes exponential during ripening. In (B) green-fruited species (*S. neorikii*, *S. arcanum*, *S. corneliomullieri*, *S. pennellii)* show a general linear increasing of C value with a low flexion between 30-50dpa. In (C), *S. huayalasense* shows a characteristic behaviour, with two separate linear trends: 10-30dpa and 40-60dpa. The black line (—) shows the trend of C value increase.

The correlation analysis between the MCV and the cell size showed different behavior between species. Based on this correlation the species could be dividen in four groups. A first group includes *WVa106, S.pimpinellifolium* (red fruit), *S.neorikii* and *S.arcanum* (green fruit), which showed a linear increase of MCV and cell size suggesting that the cell size is strongly linlked to the ploidy. A second group with *S.l.cerasiforme* and *S.corneliomulleri* respectively red and fruited species had an increase of the cell size ranging a plateau showing an increase of the mean C value without increase in cell size. Opposite situation was showed from *S.pennellii* (third group) while *S.huayalasense* showed two linear increase of cell size and MCV interrupted from a stationary phase (Fig.42).

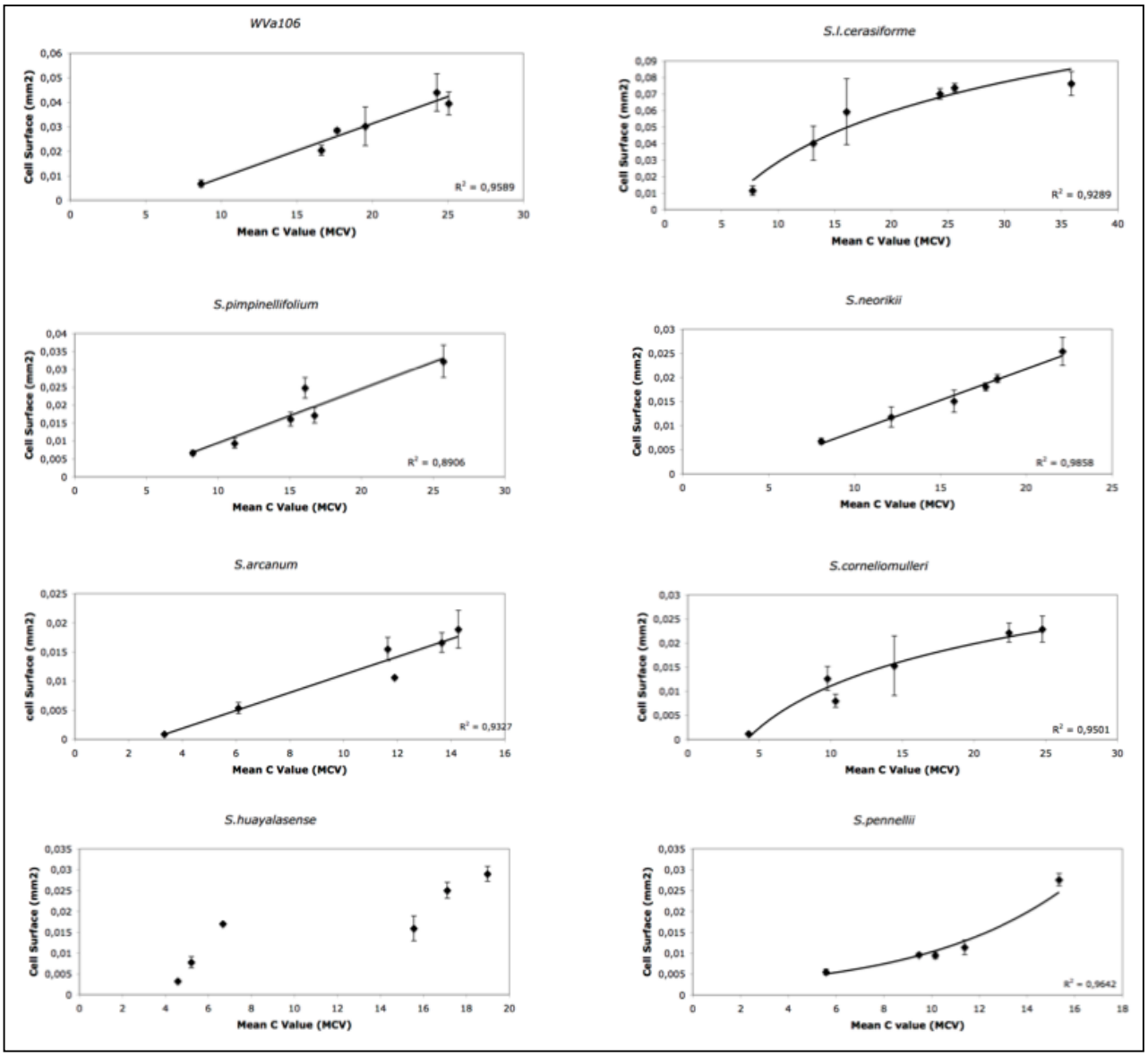

Fig.:42. Correlation between Mean C Value (MCV) and cell size in fruits. A linear correlation is present for red (WVa106, S.pimpinellifolium) and green (S.neorikii and S.arcanum) fruits. S.l.cerasiforme and S.corneliomulleri show an increase of cell size ranging a plateau while S.pennellii shows an exponential increase in cell size. In S.huayalasense shows two separate linear increase compared to its MCV.

On the other hand, the mean C value that indicates the mean DNA content (1C) per nucleus present disadvantages. A disadvantage of this parameter is the overemphasis of the high ploidy levels because of the exponential character of the different ploidy classes. For this reasons, it has been defined a "cycle value" indicating the number of endoreduplication cycle per nucleus. The cycle value and/or endoreduplication index is calculated from the number of nuclei of each represented ploidy class multiplied by the number of endoreduplication cycles necessary to reach the corresponding ploidy level. The sum of the resulting products is divided by the total number of nuclei measured (Barow and Meister, 2003):

$$EI = \frac{0xF_{2C} + 1xF_{4C} + 2xF_{8C} + 3xF_{16C} + 4xF_{32C} + 5xF_{64C} + 6xF_{128C}}{\sum_{n=2C}^{n=128C} Fx};$$

F, is the frequency of nuclei present in each C class expressed as percentage.

The results showed a linear increasing of the E.I. during fruit development of red-fruited species. As expected, *S.l. cerasiforme* had the highest E.I. value (3.17±0.14) as compared to *WVa106* (2.51±0.31) and *S. pimpinellifolium* (2.68±0.08) at the red ripe stage (Fig43A).

The EI of Green-fruited species showed a different situation. In *S. neorikii, S. arcanum* and *S. pennellii* there was a linear increase in E.I. between 10 and 40dpa and also from 50 to 60 dpa (Fig.43B), while in *S. huayalasense* and *S. corneliomulleri* EI increase in a linear manner until 20dpa (S.corneliomulleri) or 30dpa (S.huayalasense) and after a stationary phase of 10 days thy showed a second increase up to 50dpa (Fig.43C). In the green-fruited species at 60dpa the E.I. varied between 2.28±0.06 (*S. arcanum)* and 2.90±0.17 (*S. corneliomulleri)*.

Finally, a Tukey's test were performed showed that *S.l. cerasiforme* has a statistically high E.I. value compared to the other species (Tab.3) suggesting that a high amount of nuclei in this species was able to undergo up to 8 endocycles. Statistical differences were also observed between *S. arcanum* and *S. neorikii*.

Comparison	Difference	Standardized difference	Critic value	Pr > Diff	Significant
S.l.cer. vs S.arcanum	0,889	6,639	3,462	0,000	****
S.l.cer. vs S.pennellii	0,784	5,860	3,462	0,001	***
S.l.cer. vs WVa106	0,662	4,945	3,462	0,003	***
S.l.cer. vs S.huay.	0,567	4,235	3,462	0,011	***
S.l.cer. vs S.pimpin.	0,487	3,638	3,462	0,036	***
S.l.cer. vs S.neoriki	0,356	2,661	3,462	0,204	NO
S.l.cer. vs S.cornel.	0,264	1,975	3,462	0,525	NO
S.cornel. vs S.arcanum	0,624	4,664	3,462	0,005	***
S.cornel. vs S.pennellii	0,520	3,885	3,462	0,022	***
S.cornel. vs WVa106	0,398	2,970	3,462	0,122	NO
S.cornel. vs S.huay.	0,302	2,260	3,462	0,370	NO
S.cornel. vs S.pimpin.	0,223	1,663	3,462	0,709	NO
S.cornel. vs S.neoriki	0,092	0,686	3,462	0,996	NO
S.neoriki vs S.arcanum	0,532	3,978	3,462	0,019	***
S.neoriki vs S.pennellii	0,428	3,199	3,462	0,081	NO
S.neoriki vs WVa106	0,306	2,285	3,462	0,358	NO
S.neoriki vs S.huay.	0,211	1,574	3,462	0,758	NO
S.neoriki vs S.pimpin.	0,131	0,977	3,462	0,971	NO
S.pimpin. vs S.arcanum	0,402	3,001	3,462	0,116	NO
S.pimpin. vs S.pennellii	0,297	2,222	3,462	0,389	NO
S.pimpin. vs WVa106	0,175	1,307	3,462	0,883	NO
S.pimpin. vs S.huay.	0,080	0,596	3,462	0,998	NO
S.huay. vs S.arcanum	0,322	2,404	3,462	0,302	NO
S.huay. vs S.pennellii	0,218	1,626	3,462	0,730	NO
S.huay. vs WVa106	0,095	0,711	3,462	0,995	NO
WVa106 vs S.arcanum	0,227	1,693	3,462	0,691	NO
WVa106 vs S.pennellii	0,122	0,915	3,462	0,980	NO
S.pennellii vs S.arcanum	0,104	0,779	3,462	0,992	NO

Tab.3. Comparison of Endoreduplication Index in cultivated tomato and wild relatives: *WVa106, S.l. cerasiforme* (S.l. cer.), *S. pimpinellifolium* (S. pimpin.) *S. neorikii*, *S. arcanum*, *S. huayalasense* (S. huay.) and *S. pennellii*. Tukey's critic value= 4,896 at 95%; analysis Type III Sum of Squares Pr>F = 0,0001. (* Pr>F = 1%; ** Pr>F = 5%).**

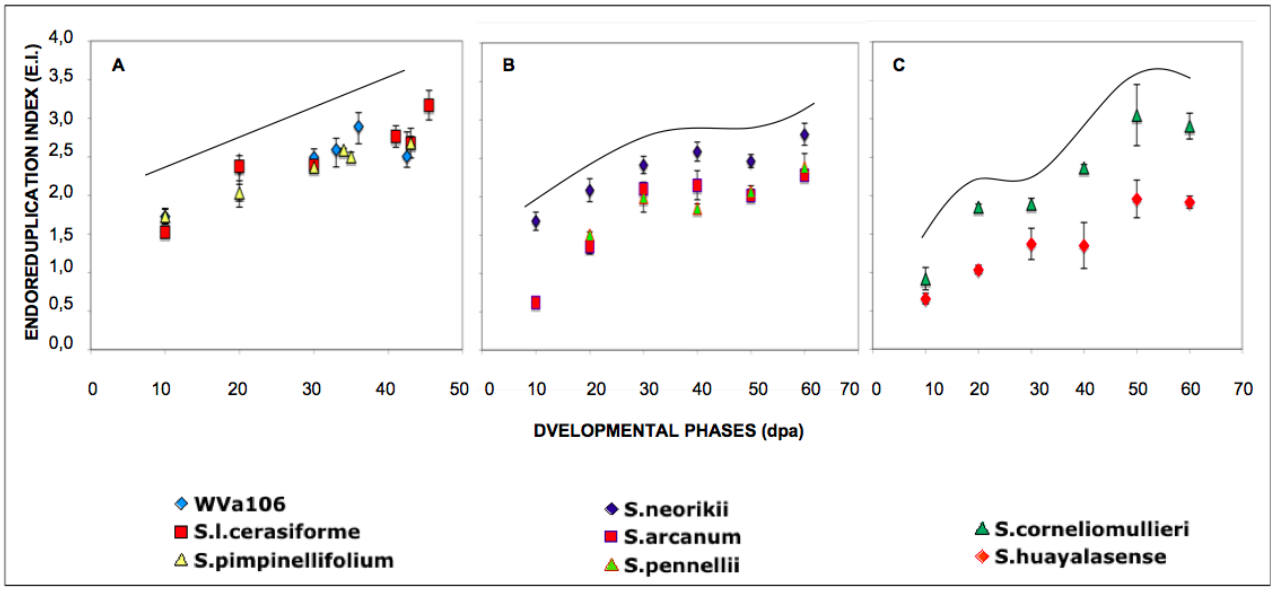

Fig.:43. Change of the Endoreduplication Index (E.I.) in pericarp, during fruit development. The species are grouped in red- (A) and green-fruited species (B and C). The red fruits show a linear increase of the E.I. during fruit development (A). On the other hand the green fruits show two types of trends. In B (*S. neorikii* LA, *S. arcanum* LA and *S. pennellii* LA) there is a first linear increase up to 30-40dpa and a second increase from 50 up to 60dpa. In C (*S. corneliomullieri* LA, and *S. huayalasense* LA) there is a linear increase of the E.I. from 10 up to 20-30dpa and from 30-40dpa up to 50dpa; then the value of E.I. decreases at 60dpa. The black line (—) shows the trend of C value increase.

Correlation analysis between fruit size and E.I and cell size and E.I. was also performed. Despite the differences in size and endoreduplication index between species, a general linear correlation between both parameters was found for most the species (Fig.44) with the exception of *S. huayalasense*. In this case, a strong increase in fruit size is observed

between 10 and 30dpa that was not associated with an increase in the endoreduplication index. Endoreduplication index was also correlated with cell surface (Fig.45). Although an approximately linear correlation between both parameters was found in most cases, species such as *S. pimpinellifolioum, S.pennellii* and *S. huayalasense* present a different situation.

In S. pimpinellifolium as *S.pennellii* fruits, the cell surface increase in a exponential way without any increase in EI. *S. huayalasense* fruits are typical in the sense that both parameters increase concomitantly in a linear in two separate phases: from 10 to 30dpa cell size increases more rapidly than The EI, and inversely between 40 and 60 dpa (Fig.45). A similar situation was observed for *S.huayalasense* when we compared its own endoreduplication levels with the increase in fruit size (Fig.44)

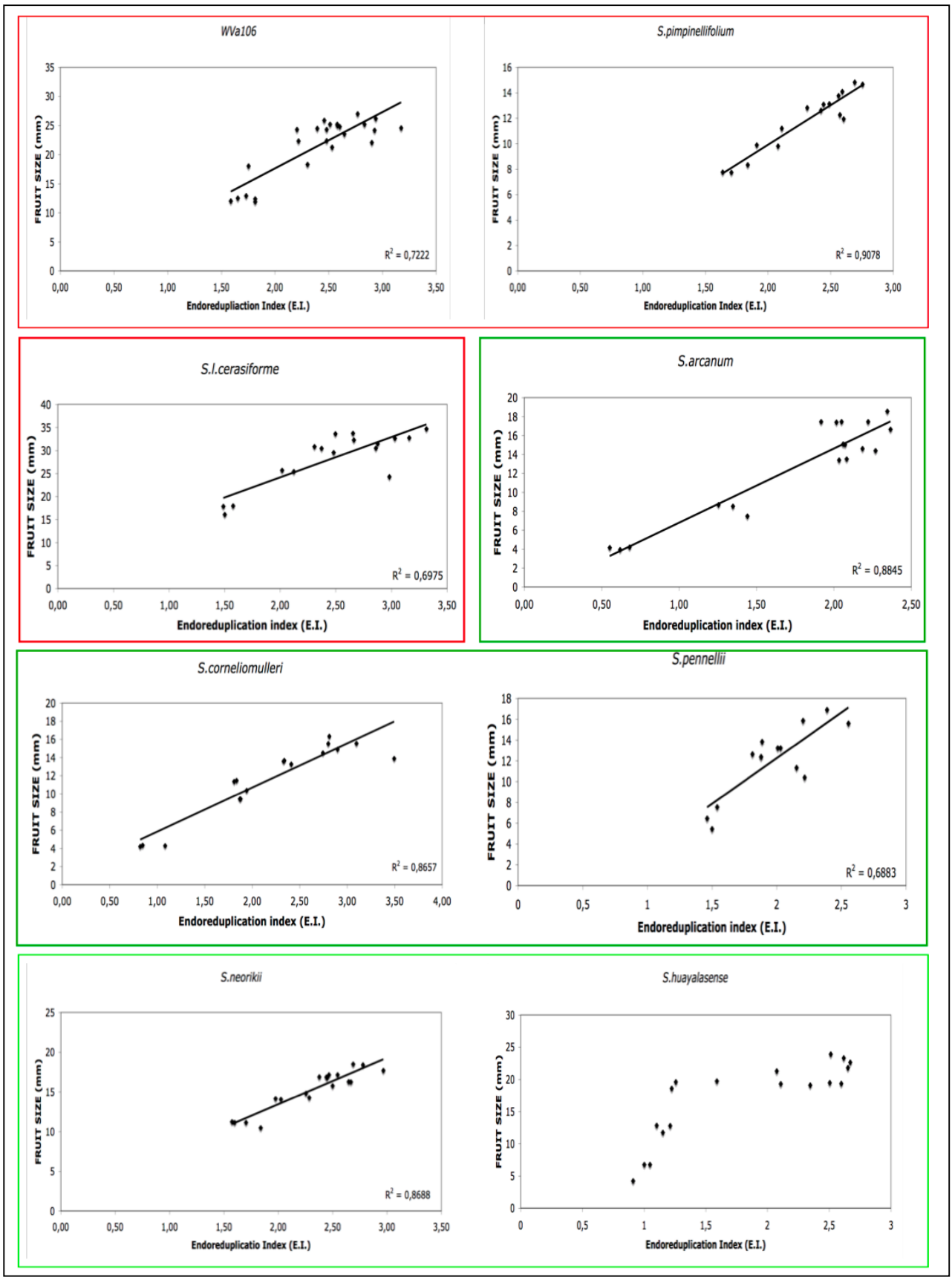

Fig.:44. Relationship between fruit size and endoreduplication index (E.I.). A general linear correlation between fruit size and endoreduplication index (E.I) can be seen for *WVa106*. Similar situation can be found for all the other species but not for *S. huayalasense* .

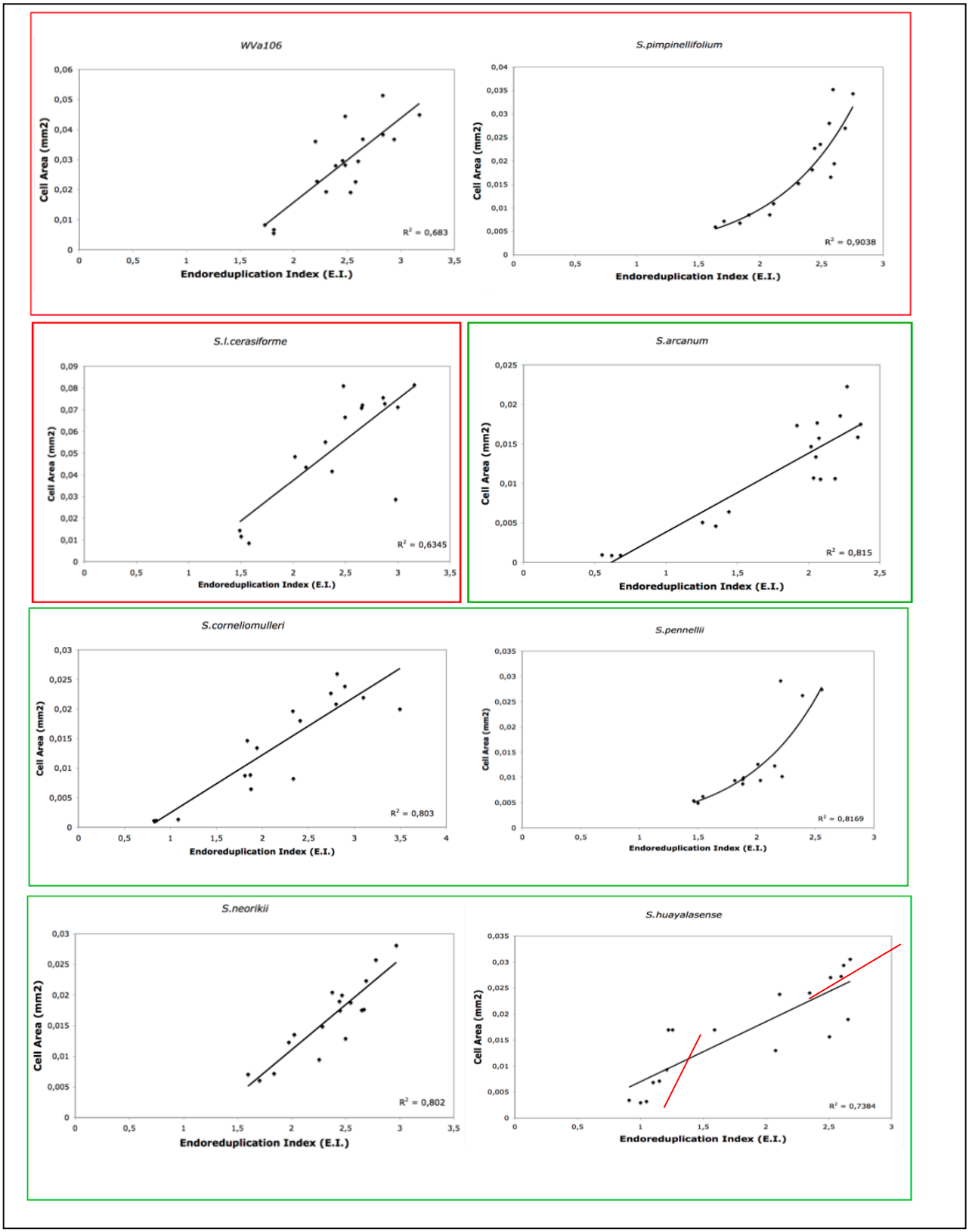

Fig.:45. Relationship between pericarp cell area and endoreduplication index (E.I.). In The red square the red-fruited species while in the green square the green ones. Although wild species such as *S. neorikii, S.corneliomulleri, S.arcanum, and S.l.cerasiforme* show a linear correlation between endoreduplication and cell area similarly to *WVa106*, other species are different. In *S. pimpinellifolium like S.pennellii* cell area increases more rapidly than the endoreduplication index while in *S. huayalasense*, even if there is linear correlation (r^2=0,7384), it seems that there are two different linear dynamics (red bars).

3.3 – VARIATION OF DNA METHYLATION IN CULTIVATED TOMATO AND WILD RELATIVES.

3.3.1 – Introduction.

Fruit development and ripening is an exclusive process during which tissues undergo physiological and metabolic changes that promote seed dispersion. Fruits are also important components of human and animal diets (Klee et al., 2011).
The development and ripening of this organ are under the control of environmental stimuli (external factor) and internal factors such as physiological and molecular processes including epigenetic mechanisms (Adams-Phillips et al.,2004).
Traditionally phenotype changes are explained through genetic variation, which may occur naturally or after specific mutagenic treatments in the nucleotide sequences. However, phenotypic variations can also be linked to inheritable epigenetic variations, which are potentially sensitive to environmental inputs. These epigenetic variations may contribute to the molecular mechanisms at the basis of complex traits such as floral symmetry, floral organ identity or fruit ripening (Manning et al., 2006) and vernalization responses (Bastow et al.,2004), therefore, in a more general way, these mechanisms help the plant to relate with the environment. In addition, the interaction between genetic and epigenetic variations makes it difficult to identify the importance of inherited epigenetic variation in phenotype diversity. Essentially, epigenetic should be viewed as a mediator between genotype and phenotype.

In this chapter we will investigate the epigenetic diversity in tomato species bearing in mind the limitations due to the genetic diversity of tomato species. For this reason different approaches will be used. Firstly, we will analyze the methylation status at the 5s rDNA locus, taken as an example of repeated DNA sequences highly conserved between cultivated tomato and wild relatives. We will then analyse the global methylation level in tomato fruits in order to investigate to which extend variations in methylation may account for changes in endoreduplication levels between species.

Although it is well described that tomato fruit cells undergo extensive endoreduplication (Cheniclet et al.,2005, Bertin et al.,2005) and are subject to changes in their methylation level during fruit development (Teyssier et al, 2008), evidence of a direct relationship between these processes is still missing.

3.3.2 – DNA methylation content in tomato.

A first pioneering work, published in 1991, has studied the DNA methylation levels in tomato (*S.L. esculentum*), (Messeguer et al.1991). Their results showed that the G+C content of the tomato genomic DNA could be estimated around 37.4%, which is the lowest, reported for any plant species. Non-coding regions have an even lower G+C content (32%). As expected, coding regions are enriched in G+C with an average level of 46% (Messeguer et al.1991). The 5-methyl cytosine (5MeC) was the only modified base observed by HPLC analysis in immature tissues and protoplast (average 20%), mature tissues (average 25%), mature pollen (average 22%) and seeds (average 27%) with an average estimated to be approximately 23% of total Cs (Immature tissues and protoplast (20%), which is high as compared to animal species, but well within the range reported for other plants (0-37%). This is consistent with a more recent work that found that m^5C content of tomato leaf genomic DNA was close to 22.3% (Teyssier et a, 2008l). However, considerable variation was also observed in the levels of methylation across different stages/tissues. Messeguer analyzed the nuclear DNA methylation of tomato *S.l. esculentum* cv VFNT, and showed that a whole green fruits had 5MeC content of the 25,5% that did not change significantly during ripening Nuclear DNA from immature leaves contained 20,3% of 5MeC while in mature leaves was higher the 5MeC content increased to 25,5%. Finally, DNA from immature stems and roots contained 20% of 5MeC. Similarly, in fruits of the cultivar *Ailsa Craig* harvested at different developmental stages, tissues and stage specific variations of DNA methylation levels were observed. Thus the 5MeC content in locular tissues was close to 20% and did not change during fruit development (Teyssier et al., 2008). Inversely, although the methylation levels of pericarp cell genomic DNA was stable at 30% during fruit development this level dropped down to 23% during fruit ripening. These results clearly indicate tissue-specific variations of the global DNA methylation during fruit development in tomato (Teyssier et al., 2008) and probably variations between cultivated tomatoes or growing conditions.

3.3.3 - Methylation analysis of leaf genomic DNA.

Genomic DNA from leaves was used to make a global methylation analysis using high performance liquid chromatography (HPLC) by Dr A. Rival (IRD, Montpellier) (Tab.4) The analysis showed all species had a similar level of 5MeC content in leaves.

