



UNIVERSITÀ DEGLI STUDI DELLA TUSCIA



DIPARTIMENTO DI SCIENZE AGRARIE E FORESTALI

Dipartimento di Scienze Agrarie e Forestali

Corso di Dottorato di Ricerca in
Scienze delle Produzioni Vegetali e Animali - XXXI Ciclo

Titolo:

Phenotypic and genotypic characterization of European tomato
(*Solanum lycopersicum* L.) landraces for the study of traits related to the
reproductive process

Tutor:

Prof. Andrea Mazzucato

Candidato:

Dr. Alessandro Riccini

Tutor del corso di dottorato:

Prof.ssa Stefania Masci

Socrate ci disse: “la vita senza ricerca non è degna di essere vissuta. Penso che alludesse alla curiosità, più che alla conoscenza. In ogni società umana in ogni momento e a tutti i livelli, i curiosi sono all'avanguardia.

(Roger Ebert)

Contents

Chapter 1

Background.....	1
1.1 The tomato - Origin, Domestication and Diffusion.....	2
1.2 The concept of landrace.....	4
Aim of the work and thesis structure.....	6
References.....	8

Chapter 2

Phenotyping analyses of Unitus and Traditom Core Collection

2.1 Introduction.....	11
2.1.1 The Italian landraces.....	11
2.1.2 The Traditom project.....	14
2.1.3 Unitus in the Traditom project.....	16
2.2 Materials and methods.....	17
2.2.1 Plant materials.....	17
2.2.2 Phenotyping.....	17
2.2.3 Statistical analyses.....	17
2.3 Results.....	19
2.3.1 The composition of the Unitus collection.....	19
2.3.1.1 Correlation analysis.....	20
2.3.1.2 Analysis of variance.....	22
2.3.1.3 Multivariate analysis of principal components.....	25
2.3.2 The composition of the Traditom Core Collection (CC).....	26
2.3.2.1 Grouping of accessions in known types.....	27
2.3.2.2 Analysis of variance.....	29
2.3.2.3 Firmness and colour analysis.....	30
2.3.2.4 Correlation analyses.....	31
2.3.2.5 Multivariate analysis of principal components.....	32
2.4 Discussion.....	35
References.....	37

Chapter 3

Evaluation of floral and reproductive traits in the Traditom Core Collection: phenotypic and association mapping analyses

3.1 Introduction.....	39
3.2 Materials and methods.....	44
3.2.1 Plant material and data collection.....	44
3.2.2 Fertility traits phenotyping.....	44
3.2.3 Selection of a sub-core collection.....	44
3.2.4 Measurement of ovule and ovary size.....	45
3.2.5 Stigma position (SP) phenotypic data.....	45
3.2.6 Seed weight (SW) phenotypic data.....	46
3.2.7 Statistical analyses.....	46
3.2.8 Genotypic data and structure analysis.....	47
3.2.9 Association mapping and linkage disequilibrium.....	48
3.3 Results.....	49
3.3.1 Fertility traits phenotyping.....	49
3.3.2 Correlation analysis.....	50
3.3.3 Dimensional analysis of reproductive organs at anthesis at fruit ripening.....	51
3.3.4 Investigations on stigma position (SP).....	52
3.3.4.1 Populations object of the study and phenotyping.....	52
3.3.4.2 Estimation of heat sensitivity by stigma position.....	53
3.3.4.3 Distribution of stigma position in different tomato types.....	53
3.3.4.4 Genotyping and linkage disequilibrium.....	55
3.3.4.5 Structure analysis of Traditom CC.....	55
3.3.5 GWAS analyses.....	57
3.3.5.1 Stigma position (SP) association mapping.....	57
3.3.5.2 Seed weight (SW) association mapping.....	60
3.4 Discussion.....	62
References.....	68

Chapter 4

Bulk RNA-seq analysis to dissect stigma position regulation in tomato

4.1 Introduction.....	74
4.2 Materials and methods.....	77
4.2.1 Plant material and data collection.....	77
4.2.2 Material collection and RNA extraction.....	78
4.2.3 RNA-sequencing, filtering and mapping.....	79
4.2.4 Enrichment analysis.....	80
4.2.5 Tissue-specific genes.....	80
4.2.6 RT-PCR and RT-qPCR.....	80
4.2.7 Statistical analysis of RT-qPCR expression data on single genotypes.....	82
4.2.8 Identification of structural variants from RNA-Seq Data.....	82
4.3 Results.....	83
4.3.1 The “stigma position” population.....	83
4.3.2 Dynamic of stamen and pistil growth in flower buds.....	88
4.3.3 Bulk RNA-seq analysis.....	91
4.3.4 Sequence reads filtering and mapping.....	91
4.3.5 Analysis of differentially expressed genes (DEGs).....	92
4.3.6 Validation of RNA-seq data by qRT-PCR.....	93
4.3.7 GO Enrichment and network analyses.....	96
4.3.8 Transcription factors listed among DEGs involved in stigma position.....	99
4.3.9 Analysis on tissue specificity.....	101
4.3.10 RT-PCR and qPCR analysis of the selected DEGs on single genotypes.....	102
4.3.11 Identification of structural variants from RNA-seq Data.....	106
4.4 Discussion.....	108
4.4.1 Phenotyping of SP in tomato landraces and selection of contrasting genotypes.....	108
4.4.2 RNA-seq analysis, specificity and enrichment.....	108
4.4.3 Validation of DEGs on single genotypes.....	110
4.4.4 Identification of structural variants from RNA-Seq Data.....	113
References.....	115
Conclusions.....	120
Supplementary materials.....	123

Abstract

This work of thesis has been carried out in the frame of the Traditom project, developed from 2015 to 2018, with the aim to foster the conservation of traditional tomato varieties and increase their competitiveness in the global and local market (www.traditom.eu). Phenotyping and genotyping data of the collection used in the project, collected in Viterbo and in other experimental fields, were used to develop a series of investigations related to fertility traits, with two specific focuses on seed size and stigma exertion of tomato.

The study of fertility traits is important in tomato and other species because they are linked to the processes that allow the development of the final product of the plant, the fruit, in order to optimize and increase the production.

Seed size is in general directly related to plant vigour and information on QTLs and candidate genes could be useful in species where the seed is the edible part.

The phenomenon of variability in stigma position consist in the definition if the stigma is under or upper the anther cone tip. The cultivated tomato normally has an inserted stigma (within the anther cone), but a number of landraces have an exerted stigma (protruding from the anther cone). In addition, elongation of the style is enhanced by heat stress.

The molecular process that in cultivated tomato underlies the stigma position is poorly known and to gather information on this aspect is very important to ensure tomato fertility (higher fruit set) and self-pollination. On the other hand, controlling the position of the stigma could be an alternative to emasculation for hybrid seed production.

The results of this thesis contribute to study and characterize the European collection of tomato landraces, particularly focusing on that factors controlling fertility trait. The results were achieved with association mapping and transcriptomic analyses, identifying candidates QTLs and genes mainly involved in the control of seed size and stigma position in tomato.

Chapter 1

Background

1.1 The tomato - Origin, Domestication and Diffusion

1.2 The concept of landrace

Aim of the work and thesis structure

References

1.1 The tomato - Origin, domestication and diffusion

The tomato (*Solanum lycopersicum* L.) is a vegetable crop widely grown around the world and it is the second most consumed vegetable after potato. All cultivated forms of tomato belong to the species *Solanum lycopersicum*, as defined by Linnaeus in 1753. Despite that, Philip Miller placed the species in its own genus, and he classified it as *Lycopersicon esculentum*. Genetic evidences showed that the Linnaeus classification in the genus *Solanum* was correct, making the tomato *S. lycopersicum* name correct (Peralta et al. 2006).

Wild species of tomatoes are native of western South America, the Andean region, including parts of Chile, Colombia, Bolivia, Ecuador and Peru (Rick, 1973; Taylor, 1986; Amadei, 1990).

When tomato was brought to Europe, it had already reached an advanced stage of domestication. Rick (1984) found a close similarity between the early European cultivars and the primitive forms of the “cherry tomato” (*S. lycopersicum* var. *cerasiforme*) found in Mexico and Central America (Ranc et al. 2008; Causse et al. 2013).

In fact, the wild forms of tomato were moved from Peruvian-Ecuadorian regions to Southern Mexico, where tomato was first domesticated and improved (Jenkins, 1948).

The introduction of tomato in Europe from Central America took place at the beginning of the sixteenth century, but at the first it was only considered a botanical curiosity. It was only in the seventeenth century that the fruit was discovered as an edible product and its cultivation rapidly diffused through the Mediterranean area (Mazzucato et al. 2008).