Specie	% 5MeC/(5MeC+C)	± STD.DEV.
WVa106	23,41	0,18
S.l.cerasiforme	21,62	
S.pimpinellifolium	23,56	0,22
S.chmielewskii	24,49	
S.neorikii	22,40	0,72
S.arcanum	22,63	0,45
S.corneliomulleri	24,01	0,67
S.huayalasense	23,57	0,40
S.chilense	23,94	0,76
S.habrochaites	22,94	1,16
S.pennellii	24,57	1,19

Tab.4. Global methylation analysis by HPLC on genomic DNA of leaves from cultivated and wild tomato species.

To identify possible differences in methylation profile between tomato species, the methylation profile at the 5s rDNA locus was analyzed by Methyl sensitive Southern blot using the methylation-sensitive enzymes *Hpa*II and *Msp*I. Both enzymes recognize the sequence -C^1C^2GG-: *Hpa*II is inhibited when a single C or both are methylated whereas *Msp*I activity is only blocked when C^1 is methylated. Therefore *Msp*I activity reflects the CNG methylation while the comparison between *Hpa*II and *Msp*I provides an evaluation of the CG methylation.

HpaII digestion did not show difference between species: in all cases a unique band of high molecular weight was detected. This indicates a high methylation levels at this locus in all species analyzed without detectable variations between species. Inversely a ladder is seen in all cases when *MspI* is used, indicative of good digestion efficiency with this enzyme. This indicates a rather low CNG methylation level, which furthermore varies between species.

Based on *MspI* southern profile, tomato species could be separated in two subgroups. The first one includes *S.neorikii, S.arcanum, S.corneliomulleri, S.huayalasense, S.chilense* and *S.habrochaites* and is characterized by a lower digestion efficiency at the 5s locus

that the second group which contains *WVa106, S.l.cerasiforme, S.chesmaniae, S.pimpinellifolium* and *S.chmielewskii*. Hence, the CNG methylation level seems to vary between species at the 5s rDNA locus (Fig.46).

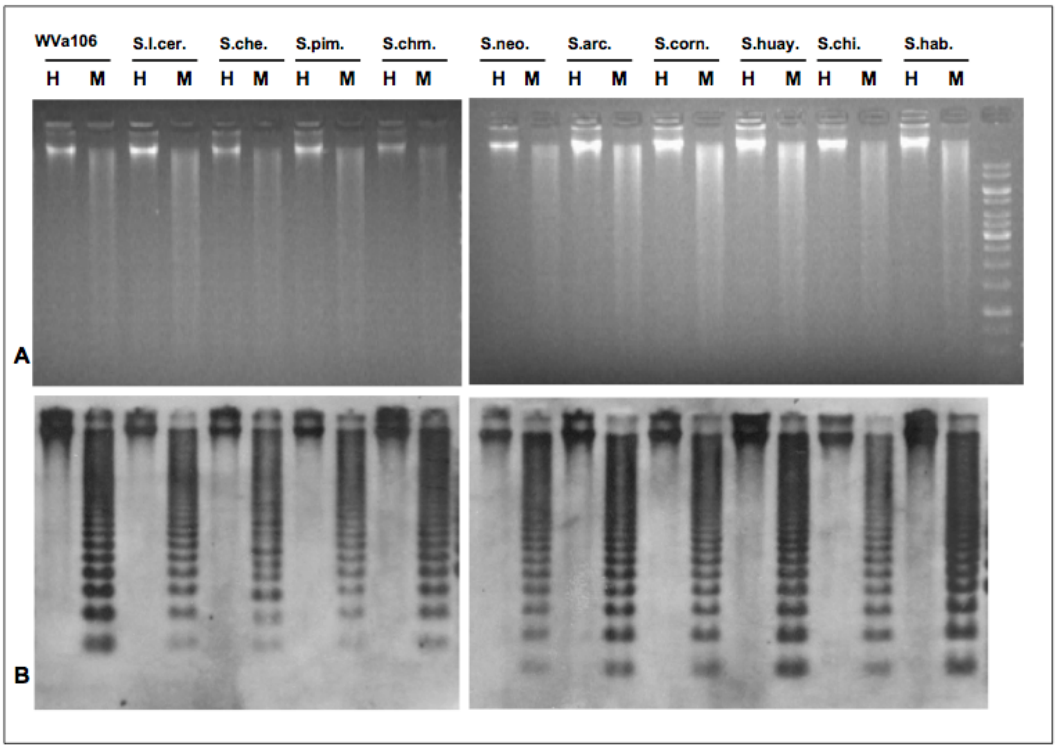

Fig.:46. Methylation analysis at repeated loci in leaves from wild tomato species. Gel was stained with ethidium bromide (A), blotted and hybridized with 5S rDNA (B).

3.3.4. - Fruit methylation analysis at the 5s locus.

Recent work have shown that the 5s rDNA locus is subjected to tissues specific changes in DNA methylation pattern during fruit development (Teyssier et al, 2008). Briefly, using the *Ailsa craig* cultivar, it has been shown that the CNG methylation increases in pericarp cells during fruit development at the 5s rDNA locus before decreasing during ripening.

Fruits samples were collected at these developmental phases: 20dpa, Breaker (Br) and Orange (Or) for the red-fruited species and 20dpa, 40dpa and 60dpa for green-fruited species. The different stages were choosing on the base of the gene expression analysis.

DNA methylation pattern was determined at the 5s rDNA locus using fruit pericarp genomic DNA prepared at different developmental stages. As shown in Fig. 47, *HpaII* did not efficiently cut genomic DNA in any sample analysed. This is consistent with a

high level of methylation at the 5S rDNA locus. However, as already observed in *A. craig* (Teyssier et al, 2008), the digestion efficiency of *MspI* varied between species and fruit developmental stages. Hence, in *S.l. cerasiforme* and *S. pimpinellifolium* the *Msp*I digestion is consistent with an increase in CNG methylation at 20dpa and breaker stage while the orange stage showed a lower level of methylation in the same sequence context. In WVa106 no difference was observed between the breaker and orange stages. A similar observation can be made for the green-fruited species the for *S. arcanum, S. corneliomulleri* and *S. huayalasense* . In these species, *MspI* digestion effeiciency is higher at 20 dpa than at 40 and 60 dpa. However there is no clear evidence of reduction in methylation level at 60 dpa *S. pennellii* showed an opposite situation with higher CNG type of methylation at the 5s rDNA locus at 20dpa compared to 40 and 60dpa.

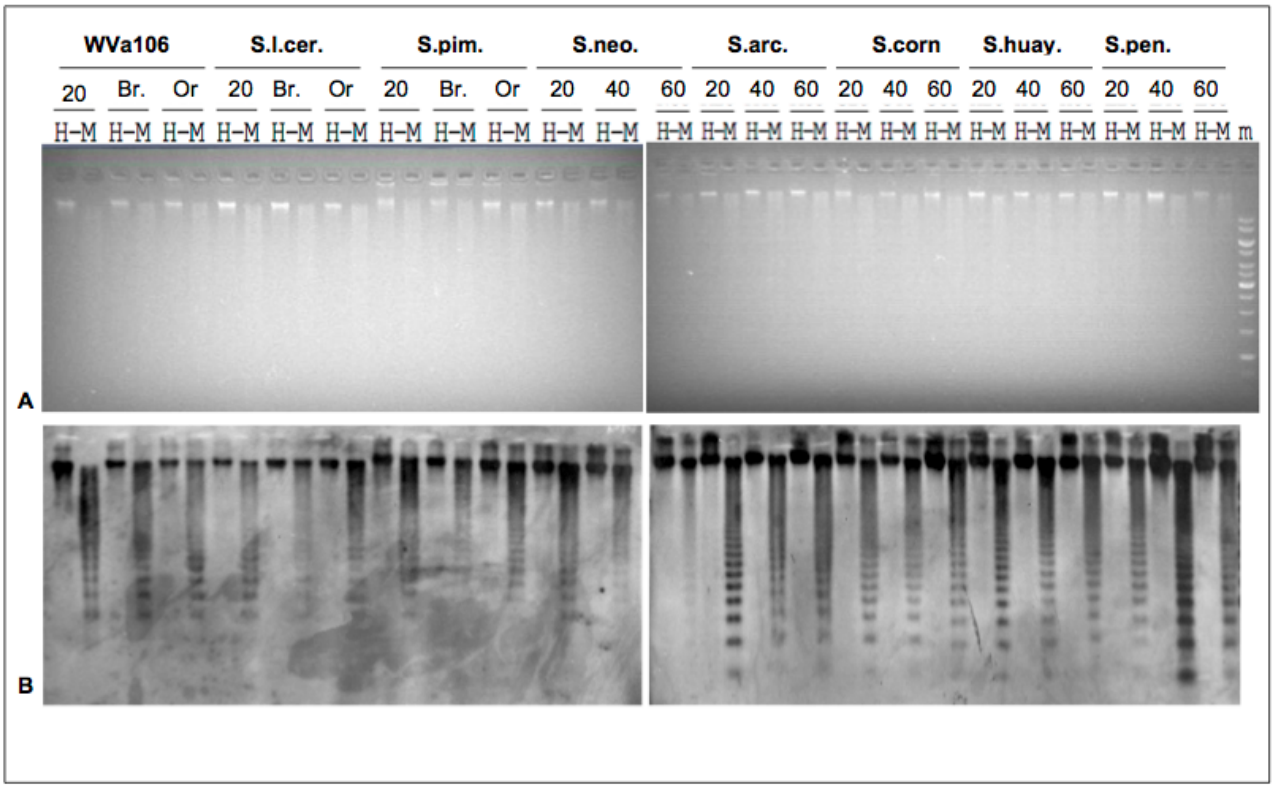

Fig.:47. Southern blot methylation analysis at 5S rDNA locus of cultivated tomato and wild relatives. Gel were stained with ethidium bromide (A) and blotted. The blot was successfully hybridized with a 5S rDNA probe. Developmental stage of fruit is indicated in days post anthesis (dpa) during fruit groth and as ripening stage breaker (Br.) and orange (Or.).

Taken together, these results clearly indicate that in all species, genomic DNA from pericarp cells is subjected to change in methylation during the different phases of fruit development and ripening.

In addition, changes in DNA methylation at this locus methylation vary between species.

3.3.5 – Analysis of possible interaction between methylation and endoreduplication.

Our previous analysis has demonstrated that although endoreduplication occurs during tomato fruit development in cultivated as well as in wild tomato species, both kinetic and extend of this phenomenon vary between species. In general, a lower level of ploidy in pericarp cell characterizes wild species with green fruits and the endoreduplication process is often delayed in these species (see Fig.40A and B) as compared to *VWA106* or to *S. pimpinelifolium*, a species closely related to the cultivated tomato.

At the time being there is some evidence that genomic DNA methylation could impact cell ploidization. Such evidence was provided by treating Chinese hamster cells with 5 aza cydine (5-azaC), a drug that leads to DNA demethylation (Mateos et al.,2005). They observeded that 5azaC was able to induce endoreduplicationin a dose-dependent fashion (Mateos et al.,2005). This led to an increase in cell ploidy level consistent with the hypothesis that methylation limit DNA endoreduplication.

We made use of the variability of DNA endoreduplication processes between tomato species to address this question. To investigate this possible relationship genomic DNA was extracted from fruits at different developmental phases and total 5MeC content was immunological measured using Imprint® Methylated DNA Quantification kit (Sigma Aldrich, MDQ1). Since the analysis required an extremely pure DNA a specific kit containing a resin able to remove all the protein and polysaccharides was employed (Nucleon phytopure).

The analysis revealed differences in red and green fruits at 10dpa. The methylation level of *WVa106* was set as control (100%, Fig.48B) while the DNA methylation level of the other species were expressed as a relative amount. *S.l. cerasiforme* appeared very similar to the cultivated species while the situation was more complicated for the other wild species.

The red-fruited species *S. pimpinellifolium* as well as the green-fruited *S. huayalasense* showed the highest relative methylation levels (250%) while the other species had a methylation level of twice the control.

In addition, when compared to the endoreduplication levels (Fig.43A) species that had similar endoreduplication index (*WVa106, S.l. cerasiforme, S. pimpinellifolium* and *S. neorikii*) showed differences in the methylation levels. To try to understand this apparent discrepancy, maximum C value of the pericarp cells was also considered. *WVa106* and *S.l. cerasiforme* showed a C value up to 256C and a lower relative methylation value compared to *S. pimpinellifolium* and *S. neorikii* that have a max C value of 128C and 64C, respectively but a higher methylation level. Similar considerations can be done for *S. arcanum, S. corneliomulleri* and *S. huayalasense*.

Finally, even if we did not find a linear correlation between endoreduplication and methylation species with a high C value (*WVa106* and *S.l. cerasiforme*) can have a low level in methylation as compared to the other species with a lower C value and a high methylation level. This observation would be in agreement with the initial hypothesis where an epigenetic mark could influence the endoreduplication process.

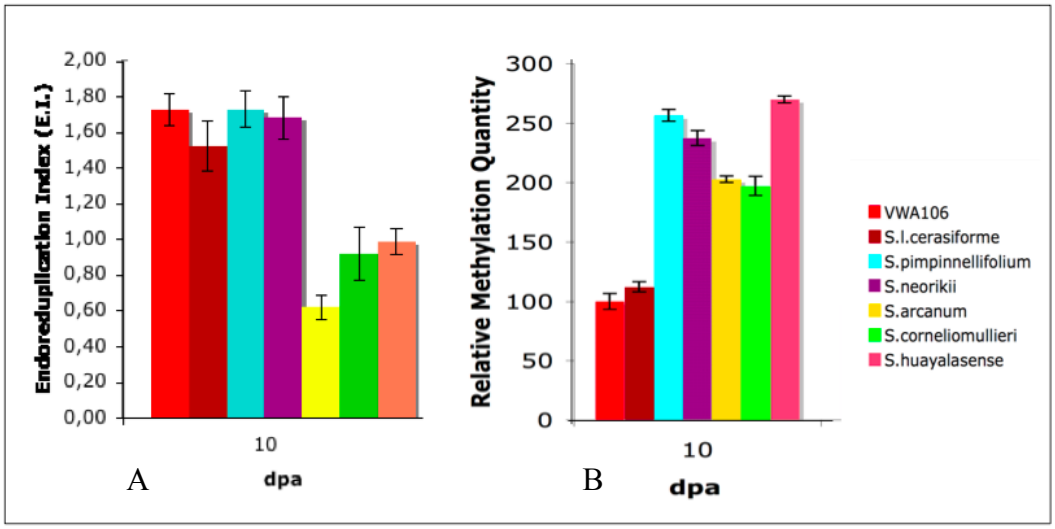

Fig.:48.Comparison of endoreduplication indexes and relative methylation quantity levels at 10 days post anthesis. In A, the endoreduplication index (E.I.) at 10dpa and in (B) the relative methylation level at 10dpa.

3.4 – DISCUSSION AND CONCLUSION.

As already shown in cultivated species, the dynamics of the pericarp thickness and structure is linked to the cell layer number but also to the size of the cells and to the ploidy levels (see chapter 1). The measurement of the cell size confirmed it. *S.l. cerasiforme* fruit perciarp contains cells with largest cell size among the species

considered in this study. When considering the kinetic of cell size increase, All red-fruited species behaved in a similar way with a first phase of cell size increase between 10 and 30dpa and a second one that took place during the ripening stage. This was also observed in green-fruited species though the second phase of cell size increase was species specific.The analysis max C value showed that the red-fruited species had a max C value comprised between 128C (*S. pimpinellifolium*) and 256C (*WVa106* and *S.l. cerasiforme*) while the green-fruited species had a max C value between 64C (*S. neorikii, S. arcanum* and *S. pennellii*) and 128C (*S. corneliomulleri* and *S. huayalasense*). In addition the analysis showed a different distribution of the nuclei within the different ploidy classes during fruit development. Main differences between species were detected at 10dpa and at 60 dpa or red ripe stage. Hence, in the red fruited species in the max C value increased during ripening between the breaker and orange stage, but in green fruited speceis, the max C value did not change after 30 and/or 40dpa. Difference in ploidy control was also evidenced by analysing the mean C value (MCV). In green fruited species the MCV increased progressively in a linear way, although two phases could be detected (before 30 dpa and after 40 dpa) that reflect different kinetic of ploidy change during fruit development (Fig.40B). *S. huayalasense* was somehow specific, it has the two phases of progressive increase in average ploidy level that are separated by a short period of rapid increase in MCV (between 30 and 40 dpa).In contrast, in red-fruited species the MCV increased in an exponential way at late developmental stages reflecting the fact that numerous nuclei were in the classes 128C and 256C (Fig.40A). In general, MCV was correlated with the cell size but the intrinsic nature of this parameter give an overestimated value of the ploidy.

The endoreduplication index (EI), which represents an accurate value of the ploidy levels due at the number of endocycles of nuclei and not to the value class, was also analysed. It can therefore be used to correlate the increase ploidy fruit and cell size. The correlation between EI and fruit size showed in all cases except *S.huayalasense* that exists a positive and linear correlation between fruit size and EI. The difference showed from *S huaylasense* was a high increase in fruit size with little changes in the endoreduplication index between 10 and 30dpa as well as between 40 and 60dpa.

Similarly, EI was positively correlated with cell size. A general linear correlation was found in almost cultivated and wild tomato species. Inversaly *S. pimpinellifolium* and *S. pennellii* showed an exponential correlation of these two parameters, this phenomenon has been also observed in *S.huayalasense* between 10 and 30dpa.

Although a linear correlation between fruit size and ploidy level has been found in all the species (excluded the specie *S.huayalasense*) a linear correlation was not found between cell size and ploidy levels with clear different situation between species. This allows the separation of the species in three principal groups. In the first group we have species that showed a linear correlation between fruit size and ploidy in addition to ploidy and cell size as *WVa106, S.l.cerasiforme, S.arcanum,S.corneliomulleri* and *S.neorikii*; in the second group we found *S.pimpinellifolium* and *S.pennellii* that showed a linear increase between fruit size and ploidy and an increase of the cell size not linked to ploidy levels. Finally, *S.huayalasense* is the only member of this group; it did not show any linear correlation between the parameters considered suggesting that different mechanisms could participate at the increase of fruit and cell size as well as ploidy level.
These data would altogether suggest that the developments of the fruit size and ploidy levels are not so linked as thought and that other factors could be players of these mechanisms.

To investigate the possibility that an epigenetic mark could play a role in the regulation of fruit development and ripening we performed a global and locus specific genomic DNA methylation analysis. Therefore the wild tomato species would represent a good model system to study this specific interplay because they behave in different way when considering ploidy parameter.
Global DNA methylation analysis done by HPLC on genomic DNA from leaves of the different species did not show particular difference between species.

On the other hand, locus specific DNA methylation analysis indicated that the 5s rDNA locus was not similarly methylated in all species analysed. Obviously in all cases the CG methylation at this locus is very high, irrespective to the tissues considered (Leaves or fruits). Hence no difference between species could be seen in this methylation context. Inversely, clear differences were detected when analysing CNG methylation In leaves. Red-fruited species and *S. chmielewskii* (green-fruited species), had similar methylation level at the 5s locus rDNA in CNG sequences and it was higher than the methylation level found in green-fruited species at this locus.

As said above, due to the very level of methylation at a CG context no variations of DNA methylation during fruit development could be detected in this sequence context in any

species analyzed. Inversely, DNA methylation in a CNG context was subjected to variations during pericarp fruit development. The methylation at CNG sequences was generally low in early stages of fruit development, and increased at the breaker stage before decreasing during ripening. This changes in the CNG methylation levels were observed in red-fruited species except *VWA106.* Green fruited-species showed changing in the methylation levels at CNG sequences different compared to *WVa106*. Compared to *WVa106,* which had a variation in DNA methylation at CNG sequences, species as *S. arcanum* and *S. corneliomulleri* and *S. pennellii* did not show a similar behaviour.
Finally, the global methylation analysis done using genomic DNA extracted from fruit pericarp at 10dpa for investigate its relationship with the ploidy levels showed controversial results.

Our results, similarly to previous results of Teyssier and al (2008), support the hypothesis that in pericarp tissue, locus specific variation in DNA methylation is taking place. In addition wild species showed different methylation profiles in pericarp during fruit development as well as different endoreduplication levels. All together, these results can suggest that the species can be separated on the base of an epigenetic information as can be the DNA methylation and also that these mechanisms could be one of the players of the different mechanisms that govern fruit development.
The wild tomato species showed in this work a series of interesting results making them powerful tools to investigate the different aspects of fruit development.

CHAPTER 4

STUDY of the EPIGENETIC DIVERSITY IN CULTIVATED AND WILD TOMATO SPECIES: THE CASE OF the ENHANCER OF ZESTE GENES, *SlEZ2* AND *SlEZ3*

4.1 – INTRODUCTION.

Despite the existence of numerous epigenetic marks and regulatory proteins, only two systems that cause changes in the epigenetic state of a cell have been described extensively in plants: one involves DNA methylation, whereas the second is controlled by the Polycomb-group proteins (PcG) (Shubert et al.,2005).

Polycomb group proteins (PcG) were first discovered in the fruit fly *Drosophila melanogaster* and their characterization has revealed that these proteins work in a complex way. Distinct complexes called PRC1 (Polycomb Repressive Complex 1), PRC2 (Polycomb Repressive Complex 2) and PhoRC (Pleiohomeotic Repressive Complex) have been identified (Schwartz and Pirrotta, 2008*)*.

Polycomg group proteins convey epigenetic inheritance of repressed transcriptional state. Although the mechanism of the action of the PcG is not completely understood, methylation of histone H3 lysine 27 (H3K27) is important in establishing PcG-mediated transcriptional repression.

In plant only the PRC2 has been clearly identified. *Drosophila* PRC2 contains three main members: Enhancer of zeste (E(z)), suppressor of zeste 12 (Su(z)12) and extra sex comb (ESC). Based on sequence homology several homologous proteins have been identified in *Arabidopsis*: three E(Z) homologues MEDEA (MEA), CURLY LEAF (CLF) and SWINGER (SWN) and three SUZ12 homologues, FERTILIZATION INDEPENDENT SEED2 (FIS2), VERNALIZATION 2 (VRN2) and EMBRYONIC FLOWER 2 (EMF2). The remaining *Arabidopsis* PRC2 proteins include two WD40 motif proteins, FERTILIZATION-INDEPENDENT ENDOSPERM (FIE) and MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) homologs of the *Drosophila* ESC and p55 proteins (Fig.27) (Reyes and Grossniklaus et al., 2003). These proteins play an important role in controlling normal plant development. Based on molecular and

biochemical evidence, at least three PRC2 complexes co-exist in *Arabidopsis (*FIS2-PRC2; EMF2-PRC2; VRN2-PRC2), harbouring different paralogs of E(z) and Su(z)12 proteins families, each complex controls a particular developmental program such as seed development, germination, vegetative growth and flowering cell identity.
The PcGs have an intrinsic histone methyltransferase activity, which is mediated by the evolutionarily conserved SET-domain. The PRC2 complex suppresses *Homeobox* (Hox) genes and other target genes by placing trimethylation marks on histone H3 lysine 27 (H3K27me3). The histone methyltransferase activity of the subunit E(Z) in the PRC2 is essential for gene silencing.

Recently three genes have been identified in *Solanum lycopersicum* by sequence homology with *Arabidopsis thaliana* which code for proteins of the Enhancer of zeste class called *SlEZ1*, *SlEZ2* and *SlEZ3* (How Kit et al.,2010). These genes correspond to two CLF-homologous sequences *SlEZ2* (*SlCLF1*) and *SlEZ3 (SlCLF2)* and one SWN-homologous sequence, *SlEZ1* (*SlSWN*).
All *SlEZ* genes the typical signature domains of E(z) proteins; EZD1, EZD2, the SANT domain, the CXC domain and the SET domain which is responsible for the methyltransferase activity (Fig.27). SlEZ3 on the other hand is thought to encode a truncated peptide lacking the SET domain.
Phylogenetic analysis using full-length cDNAs of E(z) proteins from *Drosophila*, *Arabidopsis*, maize, rice and petunia has shown that the similarity between SlCLF2 and PhCLF1, or SlSWN and PhSWN is higher than between CLF and SWN homologues within each species.

In this chapter we will investigate the evolutionary differences between the sequences of the genes *SlEZ2* and *SlEZ3* in cultivated tomato and wild relatives. In addition the characterization of a new LTR-retrotransposon, which is inserted within the promoter region of the SlEZ2 genes in a few species and its possible impact on the nearby gene, will be analyzed.

4.2 – ALLELIC DIFFERENCES IN *SlEZ2* AND *SlEZ3* GENES.

How Kit during his PhD thesis work sequenced the cDNA of *SlEZ1* (2490bp), *SlEZ2* (2744bp) coding for two proteins of 829 and 921 amino acids. In addition, he partially sequenced the *SlEZ3* cDNA showing that the coding sequence is interrupted by a stop codon at position 1256 before the catalytic SET domain (How Kit et al., 2010).

On the basis of his work and the information available on the tomato genome website (http://solgenomics.net/) we used the partial cDNA sequence of SlEZ3 to identify in the transcriptome database of *S. peruvianum* a complete coding sequence for a functional *SlEZ3* protein. We then identified the genomic position of this gene, which is located on chromosome 2 (SL2.40ch02:48708600.. 48725735), while *SlEZ2* paralogue of *SlEZ3* is located on chromosome 3 (SL2.40ch03:12923432..12932543). *SlEZ2* and *SlEZ3* genomic sequences were PCR amplified were then complete sequenced in cultivated tomato (*WVa106*) and its wild relatives (*S. pimpinellifolium*, *S. neorikii*, *S. arcanum* and *S. pimpinelifolium*).

Although *SlEZ2* and *SlEZ3* are paralogue genes, they showed important differences (Fig.49A). The comparison revealed that *SlEZ2* is shorter (9111 bp) as compared to *SlEZ3* (17135 bp) (Fig.49A) and even if *SlEZ2* showed a similar organization of introns and exons as compared to *SlEZ3*, *SlEZ3* had 5 exons more at the 3'-end. A comparison of the complete genomic sequence of *SlEZ2* and *SlEZ3* showed a similarity of the 44%. By dot plot analysis we showed that *SlEZ3* has been subjected to two insertions in the intron number 4 (Fig.49B-1a) and in the intron number 12 (Fig. 49B-1b) and to two deletions in introns 3 and 9, respectively (Fig. 49B-2a and 2b). *SlEZ3* contains a long intron (up to 6835bp) that was recognized by the Plant Repeat Database (http://plantrepeats.plantbiology.msu.edu/index.html) showing the presence of a LINE-like retrotransposon (classified as a TERT003) and a repeat sequence identified as OTOT000, unclassified.

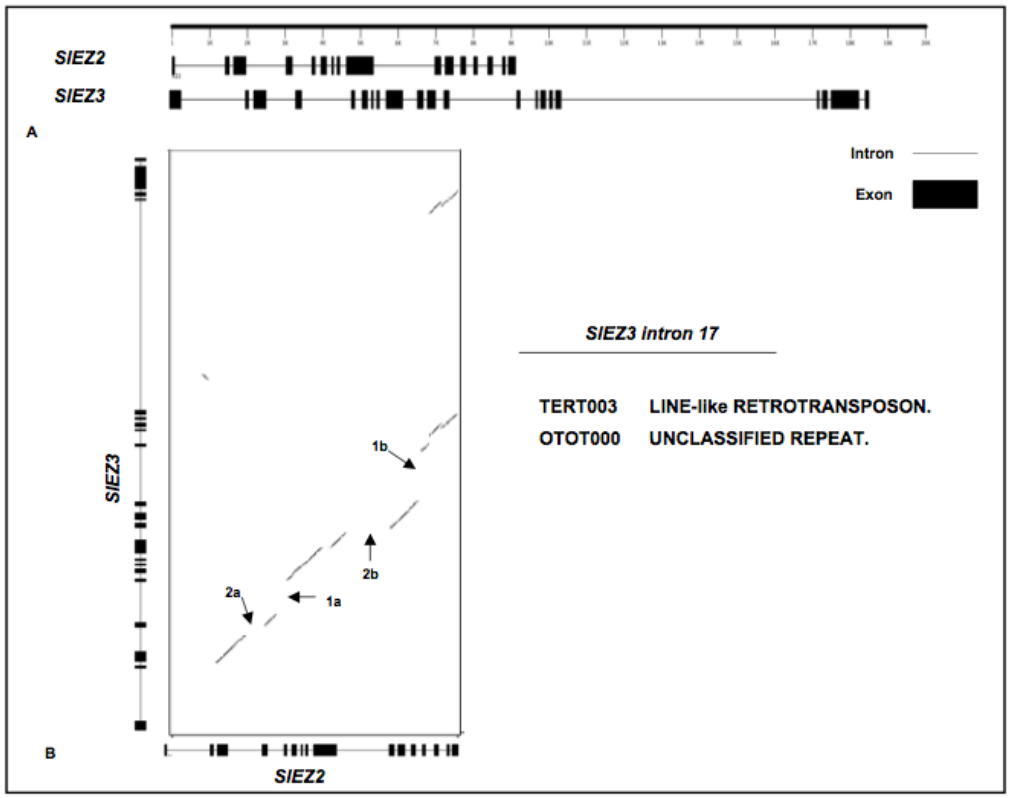

Fig.: 49.Comparison of the *SlEZ2* and *SlEZ3* genes. The genes (A) are different in size with a global length of 9kb for *SlEZ2* and 18.5Kb for *SlEZ3*. A global dot plot matrix (B) has been performed and it shows a similarity of 44%. The alignment has been done using the complete SlEZ2 gene and the first 16 exons of *SlEZ3*. In addition the dot plot matrix reveals clearly that *SlEZ3* has been subjected to two insertions in the intron number 4 (1a) and in the intron number 12 (1b) and to two deletions in the intron 3 and 9, respectively (2a and 2b). Finally, *SlEZ3* shows a long intron (up to 6835bp) containing two LINE-like retrotransposons identified using the Plant Repeat Database of the Michigan University (http://blast.plantbiology.msu.edu/plantrepeats/tmp/blastn-prdb-9179-1328111994.html - aln1).

Sequence analysis was also performed in the wild species. *SlEZ2* showed small differences in size, ranging from 9108bp (*S. arcanum*) to 9163bp (*S. neorikii*). *SlEZ3* showed more differences ranging between 16222 bp (*S. neorikii*) and 17135bp (*WVa106*). Sequence analysis revealed that the difference was due to variations in intron 17 length, which varied between 6000bp (*S. neorikii*) and 6835bp (*S. pimpinellifolium*) probably due to new insertions.

4.3 – EVIDENCE OF AN ALTERNATIVE SPLICING FOR THE *SlEZ3* GENE.

The analysis of the coding DNA of *SlEZ3* made by How Kit (How Kit et al., 2010) suggested that *SlEZ3* was a pseudogene coding for a truncated peptide lacking the SET

domain necessary for the histone methyltransferase activity suggesting that *SlEZ3* (How Kit et al.,2010). We used this sequence to rescreen the SGN database (Sol Genomics Network website). This led to the identification of a new *SlEZ3.*

A cDNA sequence (long expressed sequence tag EST of 3780bp named "a18898") was from *S. peruvianum*/*S. corneliomulleri* .
The translation of this sequence using Expasy translation tool (http://web.expasy.org/translate/) showed an open reading frame (ORF) of 2523bp coding for a 841 amino acid long protein. The amino acid sequence analysis using SMART (http://smart.embl-heidelberg.de/) showed that in addition to EZD1, and EZD2 domains it contained the SANT, CXC and SET domains characteristic of proteins with a methyltransferase activity.

Our sequence results concerning *SlEZ3* were not in a complete agreement with the work of How Kit (How Kit et al., 2010). The evidence of a cDNA encoding a truncated protein (How Kit et al., 2010) and of a cDNA encoding a completely functional protein, suggested us that *SlEZ3* might be subjected to alternative splicing or it could encode a functional protein in a selected set of species and not in other.

To discriminate between these hypotheses a SlEZ3 specific retrotransciption was performed using a primer complementary to the 3' non-coding region of the *SlEZ3* mRNA (EZ3R1 - CAACTTTACTGGAAGATCAATGG) in *WVa106* and *S. arcanum*. The amplification of the coding region of *SlEZ3* was done splitting the sequence in 4 parts with an expected length of approximately 1300bp each (Fig 50).
The results showed that between exons 4 and 9 at least three amplification products were present (Fig.50B) that confirmed the hypothesis of an alternative splicing. Cloning and sequencing of the RT-PCR products from *SlEZ3* of *WVa106* and *S. arcanum,* revealed three types of transcripts that were detectable in young fruits (10dpa) and leaves as well (Fig.50B). In addition these transcripts were present in all the species analyzed (*S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pennellii*) (Fig.50B).
The cloning and sequencing of the fragments amplified showed that *SlEZ3* gene is able to give rise to a full length mRNA coding for a functional protein (Fig.50C) and that the alternative splicing forms of *SlEZ3* mRNA either still retained the intron number 9 or the introns number 6, 7 and 8 (Fig.50C). The translated mRNA isoforms of the *SlEZ3* had

two stop codons: one within the intron 6 (isoforms containing introns 6, 7, 8) and the second one within the intron 9.

A sequence alignment between the sequence found from How Kit and the sequence analyzed in this study, showed that the isoform sequenced from How Kit was the mRNA isoform containing the intron number 8 which showed a stop codon at the position 1256.

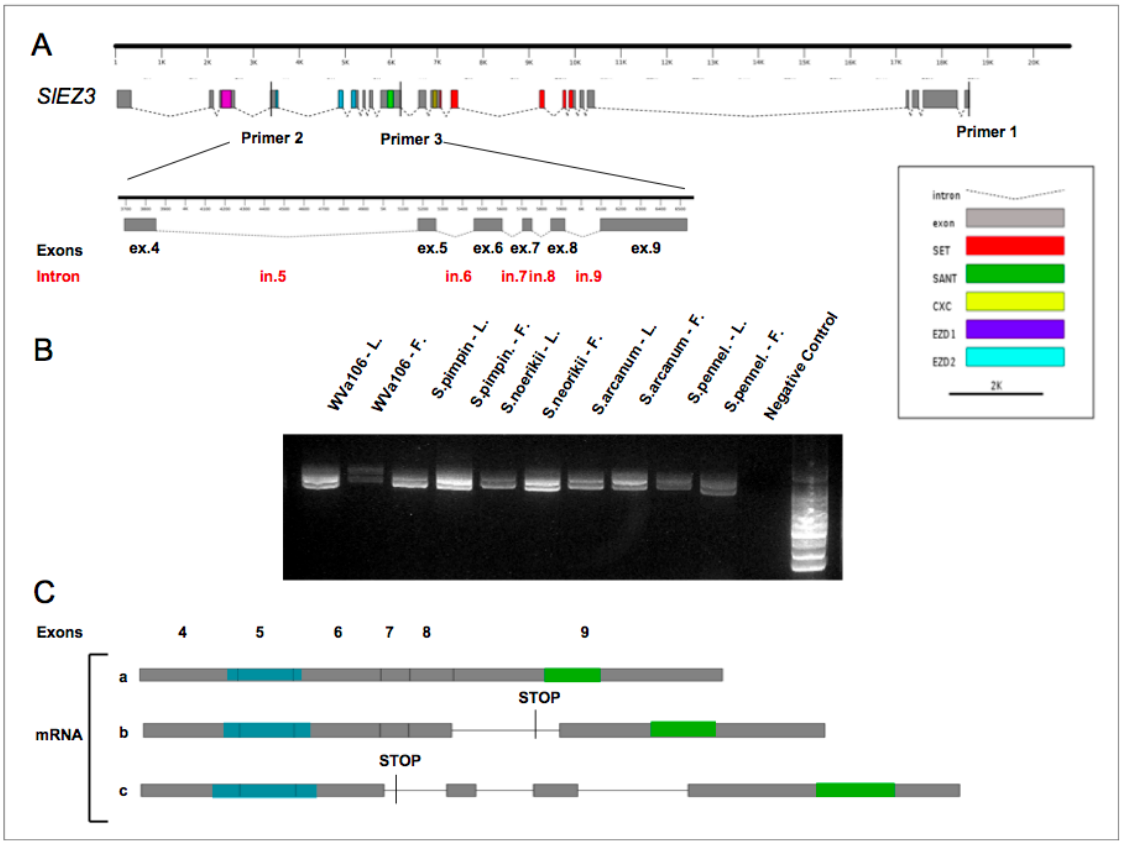

Fig.:50. Alternative splicing and structure of *SlEZ3* transcripts. In A, *SlEZ3* gene structure with the different domains and a schematic representation of the gene organization between the exons 4 and 9 were the alternative splicing has been detected. In B, the PCR amplification of the RT products obtained from leaves and young fruits (10dpa) of *WVa106, S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pennellii* that have been cloned and sequenced showing three different isoforms: a fully spliced sequence and two longer sequences still containing one or three introns thus coding for a truncated protein (C).

4.4–SEQUENCE ANALYSIS

To investigate the similarity of the *SlEZ3* gene between the tomato species, we cloned and sequenced the complete genome sequence of *SlEZ3* in *S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pimpinellifolium*. On the Other hand, the analysis of the sequence of the transcripts was done by cloning and sequencing the different fragments in

cultivated tomato var. *WVa106* (where How Kit showed that *SlEZ3* was a pseudogene) and in *S. arcanum*. Alignment of the two coding sequence (cDNA) cloned and sequenced in *WVa106* and *S. arcanum*, showed a similarity of the 98%. On the base of this observation we deducted the coding sequence of *S. neorikii*, *S. pimpinellifolium* and *S.pennellii* by alignment with the coding sequence of *WVa106*.

In addition the analysis of the alternative splicing (Fig.50) clearly showed that in all the species analyzed (*WVa106*, *S. pimpinellifolium*, *S. neorikii*, *S. arcanum* and *S. pennellii*) three isoforms with the same size were present. This observation suggested that in each species a isoform encoding a functional SlEZ3 protein was persent. Finally, the cDNA deducted (*S. neorikii*, *S. pennellii* and *S. pimpinellifolium*) presented a similarity compared with the cDNA cloned and sequenced (*S. arcanum* and *WVa106*) a 99, 98, 99 and 98% homology respectively with *WVa106 gene* (Fig.51).

Despite the high level of similarity between species, *SlEZ3* cDNA showed differences in some positions among red- and green-fruited species (Fig.51). While the 3' end sequences encoding for the SANT, CXC and SET domains were highly conserved, differences were present in the 5' end and in the central part of *SlEZ3* cDNA. These differences are observed at the positions 207, 249, 318, 593, 614, 637, 754, 762, 864, 1254, 1270, 1290, 1466 and 1558 of the cDNA. For the positions 207, 593, 614 and 1254 the purine base present in the red-fruited species are changed with a pyrimidine in the green-fruited species. An opposite situation was seen in the others positions (Fig.51).

Similar analyses were also made for *SlEZ2* (Dr. Anne Pribat and PhD. Student Lisa Boureau). The cDNA from *SlEZ2* obtained were cloned and sequenced in the wild tomato species (*S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pennellii*) and showed a similarity of the 99% compared to *WVa106.*

The *SlEZ2* cDNA sequenced in wild species, when compared to the *WVa106* cDNA, showed differences at the positions 636, 773 and 1860. The positions 636 and 773 showed a thymine in the cultivated *S.l. lycopersicum* while the wild relatives showed a cytosine; opposite situation was observed at the position 1860. *SlEZ2* in the wild specie *S. pennellii* contained an ATT sequence inserted from the position 1881 until 1884, not found in the other species (Fig.52).

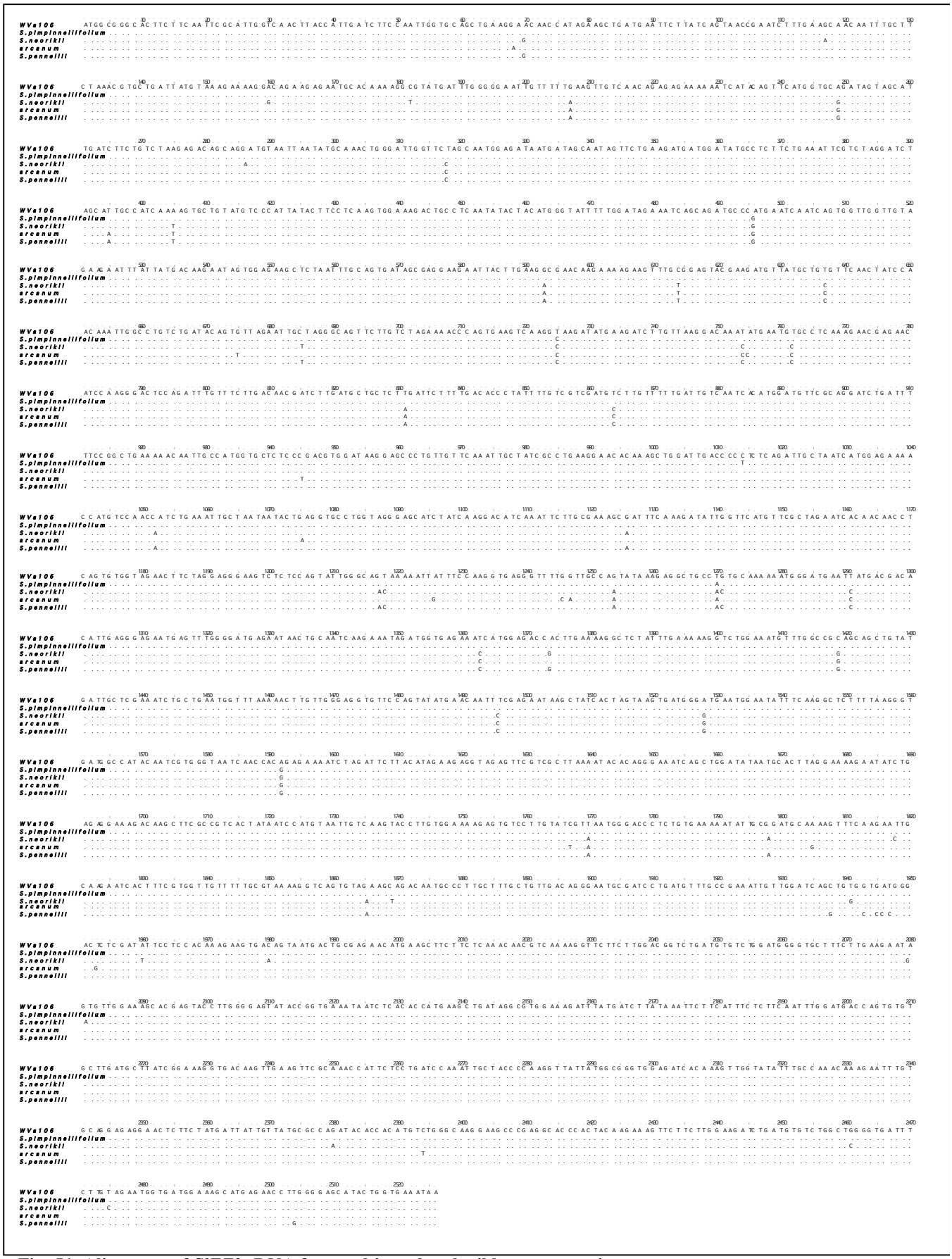

Fig.:51. Alignment of *SlEZ3* cDNA from cultivated and wild tomato species.

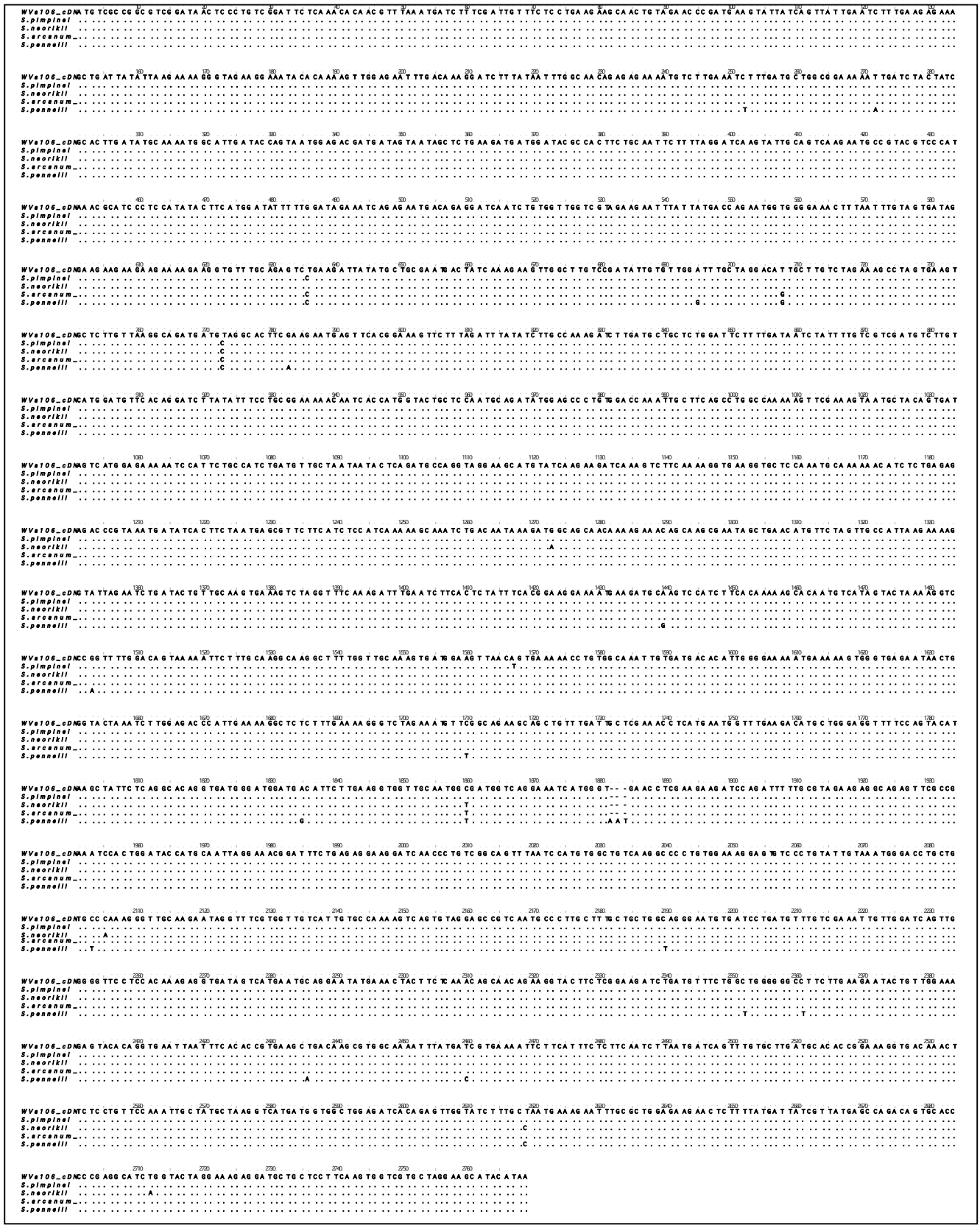

Fig.:52. Alignment of *SlEZ2* coding sequences from wild and cultivated species.

To evaluate the impact of DNA sequence modifications on the protein sequence, all sequences were translated (Expasy Translator tool) and aligned. The analyses of the amino acid sequences were performed for all species for each gene and between the two genes.

The results (Tab.5) showed a similarity between species of 99-100% for *SlEZ2* and 97-100% for *SlEZ3*, while the comparison between *SlEZ2* and *SlEZ3* showed a similarity of the 58-59%.

	S.arcanum EZ2	S.neorikii EZ2	S.pennelii EZ2	S.pimpin EZ2	WVa106 EZ2	WVa106 EZ3.	S.pimpin EZ3.	S.neorikii EZ3.	S.arcanum EZ3.	S.pennellii EZ3.
S.arcanum EZ2	100%	100%	100%	100%	100%	58%	59%	58%	59%	58%
S.neorikii EZ2	100%	100%	99%	100%	100%	58%	59%	58%	58%	58%
S.pennelii EZ2	100%	99%	100%	99%	99%	59%	59%	58%	59%	58%
S.pimpin EZ2	100%	100%	99%	100%	100%	58%	59%	58%	58%	58%
WVa106 EZ2	100%	100%	99%	100%	100%	58%	59%	58%	58%	58%
WVa106 EZ3	58%	58%	59%	58%	58%	100%	99%	97%	98%	97%
S.pimpin EZ3	59%	59%	59%	59%	59%	99%	100%	97%	98%	98%
S.neorikii EZ3	58%	58%	58%	58%	58%	97%	97%	100%	98%	99%
S.arcanum EZ3	59%	58%	59%	58%	58%	98%	98%	98%	100%	98%
S.pennellii EZ3	58%	58%	58%	58%	58%	97%	98%	99%	98%	100%

Tab.5. Similarity comparison between *SlEZ2* and *SlEZ3*. The results showed a similarity between species of 99-100% for *SlEZ2* and 97-100% for *SlEZ3*, while the comparison between *SlEZ2* and *SlEZ3* showed a similarity of 58-59%. The analyses were performed using UGENE, a free open-source cross-platform bioinformatics software (http://ugene.unipro.ru/).

The analysis of the *SlEZ3* amino acid sequences from *WVa106* and wild relatives showed a good domain conservation even though differences in the coding sequences between species were present. These differences between species were present in the EZD1 domain only which showed a methionine (M) at the position 166 of *WVa106* while a valine (V) was present in the wild relatives (Fig.53). A similar situation was observed at the position 198 where a glycine (G) was present in red fruited species while in the green fruited species it was substituted by an aspartic acid (D) (Fig.53). These two positions were located in two highly conserved regions (Fig.54). The methionine and/or valine were observed in the CEP domain (CLF, EZA1 and PHCLF) proposed from Mayama in 2003 (Mayama et al.,2003). In the position 166 in the SlEZ2, SlEZ1 and also in AtCLF,

AtSWN and AtMEDEA (Fig.54) a glutamic acid was highly conserved (E) but *WVa106* and the wild relatives showed respectively a methionine (M) and valine (V). The situation observed at the highly conserved position 198 was different, in this case *WVa106* and *S. pimpinellifolium* had a glycine instead of a glutamic acid (E) and the green-fruited species had an aspartic acid (D) which is similar to glutamic acid.

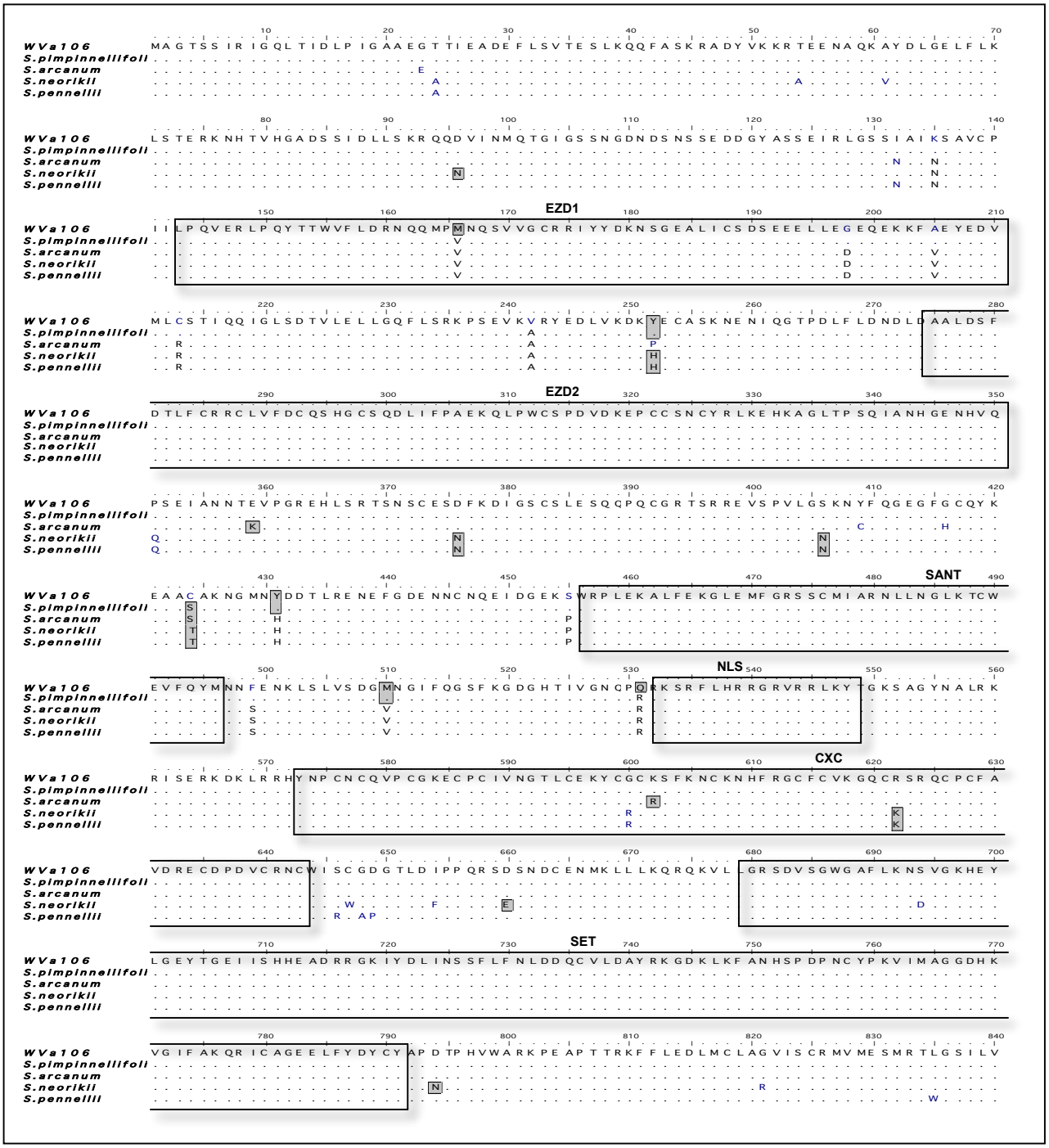

Fig.:53. Alignment of the *SlEZ3* amino acid sequences from cultivated and wild tomato species. The black lines group the different domains EZD1, EZD2, SANT, NLS, CXC and SET. Grey box shows the amino acids different between the different species as compared to the consensus sequence.