However, the introduction of tomato in Europe represented a genetic bottleneck, narrowing the genetic diversity of the cultivated tomato germplasm (Rick 1976).

The cultivation of tomato in Europe was so successful that it found a second centre for diversification and several new types were selected and reintroduced into North America in the eighteenth century (Rick, 1976; Garcia-Martinez et al. 2006; Fig. 1.1).

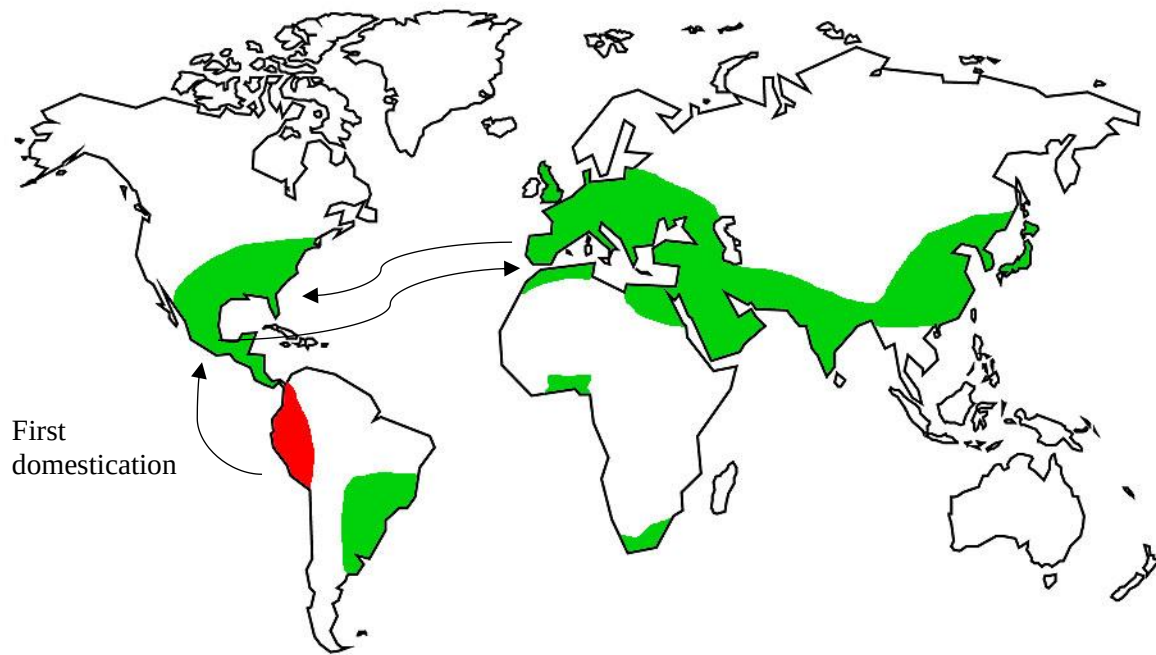


Fig. 1.1 Region of origin (red) and regions of cultivation (green) of tomato. Is indicated the first domestication in Mesoamerican regions, such as Mexico. Other arrows indicate introduction in Europe and subsequent reintroduction in North America. Image modified from W. Schuchert (Universitat Hamburg website) <http://www1.biologie.uni-hamburg.de/b-online/schaugarten/Lycopersiconlycopersicum/Tomato.html>

Between the seventeenth and nineteenth century, a selection process carried out in the Mediterranean countries by farmers, generated those that are commonly defined as landraces or traditional varieties. This means that landraces and modern cultivars were in competition for more than one-hundred years. On the other hand, other forms of tomato grown for curiosity were called “heirlooms” or “garden varieties”.

In some countries, intensive breeding started since the 20th century and a huge array of morphologically different cultivars have been created up to the development of modern tomato varieties (mainly hybrids) with various shapes, colours and sizes (Bai and Lindhout, 2007).

1.2 The concept of landrace

The definition of landrace is inherent in the process of domestication of wild plants. In fact, empirical selection by man combined with natural selection selected phenotypes and genotypes with certain desirable characteristics and suitable to be cultivated in a particular area (Zeven, 1998).

The first definition of landrace was that given by von Rümker (1908) as “*landraces are varieties, which in the region, of which they carry the name, were grown since time immemorial*”.

This definition could make landraces appearing as fixed and inalterable materials, but they have been and actually are in a constant state of evolution. In fact, movement of genetic material over short or long distances, as well as local crosses and segregations brought landraces into a state of evolution and of reciprocal competition; this process could cause disappearance, replacement or creation of new forms of landraces (Zeven, 1998).

Several definitions of landraces were adopted by many authors during the 20th century; a commonly accepted and reported as the best by Zeven’s review (1998) and others, is the definition of Mansholt’s (1909): “*a landrace is a variety (domesticated) with a high capacity to tolerate biotic and abiotic stress (locally adapted) resulting in a high yield stability and an intermediate yield level under a low input agricultural system.*”. Successively, other definitions eliminated the link with the agricultural system or the “traditional” method of selection (absence of formal breeding) (Negri et al. 2009).

Finally, an updated and “dynamic” definition of landraces was given by Casanas et al. (2017), arguing that landraces are “*plant materials consisting of cultivated varieties that have evolved and may continue evolving, using conventional or modern breeding techniques, in traditional or new agricultural environments within a defined ecogeographical area and under the influence of local human culture*”.

The last part of this definition distinguishes traditional from modern accessions. Moreover, landraces are appreciated by a niche of consumers for other elements of distinctiveness: the organoleptic properties, also conferred by the particular area or agronomic cultivation techniques, the methods of distribution and storage of the harvested fruits and finally consumers’ affection, that implies cultural aspects (Andreakis et al. 2004; Causse et al. 2010; Corrado et al. 2013).

As for most cultivated species, the needings of modern agriculture to grow resistant plants, mechanical harvesting and commercialize the final product in long temporal frames, determined the affirmation of cultivars with higher yield and defined with specific commercial parameters. Local markets were replaced by global markets, with the almost complete disappearance of landraces. The adoption of modern regulation procedures aimed to protect the breeder’s rights and certificate the genetic seed quality, further contributed to exclude landraces from the seed market and wide cultivation (Corrado et al. 2014). Beside the unique morphological and organoleptic characteristics,

landraces often present several defects (cracking, fasciation, blossom-end rot, blossom-end scar condition, lack of colour homogeneity etc.). Many times, landraces were cultivated outside their selection area, thus increasing problems of tolerance to biotic and abiotic stresses. Lack of shelf life and/or consistency is another important deficiency of landraces.

Nevertheless, conservation (reservoir of genes) and valorization (new forms) of landraces should be an important task of modern breeding activities, because they preserve a valuable genetic variability and new materials can be produced after correcting defects with the available breeding technology.

Aim of the work and thesis structure

First aim of this thesis work was to characterize and study populations of European landraces of tomato contributing to the goals of the Traditom project (www.traditom.eu) and perform a particular focus on reproduction related traits. The objective was achieved investigating a population of mainly Italian landraces belonging to the tomato germplasm collection held at the University of Tuscia and the core collection selected within the European project.

The present thesis is composed of four main chapters.

Chapter 1 represents the general introduction and provides the essential background on origin and domestication of the tomato plant.

Special attention is given to the process of selection and the development of traditional varieties in Europe.

Chapter 2, 3 and 4 are organized in the form of scientific papers and report the main results of the thesis.

Chapter 2 describes the results of the phenotypic characterization of the Italian-Unitus collection and of the Core Collection selected during the European Traditom project. After a brief introduction of the Traditom project and a description of the Italian landraces scenario, the results of phenotyping, correlation analyses and multivariate analyses on both collections are illustrated.

After a general phenotypic investigation of the landrace collection, focus was given to describe flower and other reproduction related traits in order to perform correlations and genome wide association analysis using genotypic data produced within the project.

Accordingly, **Chapter 3** deals with investigations on fertility-related traits of tomato, with a particular attention on stigma position and seed weight.

The chapter starts with an introduction on flower and other reproductive traits that have been studied in tomato. Results describe a focus on style elongation dynamics and a GWAS study on seed weight and stigma position.

A final part of this work took into consideration two groups of 13 accessions differing for stigma position. These groups were subjected to a bulk RNA-seq experiment in order to identify genes involved in the control of style elongation in tomato.

| Chapter 1

This was described in the **Chapter 4**, that is the result of a transcriptomic experiment for understanding the regulation of stigma position in tomato. A set of differentially expressed genes found after the bulk RNA-seq experiment were validated and the most interesting genes were further studied on single genotypes.