In contrast to the EZD1 domain we found that the EZD2 and SANT domains were perfectly conserved (Fig.53).
Nuclear localization signal (NLS) was predicted at the position 532 up to 548 using the freely available NLStranamus website (http://www.moseslab.csb.utoronto.ca/NLStradamus/) (Nguyen et al.,2009) (Fig.53). The CXC sequences were also detected using the plant transcription factor database (PlantTFDB freely available http://planttfdb.cbi.edu.cn/) (Zhang et al., 2011) and revealed only few differences without changes in the quantity of the cysteines. The SET domain was also perfectly conserved (Fig.53).

The *SlEZ3* amino acid sequence from *WVa106* was compared to the protein sequences of *SlEZ1* and *SlEZ2* (How Kit et al., 2010) and to the proteins *AtCURLY-LEAF, AtMEDEA* and *AtSWINGER* of *Arabidopsis thaliana* (Fig.54).
The alignment of the entire amino acid sequence of *SlEZ3* from *WVa106* with the related proteins revealed that several domains are well conserved. Although the CXC domain can been easily identified in the amino acid sequence of *SlEZ3* and it is highly conserved between wild tomato species, it appeared to have a sequence 41 amino acids that did not show any similarity with the CXC sequences of AtCURLY-LEAF, AtMEDEA, AtSWINGER and also SlEZ1 and SlEZ2 (Fig.54).

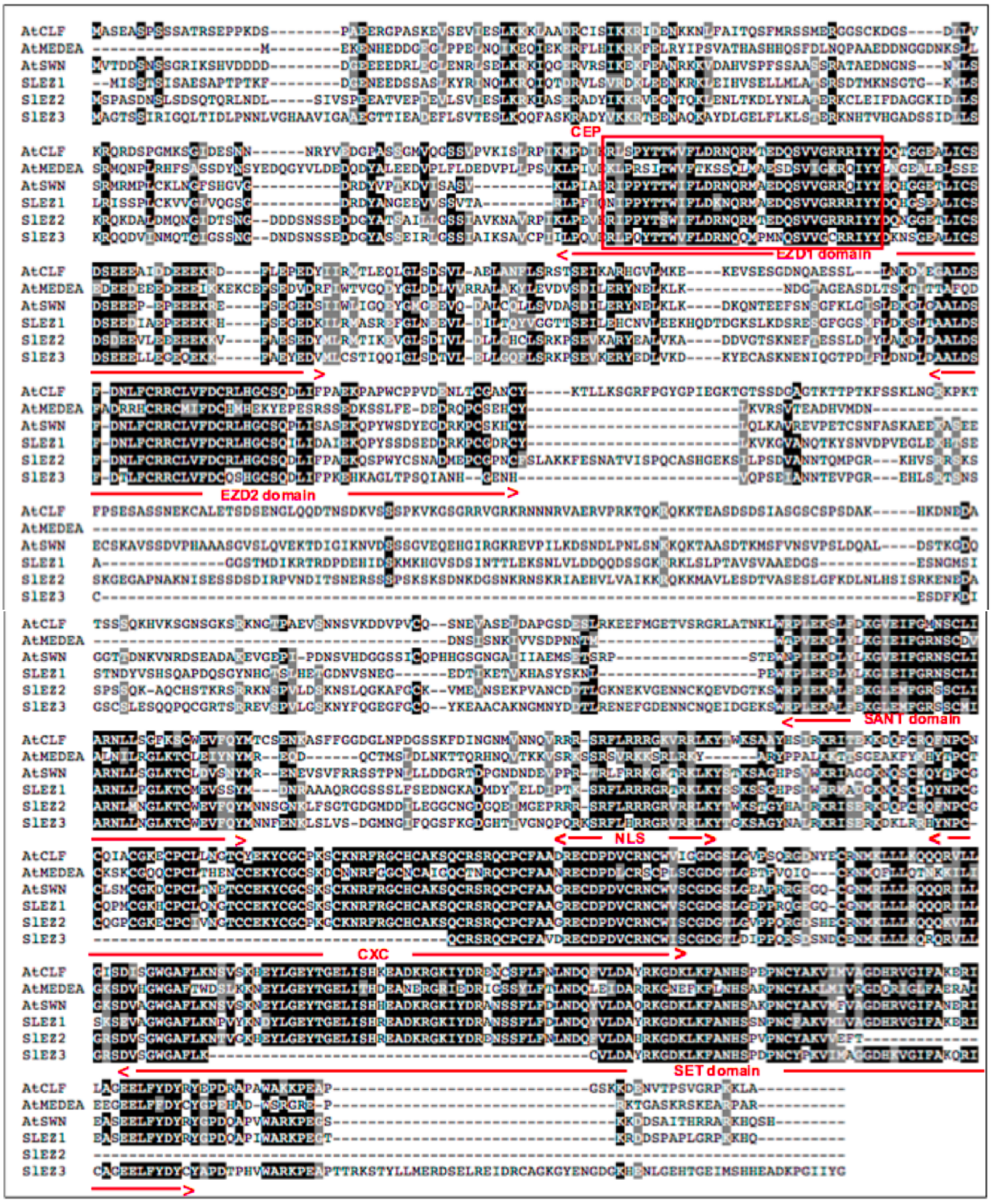

Fig.:54. Amino acid sequence of *SlEZ3* (*WVa106*), *SlEZ1* (EU057688), *SlEZ2* (EU057689), *AtCLF* (NM_127902.5), *AtMEDEA* (NM_100139.3), *AtSWN* (NM_116433.2). The sequences were aligned using ClustalW. Identical amino acids are shaded in black and conserved residues in grey. The position of EZD1, CEP, EZD2, SANT, NLS, CXC and SET domains are indicated in red.

.

4.5 – *SlEZ2* AND *SlEZ3* GENE EXPRESSION ANALYSIS DURING FRUIT DEVELOPMENT.

We analyzed the pericarp expression levels of *SlEZ2* and *SlEZ3* during fruit development and ripening of cultivated and wild species by real time quantitative PCR.

The analysis (Fig.55A and B) showed that the expression levels of the gene *SlEZ2* decrease during fruit development in all the species suggesting it to be involve in the first phase of fruit development (corresponding to active cell division) and also that its function was conserved between species.

Unlike *SlEZ2*, *SlEZ3* (Fig.56) was not highly expressed during fruit development. Despite its low expression, two peaks of expression were detected at 10dpa (*WVa106*) and orange stage in *WVa106* and *S.l. cerasiforme*. Green-fruited species showed a different situation: in *S. corneliomulleri* the highest level was detected at 10dpa, in *S. neorikii* and *S. huayalasense* the maximum expression level was reached after at 50dpa while in *S. pennellii* at 40dpa. In *S. arcanum* no differences were detected during fruit development.

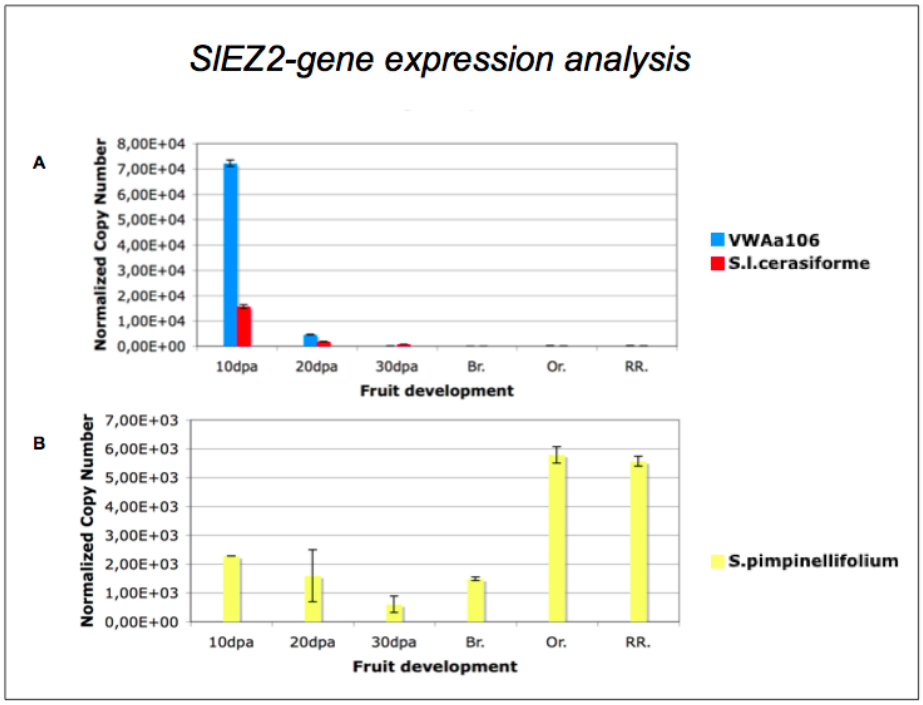

Fig.:55A. ***SlEZ2*** **expression levels. Red-fruited species (A and B) showed that** ***SlEZ2*** **expression decreases during fruit development.** ***S. pimpinellifolium*** **showed a decrease from 10dpa to breaker (Br.) stage and a further increase at orange (Or.) and red ripe (RR.). The normalized copy number is expressed in number of transcripts per microgram of RNA.**

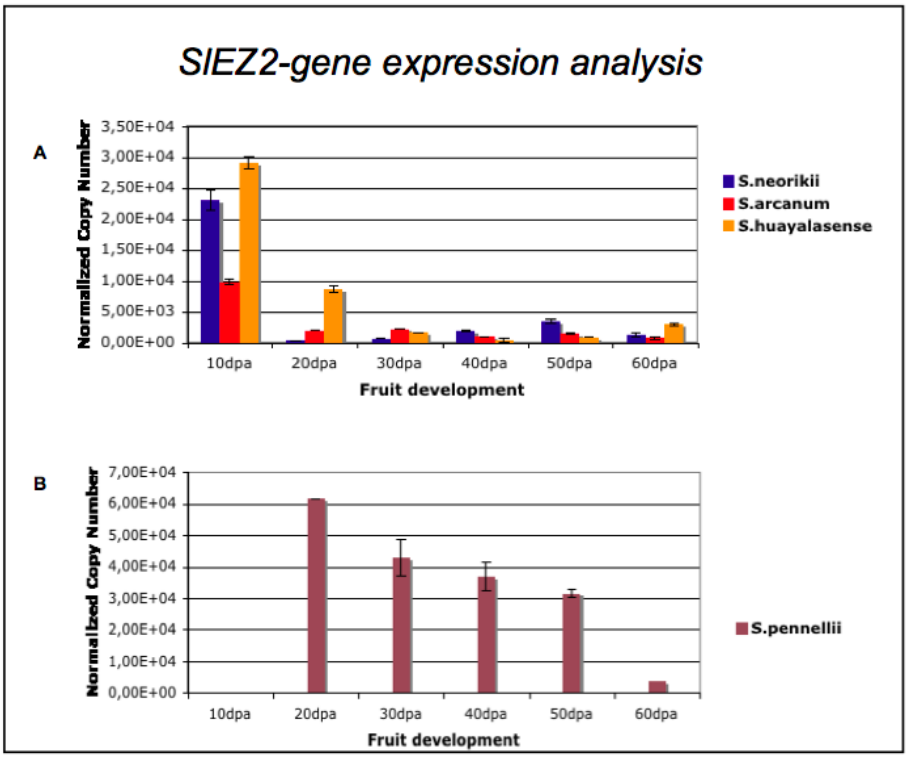

Fig.:55B. ***SlEZ2*** **expression levels. Green-fruited species (A and B) showed that** ***SlEZ2*** **expression decreases during fruit development. The normalized copy number is expressed in number of transcripts per microgram of RNA.**

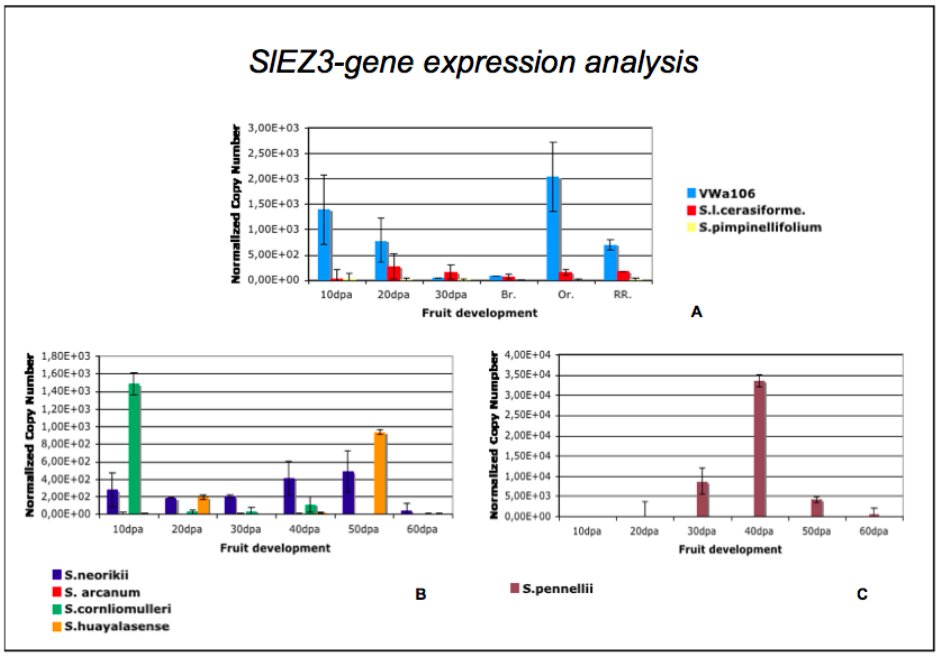

Fig.:56. *SlEZ3* expression levels. *N* red-fruited species (*WVa106* and *S.l. cerasiforme*, A) *SlEZ3* is highly expressed at 10dpa and orange (Or.) stage while in green fruits the maximum expression level was observed at 10dpa (*S. corneliomulleri*) (B) and/or 40 and 50dpa (*S. neorikii*, *S. huayalasense* and *S. pennellii*) (B and C). The normalized copy number is expressed in number of transcripts per microgram of RNA.

On the bases of all this observation we can conclude that SlEZ2 like SlEZ3 protein are highly conserved between cultivated and wild tomato species. In addition we can rehabilitate SlEZ3 protein as a functional gene and not a pseudogene as though (How Kit et al.,2010) with the suggestive hypothesis that SlEZ3 as SlEZ2 and SlEZ1 could play an a role in the fruit development although it showed a basal expression level. In addition the analysis of the E(z) domains between showed that in SlEZ3 the CXC was different compared to SlEZ1, SlEZ2, AtSWN, AtMEDEA and ATCLF suggesting a different evolution process.

4.6 – MOLECULAR CHARACTERIZATION OF A GALADRIEL RETROTRANSPOSON LOCATED IN THE PROMOTER REGION OF *SlEZ2*

Sequence of the SlEZ2 promoter region was retrieved from the SGN data base and analysed as a first step toward its functional analysis (PhD thesis, Lisa Boureau, 2011). In addition, during its thesis work, How Kit in an effor to study the functional regulation of the promoter region of *SlEZ2* made the retrotranscription of this region (PhD thesis,

Alexandre How Kit). Furthermore at that time the genome sequence was not yet completed. The analysis revealed the presence of a retrotransposon, long 5865 nt, located 774bp upstream to the ATG codon of the *SlEZ2* gene. Sequence comparison performed against the Gydb transposon database (http://gydb.org/index.php/Main_Page) revealed that this retrotransposon belongs to the a *Ty3/Gypsy*-like group, and is a member of the Galadriel family. Sequence analysis using the LTR-finder software (Zhao et al., 2007) revealed that the retrotransposon harboured two 458bp-long LTRs with the classical signature at the 5' and 3'ends (TG…CA). At the 5'end the primer binding sites (PBS) could be found as well as the polypurine tract (PPT) at the. The PBS (TTGGTATCAGAGCAAAGGTT) and PPT (TGGGGTGGGGGAGA) sequences are implicated in the retrotranscription. The coding regions for the capsid-related protein (gag) and the polyprotein (pol.), which includes the retrotranscriptase (RT), RNase H (RH), and the integrase (INT) are located between the PBS and the PTT sequences. A typical feature of the Galadriel group is also the presence of a chromodomain (CHR) that can be found in the *SlEZ2* retrotransposon at position SL2.40ch03: 129170001..12922866.

Despite the high similarity between the amino acid sequence of the *SlEZ2* retrotransposon and the amino acid sequence of Galadriel (AF119040) (Fig.57), RT_SlEZ2, when compared to all the amino acid sequences common to *Ty3/Gypsy* plant chromoviruses, appeared more similar to a retrotransposon called Monkey (AF143332) identified in banana, *Musa sp.* (Fig.58).

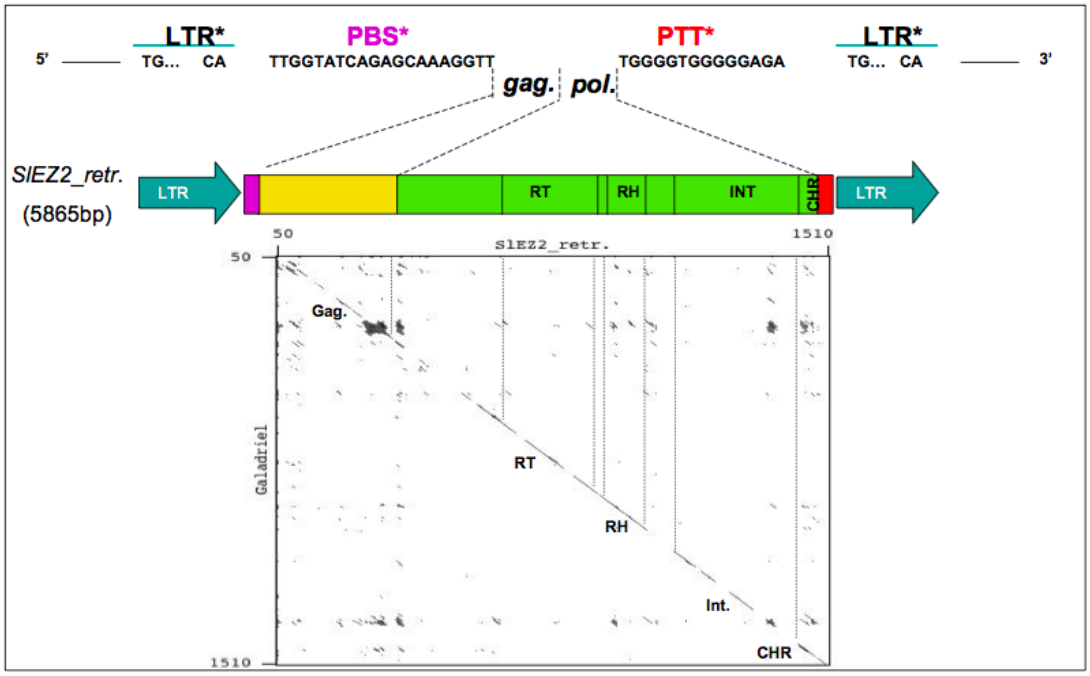

Fig.:57. Dot plot comparison between Galadriel (AF119040) and the *Ty3-Gypsy*-like retrotransposon of *SlEZ2*. The comparison of the two amino acid sequences shows high homology for the retrotranscriptase (RT), RNaseH (RH), integrase (Int) and the chromodomain (CHR) coding regions.

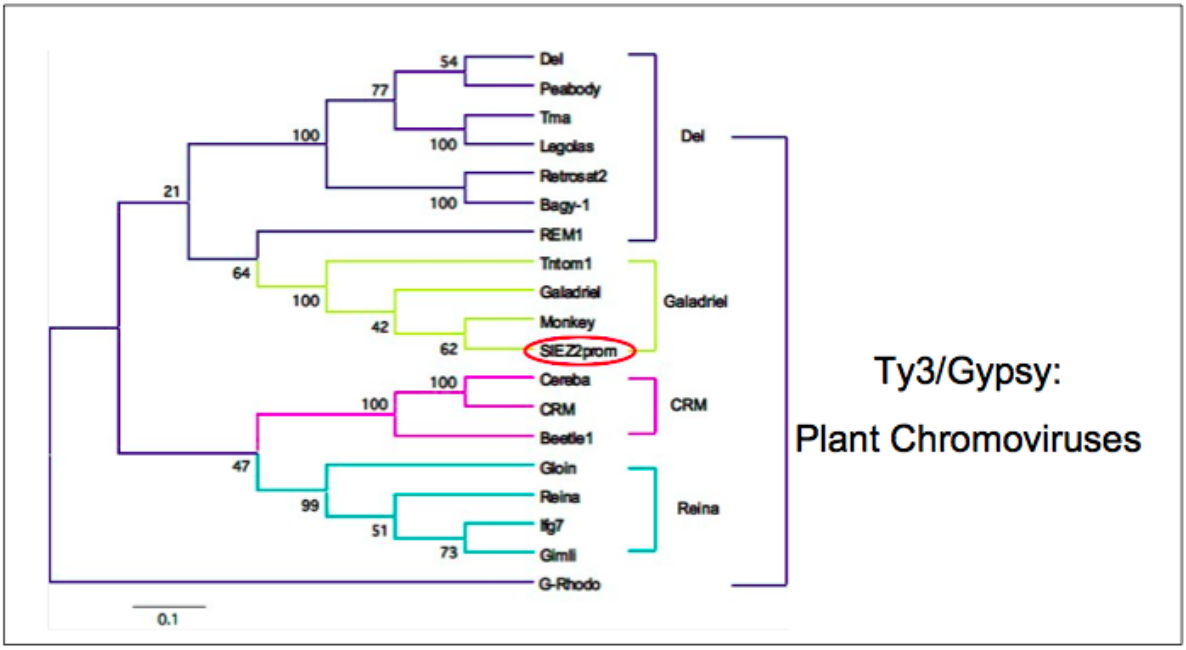

Fig.:58. Phylogenetic tree of the *Ty3/gypsy*: Plant Chromovirus Del (X13886), Peabody (AF083074), Tma (AC005398), Legolas (AC006570), Retrosat2 (AF111709), Bagy-1 (Y14573), REM 1 (29423675), Tntom 1 (AJ508603), Galadriel (AF119040), Monkey (AF143332), Cereba (AF078801), CRM (AY129008), Beetle 1 (AJ539424), Gloin (AC007188), Reina (U69258), Ifg7 (Z11866), Gimli (AL049655), G-Rhodo (114386440). Phylogenetic reconstruction was obtained on the concatenated protein product encoded by the pol internal region common to *Ty3/Gypsy*, family using MEGA 5.0 and the Neighbor joining (NJ) method of phylogenetic reconstruction with 1000 bootstrap replicates under the conditions of uniform rates among sites and pairwise deletion of gaps.

4.7 – RT_SLEZ2 DISTRIBUTION IN CULTIVATED TOMATO AND WILD RELATIVES.

The analysis of the promoter region of *SlEZ2* in wild and cultivated species was performed by PCR amplification, using genomic DNA, to test the presence or the absence of the retrotransposon at this locus. After amplification using *EF1α* primers to check DNA quality, we analyzed the promoter region using 5 primers coupled in different ways (Fig 59A). The first two primer pairs used (3G/3A and 3G/4M) gave a PCR product in each species, indicating that the sequence is sufficiently conserved between species to allow efficient primer binding. However PCR fragment presented various length suggesting that the promoter region is variable in length. To establish the presence of the retrotransposon we used a primer upstream the 3'-LTR and the primer 2M (complementary to the 3G) and the pair 3H/3A (Fig.59).

It is interesting to note that the retrotransposon was detected at the *SlEZ2* locus in species with red cherries only (*WVa106, S.l. cerasiforme,* and *S. pimpinellifolium*) and orange-fruited species as *S. cheesmaniae*. No retrotransposon was present upstream of *SlEZ2* in the green-fruited species (Fig.59).

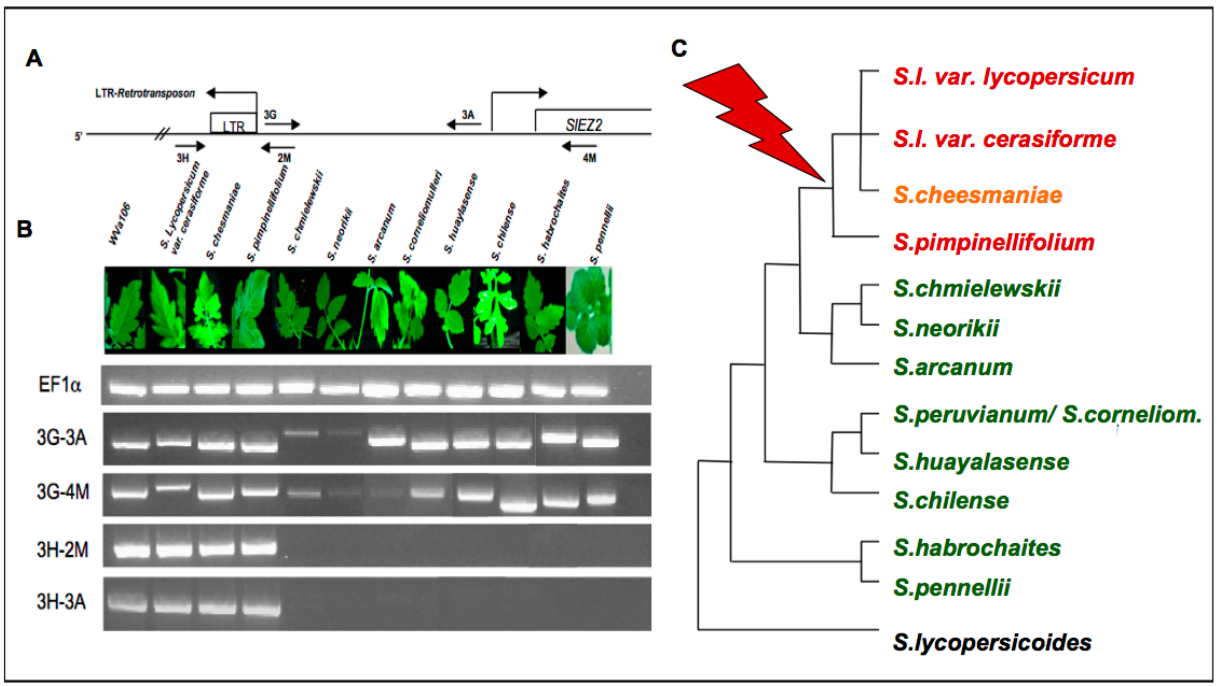

Fig.:59. *SlEZ2* retrotransposon distribution in red- and green-fruited species. In A, a schematic representation of *SlEZ2* promoter region with primers position. In B, the PCR amplification of the promoter region shows the presence of the retrotransposon in red-fruited species only (*WVa106*, *S.l. cerasiforme*, *S. pimpinellifolium*) and orange-fruited species as *S. cheesmaniae*. In C the phylogenetic tree of tomato clade with the insertion of the retrotrasposon in red-fruited species.

We analyzed the genome distribution of the RT_SlEZ2 into the genomes of *WVa106* and of the wild relatives using a strategy based on two different approaches. The strategies chosen were the Southern blot and real time quantitive PCR.

For the hybridizations, we prepared a specific DIG-labelled probe corresponding to the retrotranscriptase (RT) domain. It has already been shown that the amino acid sequences of this domain, even if conserved, show sufficient sequence divergences to use it as a specific probe.

5 micrograms of genomic DNA were extracted from leaves and digested with *Xba*I.

The digested DNA was then separated on a agarose gel, blotted and hybridized with the RT_SLEZ2 probe (Fig.60A). The results showed clearly that *S.l. cerasiforme* and *S. pimpinellifolium* were particularly rich in RT_SlEZ2 elements; *WVa106*, *S. cheesmaniae*, *S. chmielewskii*, *S. arcanum, S. habrochaites* and *S. huayalasense* showed a number of copies of approximately half the one found in the former species, whereas in *S. neorikii, S. corneliomulleri*, *S. chilense* and *S. pennellii* the abundance of RT_SlEZ2 seemed to be much lower (Fig.60B).

These observations have been subsequently confirmed by real-time PCR with primers directed on the conserved region of the RT sequence (Fig.60C). The copy number of the RT_SlEZ2 in the different species has been calculated by real-time quantitative PCR considering 95 picograms the tomato haploid genome (1C). The results showed that *S.l. cerasiforme* and *S. pimpinellifolium* have the maximum copy number (Tab.6 Fig.60) followed by *WVa106*, *S. cheesmaniae* and *S. chmielewskii. S. huayalasense, S. arcanum, S. habrochaites* and *S. neorikii* had a lower number of copies while *S. pennellii, S. corneliomulleri* and *S. chilense* had the lowest value.

Specie	Retrotransposon Copy/genome	±STD. DEV.
WVa106	3,45E+04	1,87E+04
S.l.cerasiforme	6,43E+04	8,68E+03
S.chesmaniae	2,81E+04	3,43E+03
S.pimpinellifolium	5,56E+04	9,49E+03
S.chmielewskii	2,65E+04	4,24E+03
S.neorikii	1,16E+04	2,76E+03
S.arcanum	2,00E+04	5,43E+03
S.corneliomulleri	8,58E+03	1,08E+03
S.huayalasense	2,19E+04	1,75E+03
S.chilense	5,80E+03	1,48E+03
S.habrochaites	1,72E+04	1,43E+03
S.pennellii	9,83E+03	1,71E+03

Tab.6. Copy number of the RT_SlEZ2-like retrotransposon in tomato genomes. The value has been calculated on the basis of a tomato genome of 950Mbp.

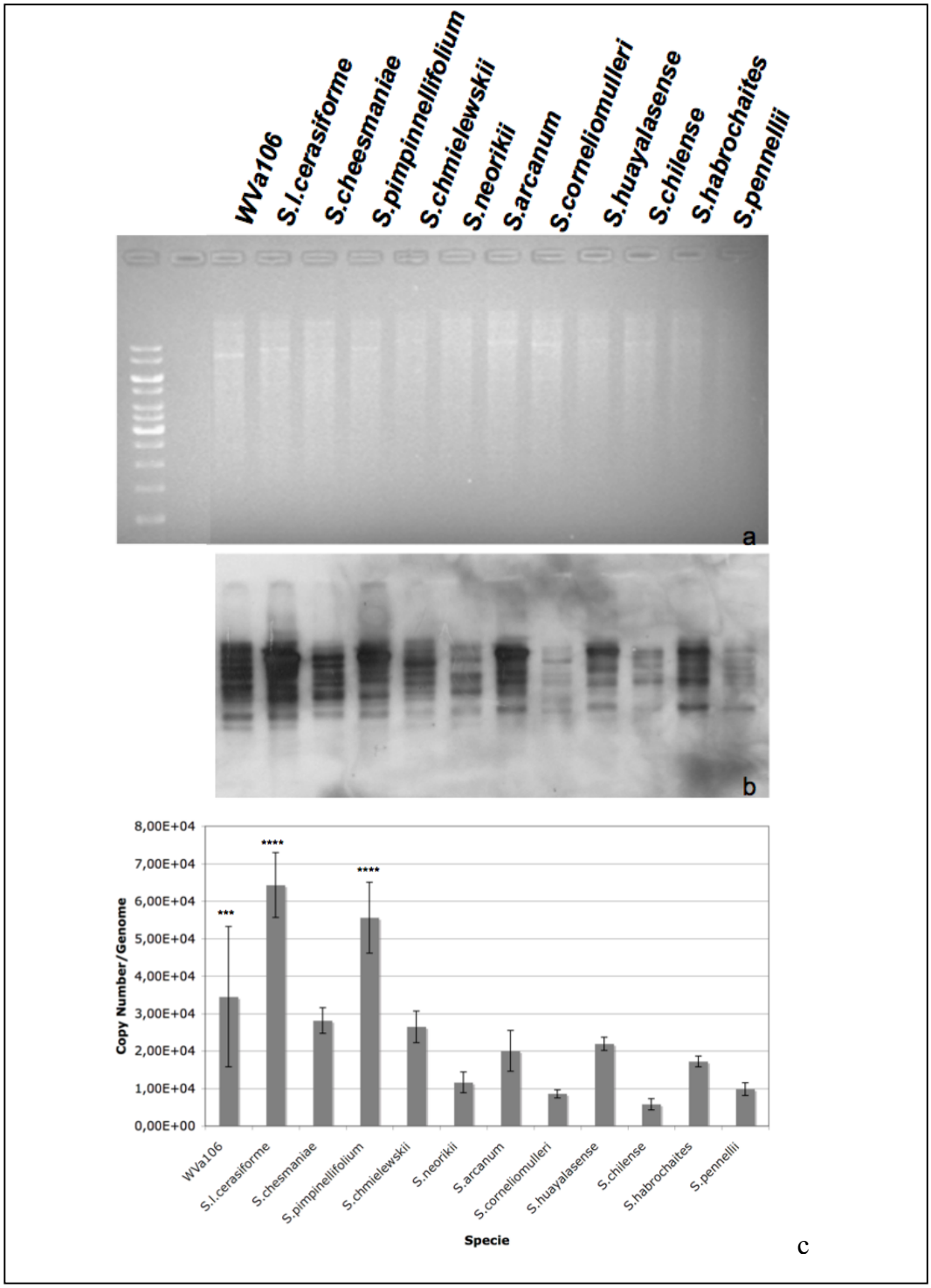

Fig.:60. Comparison of the RT_SlEZ2-like retrotransposon distribution in cultivated and wild tomato species. The distribution of the retrotransposon in the genome of cultivated tomato and wild relatives has been analyzed by Southern blot and quantitative-PCR. Southern blot analysis (b) using a specific probe for the RT domain shows that in red-fruited species the retrotransposon is more abundant as compared to green-fruited species. Real time quantitative-PCR confirmed what observed by Southern blot. The copy number has been expressed in copy for haploid genome (950Mbp). Data are means ± standard deviation. The p-values obtained by the Tukey's test lower than 5, 1, 0,1 and 0,001% are indicated with *, **, * and ****, respectively (n=48).**

4.8 – CAN THE RT_SLEZ2 REGULATE THE ACTIVITY OF SLEZ2?

Although transposable elements (TEs) play an important role in genome evolution, their incorrect mobilization represents a threat for genome integrity. To limit TEs harmful potential, host genomes have developed sophisticated mechanisms to counteract TE activation and to maintain TEs in a silent state. These mechanisms are epigenetic in nature because they do not result from genetic mutation but generate a repressive chromatin environment.

An important role in the repression of transposable elements is played by DNA methylation via the DNA methyltransferases and the RNA-directed DNA methylation (RdDM) (Melanie Rigal and Olivier Mathieu, 2011). In addition to DNA methylation, TEs are associated with various post-translational histone modifications, which are characteristic of a repressive chromatin state. In particular, two epigenetic marks seem to be involved in this silencing: the H3K9me2 and H3K27me1 (Rigal and Mathieu, 2011; Lish 2009).

LTR and non-LTR retrotransposons, as well as DNA transposons, can interfere with nearby genes expression and this effect depends on epigenetic mechanisms triggered by the transposons themselves (Slotkin and Martinsen, 2007). An example include the *agouti-viable yellow* allele in mouse in which the expression of the gene depends on a LTR retrotransposon that act as a promoter for the gene (Blewitt et al.,2005). Nonetheless, transposon insertion into a nearby gene can alter dramatically the effects of methylation on gene expression. For example, the transposons can insulate genes in close regions, while simultaneously recruiting epigenetic modifications that bring the gene under their control with co-transcription of the gene and transposon (Weill and Martinssen, 2008).

After gene expression analysis using real-time quantitative PCR (Fig.55) of *SlEZ2* gene expression in fruits of wild and cultivated tomato species with or without retrotransposon, we analysed the possible effect of the RT_SlEZ2 on *SlEZ2* gene expression

To investigate possible transcriptional read through from the RT element tomato toward the *SlEZ2* coding region, a retrotranscription using random hexamers coupled to PCR

analysis was performed. Furthermore to determine whether the presence of the RT could impact the methylation porfole at the SlEZ2 locus, a methylation analysis of the LTR and promoter region of *SlEZ2* was performed by *Mcr*BC digestion followed by PCR amplification of the regions of interest.

To investigate if the retrotransposon was co-transcribed with the *SlEZ2* gene, we performed a retrotranscription using random hexamers starting from 2 microgram of total RNA extracted from fruits at 10dpa. An identical reaction without the retrotranscriptase was used as a negative control. Random hexamers can anneal to all the transcripts presents even if not processed (Fig.61).

After an amplification using primers specific for the *ACTIN* gene to confirm the correct retrotranscription, we performed several amplification to analyze if the RT_SlEZ2 could be part of *SlEZ2* transcripts and also if other RT_SlEZ2-like retrotransposons were possibily transcribed (Fig.61).

Although the amplification of the promoter region of *SlEZ2* among the cDNAs was always negative it is worth noting that when we used the primers specific for the sequences encoding the retrotranscriptase, the RNaseH and integrase domains we were able to obtain amplification products. In particular, all the species but not *S. corneliomulleri* showed an active transcription of the sequenced involved in the retrotanscription events (Fig.61), while only *S. pimpinellifolium*, *S. neorikii*, *S. huayalasense* and *S. pennellii* showed the transcription of all the coding regions analyzed. In addition, the RT_SlEZ2-like seemed to be very active in the *S. pimpinellifolium* (Fig.61).

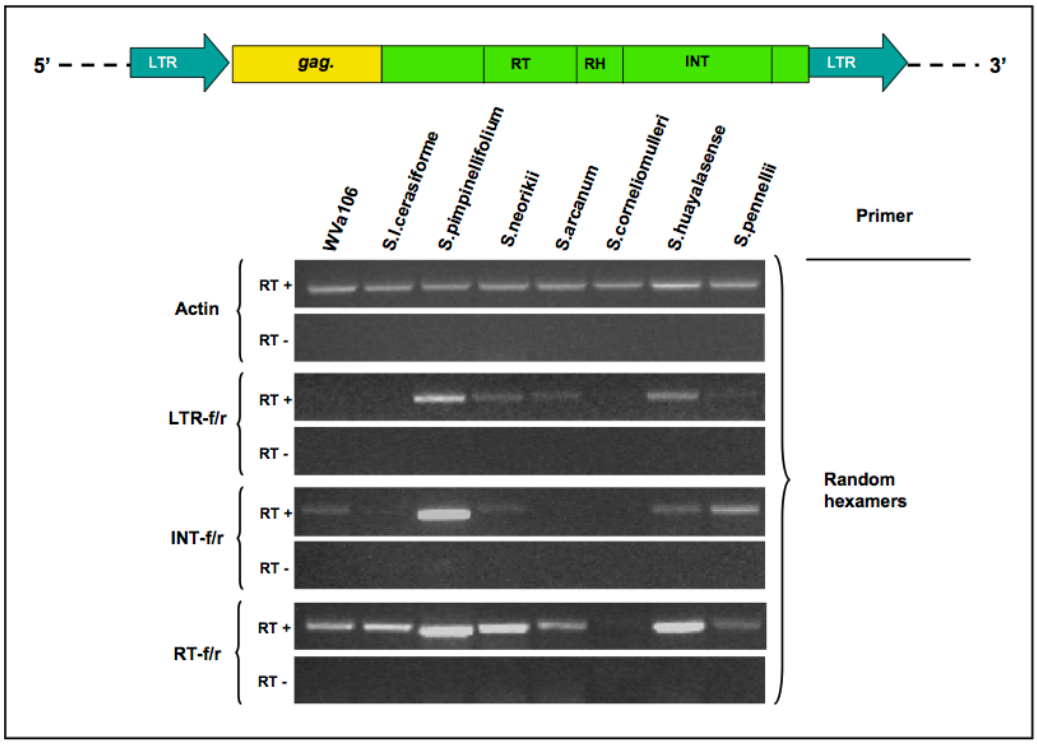

Fig.:61. Retrotransposon RT_SlEZ2-like transcription. RT-PCR using random primer was performed on total RNA extracted from fruits at 10dpa. The primers used were specific for the *ACTIN* gene, and for the *LTR, INTEGRASE* (INT.) and *RETROTRANSCRIPTASE* (RT) domains. The Monkey-like retrotransposons are actively transcribed in all the species but *S. corneliomulleri*.

As already mentioned, the DNA methylation is one way to silence a retrotransposon therefore we analysed the methylation status of *SlEZ2* promoter region and of the upstream retrotransposon via *Mcr*BC digestion and PCR amplification. The *Mcr*BC is a methylation-dependent endonuclease that recognizes and cleaves DNA containing methylcytosines proceeded by a purine ($Pu^{m}C$) on one or both strands.

The analysis will reveal the methylation variations of these sequences during fruit development and between species depending on the retrotransposon insertionwithin the SlEZ2 promoter region.

As shown in Fig 61, in *WVa106* the PCR amplification was more efficient after McrBC treatment of fruit genomic DNA at 10 dpa than at later developmental stages. This is consistent with an increase in the methylation level of the SlEZ2 promoter region during fruit development, observation that is strongly correlated with the reduction of *SlEZ2* gene expression previously analyzed (Fig.62).

A similar situation can be seen in the promoter region of *S. pimpinellifolium* while in *S.l. cerasiforme* the methylation level remained high during all fruit development. Also in these two cases the epigenetic profile is in agreement with the *SlEZ2* expression profile.

The expression levels of *SlEZ2* are, indeed, high during the first phases of development and then decrease in the cultivated tomato (*WVa106*) (Fig.55). On the other hand, in *S.l. cerasiforme*, at 10dpa, *SlEZ2* was less expressed than in *WVa106* at the same stage consistent with the observation that the *SlEZ2* promoter region is more methylated and finally more digested by the enzyme.

S. pimpinellifolium showed an increase of *SlEZ2* expression in the last phases of fruit ripening that are correlated with a demethylation of the promoter region as analyzed by *Mcr*BC.

No significant differences in the methylation levels were observed at the LTR sequence among the species analysed. In all cases, the amplification efficiency of the LTR sequence increases during fruit development and ripening, consistent with a reduction of the methylation level.

As expected, green-fruited species (*S. corneliomulleri* and *S. pennellii*) that do not contain any retrotransposon in the SlEZ2 promoter region, did not show any difference in methylation level during fruit development at this locus. Thus the *SlEZ2* promoter region was only methylated when the RT is inserted, suggesting that this insertion may impact *SlEZ2* gene regulation.

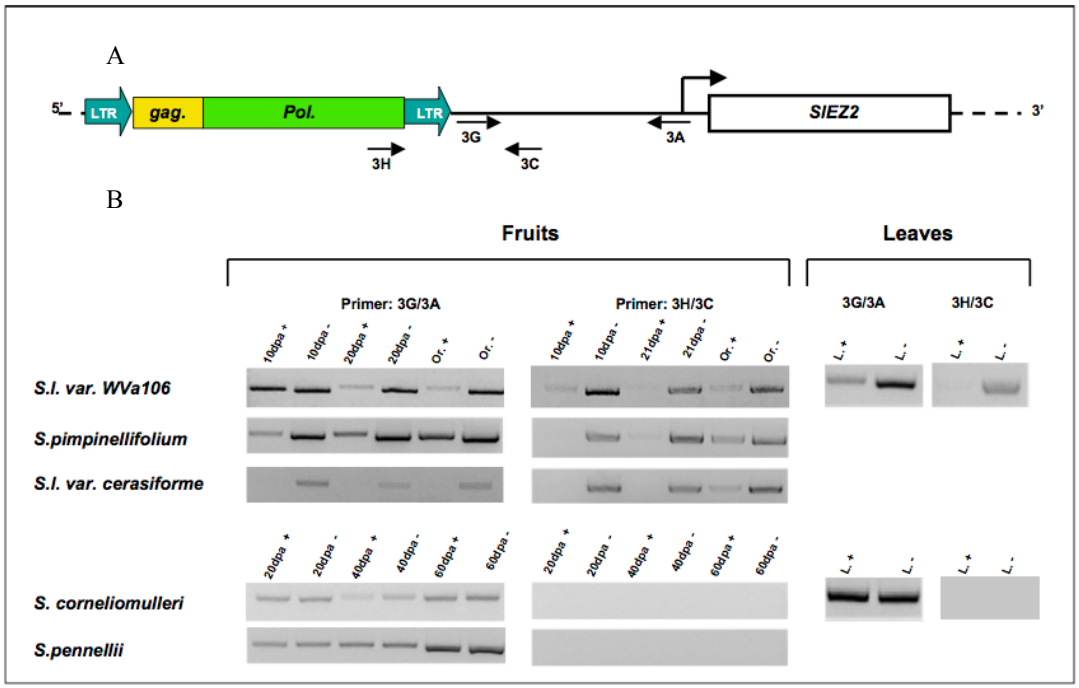

Fig:62. Methylation analysis in the promoter region of *SlEZ2* during fruit development. The analysis was performed on the promoter region of *SlEZ2* and on the LTR of the upstream retrotransposon (A). The sequence specific methylation analysis was performed on fruits at 10, 20 and orange stage for *WVa106*, *S.l. var. cerasiforme* and *S. pimpinellifolium*; for green fruited species, *S. corneliomulleri* and *S. pennellii*, we used genomic DNA from fruits at 20, 40 and 60dpa. In addition, the digestion was also performed on genomic DNA from leaves of *WVa106* and *S. corneliomulleri* as control.

4.9 – DISCUSSION AND CONCLUSION

As recently showed, the tomato proteins of the class Enhancer of zeste are encoded by a multigenic family which includes three members: *SlEZ1, SlEZ2* and *SlEZ3* (How Kit et al., 2010). In previous work, How Kit et al (2010) has shown that *SlEZ1* is orthologuous to the *SWINGER* gene of *Arabidopsis thaliana* (*AtSWN*) while *SlEZ2* and *SlEZ3* are orthologuous to the CURLY-LEAF (*AtCLF*) gene. These genes are therefore paralogue probably arising following a recent duplication of *SlEZ2* from *SlEZ3* (How Kit et al.,2010). In addition *SlEZ3* is likely to be subjected to alternative splicing as we could identify three different mRNA forms, only one of them encoding a 841 aa long EZ protein that present all features of a functional protein. A similar phenomenon has been observed in *Petunia hybrida,* which encodes three *CLF* like genes namely *PhCLF1, PhCLF2,* and *PhCLF3* (Mayama et al., 2003). *PhCLF2* is also subjected to an alternative splicing, which is not tissue and organ specific. *PhCLF2* gives rise to 6 mRNA forms only one of them encoding a functional protein of 922 amino acids (Mayama et al., 2003). We have not evidence that *SlEZ3* alternative splicing is organ specific as all *SlEZ3* mRNA forms were detected in leaves and young fruits. Furthermore, all wild tomato species presented the same phenomenon consistent suggesting that this is an evolutionary conserved trait that was initiated before the speciation of the tomato clade. In all cases a fully sliced mRNA form was identified that encoded a protein with all EZ characteristics. Since a similar situation is also found in *Petunia hybrida* (Mayama et al., 2003), it is probable that the specific characteristic of *SlEZ3* will be found in other solanaceae. A similar situation is found in *Zea mays* which contains three genes encoding enhancer of zeste protein named *MEZ1, MEZ2* and *MEZ3* (Springer et al.,2003). The *MEZ2* gene is ortologuous to *AtSWN* as is the tomato *SlEZ1* gene, is also subjected to alternative splicing. Thus the alternative splicing genes encoding Enhancer of zeste proteins might be a conserved mechanism in plants that may lead to the acquisition of new function for these genes.

Since *SlEZ3* probably encodes a functional protein (see above and part) we analyzed the expression of this gene during fruit development in all cultivated and wild tomato relatives in parallel with the *SlEZ2* gene.

In red-fruited species (excluded *S. pimpinellifolium*) *SlEZ2* appeared in most cases to be highly expressed during the first phases of fruit development at 10dpa (*WVa106*) or before as showed from *S.l. cerasiforme* that at 10dpa had an expression level lower than *WVa106*.

Green-fruited species (*S. neorikii, S. arcanum, S. huayalasense* and *S. pennellii*) showed the highest expression level at 10dpa like WVa106, which decreases during fruit development.

The *SlEZ3* expression level appeared lower than *SlEZ2* during fruit development. Despite its reduced mRNA abundance, *SlEZ3* expression profile in cultivated species was characterized by two peaks at 10dpa (*WVa106*) and orange stage (*WVa106* and *S.l. cerasiforme*). Green-fruited species revealed different expression profiles with maximum expression levels at 50dpa or 40dpa (*S. pennellii*) in or at 10dpa for *S. corneliomulleri*. Finally in *S. arcanum* a low expression level was detected during fruit development.

SlEZ2 promoter region leaded to the identification of a retrotransposon in the promoter region of *SlEZ2*. Its sequence analysis revealed that it is a *Ty3/gypsy*-like retrotransposon, member of the *Galadriel* family (AF119040). A comparison of its amino acid sequence with the other of the *Ty3/gypsy* plant chromovirus showed that it was similar to Monkey (AF143332) a retrotransposon identified in *Musa sp*. The distribution of this mobile element varies dramatically even between closely related species (Kumar and Bennetzen,1999). By PCR using genomic DNA we analyzed the presence of this retrotransposon upstream to the *SlEZ2* promoter region and found that it was present at that particular locus only in red (*WVa106, S.l. cerasiforme* and *S. pimpinellifolium*) and orange-fruited species (*S. cheesmaniae*). This indicates that RT_SLEZ2 get inserted within the SlEZ2 promoter region in a common ancestor of all red-fruited species, as show in Fig.58. The distribution of the RT-SlEZ2-like retrotransposon in the genome showed important differences between species (Fig.60). Species with the higher copy number of RT_SlEZ2-like were the red-fruited species *S.l. cerasiforme*, *S. pimpinellifolium* while *WVa106* is more similar to the green-fruited species *S. cheesmaniae*. The genomic DNA of the green-fruited species contained a lower copy number of RT_SlEZ2 like retrotransposon and the species can be separated in two group. One group contain *S. chmielewskii, S. arcanum, S.huayalasense* and *S. habrochaites* that showed a higher copy number compared to *S. neorikii, S. corneliomulleri, S. chilense* and *S. pennellii*. This observation was then confirmed by real-time quantitative PCR.

Although an approximately agreement, there are some differences between Southern and Q PCR analysis; this is probably due to the difference of specificity of the technique. Although Southern blot analysis use a probe directed to the RT_SlEZ2 retrotransposon the technique uses lower stringency conditions compared to the quantity PCR where the two primes used were directed in the more divergent sequence.

We performed a detailed expression analysis of the *SlEZ2* gene (Fig.55) during fruit development in wild and cultivated tomato species to analyze if the presence of the retrotransposon coul be involved in the regulation of the transciption of the SlEZ2 gene.
A retrotranscriptase using random hexamers and PCR reactions were performed to analyze the possibility that retrotransposon were transcripted with the *SlEZ2* gene. Although no direct link between *SlEZ2* transcription and its retrotransposon was shown, using RT_SlEZ2 specific primers we revealed that some of its coding region was transcribed (Fig.61). We found that the retrotranscriptase domain (RT) was actively transcribed in all the species (excluded *S.corneliomulleri*), moreover *S. pimpinllifolium* showed the highest expression level. In addition all the coding region analyzed were expressed at different levels in the different species. S. pimpinellifolium, S.neorikii and S.huayalasense showed the expression of al the regions analyzed (LTR, Integrase and retrtotranscriptase) while in species as *WVa106* and *S.pennellii* were only transcribed the integrase an retrotranscriptase domain.

The retrotransposons are silenced by epigenetic mechanisms among which RNA-directed DNA methylation has been shown to play a major role. The analysis of the methylation status at the *SlEZ2* locus by *Mcr*BC digestion/ PCR analysis showed differences between species depending on the presence of the retrotransposon. Variations in DNA methylation levels were detected at the *SlEZ2* promoter region in the red-fruited species and they were closely linked to the variations observed in *SlEZ2* gene expression in *WVa106*, *S.l. cerasiforme and S.pimpinellifolium*). Hence, in *S pimpinellifolium*, the slight increase of mRNA level observed during ripening is correlated to a reduced methylation of the SlEZ2 promoter at this stage. On the other hand, in the green-fruited species lacking the retrotransposon there was no change in the methylation of *SlEZ2* promoter region. This indicates that the intrinsic regulation of *SlEZ2* in these species is independent from the DNA methylation process. Inversely in red-fruited species, it is likely that methylation may impact the regulation of this gene expression.