Each chapter reports relative discussion and the final part of the thesis reports general conclusion and future perspectives.

References

- Amadei, G., Trentini, L., & Soressi, G. P. (1990). *Il pomodoro*. Gruppo Enimont.
- Andreakis, N., Giordano, I., Pentangelo, A., Fogliano, V., Graziani, G., Monti, L. M., & Rao, R. (2004). DNA fingerprinting and quality traits of Corbarino cherry-like tomato landraces. *Journal of agricultural and food chemistry*, 52(11), 3366-3371.
- Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of botany*, 100(5), 1085-1094.
- Casañas, F., Simó, J., Casals, J., & Prohens, J. (2017). Toward an evolved concept of landrace. *Frontiers in plant science*, 8, 145.
- Causse, M., Friguet, C., Coiret, C., Lépicier, M., Navez, B., Lee, M., ... & Grandillo, S. (2010). Consumer preferences for fresh tomato at the European scale: a common segmentation on taste and firmness. *Journal of Food Science*, 75(9), S531-S541.
- Causse, M., Desplat, N., Pascual, L., Le Paslier, M. C., Sauvage, C., Bauchet, G., ... & Bouchet, J. P. (2013). Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC genomics*, 14(1), 791.
- Corrado, G., Piffanelli, P., Caramante, M., Coppola, M., & Rao, R. (2013). SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. *BMC genomics*, 14(1), 835.
- Corrado, G., Caramante, M., Piffanelli, P., & Rao, R. (2014). Genetic diversity in Italian tomato landraces: Implications for the development of a core collection. *Scientia Horticulturae*, 168, 138-144.
- García-Martínez, S., Andreani, L., Garcia-Gusano, M., Geuna, F., & Ruiz, J. J. (2006). Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars. *Genome*, 49(6), 648-656.
- Jenkins, J. A. (1948). The origin of the cultivated tomato. *Economic Botany*, 2(4), 379-392.
- Mansholt, U. J. (1909). Van Pesch Plantenteelt, beknopte handleiding tot de kennis van den Nederlandschen landbouw. pt 2. *Plantenteelt*. Zwolle.
- Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., ... & Veronesi, F. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theoretical and Applied Genetics*, 116(5), 657-669.
- Negri, V., Maxted, N., & Veteläinen, M. (2009). European landrace conservation: an introduction. *European Landraces: On-farm Conservation, Management and Use*. Rome IT: Bioversity International, 1-22.
- Peralta, I. E., Spooner, D. M., Razdan, M. K., & Mattoo, A. K. (2006). History, origin and early cultivation of tomato (*Solanaceae*). *Genetic improvement of solanaceous crops*, 2, 1-27.
- Ranc, N., Muños, S., Santoni, S., & Causse, M. (2008). A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (*solanaceae*). *BMC plant biology*, 8(1), 130.

- Rick, C. M. (1973). Potential genetic resources in tomato species: clues from observations in native habitats. In *Genes, enzymes, and populations* (pp. 255-269). Springer, Boston, MA.
- Rick, J. W. (1976). Downslope movement and archaeological intrasite spatial analysis. *American Antiquity*, 41(2), 133-144.
- Rümker, K. V. (1908). Die systematische Einteilung und Benennung der Getreidesorten für praktische Zwecke. *Jahrbuch der Deutschen landwirtschafts-Gesellschaft*, 23, 137-167.
- Taylor, I. B. (1986). Biosystematics of the tomato. In *The tomato crop* (pp. 1-34). Springer, Dordrecht.
- Zeven, A. C. (1998). Landraces: a review of definitions and classifications. *Euphytica*, 104(2), 127-139.

Chapter 2

Phenotyping analyses of Unitus and Traditom Core Collection

2.1 Introduction

2.1.1 The Italian landraces

2.1.2 The Traditom project

2.1.3 Unitus in the Traditom project

2.2 Materials and methods

2.2.1 Plant materials

2.2.2 Phenotyping

2.2.3 Statistical analyses

2.3 Results

2.3.1 The composition of the Unitus collection

2.3.1.1 Correlation analysis

2.3.1.2 Analysis of variance

2.3.1.3 Multivariate analysis of principal components

2.3.2 The composition of the Traditom Core Collection (CC)

2.3.2.1 Grouping of accessions in known types

2.3.2.2 Analysis of variance

2.3.2.3 Firmness and colour analysis

2.3.2.4 Correlation analyses

2.3.2.5 Multivariate analysis of principal components

2.4 Discussion

References

2.1 Introduction

2.1.1 The Italian landraces

Italy is one of the largest producers of fresh tomatoes of the world. With an amount of more than 6 millions of tonnes in 2017, it is the seventh world producer and the first in Europe (FAOSTAT data). This success was possible because Italy and other Mediterranean countries, as Spain, have the best natural conditions for the growth and production of tomatoes, becoming a second centre of diversification for tomato from its arrival in Europe in the 16th century.

This was the initial step that led up to the development of a great number of landraces or traditional varieties (Mazzucato et al. 2008).

From that moment, and until the 20th century, a process of selection gave rise to a great number of varieties, with a high degree of differentiation in size and shape (Garcia-Martinez et al. 2006).

From the North to the South of Italy, traditional varieties are characterized by differences in fruit shape and size. In fact, in the North bigger or flattened-ribbed types (Costoluto Genovese, Cuore di Bue d'Albenga) predominate. In the Centre of Italy are prevalent medium flattened and globular types (Spagnoletta, Mezzo Tempo, Scatolone), whereas smaller oval, round or elongate types (San Marzano, Piennolo, Ciliegino, Corbarino) were mainly diffused in the Southern regions (Soressi, 1969; Andreakis et al., 2004; Mazzucato et al., 2008).

Fig. 2.1 describes some examples of the several traditional varieties grown in Italy; some of them have a specific area of cultivation and the name was derived from the small town of original selection or cultivation (Cuore di Bue d'Albenga, Scatolone di Bolsena, San Marzano, Costoluto di Pachino etc.). Other varieties are more common types, spread in one or more regions and hardly ascribed to a “communal” boundary (Ciliegino, Pizzutello, Cuore di Bue, Da Serbo).

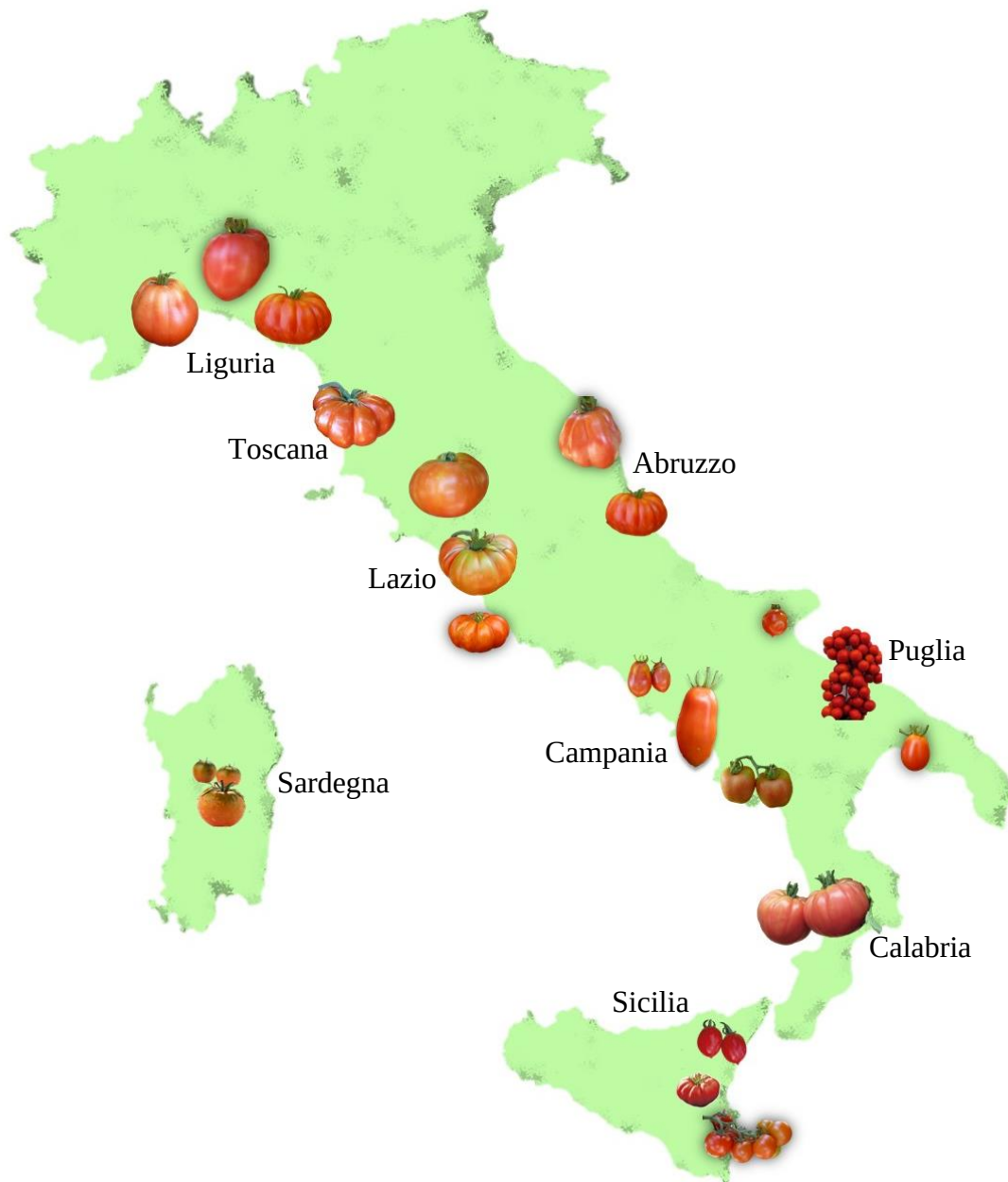


Fig. 2.1 From the North of Italy until the South are illustrated some examples of Italian landraces. Some of these have a specific origin and area of cultivation, others are generic types of an areal or region. Sample varieties are reported from Liguria: *Cuore di bue d'Albenga*, *Cuore di bue*, *Costoluto genovese*; Tuscany: *Costoluto fiorentino*; Abruzzo: *Pera d'Abruzzo*, *Mezzotempo*; Latium: *Scatolone di Bolsena*, *Pantano romanesco*, *Spagnoletta di Formia e Gaeta*; Campania: *Corbarino*, *San Marzano*, *Da Serbo*; Apulia: *Pizzutello di Apricena*, *Pomodoro Regina Torre di canne*, *Pomodoro di Manduria*; Calabria: *Pomodoro di Belmonte*; Sicily: *Pizzutello*, *Costoluto di Pachino*, *Cilieginio*; Sardinia: *Camone*.

In add, agronomic practices are also different throughout the country, for example, small varieties of the South of Italy are grown with very limited or no watering supply, thus exalting the sensorial properties of these types (Sacco et al. 2017).

Therefore, the noticeable characteristics of many traditional varieties often consist in the interaction of the genetic background with the peculiarities of cultivated soils and climatic conditions.

Some Italian landraces have been extensively characterized in the previous years, such as Scatolone di Bolsena (Mazzucato et al. 1998), Pera Abruzzese (Mazzucato et al. 2010), S. Marzano (Monti et al. 2004; Rao et al. 2006; Ercolano et al. 2014) and Pomodorino del Piennolo (Sacco et al. 2017).

To date, commercial interests on traditional tomato varieties in Italy are very poor and remain limited to local markets. Only two Italian tomato landraces possess a mark protection through PDO (Protected Designation of Origin); it is the case of “Pomodoro S. Marzano dell'Agro Sarnese-Nocerino” and the “Pomodoro del Piennolo del Vesuvio”. The “Pomodoro di Pachino”, instead, has been granted through the PGI (Protected Geographical Indication) denomination and includes different tomato types that genetically do not represent autochthonous types.

Despite the conspicuous number of landraces estimated in Italy (about 45 “named” landraces are listed in regional inventories; A. Mazzucato, personal communication), the breeding programs prevalent in the past years for the professional seed market selected varieties focusing on yield, adaptation to processing, resistance to biotic stresses and improved shelf life (Diez and Nuez, 2008; Rao et al. 2009; Causse et al. 2010; Blanca et al. 2015).

More recently, breeding implemented the selection also on different shapes and colours, with the consumer preferences that were oriented towards small and sweet tomatoes. Furthermore, climate changes increasing heat stress and water shortage drove the breeding toward the research on abiotic stresses (Grillo and Leone, 2012; Albert et al. 2016). In reality, the lack of “taste and flavour” in modern varieties is due to a complex combination of reasons, from the difficulty in selecting this complex quantitative trait, to the conflicts in harmonizing organoleptic characteristics and the marketing strategies of tomato that require long shelf-life and delayed ripening (Cocaliadis et al. 2014).

Developing of new improved forms of traditional varieties could give the possibility to many small farmers to enter in a niche market of their own region; at the same time, the consumers of these varieties will be aware to enhance the excellence of the territory in a socio-cultural improved context.

2.1.2 The Traditom project

TRADITOM (TRADItional TOMato) is a three-year project (Grant agreement number 634561) funded by the European Union's Horizon 2020 research and innovation programme. The extended title of the project is: "Traditional tomato varieties and cultural practices: a case for agricultural diversification with impact on food security and health of European population".

The TRADITOM project focused on four main objectives: information and conservation, assessment, improvement, valorisation.

Concerning the information and conservation task, the aim was an accurate description of European landrace collections, of their methods of cultivation, of the environmental characteristics of regions where they have been traditionally cultivated. Germplasm of the collection was stored in a repository at COMAV (Valencia, Spain) and made available for stakeholders. The amount of information generated during the three years project, are available through a website repository (www.traditom.eu).

The second objective of TRADITOM was to assess the genotypic and phenotypic variability present in the collection. An evaluation of traditional and modern varieties was carried out in order to enable transfer of materials and information from the project to the community of breeders, growers and consumers.

Improvement was the third intention of the project. Starting from the concept that traditional varieties need to be improved for their use in a modern context, new selections of them could be developed improving those main defects such as low yield and disease resistance without affecting their good sensory characteristics.

The last point addressed by Traditom was valorisation. To achieve this goal, a series of evaluations were required; a thorough characterization of nutritional and health compound contents, the identification of consumers' preferences, the evaluation of socio-economic factors limiting market diffusion, and the protection through PDO or PGI denominations.

Nowadays there is an increasingly promoted commercial message that holds that landraces are generally endowed with superior nutritional and sensory properties ("a flavor of the past" belief), and this has influenced the concept of landrace (Casanas et al. 2017). Effectively, objectives of breeding are mostly centred on yield, shelf-life, resistance to pests and pathogens, neglecting other characteristics such as tasty and flavour. In addition, the present methods of distribution, harvesting green/unripe fruits, certainly affect the sensorial quality.

The project has involved scientific experts on different and complementary disciplines, such as genetics, genomics, metabolomics, agronomy, sensorial and socio-economic aspects.

Is it possible to restore the appealing aspects of landraces in new versions accepted by the market minimum requirements? This was the challenge that Traditom wants to achieve. The way to obtain this improved form of landraces is to follow an accurate selection, using the tools now available for genetic improvement, agronomy and biotechnology, which is nothing more than a new generation of local varieties for a new period of agriculture (Casanas et al. 2017).

Novel hybrids could be generated, combining the quality of landraces and yield and disease resistance traits of modern accessions.

The Traditom project involved 16 main partners in Europe and 14 associate partners, coming from nine different countries.

The project assembled the largest collection of European traditional tomato varieties on which phenotypic and genotypic data were collected. The actual Traditom collection exceeds 1,800 different varieties (during the project other varieties were added to the 1472 accessions at the start of Traditom). Quantitative, qualitative and morphometrics data characteristics were evaluated; information obtained included agronomic aspects, inflorescence, fruit, quality and post-harvest performances. Collections, belonging mainly to four European countries (France, Greece, Italy and Spain), were cultivated in eight different fields in Europe; phenotypic and genotypic data collected from the entire collection were used to develop a Core Collection of 218 accessions. The Traditom Core Collection was evaluated in four sites (Italy and Spain in 2016; France and Israel in 2017).

2.1.3 Unitus in the Traditom project

This Chapter describes results of the Traditom trials carried out at the University of Tuscia (Viterbo, Italy) during the seasons 2015 and 2016.

The collection of Italian traditional varieties of the Laboratory of Biotechnology of Vegetable Crops was provided to the Traditom Consortium and grown in Viterbo during the summer 2015. Phenotyping activities were performed and collected data checked, analysed and send to a central database.

During the summer 2016, the Traditom Core Collection was grown together with a selection of commercial hybrids. Phenotypic data of all accessions were collected and analysed, yield comparisons and other commercial parameters were evaluated.

2.2 Materials and methods

2.2.1 Plant materials

The Unitus collection of 206 accessions and the Traditom CC of 218 accessions of tomato traditional varieties were grown respectively in 2015 and in 2016 at the University of Tuscia's experimental farm, Viterbo, Italy (42°260'N, 12°040'E) in open field growth conditions.