CHAPTER 5

GENERAL CONCLUSIONS AND PERSPECTIVE

Solanum Sect. *Lycopersicon* is a relatively small monophyletic clade that consists of 14 closely related species including the domesticated tomato, *Solanum lycopersicum* (formerly *L. esculentum*).

Tomato and its wild relatives are native of western South America along the coast and high Andes from central Ecuador, through Peru and northern Chile. One species is also found in Mexico and one is endemic to the Galapagos Islands (Nakazato et al, 2010).

The classification and phylogeny of *Solanum* section *Lycopersicon* is a complex issue that has not yet reached a widely accepted consensus. Different works using different approaches based on morphology, gene sequence analysis and metabolomic have tried to characterize wild tomato species (Peralta and Spooner 2007; Spooner, 2005; Zuriaga et al.,2009; Steinhauser M.C. et al., 2010).

Aim of our study is to characterize wild tomato species by a fine characterization of different aspects of fruit developmental process, including fruit size, cytological characterization of pericarp development, ploidy level analysis and gene expression measurement. In addition, since it was already demonstrated that genomic DNA is subjected to tissue specific changes in DNA methylation levels and patterns during fruit development (Teyssier at al, 2008), we have investigated various epigenetic parameters in fruit of wild tomato species: these include DNA methylation analysis and polycomb gene characterization.

Among a set of more than 1160 accessions of wild tomato species (Knapp et al., 2004), we have chosen a representative accession for each species to study the different mechanisms that govern fruit development and ripening.

5.1 – CHARACTERIZATION OF WILD TOMATO SPECIES.

The analysis of fruits from red and green-fruited species using the accession received from the UC Davies revealed morphological and physiological variations in size, colour

and morphology in agreement with the phenotypes already described in Spooner et al. (2005).

The developmental kinetics were very similar in most species, irrespective of the colour of the cherries. Hence there was an approximately linear increase in weight during fruit development except in *S. huayalasense* and *S. pennellii* which were characterized by a bimodal increase in weight, and size, between 10 and 30 dpa and then between 50 and 60dpa. We also found differences in the dynamics of ripening in *S.l. cerasiforme* as compared to the other red species analysed. While *WVa106* and *S. pimpinellifolium* fruits change their colour from green to orange simultaneously in pericarp and locular tissue, *S.l. cerasiforme* showed a locular tissue coloured before the pericarp.

The ripening process in red and green-fruited species has also been analyzed. Although the ripening process in red fruits is characterized by change in colour from green to red, no clear change of colour was visible in green-fruited species fruits. It was therefore necessary to identify other parameters to characterize the ripening process. The analysis of RIPENING INHIBITOR (*RIN)* and PHITOENE SYNTASE (*PSYI*) genes acting upstream and downstream, respectively, of the ethylene-dependent ripening process has helped the comparison of the different species and their different developmental phases. On the basis of *RIN* gene expression profile, we can propose that the breaker, orange and red ripe stages in red fruit could be approximately compared to 40, 50 and 60dpa, respectively, in green-fruited species. This observation can only be partially confirmed by the expression analysis of *PSYI* gene, which is expressed at basal levels in most of the green fruits.

In addition to gene expression analysis other parameters of fruit development and ripening were analyzed. Previous observations on cultivated species showed that the pericarp and its cells are involved in the developmental process (Cheniclet et al., 2005; Bertin et al., 2005).

The characterization of fruit development required also the analysis of the pericarp structure. Clearly, in all species, pericarp increases its thickness in a linear way during fruit development and still contains inner and outer sub-epidermal cell layers. It has been shown that these cells are involved in the determination of the total number of cell layers in pericarp (Cheniclet et al 2005) which is determined at the beginning of the fruit development (Cheniclet et al., 2005 Bertin et al., 2005). Our analysis showed an increase of the pericarp number of cell layers suggesting that the inner and outer sub-epidermal cell layer are in active cell division during the whole fruit development and ripening in all

the species (except *WVa106* and *S. neorikii*). Hence, in these wild species the number of cell layers increases even at late developmental stages, suggesting that the final number of cell layers is not determined during the first days after anthesis as was demonstrated in *WVa106 (Cheniclet et al., 2005; Bertin et al., 2005)*.

Finally, important differences between species were found not only at the morphological level but also in the dynamics of development reflecting the diversity of the tomato clade. These observations suggest that, while in *WVa106* the increase in the pericarp thickness correlates with cell size increase via cell expansion, in the wild tomato species the cell division plays an essential function during the whole process of fruit development. These results have been confirmed by the analysis of the relationship between cell surface and pericarp thickness.

Moreover, the use of specific genes as markers of fruit development timing might represent a useful tool to compare species that do not show the same dynamic of development and ripening. It will be also important to understand how these mechanisms have evolved in the different species.

5.2 – ENDOREDUPLICATION AND DNA METHYLATION VARIATION IN WILD TOMATO SPECIES.

Many works indicate that endoreduplication in plants is connected with fruit size and it is closely correlated to cell growth ad size (Bourdon et al, 2010, Cheniclet et al., 2005; Bertin et al., 2005; Nafati et al., 2010). We analysed this relationship in the wild tomato species by measuring the endoreduplication level in relationship with fruit size, fruit pericarp thickness and cell size. Indeed, max C value in fruits differed between the species under study. Red-fruited species had a maximum C value varying between 128C (*S. pimpinellifolium*) and 256C (*WVa106* and *S.l. cerasiforme*) while green-fruited species had a C value between 64C (*S. neorikii, S. arcanum* and *S. pennellii*) and 128C (*S. corneliomulleri* and *S. huayalasense*). Similar conclusions were obtained when the mean C Value (MCV) was calculated. However, when the endoreduplication index (E.I.), which indicates the average endocycle number per nucleus, was considered we were able to highlight other important differences between species. A linear and positive correlation was found between E.I. and fruit size for almost all the species but not in *S. huayalasense*.

A linear correlation was also found, as expected, between EI and cell size in almost all the species, except *S. pimpinellifolium*, *S. pennellii* and *S. huayalasense* which behaved in a different way. In these cases cell size increase was not correlated to an increase in endoreduplication level. This contrasts with the general observation that the endoreduplication process occurs in cells concomitantly with an increase in cell size (Cheniclet et al., 2005), and may therefore reflect the existence of different control mechanisms. These data would altogether suggest that the variations in cell size and ploidy levels are not so linked as thought and that other factors play an important. Our results on tomato wild species therefore confirmed what was already known in literature on the cultivated ones (*WVa106*), even if we could identify some controversial situations.

We further analysed the DNA methylation profiles in pericarp DNA at repetitive sequences. This analysis showed (as already shown by Teysser et al., 2008) that pericarp and locular tissues are characterized by tissue-specific variations in DNA methylation that were correlated with tissue specific changes in ploidy levels. These results suggested a close relationship between the control of DNA methylation and cell ploidy levels. We therefore investigate variations in methylation in fruits and leaves of the wild species. We did not observe major variations in the global methylation level or locus specific methylation in leaf genomic DNA. However, the methylation profile of fruit pericarp genomic DNA at the 5S rDNA differed between species.

In pericarp tissues the DNA methylation at CNG context changes during fruit development with three different profiles: *WVa106, S. arcanum, S. corneliomulleri, S. huayalasense* showed a low DNA methylation at 20dpa with an increase at breaker stage (or 40dpa in green-fruited species) which was maintained till the orange stage (or 60dpa in green-fruited species); *S.l. cerasiforme* and *S. pimpinellifolium* showed an increase in CNG methylation at 5s DNA during all fruit development while in *S. pennellii* no difference was observed during development.

These groups do not correspond to the different groups defined on the basis of the endoreduplication levels suggesting that DNA methylation at repeated sequences is not correlated with the endoreduplication level.

Finally we compared the global DNA methylation levels with the level of endoreduplication. Even if we did not find a linear correlation between endoreduplication and global methylation it is surprising to note that species with a high C value (*WVa106* and *S.l. cerasiforme*) can have a low level in global methylation as compared to other species that showed a lower C value and a high methylation level.

Our results confirmed what was already known about the cultivated tomato. We also showed that even if the endoreduplication plays an important role in fruit development, its real contribution is far to be clear.

Finally on the basis of these observations we can conclude that the wild tomato species represent an important tool for the analysis of endoreduplication during fruit development and of the epigenetic mechanisms that could be acting.

5.3 – THE STUDY OF SLEZ2 AND SLEZ3 GENES IN CULTIVATED AND WILD TOMATO SPECIES.

As recently showed, the tomato proteins of the Enhancer of zeste class are encoded by a multigenic family which includes three members: *SlEZ1, SlEZ2* and *SlEZ3* (How Kit et al., 2010). In his work, How Kit showed that *SlEZ1* is an orthologue of the *Arabidopsis thaliana SWINGER* gene (*AtSWN*) while *SlEZ2* and *SlEZ3* are both orthologues of the *CURLY-LEAF* gene (*AtCLF*). These two latter genes are therefore paralogues that probably arose following a recent duplication (How Kit et al., 2010).

In addition, this situation has already been observed in *Petunia hybrida (member of the solanacee family),* which contains three CLF genes (*PhCLF1, PhCLF2,* and *PhCLF3*) and this might also apply to other solanaceae such as potato.

We showed for the first time that *SlEZ3* is subjected to alternative splicing, which produces three transcripts, one of which encodes a *SlEZ3* protein 841 aa-long which presents all characteristics of a functional EZ protein. The two other mRNA forms encode truncated proteins lacking the SET domain. We also showed that the alternative splicing mechanism is conserved between the different species analyzed (*WVa106, S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pennellii*) and that the functional domains of the protein are well conserved. Similar analysis made on *SlEZ2* gene showed that it was not subjected to alternative splicing and that the functional domains were conserved as well between species.

In addition, expression analysis of *SlEZ2* and *SlEZ3* during fruit development showed that *SlEZ2* is involved, in almost all the species (except *S. pimpinellifolium* that showed an increase of *SlEZ2* expression during fruit development), in the early phases of fruit development when cell division is predominant while *SlEZ3* is more expressed at 40 and/or 50dpa, even if at very low levels.

Differences between *SlEZ2* and *SlEZ3* were also found at the genomic level even if it has been hypothesized that *SlEZ2* is a duplication of *SlEZ3*. The entire genomic sequence of *SlEZ3* and *SlEZ2* genes in cultivated and wild tomato species (*WVa106, S. pimpinellifolium, S. neorikii, S. arcanum* and *S. pennellii)* revealed that *SlEZ2* is 9.5Kb long and composed of 16 exons while *SlEZ3* is 17.5kb long harbouring 21 exons and a

long intron where we identified a LINE-like retrotransposon. Since *PhCLF1*, orthologue of *SlEZ3*, is also subjected to alternative splicing (Mayama et al.,2003) it could be interesting to analyse if the same genomic organization and regulation are conserved in other solanaceae.

The analysis of the gene sequences did not show important differences between species, except for the presence of a *Ty3/gypsy*-like retrotransposon (member of the Galadriel family) in the promoter region of *SlEZ2* in red/orange-fruited species only (*WVa106, S.l. cerasiforme, S. cheesmainae* and *S. pimpinellifolium*). Furthermore, the retrotransposon of this class is more abundant in the genome of the red-fruited species than in the green-fruited species.

We found that this transposable element has the capacity to code a polyprotein with a chromodomain and it is actively transcribed in all the species (except *S. corneliomulleri*), moreover *S. pimpinellifolium* showed the highest expression level.
The retrotransposons are silenced by epigenetic mechanisms among which RNA-directed DNA methylation has been shown to play a major role (Rigal et al., 2011). The analysis of the methylation status at the *SlEZ2* locus by *Mcr*BC digestion/PCR analysis showed differences between species depending on the presence of the retrotransposon. Red-fruited species showed locus specific variations in DNA methylation in agreement with the expression profile of the SlEZ2 gene. On the other hand green-fruited species without the RT_SlEZ2 retrotransposon did not show a locus specific change in DNA methylation in agreement with gene expression. This indicates that the intrinsic regulation of *SlEZ2* in these species is independent from the DNA methylation process. Inversely in red-fruited species, it is likely that methylation may impact the regulation of this gene expression.

CHAPTER 6

MATERIALS AND METHODS

6.1 - PLANT MATERIAL AND GROWTH CONDITIONS.

Tomato plants *Solanum lycopersicum* cv.Weat Verginia a 106 (WVa106) and wild relatives were grown in a greenhouse during the spring season with a photoperiod of 12.5h under a minimum of 500lux. In average the temperature was between 23 and 26°C during the day and between 18 and 19°C during the night. The humidity approximately to 70%.

Representative accession for each species was provided by C.M. Rick Tomato Genetics Resource Center (TGRC, http://tgrc.ucdavis.edu/) of the University of Davis, California. These include: *WVa106* (cherry tomato), *S.l. var. cerasiforme* (LA1226), *S. cheesmaniae* (LA0930), *S. pimpinellifolium* (LA0722), *S. chmielewskii* (LA1330), *S. neorikii* (LA1326), *S. arcanum* (LA2152), *S. corneliomullieri* (LA0103), *S. huayalasense* (LA1982), *S. chilense* (LA1930), *S. habrochaites* (LA1353), *S. pennellii* (LA1926).

6.2 - NUCLEIC ACIDS PURIFICATION AND ANALYSIS.

6.2.1 - Genomic DNA extraction.

6.2.1.1 - Genomic DNA extraction from tomato leaves using hexadecyltrimethylammonium bromide (CTAB).

Leaves tissue was grinded in liquid nitrogen using a mortar and pestle.
One gram of leaf powder was dissolved in 5ml of CTAB buffer (2% hexadecyltrimethylammonium bromide (CTAB), 100mM Tris-HCl pH 8.0, NaCl 1.4M,

ethylenediaminetetraacetic acid (EDTA) 20mM, 2-mercaptoethanol 0.2% v/v) preheated at 60°C.
The mix was incubated at 60°C for 30 minutes, than an equal volume of chlorophorm-isoamyl alcohol (24:1 v/v) was added and mixed gently for 30 minutes.
After centrifugation for 15 minutes to 1300g the upper phase containing the genomic DNA extracted was transferred to a new tube and an equal volume of chloroform isoamyl-alcohol was added to remove all the contaminants.
The samples were gently mixed for 30 minutes and, after a new step of centrifugation, the aqueous upper phase was transferred into a new tube where 0.7 vol. of cold isopropanol were added to precipitate nucleic acid.
After a new step of centrifugation the isopropanol was removed and the pellet formed was washed several times with ethanol 75% v/v.
Finally the pellet was dried at room temperature and resuspended in a suitable volume of TE buffer (10mM Tris-HCl pH 7.4, 1mM EDTA, pH 8.0).

To remove RNA contaminations, an RNA digestion was performed at 37°C for 30 minutes using RNaseA (Sigma, prod. R5125) to a final concentration of 10 μg/ml.
The enzyme was removed by DNA precipitation with 1/10 volume of sodium acetate 3M and 2.5 volume of absolute ethanol.
Finally the pellet was washed in ethanol 75% v/v and dried to room temperature until its resuspension in a suitable volume of TE buffer.

6.2.2.2 - Genomic DNA extraction from tomato fruit pericarp.

Cherry tomatoes during their developmental phases increase the amount of many different metabolic components, which play an important role in ripening: polysaccharides represent one of them.
While most plant DNA extraction techniques are effective in removing proteins they are usually less successful with polysaccharides.
Furthermore, polysaccharides are very common contaminants in plant nucleic extracts and they often interfere with DNA precipitation, generating “slimy” DNA pellets difficult to handle. They could also have a negative effect in several steps such as enzymatic digestions, PCR reactions, etc.

To avoid and/or reduce the presence of polysaccharides genomic DNA preparations, the kit "Nucleon Phytopure Genomic DNA Extraction Kit" (GE Healthcare, RPN8510, RPN8511) was employed. This kit uses a resin whose particles contain free boric acid groups (-B(OH)2) able to bind the polysaccharides yielding cyclic boric acid esters and therefore removing them from the sample.
To 0.1g (fresh weigh) of plant tissue grinded in liquid nitrogen was added a cell lysis solution with potassium/SDS and the DNA was than extracted with Nucleon PhytoPure resin and chloroform. The genomic DNA was precipitated and washed in ethanole (70%v/v).
A RNA digestion was performed to remove RNA contaminants.

The quantity and quality of genomic DNA extraction was measured using a NanoVue spectrophotometer at 260nm and 280nm and on agarose gel by electrophoresis.

6.3 - PLASMID DNA EXTRACTION.

The plasmid DNA extraction was performed using the PureYield™ PLASMID Miniprep System (Promega cat.#A1223) from a bacterial colture (5ml) grown in Luria Bertani broth added with the selective antibiotic.

6.4 - RNA EXTRACTION.

6.4.1 - Total RNA extraction from leaves and tomato pericarp.

Total RNA extraction was performed using TRI Reagent® RNA Isolation Reagent (Sigma-Aldrich T9424).
100 mg of leaf tissue were grinded in liquid nitrogen and transferred to a microcentrifuge tube. 1ml of TRI Reagent® was added to each sample, centrifuged at 12000g at 4°C for 10 minutes and the upper phase collected into a new tube. 200µl of chloroform/isoamyl alcohol were then added.

The samples were than mixed using a vortex, kept at room temperature for 5 minutes and subsequently centrifuged at 12000g for 15 minutes at room temperature.
After this centrifugation three phases can be seen: a lower phase with all the proteins, an intermediate phase, which contains the DNA, and an upper phase where all the RNAs are present. 400µl of this phase were transferred into a new microcentrifuge tube and then an equal volume of isopropanol was added. After a gently mix by inversion, the samples were kept at room temperature for 20 minute and then centrifuged at 12000g for 10 minutes.
The pellet was washed 3 times with 500µl of ethanol 75% v/v in DEPC water, dried for 5-10 minutes and resuspended in 50µl of DEPC water.
2µl of RNA were then used for a quick and accurate quantification of nucleic acids and protein using a NanoVue spectrophotometer.

A DNase treatment was performed on the purified RNA to remove any genomic DNA contamination from the samples. A Turbo™ DNase *free* Kit (Ambion) was used in accordance with the manufacturer's instructions.
The absence of genomic DNA contamination was confirmed by PCR.

6.5 - QUALITY AND QUANTITY CONTROL OF THE NUCLEIC ACID.

All the nucleic acid (gDNA, Plasmid and RNA) extracted were quantified at 260nm and 280nm using the NanoVue spectrophotometer (GE healthcare) and then checked on agarose gel by electrophoresis.

6.6 - METHYLATION ASSAYS ON gDNA.

6.6.1 - Methyl-sensitive digestions using McrBC endonuclease.

McrBC (New England Biolabs®, M0272L) is a tool for determining the methylation state of CpG dinucleotides.

It is a methylation-dependent endonuclease from *Escherichia coli* K-12 encoded by two genes: *mcrB* and *mcrC*. It recognizes and cleaves DNA containing methylcytosine (5-methylcytosine, 5-hydroxymethylcytosine or N4-methylcytosine) preceded by a purine (Pu^mC) on one or both strands.

The very short half-site consensus sequence (Pu^mC) allows a large proportion of the methylcytosines present to be detected, even in DNA, which is not heavily methylated.

McrBC detects a high proportion of methylated CpGs but it does not recognize *Hpa*II/*Msp*I sites (CCGG) in which the internal cytosine is methylated.

The DNA digestion was done in a final volume of 50 µl using NEBuffer 2 1x (50 mM NaCl, 10 mM Tris-HCL, 10 mM $MgCL_2$, 1 mM DTT) supplemented with 100 µg/ml BSA, 1mM GTP and 20u of enzyme for 1µg of genomic DNA. The reactions were incubated at 37°C for 2 hours. The enzyme is then inactivated by incubation at 65°C for 20 minutes.

6.6.2 - Methyl-sensitive digestion using HpaII/MspI endonuclease.

This is a classical method for methylation analysis based on the property of some restriction enzymes to cut or not cut methylated DNA.

In this study two classical enzymes were used: the isoschizomer *Hpa*II and *Msp*I which both recognize the sequence CCGG. When the external C in the sequence CCGG is methylated, *Msp*I and *Hpa*II cannot cleave. When the internal C residue is methylated, only *Msp*I can cleave the sequence whereas *Hpa*II does not cut.

6.6.3 - Methylated DNA quantification.

The global DNA methylation analysis was performed using the "Imprint® Methylated DNA quantification kit" (Sigma, Catalog Number. MDQ1), in accordance with the manufacturer's instructions. The quantification strategy is based on an ELISA format: the genomic DNA is bound to the wells of the plate and then a first antibody is used to detect the methyl-cytosines. This first antibody is then detected by a second antibody conjugated to an enzyme able to convert an added substrate to a product, which is read at 450nm.

6.7 - RETROTRANSCRIPTION OF RNA TO cDNA.

In this study, the Moloney Murine Leucemia Virus Reverse Transcriptase (M-MLV RT (H-)) was used. This enzyme is a RNA-dependent DNA polymerase that can be used in cDNA synthesis with long RNA templates (>5kb). Moreover the enzyme lacks the RNase H activity that can start to degrade RNA templates when the incubation times are too long, as they may when making long cDNA.

The reaction uses 1 or 2μg of total RNA in a final volume of 30μl.

In the first step the RNA was added with 0,5μl of specific primer or poly$[dT]_{18}$ primers to a final concentration of 1,6μM, in DEPC water at the final volume of 17μl.

This solution was then incubated for 5 minutes at 70°C and then cooled on ice, these passages are necessary to melt secondary structures within the template and to prevent them from reforming.

In each sample 10,5 μl were added of a solution containing: M-MLV buffer 1x, DTT 1,6mM, dNTPs 1,5mM each, RNase Ribonuclease Inhibitor (Promega, catalog.# N2511) 0,8 U/μl and water for the total volume of 30μl.

The samples were then incubated 2 minutes at 42°C and then 2μl of M-MLV reverse transcriptase was added. Retrotranscription was performed at 42°C for 1 hour, the inactivation of the enzyme was done at 80°C for 5 minutes, than the samples could be used or stored at -20°C.

6.8 - GENETIC AMPLIFICATION USING POLYMERASE CHAIN REACTION (PCR)

The polymerase chain reaction (PCR) is a technique used to increase the copy number of DNA target using two start points called primer.

It includes a denaturation phase where the double stranded DNA (dsDNA) is heated until 95-98°C and converted to single stranded DNA (ssDNA); a second phase called primers annealing where the primers bind the ssDNA template at the homologous sequences, the third phase is an extension phase where the polymerase starts the new dsDNA synthesis from the primes.

6.8.1 - The DNA template.

The reaction of PCR was performed on genomic DNA after extraction or directly on single bacterial colony.
When possible, the DNA was quantified and diluted, before the reaction, to 10-100 ng/µl.

6.8.2 - Reaction conditions and thermal cycles.

6.8.2.1 - Reaction of PCR using Taq Polymerase

The reaction of PCR were performed using GoTaq® Flexi DNA Polymerase (Promega, REF. M8305) in a final volume of 50µl (1x buffer, 2mM $MgCl_2$, dNTPs 200µM each, 0.5µM of primer forward and reverse, 1.25U of GoTaq® DNA Polymerase)
The thermal cycles included an initial denaturation to 95°C for 2 minutes followed from 25-35 cycles including a denaturation step to 95°C for 30 seconds, 30 seconds at the annealing temperature specific for the primers and an extension time proportional to the length of the fragment amplified and the rate of polymerization of the enzyme. Finally an extension step at 72°C for 3-5 minutes.

6.8.2.2 – Reaction of PCR using Taq Fidelity

The amplification reaction were performed using the Phusion® High-Fidelity DNA Polymerase (Fynnzymes, F-530L) in a final volume of 50µl (1x HF buffer, dNTPs 200µM each, 0.5µM of primer forward and reverse, 0.02 U/µl of Phusion® High-Fidelity DNA Polymerase).
The thermal cycles included an initial denaturation at 98°C for 30 seconds followed by 25-35 cycles including a denaturation step at 98°C for 10 seconds, 30 seconds at the annealing temperature specific for the primers (measured on the Finnzymes website: http://www.finnzymes.com/tm_determination.html) and an extension time proportional to the length of the fragment amplified and the rate of polymerization of the enzyme. Finally, an extension step at 72°C for 3-10 minutes.

6.9 - REAL-TIME QUANTITATIVE REVERSE-TRANSCRIPTION PCR (qRT-PCR)

Gene expression analysis by qRT-PRC has been chosen for high-throughput and accurate expression profiling of selected genes.

There are two different methods for analyzing data from real-time: absolute and relative quantification. Absolute quantification determines the input copy numbers of the transcript of interest by relating the PCR signal to a standard curve. Relative quantification describes the change in expression of the target gene relative to some reference group such as an untreated control or a sample at time zero in time-course study. In relative quantifications the results are expressed using the $2^{-\Delta\Delta t}$ *(Livak J. et al. 2001)*.

In both of cases an internal housekeeping gene must be used, whose expression profile does not change during the study. This gene is used to normalize the expression data of the target gene.

The amplification of cDNA was performed using the iQ™ SYBR® Green Supermix (Bio-Rad 170-8880) in a final volume of 20µl (1x iQ™ SYBR® Green Supermix, 0,2 µM primer forward and reverse and water to a final volume). Reactions were run in a Bio-Rad CFX-96 system thermalcycler.

The absolute quantification of the transcripts of interest was done by relating the PCR signal to a standard curve after normalization on actin expression (Perikless Simon, 2003).

7.0 – ELECTROPHORESIS.

Nucleic acid and PCR products were analyzed on agarose gel (Euromedex) at a concentration between 0.8 and 2% (w/v) diluted in TAE buffer 0.5x (20mM Tris, 35mM acetic acid, 0.25M EDTA, pH 8.0). The voltage applied was between 25 and 100 volt. Gel green™ (FluoProbes®) was added to the gel to visualize DNA.

The samples were mixed with a loading buffer 1x (Promega, Blue/Orange 6x Loading Dye, G190A).
Amplification fragments were then detected under exposure to UV rays using a Bio-Rad Gel Doc 2000.

7.1 – CLONING.

7.1.1 - Classical molecular cloning

A classic molecular cloning allows the insertion of a DNA fragment into a cloning vector and then its multiplication inside host bacteria. Subsequently the vector containing the DNA fragment can be extracted and used for molecular analysis.