In 2015, four plants for each accession were transplanted in twin rows (100 cm between twins, 60 cm between rows and 50 cm between plants within the row). Standard agronomic tomato management practices for indeterminate tomatoes were applied, using tutors and weekly removal of lateral shoots. The trial carried out in 2016 was repeated with identical methods.

2.2.2 Phenotyping

The variables considered for characterizing the Unitus collection in 2015 and the Traditom CC in 2016 differed for some aspects. For this purpose, Phenotyping kit I used for the Unitus collection is reported in Supplementary material 2.1 and Phenotyping kit II used for traits analysed in the CC is reported in the Supplementary material 2.2.

Firmness and colour parameters were evaluated in the Traditom CC using a Turoli durometer and Chroma Meter CR-400/410 (Konica Minolta) colorimeter. Brix degree was determined by a digital refractometer (Milwaukee MA871).

2.2.3 Statistical analyses

The variables described in the Phenotyping kit I and II were analysed by correlation analysis using the PROC CORR procedure of the SAS University Edition software. Correlations were represented by heat-maps using the "Gtools V.2.2.2" software (Perez-Llamas and Lopez-Bigas 2011).

In the two collections, accessions were assigned to a specific "Type" representing categories of "named" landraces or groups homogeneous for fruit typology. Differences among types were analysed by Analysis of Variance (ANOVA) using the PROC GLM (General Linear Model) procedure of the SAS University Edition software. Type levels for the Unitus collection were 22: "Pantano romanesco", "Mezzo tempo", "Scatolone di Bolsena", "Costoluto Fiorentino", "Quarantino", "Marmande", "Spagnoletta", "Money Maker", "Da industria", "Pera Abruzzese", "Canestrino", "Albenga", "Cuor di bue", "Bell Pepper", "A Corno", "San Marzano Scatolone", "San Marzano", "San Marzano Nano", "Corbarino", "Piennolo", "Invernale", "Da Serbo".

Type levels for the Traditom CC were 24: “Bell Pepper”, “Banana-type”, “Cherry”, “Karambola”, “Marmande-like”, “Marmande pink”, “M82-like”, “Moneymaker”, “Muchamiel”, “Oxheart”, “Palosanto”, “Pera Albenga”, “Pera Girona”, “Pimpi-like”, “Ramellet”, “San Marzano”, “Yellow Pear” “Big flat” (BF), “Big-round” (BR), “Medium-elliptic” (ME), “Medium flat” (MF), “Medium round” (MR), “Small oval” (SO) and “Small elliptic” (SE).

Significant differences among ”Types” were estimated by Duncan multiple range test, applied to separate means.

Principal components analysis (PCA) was carried out for the Unitus and CC collections using the principal component (PROC PRINCOMP) SAS University Edition software procedure.

2.3 Results

2.3.1 The composition of the “Unitus” collection

The collection of tomato traditional varieties grown in 2015 comprised 25 control accessions (including 6 of the Unitus collection) and other 181 accessions belonging to the Unitus collection, for a total of 206.

Of the Unitus collection, the majority were accessions collected from local farmers in different Italian regions (68,5%). Lazio, Campania, Abruzzo, Marche and Toscana were the regions with the higher number of accessions (Fig.2.2).

Seven accessions (3.7%) came from abroad and 52 (27.4%) were without a specific local origin because coming from the market, spread throughout Italy (e.g Cuor di Bue), or representing selected cultivars.

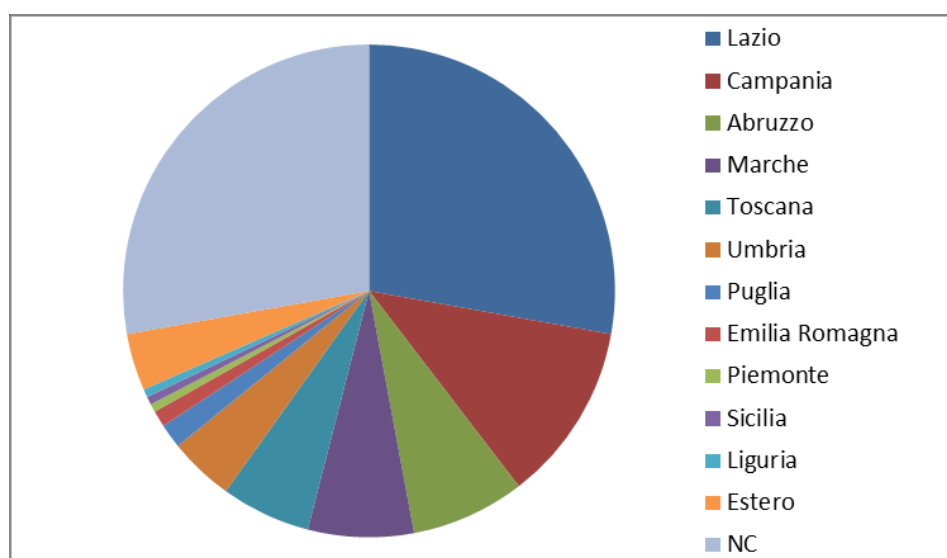


Figure 2.2. Distribution of the accessions of the Unitus collection divided by geographic origin.

Almost the total of the collection was composed by traditional varieties (97.9%); the residual part was represented by three cultivars (Ailsa Craig, Monalbo, M82) and a line originated from the segregation of an Umbrian accession (Arancione).

The varietal composition of the collection was very heterogeneous; 151 accessions out of 190 were classified in known types; 27 (14.4%) were referred to as Scatolone di Bolsena, 13 (7%) as San Marzano Scatolone, 11 (6%) as Pera Abruzzese, Da Serbo and Piennolo.

The San Marzano type, not including San Marzano Scatolone, was represented by 10 accessions (5.4%) and only 2 accessions (1.1%) represented Quarantino, Da Industria (processing), Money Maker and San Marzano Nano.

Considering fruit shape, flat types were the most represented with 64 accessions (33.7%), followed by round types (28.2%), long types (22.1%) and pear-shaped types (13.7%). Only 6 accessions (3%) represented the oxheart and bell pepper types.

The types showed in Fig. 2.3 were used to perform the ANOVA analysis.

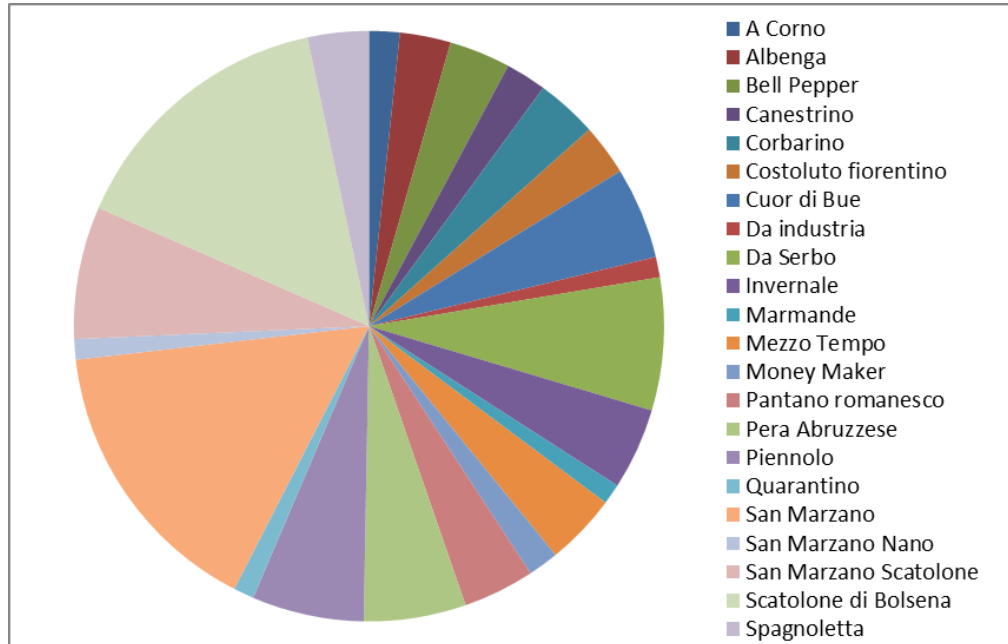


Figure 2.3 Distribution of accessions of Unitus collection divided by types

2.3.1.1 Correlation analysis

A total of 22 qualitative descriptors were considered for characterizing the accessions described above. Traits related to the plant, the leaf, the flower and the fruit were collected. Of them, 18 were ordinal variables and 4 were categoric variables (Leaf shape, Fruit color, Fruit shape, Scar shape). Fruit weight, Locule number, Brix degree and Shape Index were added to the 18 variables to perform the parametric statistics (correlation analysis, ANOVA, PCA).