7.1.1.1 - Cloning into pGEM-Teasy Vector.

The pGEM®-T Easy vector system (Promega, cat.#A1360) is a convenient system for the cloning of PCR products.
The PCR fragments were directly purified after PCR or after excision of the amplified band from agarose gel using the kit Wizard® SV Gel and PCR Clean-Up System (Promega cat.#A9281) and then quantified at the NanoVue.
The suitable quantity (ng) of product to use during the ligation reaction was estimated in accordance with the manufacturer's instructions.
The ligation was performed in 10µl (1x ligation buffer, 50ng of vector, 3U of T4 DNA ligase with a suitable quantity of insert, previously measured) overnight at 4°C or 1 hour at 25°C.

7.1.1.2 - Preparation of thermocompetent cells of *Escherichia coli (E. coli*).

To obtain thermocompetent cells (*E. coli DH5α F'*) a bacterial colture was grown overnight at 37°C in 5ml of liquid LB broth (1% tryptone,0.5% yeast extract, 1% NaCl, pH 7.0).

The day after the overnight colture was inoculated into a new liquid LB broth (1/10 v/v) and incubated at 37°C to obtain a spectrophotometric measure of the $OD_{600\ nm}$ close to 0.5-0.6.

Subsequently, all the steps were done on ice and /or at 4°C.

The bacterial colture was centrifuged at 750 g for 10 minutes at 4°C and the pellet was gently resuspended in 2 ml of cold $CaCl_2$ 0.1M. It was incubated on ice for 30 minutes, centrifuged a second time at 4°C for 5 minutes at 4000rpm (g?).

The new pellet formed was resuspended in 800µl of cold $CaCl_2$ 0.1M, glycerol 10% (v/v) and divided in 50µl aliquots that were immediately frozen in liquid nitrogen and stored at -80°C.

7.1.1.3 - Transformation of *E. coli* cell's.

DH5α thermocompetent cells were defrosted on ice and then few microliter of plasmid or 10µl of ligation mix were added and incubated 30 minutes on ice. A heat-shock step was done at 42°C for 35-40 seconds, followed by an incubation on ice for 2 minutes. 250µl of SOC (2% bacto-triptone, 0.5% yeast extract, 10mM $NaCl_2$, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose) were then added and the mix incubated at 37°C for 1,5 hours. 100µl were distributed on a plate containing solid LB (1% tryptone, 0.5% yeast extract, 1% NaCl, 3.75g of agar, pH 7.0) and the antibiotic necessary for the selection. In this study the pGEM T-easy vector was used, thus the selection after transformation for blue/white colonies was obtained with ampicillin (100mg/ml), IPTG (100mM) and X-GAL (40 mg/ml). The plates were then incubated overnight in a thermostat at 37°C.

7.2 - DNA SEQUENCING.

The nucleic acid from plasmid and/or PCR products were sequenced using Beckman Coulter Genomics services.

7.3- SOUTHERN BLOT ANALYSIS

7.3.1 Transfer of digested DNA from agarose gel to a nylon membrane.

The DNA blot is a technique that allows the binding of DNA to a nylon membrane from an agarose gel by capillarity transfer. After the separation of the digested DNA the agarose gel is firstly subjected to a depurination step where the gel was completely covered with a depurination solution (0.125M HCl) for 10-20 minutes and gently agitated. It was then rinsed in distilled water for 10 seconds and incubated for 30 minutes with gentle agitation with a denaturating solution (1.5M NaCl, 0.5M NaOH). After a brief rinse with distilled water, the gel was submerged in neutralization buffer (1.5M NaCl, 0.5M Tris pH 7.5) for two times 15 minutes. The capillarity blot was set overnight with SSC 20x solution (3M NaCl, 0.3M sodium citrate pH 7). The DNA was transferred on a nylon membrane Hybond-N+ (amersham). The day after, the nylon membrane was dried and exposed to UV light for 3minutes to fix the DNA.

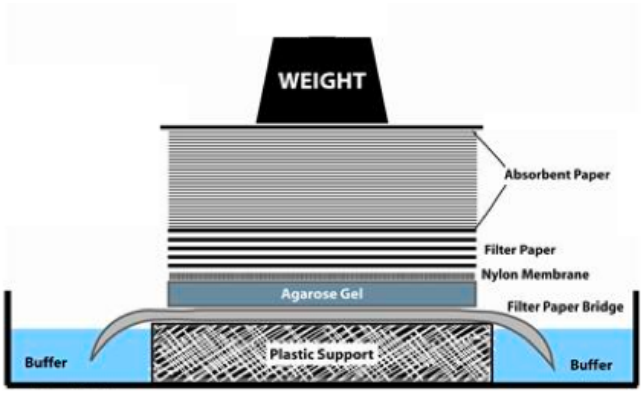

Fig.:63.Southern blot.

7.3.2 - Probe preparation.

The probes used for the hybridization of the Southern blot were labeled with Digoxigenin in a final volume of 20µl [1x buffer, 0.25µM of each primer, 0.025mM of DIG-UTP, 0.17mM TTP, 0.2mM GAC, 0.2U REDTaq® DNA polymerase (Sigma D4309-50UN)]. The probe was precipitated with 1/10 (v/v) of LiCl 4M and 3 volumes of absolute ethanol. After centrifugation at 15000g for 30 minutes, the pellet was washed 3 times with 75% (v/v) ethanol, dried and resuspended in 20µl of distilled water. Before use, the probe was boiled for 2 minutes and then diluted in the hybridization buffer [50% formamide, 2% of blocking reagent (caseine), 0.02M maleic acid, 0.03M NaCl, 5x SSC, 0.2% SDS).

7.3.3 - Hybridization of probe to target.

Nylon membranes were hybridized with specific probes overnight to 42°C and subsequently washed two times 5 minutes with 2x SSC and 0.1% SDS and then two times15 minutes with 0.2x SSC and 0.1% SDS preheated to 68°C.

7.3.4 - Immunological detection of DIG-labeled nucleic acids.

The immunological detection of the probe was done using an anti-digoxigenin antibody conjugate to the alkaline phosphatase which is able to dephosphorylate the chemiluminescent substrate CSPD (Roche Cat. No. 1 755 633).
Enzymatic dephosphorylation leads to the metastable phenolate anion which decomposes and emits light at a maximum wavelength of 477nm. The luminescent light emission is recorded on an X-ray film.
Immunological detection was done in accordance with the manufacturer's instructions (Roche Cat. No. 1 755 633).

7.4 - MORPHOLOGICAL AND PHYSIOLOGICAL ANALYSIS OF FRUIT.

Tomato fruits from different species were harvested at different developmental stages and immediately measured in weight and size.
Pictures were also taken of the whole fruit and of the internal structure of the fruit to show the ripening grade and evolution during the different phases.

7.5 - CYTOLOGICAL ANALYSIS.

7.5.1 - ***Analysis of the ploidy levels.***

The flow cytometry is a technique that can analyze the ploidy levels of the cell nuclei and measure their DNA content.
In this study the flow cytometer was used to analyze the ploidy levels of the pericarp cells of the tomato fruit.
Fruits from the different species were harvested at different developmental stages, the locular tissue removed and the pericarp reduced in small fragments using a scalpel blade. The pericarp fragments were immersed in 500µl of DAPI (Partec), which is a fluorescent dye (excitation at 344nm and emission at 466nm) that binds the DNA. The mix was then filtered in a specific filter with a diameter of 100µm and cells collected into a specific tube for the PARTEC Ploidy Analyzer (Partec-GmbH) which is a flow cytometer especially developed for ploidy determination in plants.
The ploidy analyzer measures the fluorescence of the DAPI, which is directly proportional to the DNA content of the cell.
For cell counting the sample-volume-detector measures exactly 200µl of the sample. Each fluorescent cell in this volume is counted and the resulting histogram shows how many cells are measured for each quantity class.

7.5.2 - Analysis of histogram of ploidy levels.

The data from the cytometer were analyzed using FlowMax® software (Partec, GmbH) , converted into an image and then analyzed using Image-Pro Plus (Media Cybernetics).
With this procedure the area of each peak was transformed in a quantity of nuclei at the different ploidy levels present in each class.
Finally the number of nuclei for each class was expressed as percentage of the total nuclei for each stage of fruit development. Three fruits were measured for each stage.

7.5.3 - Analysis of tomato fruit pericarp.

Tomatoes fruit pericarp analysis was performed using a Leica MZFLIII stereomicroscope carrying a DC300F-Imaging camera.

Observations and pictures were taken on 3 fruits for each developmental stage for all the species. To highlight pericarp structure Toluidine blue (Sigma, T3260- Technical grade) 0,5% (v/v) for 30-60 second was used as dye. The pictures were then analyzed using Image-Pro Plus® (Media Cybernetics) to measure pericarp thickness, number of cells layer and to estimate the average cell size.

7.6- STATISTICAL ANALYSIS

7.6.1-Statistical Analysis.

The statistical analysis of all the observations was done using a Tukey's HSD (Honestly Significant Difference) test.
Tukey's multiple comparison test is one of several tests (Sheffe's test or Dunnett's test) that can be used to determine which mean among a set of means differs from the rest.
In fact, when we have more than two groups it is inappropriate to simply compare each pair using a t-test.
The correct way to do the analysis is to use a one-way analysis of variance (ANOVA) to evaluate whether there is any evidence that the means of the population differ. If this evidence exists the Tukey's test can investigate which mean is different by comparing the difference between each pair of means with appropriate adjustment for the multiple testing.
The Tukey's multiple comparison test, like the ANOVA, assumes that the data from each group have a normal distribution and that each group has the same standard deviation.
The ANOVA and the Tukey' test were done using Excel stat.

7.7 - BIOINFORMATIC ANALYSIS

7.7.1 - Alignment of sequences and dendrograms

7.7.1.1 - MultAlin.

MultAlin (http://multalin.toulouse.inra.fr/multalin/) is a free and versatile web software for multiple sequence alignment of protein or nucleic acid.

7.7.1.2 - Blast sequences.

The Basic Local Alignment Search Tool (BLAST) allows a fast research sequence inside sequence databases and it finds regions of similarity between biological sequences.
In this study the BLAST used was present on the web site of the Sol genomics network (http://solgenomics.net/tools/blast/index.pl) and of the National Center for Biotechnology Information (http://blast.ncbi.nlm.nih.gov/).

7.7.1.3 - Gepard

The alignment between two similar sequences was performed using the Gepard free software (http://www.helmholtz-muenchen.de/en/mips/services/analysis-tools/index.html) Gepard is a rapid and sensitive tool for creating dot plots on genome scale (Krumsek et al. 2007)

7.7.2 - Phylogenetic trees.

7.7.2.1 - MEGA5 software.

MEGA is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining web-based databases, estimating rates of molecular evolution, inferring ancestral sequences, and testing evolutionary hypotheses (Tamura et al. 2011) .

7.8 - PRIMER LIST.

7.8.1 - Primer used for the analysis of the alternative splicing in SlEZ3 transcripts.

Name	Sequence
EZ3P1	GGTGCTGCCCTTATCTCCG
EZ3P2	CCATTGATCTTCCAGTAAAGTTC
EZ3P3	CAACCATAGAAGCTGATGAATTC
EZ3P4	GTAAAGGTGTGTTTTATGAATTG
EZ3P5	AAAGGCGTATGATTTGGGGG
EZ3P6	TTGGATAGGTATGTTAAGG
EZ3P7	TCAGCAGATGCCCGTGAATC
EZ3P9	TTCAACTATCCAACAAATTGGCC
EZ3P10	GTGAAGTCAAGGTAAACATTTCG
EZ3P11	GAACGAGAACATCCAAGGG
EZ3P12	CTGAAAAACAATTGCCATGGTGC
EZ3P13	GCTATCGCCTGGTATCTTG
EZ3P14	GGAACACAAAGCTGGATTGACC
EZ3P15	CGCAGCAGGTCAGTTAACTTC
EZ3P16	CTGTATGATTGCTCGAAATCTG
EZ3P18	GAAGAGGTAGAGTTCGTCGCTT
EZ3P19	GGTGTAGAAGCAGACAATG
EZ3P20	GAGTCTCGATATTCCTCCAC
EZ3P21	GATCTGATGTGTCTGGCTGG
EZ3R1	CAACTTTACTGGAAGATCAATGG
EZ3R2	TATTGGATCTAACATAGACAAACC
EZ3R3	GTGGGTGCCTCGGGCTTCC
EZ3R4	TTTGACGTTGTTTGAGAAGAAGC
EZ3R5	CTGATCCAACAATTTCGGCA
EZ3R6	CACTTACTAGTGATAGCTTATTCTC
EZ3R7	CATTCTCCCTCAATGTGTCGTCAT
EZ3R8	ATCCAGCTTTGTGTTCCTTCTTG
EZ3R9	CCTTGACTTCACTGGGTTTTC
EZ3R10	CAGCATAACATCTTCGTACTC
EZ3R11	GTATATTGAGGCAGTCTTTCC
EZ3R12	TAAGAATTCATCAGCTTCTATGG

7.8.2 - Primer used for sequencing of SlEZ3 gene

Name	Sequence
EZ3-1ATG	GATTCCTCTCACTCTCTC
EZ3-2ATG	AAGCCGCAGCGGTGGTGC
EZ3-2ATGR	GAAGGAGATGGGAGTCCTAC
EZ3r15a	CTAGGTCTTAGAATAAGACTACG
EZ3d15a	GGGTTTTATACGCGGTGCATA
EZ3r15	CATGTTCTCGCAGTCATTACT
EZ3d16	GTGGTGATGGGACTCTCGATA
EZ3r16	CCTCTCATGCCTGGTCAT
EZ3d17	CAGAATAGTGTTGGAAAGCACGAG
EZ3r17	TGCAAGCTTACTTTCTTGTAGTGG
EZ3d18b	CAAAGTTGGTATATTTGCCAAAC
EZ3d18	ATACACCACATGTCTGGGCAA
EZ3r18	GGGTCCGCCCTTCCACC
EZ3d19	CATGCTAAGCGAAGCCAGTG
EZ3r2	GGCGTGAGGTGGTGCGTT
EZ3d2	CATTGAAGTAGATACTGCATG
EZ3r1	GATTTCTTACAGAAACTTGGTT
EZ3d20	CTGGCTGGGGTGATTTCTTG
EZ3r20	CTTCTATCCTTGATCTATTGCAC
EZ3d21	AGGGAAGTTATAGATGAGTTTTGG
EZ3r21	GCTGTCAATAGCTGTAGCATG
EZ3d22	TAGAGGTAAGCACTTACCCAC
EZ3r22	GTATGCTCCCCAAGGTTCTC
EZ3d23	GCACTTTTGAAGATACAATTGCC
EZ3r23	AGATACGTGGATCTTACCAGG
EZ3d24	GATATGGTGGATAGCTATTCCTT
EZ3r24	GCAGTATGTATTGGTAATGTTGG
EZ3d25	CAGAAGCTCACTGGGAAGAC
EZ3r25	TTCAGAGGTGATGATAAGTAAGC
EZ3d26	GTGTTAGGCGTAAGTCACACC
EZ3r26	GATAGTAGCCTCTGGACTAGAG
EZ3r19	CCATTGTGGGATCTAGCACAC
EZ3d19b	GATCATGGGATATTGACCCAG
EZ3d5	TACTCGCATCATTTAGTGCAAT
EZ3d4	GAGATCAGTGATCACAACTAG
EZ3d6	GAAGGAACAACCATAGAAGCT
EZ3d8	ACAAATTGGCCTGTCTGATAC
EZ3d3	CCTACTAACCTTCTGACGT
EZ3d10	TCAGCAGATGCCCATGAATC
EZ3d12	CAGATTGGTGCAGCTGAAGG
EZ3d13	GACTTCCTCGCATCCACAAC
EZ3d1	GACATGATTGGTGAATTGCATAG
EZ3d2	GTCTCCTTAAACGTATCCTC
EZ3r4	ATAACTAATTATCGAAATCAAATTAC

REFERENCES

Abushita A.A., Hebshi E.A., Daood H.G. Biacs P.A. (1997). Determination of antioxidant vitamins in tomatoes. *Food Chem*., 60: 207-212

Abushita A.A., Daood H.G., Biacs P.A. (2000). Change in carotenoids and antioxidant vitamins in tomato as a function of varietal and technological factors. *Agric. Food Chem*., 48: 2075-2081

Adams-Phillips L., Barry C., Giovannoni J. (2004). Signal transduction sys- tems regulating fruit ripening. *Trends Plant Sci*., 9: 331–338

Arumuganathan K., Earle E. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep*., 9: 208-18

Bai Y., Lindhout P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Ann. Botany*, 100: 1085-1094

Ballestar E., Wolffe A.P. (2001). Methyl-CpG-binding-protein. *Eur. J. Biochem*., 1-6.

Barone A., Chiusano M.L., Ercolano M.R., Giuliano G., Grandillo S., Frusciante L. (2008). Structural and functional genomics of tomato. *Int. J. Plant Genomics*, Article ID: 820274

Barrow M., Meister A. (2003). Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. *Plant Cell Environ*., 26: 571–584

Barry C.S., Liop-Tous M.I., Grierson D. (2000). The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato. *Plant Physiol*., 123: 979-986

Bastow R.M., Mylne J.S., Lippman Z., Martienssen R.A., Dean C.(2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature,* 427**:** 164 – 167.

Baumbusch L.O., Thorstensen T., Krauss V., Fischer A., Naumann K., et al. (2001). The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res*., 29: 4319-4333

Bedford M.T. (2007). Arginine methylation at a glance. *J. Cell. Sci.*, 120: 4243-4246

Berg A., Meza T.J., Mahic M., Thorstensen T., Kristiansen K., Aalen R.B. (2003). Ten members of the Arabidopsis gene family encoding methyl-CpG-binding domain proteins are transcriptionally active and at least one, AtMBD11, is crucial for normal development. *Nucleic Acids Res*., 31: 5291-5304

Berger S.L. (2007). The complex language of chromatin regulation during transcription. *Nature*, 447: 407-12

Bernacchi D., Beck-Bunn T., Emmatty D., Eshed Y., Inai S., Lopez J., Petiard V., Sayama H., Uhlig J., Zamir D., Tanksley S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum*. *Theor. Appl. Genet*., 97: 381-397

Bernacchia G., Primo A., Giorgetti L., Pitto L., Cella R. (1998). Carrot DNA-methyltransferase is encoded by two classes of genes with differing patterns of expression. *Plant J.*, 13: 317-329

Bernatavichute Y.V., Zhang X., Cokus S., Pellegrini M., Jacobsen S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana. PLoS ONE*, 3: e3156

Bertin N. (2005). Analysis of the tomato fruit growth response to temperature and plant fruit load in relation to cell division, cell expansion and DNA endoreduplication. *Ann*. *Botany*, 95: 439–447

Bertin N., Lecomte A., Brunel B., Fishman S., Genard M. (2007). A model describing cell polyploidization in tissues of growing fruit as related to cessation of cell proliferation. *J. Exp. Bot.*, 58: 1903–1913

Bestor T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *The EMBO J.*, 11: 2611–2617

Bestor T.H., Verdine G.L. (1994). DNA methyltransferases. *Curr. Opin. Cell Biol.*, 6: 380-389

Bird A.P., Wolffe A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. *Cell*, 99(5): 451-4

Blewitt M.E., Vickaryous N.K., Hemley S.J., Ashe A., Bruxner T.J., Preis J.I., Arkell R., Whitelaw E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci. USA*, 102(21): 7629-34

Bhutani N., Burns D.M., Blau H.M. (2011). DNA demethylation dynamics. *Cell*, 146(6): 866-72

Bohner J. and Bangerth F. (1988). Cell number, cell size and hormone levels in semi-isogenic mutants of *Lycopersicon pimpinellifolium* differing in fruit size. *Physiol Plant.*, 72: 316–320

Boudolf V., Vlieghe K., Beemster G.T., Magyar Z., Acosta J.A., Maes S., Schueren E., Inzé D., Veylder L. (2004). The plant-specific Cyclin-Dependent Kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreduplicating cells in *Arabidopsis*. *Plant Cell*, 16: 2683-2692

Bourdon M., Frangne N., Mathieu-Rivet E., Nafati M., Cheniclet C., Renaudin J.P., Chevalier C. (2010). Endoreduplication and growth of fleshy fruits. *Progress in Botany*, 71: 101-132

Bratzel F., López-Torrejón G., Koch M., Del Pozo J.C., Calonje M. (2010). Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.*, 20(20): 1853-9

Brero A., Leonhardt H., Cardoso M.C. (2006). Replication and translation of epigenetic information. *Curr. Top Microbiol. Immunol.*, 301: 21-44

Brown J.L., Mucci D., Whitely M., Dirksen M. L., Kassis J.A. (1998). The Drosophila Polycomb group gene pleiohomeotic encodes a DNA

binding protein with homology to the transcription factor YY1. *Mol. Cell*., 1: 1057-1064

Budiman M.A., Mao L., Wood T.C., Wing R.A. (2000). A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res*., 10(1): 129-36

Cao X., Aufsatz W., Zilberman D., Mette M.F., Huang M.S., et al. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol*., 13: 2212-17

Cao X., Jacobsen S.E. (2002). Locus- specific control of asymmetric and CpNpG methylation by DRM and CMT3 methyltrasferase genes. *Proc. Natl. Acad. Sci. USA*, 99: 16491-16498

Carter C.W. Jr. (1978). Histone packing in the nucleosome core particle of chromatin. *Proc. Natl. Acad. Sci. USA*, 75: 3649-3653

Chan S.W., Zilberman D., Xie Z., Johansen L.K., Carrington J.C., Jacobsen S.E. (2004). RNA silencing genes control *de novo* DNA methylation. *Science*, 303: 1336

Chen K.Y., Cong B., Wing R., Vrebalov J., Tanksley S.D. (2007). Changes in regulation of a transcription factor lead to autogamy in cultivated tomatoes. *Science*, 318: 643-645

Cheniclet C., Rong W.Y., Causse M., Frangne N., Bolling L., Carde J.P., Renaudin J.P. (2005). Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol.*, 139: 1984-1994

Chevalier C. (2007). Cell cycle control and fruit development. *In*: Inzé D. [ed.] Cell cycle control and plant development, *Annual Plant Reviews*. Oxford: Blackwell Publishing, 32: 269-293 / doi: 10.1002/9780470988923.ch12

Chevalier C., Nafati M., Mathieu-Rivet E., Bourdon M., Frangne N., Cheniclet C., Renaudin J.P., Gévaudant F., Hernould M. (2011). Elucidating the functional role of endoreduplication in tomato fruit development. *Ann. Botany*, 107: 1159–1169

Clark K.L., Larsen P.B., Wang X., Chang C. (1998). Association of the *Arabidopsis* CTR1 Raf-like kinase with the ETR1 and ERS ethylene receptors. *Proc. Natl. Acad. Sci. USA*, 95: 5401-5406

De Block M., Van Lijsebettens M. (2011). Energy efficiency and energy homeostasis as genetic and epigenetic components of plant performance and crop productivity. *Curr. Opin. Plant Biol*., 14(3): 275-82

Eamens A., Vaistij F. E., Jones L. (2008). NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in *Arabidopsis*. *Plant J.*, 55: 596-606

Eriksson E.M., Bovy A., Manning K., Harrison L., Andrews J., De Silva J., Tucker G.A., Seymour G.B. (2004). Effect of the Colorless non-ripening (*Cnr*) mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. *Plant Physiol*., 136: 4184-4197

Eshed Y., Gera G., Zamir D. (1996). A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. *Theor. Appl. Genet*., 93: 877-886

Faurobert M., Mihr C., Bertin N., Pawlowski T., Negroni L., Sommerer N., Causse M. (2007). Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiol*., 143(3): 1327-46

Finnegan E.J., Dennis E.S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res*., 21: 2383-2388

Finnegan E.J., Genger R.K., Peacock W.J., Dennis E.S. (1998). DNA methylation in plant. *Mol. Biol*., 49: 223-247

Finnegan E.J., Kovac K.A. (2000). Plant DNA methyltransferases. *Plant Mol. Biol*., 43: 189-201.

Finnegan E. J., Matzke M.A. (2003). The small RNA world. *J. Cell. Sci.*, 116: 4689-4693

Foolad M.R. (2007). Genome Mapping and Molecular Breeding of Tomato. *Int. J. Plant Genomics*, 2007: 1–52

Francis, N.J. and Kingston, R.E. (2001). Mechanisms of transcriptional memory. *Nat. Rev. Mol. Cell Biol*., 2: 409-421

Frary A., Doganlar S., Frampton A, Fulton T., Uhlig J., Yates H., Tanksley S. (2003). Fine mapping of quantitative trait loci for improved fruit characteristics from *Lycopersicon chmielewskii* chromosome 1. *Genome*, 46: 235-243

Fujisawa M., Shima Y., Higuchi N., Nakano T., Koyama Y., Kasumi T., Ito Y. (2011). Direct targets of the tomato-ripening regulator RIN identified by transcriptome and chromatin immunoprecipitation analyses. *Planta*, DOI : 10.1007/s00425-011-1561-2

Galbraith D.W., Harkins K.R., Knapp S. (1991). Systemic endopolyploidy in *Arabidopsis thaliana*. *Plant Physiol.*, 96: 985-989

Garewal H. (1995). Antioxidants in oral cancer prevention. *Am. J. Clin. Nutr*., 62: 1410S-1416S

Gary J.D., Clarke S. (1998). RNA and protein interactions modulated by protein arginine methylation. *Progr. Nucl. Acid Res. Mol. Biol.*, 61: 65-131

Genger R.K., Kovac K.A., Dennis E.S., Peacock W.J., Finnegan E.J. (1999). Multiple DNA methyltransferase genes in *Arabidopsis thaliana*. *Plant Mol. Biol*., 41: 269-278

Gillaspy G., Ben-David H., Gruissem W. (1993). Fruits: a developmental perspective. *Plant Cell*, 5: 1439-1451

Giorio G., Stigliani A.L., D'Ambrosio C. (2008). Phytoene synthase genes in tomato (*Solanum lycopersicum* L.) – new data on the structures, the deduced amino acid sequences and the expression patterns. *FEBS J.*, 275: 527-535

Giovannoni J.J. (2004). Genetic regulation of fruit development and ripening. *Plant Cell*, 16: S 170-180

Giovannoni J.J. (2007). Fruit ripening mutants yield insights into ripening control. *Curr. Opin. Plant Biol*., 10: 283-289

Goodrich J., Puangsomlee P., Martin M., Long D., Meyerowitz E.M., Coupland G. (1997). A Polycombgroup gene regulates homeotic gene expression in *Arabidopsis*. *Nature*, 386: 44-51

Guillet C., Just D., Bénard N., Destrac-Irvine A., Baldet P., Hernould M., Causse M., Raymond P., Rothan C. (2002). A fruit-specific phospho enolpyruvate carboxylase is related to rapid growth of tomato fruit. *Planta*, 214(5): 717-26

Gutierrez C. (2009). The Arabidopsis cell division cycle. *Arabidopsis Book*, 7: e0120

Grossniklaus U., Spillane C., Page D.R., Köhler C. (2001). Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Plant Biol*., 4(1): 21-7

Hauben M., Haesendonckx B., Standaert E., Van Der Kelen K., Azmi A., Akpo H., Van Breusegem F., Guisez Y., Bots M., Lambert B., Laga B., De Block M. (2009). Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proc. Natl. Acad. Sci. U S A*., 106(47): 20109-14

Henikoff S., Comai L. (1998). A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics*, 149: 307-318

Hennig, L., Derkacheva, M. (2009). Diversity of PcG complexes in plants: same rules, different players? *Trends in Genetics*, 25: 414-423

Herr A.J., Jensen M.B., Dalmay T., Baulcombe D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science,* 308: 118-120

Hobolth, A., Nielsen, R., Wang, Y., Wu, F. and Tanksley, S.D. (2006). CpG+CpNpG analysis of protein coding sequences from tomato. *Mol. Biol. Evol.*, 23: 1318-1323

Horard B., Tatout C., Poux S., Pirrotta V. (2000). Structure of a Polycomb response element and *in vitro* binding of Polycomb group complexes containing GAGA factor. *Mol. Cell. Biol*., 20(9): 3187-3197

How Kit A., Boureau L., Stammitti-Bert L., Rolin D., Teyssier E., Gallusci P. (2010). Functional analysis of SlEZ1 a tomato Enhancer of zeste (E(z)) gene demonstrates a role in flower development. *Plant Mol. Biol.,* 74(3): 201-213

Jackson J.P., Lindroth A.M., Cao X., Jacobsen S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, 416: 556-60

Jacob Y., Feng S., Leblanc C.A., Bernatavichute Y.V., Stroud H., Cokus S., Johnson L.M., Pellegrini M., Jacobsen S.E., Michaels S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat. Struct. Mol. Biol*., 16: 763–768.