The categoric variables were excluded from parametric statistics.

Mean of the variables were considered for the correlation analysis (Fig. 2.4).

Among the correlations of interest, the variable Fruit weight was positively related with Locule number. Fruit weight was also correlated with a depressed Shoulder shape, Fasciation, Puffiness and all traits describing plant height and distance between inflorescence (Fig. 2.4).

Fruit weight was negatively correlated with Shape index, Inflorescence number and Fruit load. Plants with bigger fruits were higher, but with a lower number of inflorescences and fruits.

The trait Brix degree had a strong correlation with the height of plants and Inflorescence distance.

The lack of correlation between Brix degree and Fruit weight ($r = -0.075$, not significant) showed that big fruits, as the smaller ones, could have a high Brix degree.

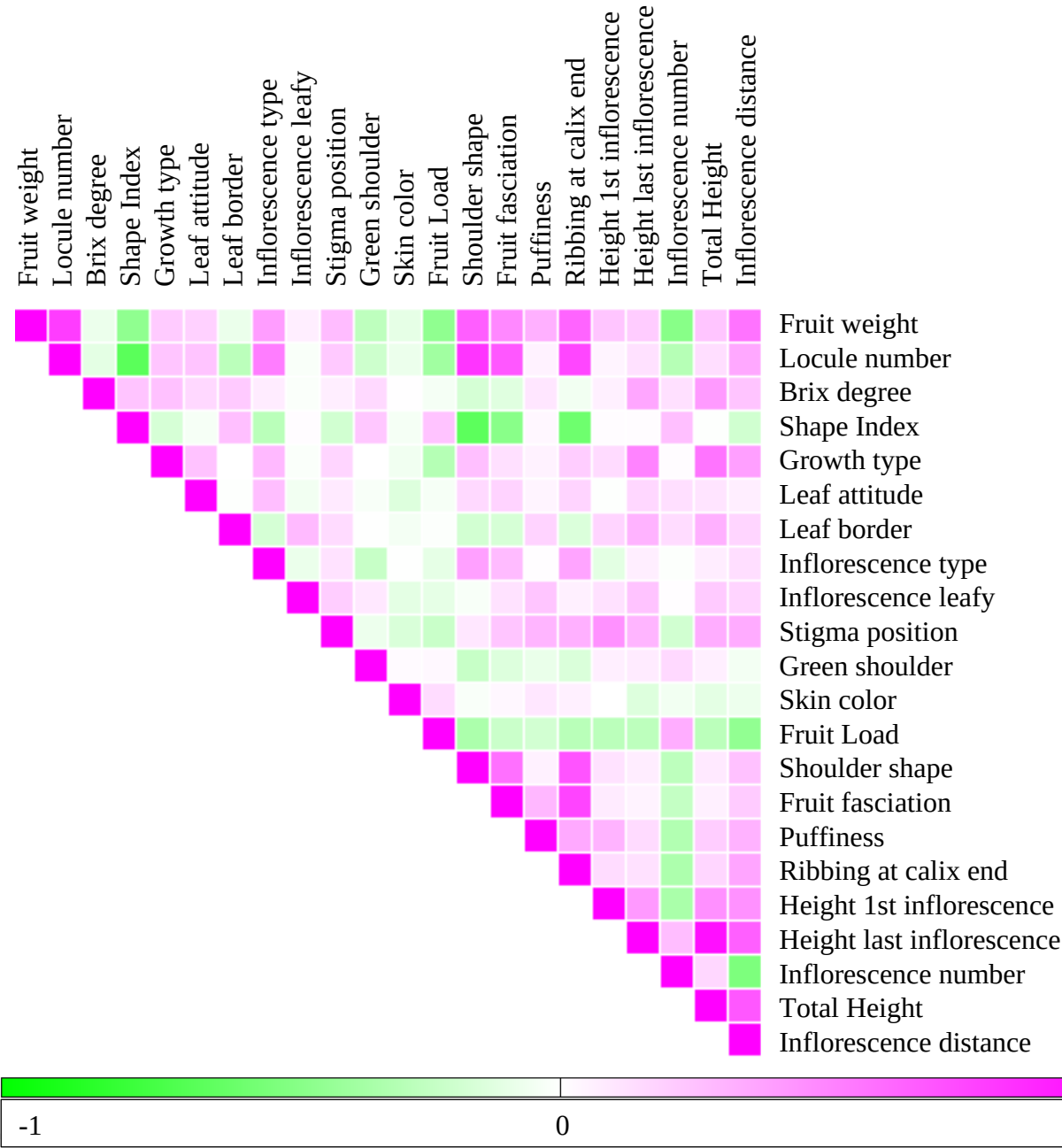


Figure 2.4 Heat-map showed the correlation among the 22 traits considered. Each trait was described in the Phenotyping kit of Traditom project (Supplementary materials 2.1).

2.3.1.2 Analysis of variance

The ANOVA test was performed for the 22 variables evaluated during phenotyping activities, considering the grouping of accessions into known typologies.

Six of the variables considered were reported for the agronomic interest (Table 2.1).

The accessions with biggest fruits resulted among flattened (Pantano, Mezzo tempo, Scatolone di Bolsena) and the globular category (Pera Abruzzese, Canestrino, Albenga, Cuor di Bue, with the exception of Bell Pepper).

Types as Corbarino and Piennolo were the smallest, despite there was no significant size difference from others medium-small types.

The typology with the highest number of locules was Pantano Romanesco, while the long types had as a rule the lowest number of locules.

The soluble solids content (Brix degree) did not show a large variability in the accessions. The type showing the highest value was named A corno (horn-shaped), but its brix value was not significantly different from that of many other varieties. On the other hand, cultivars showed the lowest values of soluble solids content (Table 2.1).

The Stigma position was stably inserted in Marmande, Da industria, A corno, Corbarino and Invernale types. A stigma exserted was noted on types such as Pera Abruzzese, Canestrino, Albenga, Cuor di bue, Bell Pepper, but also in some flattened types such as Mezzo tempo, Costoluto fiorentino and Spagnoletta.

The Piennolo type revealed the highest fruit load, while Cuor di bue had the lowest. The type with the major distance among inflorescence was Albenga, whereas Da industria was the category with the shortest distance between inflorescences (Table 2.1).

Table 2.1 Duncan's test for comparison of means referred to six variables selected for the agronomic interest. Accessions are grouped for known type and for different categories. Means with the same letter are not significantly different ($p > 0.05$).

Category	Type	Fruit weight		Locule number		Brix degree		Stigma position		Fruit load		Inflorescence distance
Flattened	Pantano romanesco	225.15	a	13.31	a	5.18	a - e	1.33	c d	4.33	a - e	28.23 b c
	Mezzo tempo	191.12	a b	10.25	b	5.04	a - e	2.57	a b	3.57	c - e	28.52 b c
	Scatolone di Bolsena	182.56	a - c	8.75	b - d	4.42	c - g	1.19	c d	2.93	d e	20.46 c e
	Costoluto Fiorentino	148.01	b - e	10.44	b	4.89	a - f	2	a - d	7	a b	24.44 b - d
	Quarantino	121.25	c - g	7.63	c - e	4.15	e - g	1.5	b - d	6	a - d	17.12 d - f
	Marmande	118.09	c - h	7.25	d e	4.55	c - g	1	d	6	a - d	15.19 e f
	Spagnoletta	98.56	d - i	8.79	b - d	4.87	b - f	1.83	a - d	6.33	a - c	21.63 c - e
Cultivar	Money Maker	82.84	e - i	2.25	f	3.95	f - g	1.5	b - d	6	a - d	24.55 b - d
	Da industria	86.49	e - i	2	f	3.78	g	1	d	6	a - d	11.67 f
Globular	Pera Abruzzese	227.56	a	6.24	e	5.45	a - c	2.18	a - d	3.73	b - e	31.54 b
	Canestrino	220.74	a	6.63	e	5.8	a b	2.75	a	3.5	c - e	31.32 b
	Albenga	218.96	a	6.53	e	5.37	a - d	2.33	a - c	3.33	c - e	42.06 a
	Cuor di bue	197.98	a b	9.11	b - d	5.12	a - e	2.33	a - c	2.33	e	30.16 b
	Bell Pepper	101.12	d - i	2.88	f	5.31	a - d	2.83	a	4.33	a - e	31.62 b
Long	A Corno San	163.46	a - d	3.83	f	5.92	a	1	d	3	c - e	29.88 b
	Marzano Scatolone San	92.95	e - i	2.25	f	5.18	a - e	1.54	b - d	5.15	a - e	21.23 c - e
	Marzano San	65.39	f - i	2.2	f	4.94	a - f	1.2	c d	5.8	a - d	21.08 c - e
	Marzano Nano	58.09	g - i	2	f	4.33	d - g	2	a - d	7	a b	13.31 e f
	Corbarino	36.36	i	2.38	f	5.03	a - e	1	d	5	a - e	20.5 c - e
Long shelf life	Piennolo	40.45	i	2.11	f	5.17	a - e	1.27	c d	7.18	a	20.45 c - e
	Invernale	134.54	b - f	9.53	b c	4.42	c - g	1	d	3.67	c - e	23.93 b - d
	Da Serbo	48.49	h i	2.39	f	4.77	b - g	1.18	c d	5	a - e	21.28 c - e

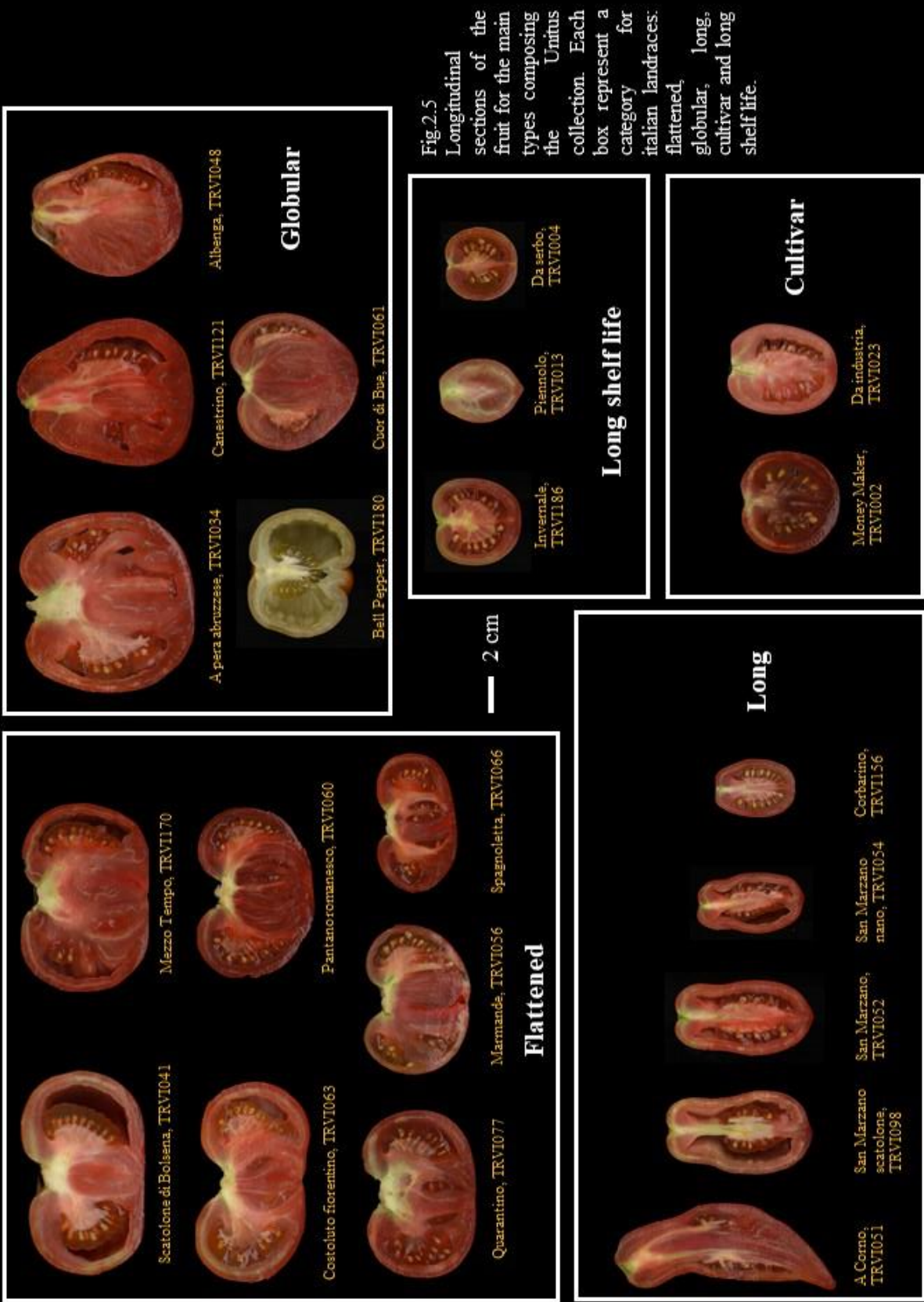


Fig.2.5 Longitudinal sections of the fruit for the main types composing the Unitus collection. Each box represent a category for italian landraces: flattened, globular, long, cultivar and long shelf life.

2.3.1.3 Multivariate analysis of principal components

Principal components analysis (PCA) was carried out including the previously described traits (22 variables). The analysis explained about 50% of variability with the first three components (PC1-PC3). Variables that loaded the PC1 were, in order, Fruit weight, Locule number, Shoulder shape and Ribbing at calix end. The PC2, was positively loaded by Total height and Height of last inflorescence. Total number of inflorescences, Leaf attitude and Inflorescence type positively explained the PC3; in contrast, PC3 was negatively loaded by Puffiness and Height of first inflorescence (Table 2.2).

Table 2.2 Coefficients for the 22 morphological variables in the first three principal components (PC1-PC3) of the PCA. Coefficients with an absolute value >0.3 are in bold.

Variable	PC1	PC2	PC3
Fruit weight	0.346	-0.076	-0.070
Locule number	0.348	-0.204	0.139
Brix degree	0.012	0.282	0.206
Shape index	-0.253	0.235	-0.054
Growth habit	0.186	0.189	0.317
Leaf attitude	0.101	0.024	0.333
Leaf border	-0.028	0.297	-0.111
Inflorescence type	0.189	-0.121	0.327
Leafy inflorescence	0.055	0.165	-0.194
Stigma position	0.189	0.157	-0.169
Green shoulder	-0.095	0.138	0.016
Skin color	-0.035	-0.074	-0.097
Fruit load	-0.227	-0.078	0.099
Shoulder shape	0.312	-0.206	0.100
Fruit fasciation	0.279	-0.169	-0.020
Puffiness	0.150	0.118	-0.341
Ribbing at calix end	0.334	-0.143	-0.020
Height of 1st inflorescence	0.154	0.247	-0.305
Height of last inflorescence	0.174	0.419	0.198
Total n of inflorescence	-0.182	0.112	0.461
Height total	0.191	0.423	0.160
Inflorescence distance	0.283	0.253	-0.129
% Variance	0.258	0.157	0.088
Cumulative % Variance	0.258	0.414	0.502

Fig. 2.6 shows the overall distribution of the accessions classified in known types according to the first two components. The grouping of accessions referred to the same type is clearly recognisable. The Moneymaker accessions were overlapped on the X-axis, as the Quarantino and Da Industria showed a similar close position, both in the third quadrant of the plane. Other types, such as San

Marzano and Mezzo Tempo, were distributed on the second and third quadrant, denoting the variability along the Y-axis (PC2).