Jones L., Hamilton A.J., Voinnet O., Thomas C.L., Maule A.J., Baulcombe D.C. (1999). RNA-DNA interaction and DNA interactions in post-transcriptional gene silencing. *Plant Cell*, 11: 2291-2302

Joubès J., Chevalier C. (2000). Endoreduplication in higher plants. *Plant Mol. Biol*. 43: 735–745

Joubès J., Chevalier C., Dudits D., et al. (2000a). Cyclin-dependent kinases related protein kinases in plants. *Plant Mol. Biol*., 43: 607–621

Kakutani T. (1998). Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in *Arabidopsis thaliana*. *Plant J*., 12: 1447-1451

Kalisz S., Vogler D.W., Hanley K.M. (2004). Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. *Nature*, 430: 884–87

Kanno T., Mette F., Kreil D.P., Aufsatz W., Matzke M., Matzke A. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol*., 14: 801-805

Kanno T., Huettel B., Mette M.F., Aufsatz W., Jaligot E., et al. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet*., 37: 761-65

Kirnos M.D., Aleksandrushkina N.I., Vanyushin B.F. (1981). 5-Methylcytosine in pyrimidine sequences of plant and animal DNA: specificity of DNA methylation. *Bio-chemistry (Mosc),* 46: 1458-1474

Klee H.J., Giovannoni J.J. (2011). Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet*., 45: 41-59

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., Muller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev*., 20(9): 1110-1122

Knapp S.L., Bohs M.N,. Spooner D.M. (2004). Solanaceae: a model for linking genomics and biodiversity. *Comp Funct Genomics* 5:285–291.

Koch-Nolte F., Haag F., Guse A.H., Lund F., Ziegler M. (2009). Emerging roles of NAD+ and its metabolites in cell signaling. *Sci. Signal.,* 10;2(57):mr1

Köhler C., Grossniklaus U. (2002). Epigenetic inheritance of expression states in plant development: the role of Polycomb group proteins. *Curr. Opin. Cell. Biol*., 14(6): 773-9

Komachi K., Redd M. J., Johnson A.D. (1994). The WD repeats of Tup1 interact with the homeo domain protein alpha 2. *Genes Dev*., 8: 2857-2867

Kornberg R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science,* 184: 868-871

Koryakov D.E. (2006). Histone modification and regulation of chromatin Function. *Genetics*, 42: 970-984

Kouzarides T. (2002). Histone methylation in transcriptional control. *Curr. Opin. Genet. Dev*., 12: 198-209

Kouzarides T. (2007). Chromatin modifications and their function. *Cell*, 128: 693-705

Kumar A. (2004). Retrotransposons and their contributions to plant genome and gene evolution. In *Encyclopedia of Plant and Crop Sci.,* ed. Goodman R.M., 1-4. New York: Marcel Dekker Inc.

Kumar A., Bennetzen J.L. (1999). Plant retrotransposons. *Annu. Rev Genet.,* 33: 479–532.

Larkins B.A., Dilkes B.P., Dante R.A., Coelho C.M., Woo Y.M., Liu Y. (2001). Investigating the hows and whys of DNA endoreduplication. *J. Exp. Bot.,* 52: 183–192

Latzko E., Kelly G.J. (1983). The many-faceted function of phosphoenolpyruvate carboxylase in C3 plants. *Physiol. Veg*., 21: 805–815

Lawrence R.J., Earley K., Pontes O., Silva M., Chen Z.J., Neves N., Viegas W., Pikaard C.S. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. *Mol Cell*., 13: 599-609

Lee M.G., Villa R., Trojer P., Norman J., Yan K.P., Reinberg D., Di Croce L., Shiekhattar R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science*, 318(5849): 447-50.

Lelièvre S.A., Weaver V.M., Nickerson J.A., Larabell C.A., Bhaumik A., Petersen O.W., Bissell, M.J. (1998). Tissue phenotype depends on reciprocal interactions between the extracellular matrix and the structural organization of the nucleus. *Proc. Natl. Acad. Sci. USA*, 95: 14711-14716.

Li L., Zhu B., Yang P., Fu D., Zhu Y., Luo Y. (2011). The regulation mode of RIN transcription factor involved in ethylene biosynthesis in tomato fruit. *J. Sci. Food Agric*., 91*:* 1822-1828

Lindroth A.M., Cao X., Jackson J.P., Zilberman D., McCallum C.M., et al. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science,* 292: 2077-2080

Lisch D. (2009). Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol*., 60: 43-66

Li E., Bestor T.H., Jaenisch R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69: *915-926*

Loidl P. (2004). A plant dialect of the histone language. *Trends Plant Sci*., 9: 84-90

Luckwill L.C. (1943). The genus *Lycopersicon*: an historical, biological and taxonomical survey of the wild and cultivated tomatoes. *Aberdeen Univ. Stud.,* 120: 1-44

Makarevich G., Leroy O., Akinci U., Schubert D., Clarenz O., Goodrich J., Grossniklaus U., Köhler C. (2006). Different Polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep*., 7: 947-952

Manning K., Tör M., Walley P., Poole M., Hong Y., Thompson A.J., King G.J., Giovannoni J.J., Seymour G.B. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Gen*., 38: 948-952

Mathieu O., Probst A.V., Paszkowski J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis. The EMBO J.*, 24: 2783 – 2791

Mateos S., Domínguez I., Pastor N., Cantero G., Cortés F. (2005). The DNA demethylating 5-azaC induces endoreduplication in cultured Chinese hamster cells. *Mutat Res.,* 578(1-2): 33-42

Mayama T., Ohtsubo E., Tsuchimoto S. (2003). Isolation and expression analysis of petunia CURLY LEAF-like genes. *Plant Cell Physiol.,* 44: 811–819

Messeguer R., Ganal M.G., Steffens J.C., Tanksley S.D. (1991). Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. *Plant Mol. Biol.,* 16: 753-770.

McBride A.E., Silver P.A. (2001). State of Arg: protein methylation at arginine comes of age. *Cell*, 106: 5-8.

McClintock B. (1967). Genetic systems regulating gene expression during development. *Dev. Biol. Suppl*., 1: 84-112

McMeekin D. (1992). Representations of pre-Columbian spindle whorls of the floral and fruit structure of economic plants. *Econ. Bot*., 46: 171–180

Morales V., Giamarchi C., Chailleux C., Moro F., Marsaud V., Le Ricousse S., Richard-Foy H. (2001). Chromatin structure and dynamicsfunctional implications. *Biochimie,* 83: 1029–1039

Moyle L.C. (2008). Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). *Evolution,* 62: 2995-3013

Muller C.H. (1940). A revision of the genus *Lycopersicon*. *USDA Misc. Publ.,* 382: 1-29

Müller J., Hart C.M., Francis N.J., Vargas M.L., Sengupta A., Wild B., Miller E.L., O'Connor M.B., Kingston R.E., Simon J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb Group repressor complex. *Cell,* 111: 197–208

Murfett J., Wang X.J., Hagen G., Guilfoyle T.J. (2001). Identification of Arabidopsis histone deacetylase HDA6 mutants that affect transgene expression. *Plant Cell.,* 13(5): 1047-61

Nafati M., Frangne N., Hernould M., Chevalier C., Gévaudant F. (2010). Functional characterization of the tomato Cyclin-Dependent Kinase inhibitor SlKRP1 domains involved in protein–protein interactions. *New Phytologist,* 188: 136-149

Nakazato T., Warren D.L., Moyle L.C. (2010). Ecological and geographic modes of species divergence in wild tomatoes. *Am. J. Bot.*, 97: 680-693

Naumann K., Fischer A., Hofmann I., Krauss V., Phalke S., Irmler K., Hause G., Aurich A.C., Dorn R., Jenuwein T., Reuter G. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis. The EMBO J.*, 24: 1418-29

Nekrasov M., Wild B., Müller J. (2005). Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Rep.* 6(4): 348-53

Nesbitt T.C., Tanksley S.D. (2002). Comparative sequencing in the genus *Lycopersicon*. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics,* 162: 365-379

Niu L., Lu F., Pei Y., Liu C., Cao X. (2007). Regulation of flowering time by the protein arginine methyltransferase AtPRMT10. *EMBO Rep.*, 8: 1190-1195

Niu L., Zhang Y., Pei Y., Liu C., Cao X. (2008). Redundant requirement for a pair of PROTEIN ARGININE METHYLTRANSFERASE4 homologs for the proper regulation of *Arabidopsis* flowering time. *Plant Physiol.*, 148: 490-503

Nunes-Nesi A., Carrari F., Lytovchenko A., Fernie A.R. (2005a). Enhancing crop yield in Solanaceous species through the genetic manipulation of energy metabolism. *Biochem. Soc. Trans.,* 33(Pt 6): 1430-4

Nunes-Nesi A., Carrari F., Lytovchenko A., Smith A.M., Loureiro M.E., Ratcliffe R.G., Sweetlove L.J., Fernie A.R. (2005b). Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. *Plant Physiol.,* 137(2): 611-22

Oakeley E.J., Jost J.P. (1996). Non-symmetrical cytosine methylation in tobacco pollen DNA. *Plant Mol. Biol*., 31: 927-930

Ohad N., Yadegari R., Margossian L., Hannon M., Michaeli D., Harada J.J., Goldberg R.B., Fischer R.L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell,* 11: 407-16

Onodera Y., Haag J.R., Ream T., Nunes P.C., Pontes O., Pikaard C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell,* 120: 613-622

Pal S., Sif S. (2007). Interplay between chromatin remodelers and protein arginine methyltransferase. *J. Cell. Physiol.*, 213: 306-315

Palma J.M., Corpas F.J., del Rio L.A. (2011). Proteomics as an approach to the understanding of the molecular physiology of fruit development and ripening. *J. of Proteomics*, 74: 1230-1243

Pandey D.K., Shekelle R., Selwyn B.J., Tangney C., Stamler J. (1995). Dietary vitamin C and β-carotene and risk of death in middle-age men. *Amer. J. Epid.,* 142: 1269-1278

Pandey R., Muller A., Napoli C.A., Selinger D.A., Pikaard C.S., Richards E.J., Bender J., Mount D.W., Jorgensen R.A. (2002). Analysis of histone acetyltransferase and histone deacetylase families of

Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res*., 30: 5036-5055

Papa C.M., Springer N.M., Muszynski M.G., Meeley R., Kaeppler S.M. (2001). Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. *Plant Cell,* 13: 1919-1928

Penterman J., Uzawa R., Fischer R.L. (2007). Genetic Interactions between DNA Demethylation and Methylation in *Arabidopsis*. *Plant Physiol.,* 145: 1549-1557

Peralta I.E., Spooner D.M. (2001). Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). *Am. J. Bot*., 88: 1888-1902

Peralta I.E., Knapp S., Spooner D.M. (2005). New species of wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae) from Northern Peru. *Syst. Botany,* 30: 424-434

Peralta I.E., Spooner D.M. (2007). History, origin and early cultivation of tomato (Solanaceae). *In*: Razdan M.K. and Mattoo A.K. [ed.] Genetic Improvement of Solanaceous Crops, Vol. 2. Enfield, USA: Science Publishers. p. 1-27

Perikles S. (2003). Q-Gene processing quantitative real-time RT-PCR data. *Bioinformatics,* 19: 1439–1440

Peterson D., Price H., Johnston J., Stack S. (1996). DNA content of hetrochromatin and euchromatin in tomato (*Lycopersicum esculentum*) pachytene chromosomes. *Genome,* 39: 77–82

Pien S., Grossniklaus U. (2007). Polycomb group and trithorax group proteins in Arabidopsis. *Biochim. Biophys. Acta,* 1769(5-6): 375-82

Pirrotta V. (2003). Transcription. Puffing with PARP *Science,* 299(5606): 528-9

Pontes O., Li C.F., Nunes P.C., Haag J., Ream T., Vitins A., Jacobsen S.E., Pikaard C.S. (2006). The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell,* 126: 79-92

Poux S., McCabe D., Pirrotta V. (2001a). Recruitment of components of Polycomb Group chromatin complexes in Drosophila. *Developmen,* 128(1): 75-85

Poux S., Melfi R., Pirrotta V. (2001b). Establishment of Polycomb silencing requires a transient interaction between PC and ESC. *Genes Dev.,* 15(19): 2509-14

Reyes J.C., Grossniklaus U. (2003). Diverse functions of Polycomb group proteins during plant development. *Semin Cell Dev Biol*., 14(1): 77-84

Rick C.M. (1963). Barriers to interbreeding in *Lycopersicon peruvianum*. *Evolution,* 17: 216-232

Rick C.M. (1973). Potential genetic resources in tomato species: clues from observations in native habitats. *In:* Hollaender A. and Srb A.M. [ed.], Genes, Enzymes and Populations. *Plenum Press*, New York, USA : 255-269

Rick C.M. (1976). Tomato (family *Solanaceae*). *In*: Simmonds N.W. [ed.] Evolution of Crop Plants. *Longman Publications*, UK : 268-273

Rick C.M. (1982). Genetic relationships between self- incompatibility and floral traits in the tomato species. *Biol. Zentralbl*. 101: 185–198.

Rick C.M., Chetelat R.T. (1995). Utilization of related wild species for tomato improvement. *In*: Munoz-Fernandez R., Cuartero J., Gomez-Guillamon M.L. [ed.] I Inter. Symposium on *Solanacea* for Fresh Market. *Acta Hort.,* 412: 21-38

Rigal M., Mathieu O. (2011). A "mille-feuille" of silencing: epigenetic control of transposable elements. *Biochim. Biophys. Acta,* 1809: 452-8

Ronemus M.J., Galbiati M., Ticknor C., Chen J., Dellaporta S.L. (1996). Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science,* 273: 654-657

Saurin A.J., Shao Z., Erdjument-Bromage H., Tempst P., Kingston R.E. (2001). A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature,* 412(6847): 655-60

Sanchez-Pulido L., Devos D., Sung Z.R., Calonje M. (2008). RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. *BMC Genomics*, 9: 308.

Scebba F., Bernacchia G., De Bastiani M., Evangelista M., Cantoni R.M., Cella R., Locci M.T., Pitto L. (2003). *Arabidopsis* MBD protein show different binding specificities and nuclear localization. *Plant Mol. Biol*., 53: 715-731.

Scebba F., De Bastiani M., Bernacchia G., Andreucci A., Galli A., Pitto L. (2007). PRMT11: a new *Arabidopsis* MBD7 protein partner with arginine methyltransferase activity. *Plant J*., 52: 210-222.

Schauer N., Zamir D., Fernie A.R. (2005). Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.,* 56: 297–307

Schmitz R.J., Sung S., Amasino R.M. (2008). Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A-Plant Biology,* 105: 411-416

Schubert D., Clarenz O., Goodrich J. (2005). Epigenetic control of plant development by Polycomb-group proteins. *Curr. Opin. Plant Biol*., 8: 553–561

Schwartz Y.B., Pirrotta V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet*., 8: 9–22

Schwartz Y.B., Pirrotta V. (2008). Polycomb complexes and epigenetic states. *Curr. Opin. Cell. Biol.,* 20: 266–273

Sengupta A.K., Kuhrs A., Müller J. (2004). General transcriptional silencing by a Polycomb response element in Drosophila. *Development, 131:* 1959-1965

Seymour G., Poole M., Manning K., King G.J. (2008). Genetics and epigenetics of fruit development and ripening. *Curr. Opin. Plant Biol.,* 11: 58–63

Sims R.J., Nishioka K., Reinberg D. (2003). Histone lysine methylation: a signature for chromatin function. *Trend. Genet.,* 19: 629-639

Slotkin R.K., Martienssen R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.,* 8(4): 272-85

Smith J.J., Rucknagel K.P., Schierhorn A., Tang J., Nemeth A., Linder M., Herschman H.R., Wahle E. (1999). Unusual sites of arginine methylation in poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3. *J. Biol. Chem.,* 274: 13229-13234

Snyder R.G. (1996). Greenhouse tomatoes – The basics of successful production. *Greenhouse Tomato Sem.* ASHS Press, *Am. Soc. of Hortic.* Science, Alexandria, VA, USA, 3-6

Spooner D.M., Anderson G.J., Jansen R.K. (1993). Chloroplast DNA evidence for the interrelathionships of tomatoes, potatoes and pepinos (*Solanaceae*). *Am. J. Bot.,* 80: 676-688

Spooner D.M., Peralta I.E., Knapp S. (2005). Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes *Solanum* L. section *Lycopersicon* (Mill.) Wettst. *Taxon,* 54: 43-61

Springer N.M., Napoli C.A., Selinger D.A., Pandey R., Cone K.C., Chandler V.L., Kaeppler H.F., Kaeppler S.M. (2003). Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol.,* 132: 907–925

Steinhauser M.C., Steinhauser D., Koehl K., Carrari F., Gibon Y., Fernie A.R., Stitt M. (2010). Enzyme activity profiles during fruit development in tomato cultivars and Solanum pennellii. *Plant Physiol.,* 153(1): 80-98

Stevens M.A., Rick C.M. (1986). Genetics and breeding. *In* : Atherton J.G. and Rudich J. [ed.] The tomato crop. *Chapman and Hall, NY :* 35-109

Tariq M., Paszkowski J. (2004). DNA and histone methylation in plant. *Trends Genet.,* 20: 244-251

Teyssier E., Bernacchia G., Maury S., How Kit A., Stammitti-Bert L., Rolin D., Gallusci P. (2008). Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. *Planta,* 228: 391–399

Tian L., Chen Z.J. (2001). Blocking histone deacetylation in *Arabidopsis* induces pleiotropic effects on plant gene regulation and development. *Proc. Natl. Acad. Sci. USA,* 98: 200-205

Tie F., Furuyama T., Prasad-Sinha J., Jane E., Harte P.J. (2001). The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development,* 128: 275–286

Tieman D.M., Klee H.J. (1999). Differential expression of two novel members of the tomato ethylene-receptor family. *Plant Physiol.,* 120: 165–172

Tieman D.M., Ciardi J.A., Taylor M.G., Klee H.J. (2001). Members of the tomato LeEIL (EIN3-like) gene family are functionally redundant and regulate ethylene responses throughout plant development. *Plant J.,* 26: 47-58

Tompa R., McCallum C.M., Delrow J., Henikoff J.G., van Steensel B., Henikoff S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.,* 12: 65-68

Tsaftaris A.S., Polidoros A.N. (2000). DNA methylation and plant breeding. *Plant Breed. Rev.,* 18: 87-176

USDA – NASS : United States Department of Agriculture – National Agricultural Statistics Service. (2003) //
http://www.nass.usda.gov/Publications/Ag_Statistics/2003/index.asp

Vanyushin B.F. (2006). DNA methylation in plants. *Curr. Top. Microbiol. Immunol.,* 301: 67-122. Review.

Vrebalov J., Ruezinsky D., Padmanabhan V., White R., Medrano D., Drake R., Schuch W., Giovannoni J.J. (2002). A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science,* 296: 343-346

Wang Y., Tang X., Cheng Z., Mueller L.A., Giovannoni J., Tanksley S.D. (2006). Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. *Genetics,* 172: 2529-2540

Wang L., Brown J.L., Cao R., Zhang Y., Kassis J.A., Jones R.S. (2004). Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell.,* 14: 637–646

Weil C., Martienssen R. (2008). Epigenetic interactions between transposons and genes: lessons from plants. *Curr. Opin. Genet. Dev.,* 18(2): 188-92

Wilkinson J.Q., Lanahan M.B., Yen H.C., Giovannoni J.J., Klee H.J. (1995). An ethylene-inducible component of signal transduction encoded by never-ripe. *Science,* 270: 1807-1809

Yamamoto K., Sonoda M., Inokuchi J., Shirasawa S., Sasazuki T. (2004). Polycomb group suppressor of zeste 12 links heterochromatin protein 1alpha and enhancer of zeste 2. *J. Biol. Chem.,* 279: 401-406

Yan D., Zhang Y., Niu L., Yuan Y., Cao X. (2007). Identification and characterization of two closely related histone H4 arginine 3 methyltransferases in *Arabidopsis thaliana*. *Biochem. J.,* 408: 113-21

Yoshida N., Yanai Y., Chen L., Kato Y., Hiratsuka J., Miwa T. et al. (2001). EMBRYONIC FLOWER2, a novel Polycomb group protein homolog, mediates shoot development and flowering in Arabidopsis. *Plant Cell,* 13: 2471–2481

Zhang X., Zhou L., Cheng X. (2000). Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *The EMBO J.,* 19: 3509-3519

Zhang H., Jin J.P., Tang L., Zhao Y., Gu X.C., Gao G, Luo J.C. (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res.*, 39: D1114-D1117.

Zhao Z., Shen W.H. (2004). Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. *Ann. NY Acad. Sci.,* 1030: 661-69

Zhao X., Hao W. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*. 35(Web Server issue): W265-W268.

Zhou D., Mattoo A., Tucker M. (1996). Molecular cloning of a tomato cDNA encoding an ethylene receptor. *Plant Physiol.,* 110: 1435–1436

Zhou J., Wang X., He K., Charron J.B., Elling A.A., Deng X.W. (2010). Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol Biol.,* 72: 585-595

Zhu J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet*., 43: 143-166

Zilberman D., Gehring M., Tran R.K., Ballinger T., Henikoff S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.,* 39(1): 61-9

Zuriaga E., Blanca J., Nuez F. (2009). Classification and phylogenetic relationships in *Solanum* section *Lycopersicon* based on AFLP and two nuclear gene sequences. *Genet. Resour. Crop Evol.,* 56: 663-678

ABSTRACT

Tomato (Solanum lycopersicum) which forms a small monophyletic clade within the large Solanaceae family has been chosen as a model system for studying the Solanaceae genome, fruit development and ripening. At that time, many efforts have been devoted to the analysis of the genetic diversity of tomato species, little work has focused on the analysis epigenetic diversity in this clade, although there is a general agreement that epigenetic processes play essential role in the phenotypic diversity in animal and plant system. As first step, DNA methylation level was analyzed in leaves and fruits of various wild and cultivated tomato species.Additionally, the Enhancer of zest (E(z)) gene family has been analyzed. In tomato, the E(z) family consists in two functional genes (SlEZ1, SlEZ2) and in a pseudogene (SlEZ3). In addition, the epigenetic stability is an important consideration that could have a significant on strategies for crop breading. Finally, we made a fine characterization of the different aspects of fruit development and ripening.

RIASSUNTO

All’interno della grande famiglia delle Solanacee è stato scelto il pomodoro (Solanum lycopsersicum) come sistema modello per studio dello sviluppo e maturazione del frutto. Molti sforzi sono stati fatti per analizzare la diversità genetica delle specie di pomodoro, pochi lavori invece riguardano l’analisi della diversità epigenetica, sebbene ci sia accordo sul fatto che processi epigenetici giochino un ruolo essenziale nella diversità fenotipica dei sistemi animali e vegetali. Inizialmente è stato analizzato il livello di metilazione del DNA in foglie e frutti delle diverse specie di pomodoro selvatico e coltivato. Inoltre, è stata analizzata la famiglia genica Enhancer of Zeste (E (z)). In pomodoro la famiglia E(z) consiste di 2 geni funzionali SlEZ1, SlEZ2 e di uno pseudogene SlEZ3. Inoltre la stabilità epigenetica è importante in quanto può avere un impatto sulle strategie di miglioramento genetico delle specie coltivate. Infine è stata condotta una attenta caratterizzazione dei meccanismi cellulari dello sviluppo del frutto e della sua maturazione.

RÉSUMÉ

La tomate (Solanum lycopsersicum), qui forme un clade monophylétique restreint au sein de la large famille des Solanacées, est utilisée comme modèle pour l’analyse du génome, et le développement du fruit. A ce jour, de nombreux efforts ont été consacrés à l'analyse de la diversité génétique des espèces de tomate. Cependant peu de travaux ont porté sur l'analyse de la diversité épigénétique, alors qu’il est aujourd’hui admis que les processus épigénétiques jouent un rôle essentiel dans la diversité phénotypique. Dans un premier temps, le niveau de méthylation de l'ADN a été comparé dans les feuilles et les fruits de différentes variétés de tomates sauvages et cultivées. Puis la famille des gènes Enhancer of zeste (E (z)) a été analysée. Chez la tomate, cette famille comprend deux gènes fonctionnels ainsi qu’un pseudogène. Finalement la stabilité épigénétique reste un facteur majeur pouvant avoir un impact essentiel sur les stratégies de sélection végétales. En outre nous avons fait une caractérisation fine des différents aspects du développement du fruit et de la maturation.