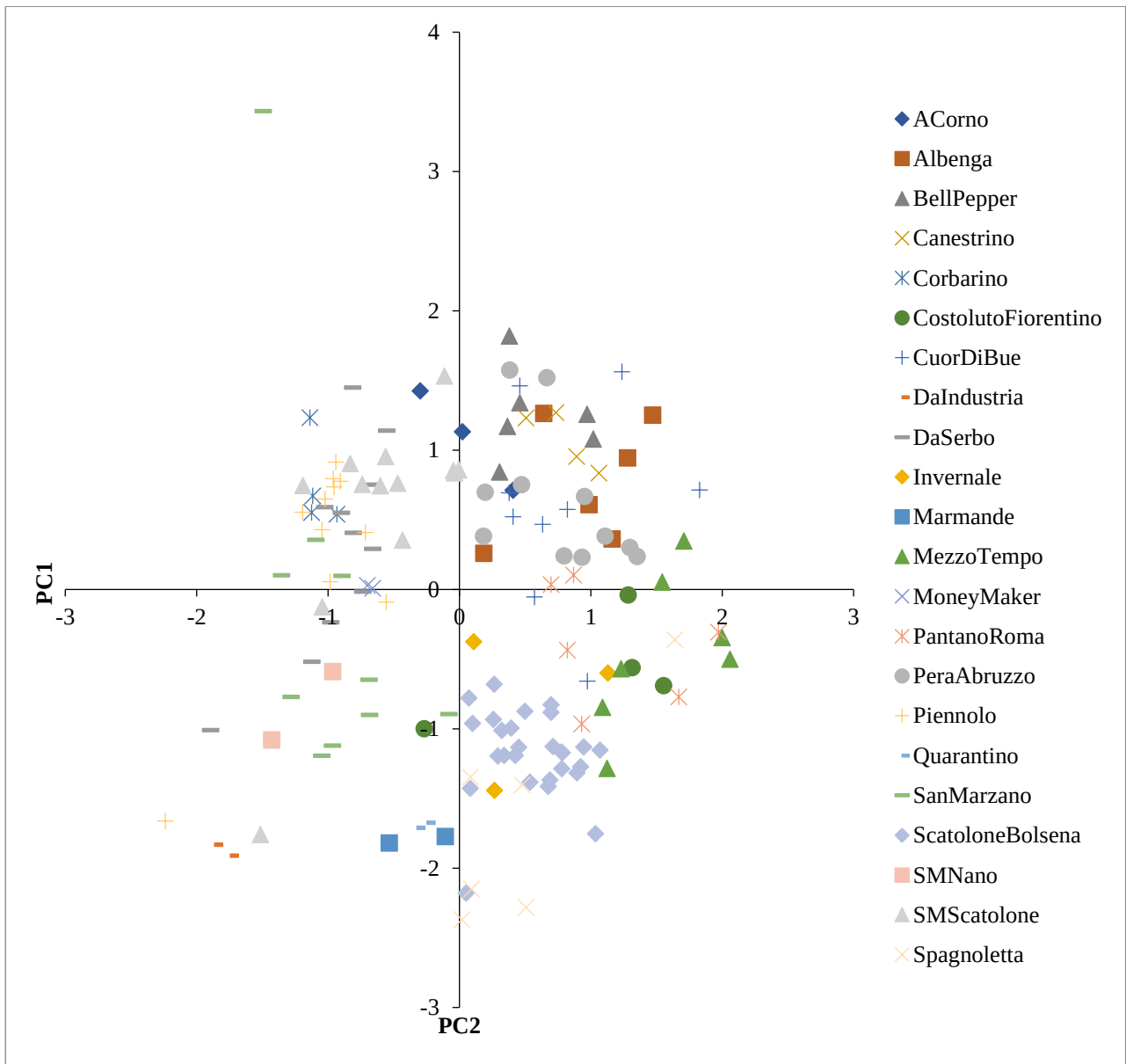


Figure 2.6 Distribution of the accessions belonging to known types along the X- and Y-axis referred to the first two principal components (PC1-PC2).

2.3.2 The composition of the Core Collection

The Core Collection (CC) was selected in the Traditom project starting from a total of 1472 accessions coming from the partners donor of collections in order to maximize phenotypic and genotypic diversity.

The morphological and molecular data that drove the selection were obtained by the Traditom Consortium during the first screening of European tomato traditional variety germplasm.

The CC population was composed of 218 traditional varieties; 98 were from Spain (21 from the Balearic, 22 from the Catalunya and 55 from the Valencia collection), 60 from Italy (4 from the Palermo, 26 from the Portici and 30 from the Viterbo-Unitus collection), 31 from Greece, 26 from France and 3 from Israel).

A total of 26 qualitative and quantitative traits were evaluated during the growing season.

Other 49 accessions representing modern varieties (mostly typologies equivalent to traditional varieties) were included in the phenotyping activities and used as comparison.

2.3.2.1 Grouping of accessions in known types

According to their provenience and phenotype, the accessions have been classified in a total of 24 different categories. Some of the categories included recognized types, such as Bell Pepper, Banana-type, Cherry, Karambola, Marmande-like, Marmande pink, M82-like, Moneymaker, Muchamiel, Oxheart, Palosanto, Pera Albenga, Pera Girona, Pimpi-like, Ramellet, San Marzano, Yellow Pear. Other accessions not recognizable as known type, were classified in more generic classes based on fruit size and shape, such as Big-flat (BF), Big-round (BR), Medium-elliptic (ME), Medium-flat (MF), Medium-round (MR), Small-oval (SO) and Small-elliptic (SE).

The classes most represented were MR and BR, with 30 and 26 accessions respectively. The M82-like class was represented by 20 accessions, while the Cherry type by 15. All other classes were represented by a smaller number of accession until the SE class that contained only one accession (Fig. 2.7).



2.3.2.2 Analysis of variance

Analysis of variance was distinctly performed for the traditional (TRAD) and modern (MOD) accessions. Categorical traits (Fruit color, Fruit shape, Shape of scar) and the almost monomorphic variable Macrocalix were excluded from the analysis.

For TRAD accessions, all variables showed significant differences among Types except for Ripening stage. The significance level was of $P \leq 0.001$ for all varieties, except than for Foliage density ($P \leq 0.05$). On the other hand, only six variables of the MOD accessions showed the maximum level of significance, while eight did not show significant differences (Table 2.3).

Table 2.3 ANOVA result for types of 23 variables morphological and agronomical divided for traditional and modern varieties. If statistical significance was found, a comparison was showed

Variable	TRAD			MOD			TRAD/ MOD comparison	
	Among accession variance (σ^2)	F	Mean	Among accession variance (σ^2)	F	Mean	P	Comparison
Brix degree	3.57	4.83***	6.05	2.59	1.79	6.34	ns	
Easiness to detach	10.07	7.23***	4.69	3.97	2.43*	3.84	***	TRAD>MOD
Flowering	61.47	2.87***	60.79	24.41	4.53***	55.02	***	TRAD>MOD
Foliage density	3.39	1.73*	5.04	1.53	1.16	5.27	ns	
Fruit homogeneity	10.79	5.50***	4.92	6.89	4.25***	5.18	ns	
Fruit load	21.96	6.67***	5.41	4.84	3.06**	5.40	ns	
Fruit weight	28343	21.03***	105.13	14798.4	5.94***	121.08	ns	
Green shoulder	3.47	4.89***	2.17	1.41	1.51	2.27	ns	
Growth habit	2.46	6.62***	2.59	0.19	0.54	2.80	ns	
Inflorescence leafy	0.59	3.21***	1.27	0.05	0.37	1.11	*	TRAD>MOD
Inflorescence type	3.23	3.57***	1.78	2.7	10.32***	1.93	ns	
Number of fruits	33100.7	35.77***	51.35	7338.64	13.19***	50.27	ns	
Ribbing at calix end	6.04	16.11***	2.15	3.79	9.12***	2.07	ns	
Ripening	77.08	1.17	108.95	28.62	2.28*	101.67	***	TRAD>MOD
Seed weight	1.12	5.82***	2.91	0.71	3.84**	2.68	*	TRAD>MOD
Stigma position	1.23	2.59***	1.73	0.23	0.52	1.50	*	TRAD>MOD
Transversal section	3.09	6.15***	1.51	0.8	1.33	1.53	ns	
Yield	2.66	3.46***	2.91	1.4	0.93	3.11	ns	

Comparison between TRAD and MOD groups of accessions evidenced a significant difference in the following variables: Easiness to detach, Flowering, Inflorescence leafy, Ripening, Seed weight and Stigma position. In all cases TRAD varieties showed a higher value than MOD. Thus, TRAD were more difficult to detach from the pedicel; whereas Flowering and Ripening were more precocious in the MOD varieties. Leafy inflorescence was present in the TRAD group more than in MOD. TRAD also had a bigger Seed weight. As expected, Stigma position was found significantly more exerted in the traditional varieties (Table 2.3).

2.3.2.3 Firmness and colour analysis

Fruits of CC accessions were also characterized for colour parameters and Firmness. In this case, the ANOVA was carried out among “Accessions”. All variables showed high significant differences ($P \leq 0.001$) among the accessions in TRAD and MOD varieties (Table 2.4). MOD varieties revealed higher variance and values of Firmness and lower variance for colorimetric variables. The Brightness component, L(D65), was significantly higher in the MOD varieties whereas the a and b variables did not differ between the two groups.

Table 2.4 Results of ANOVA for each group of accessions and comparison of the two groups relative to Firmness and colorimetric variables

Variable	TRAD		MOD		TRAD/MOD Comparison	
	Among accession variance (σ^2)	F	Among accession variance (σ^2)	F	P	Comparison
Firmness	1595	27,2***	1724	34,9***	***	TRAD < MOD
L(D65)	320	61,8***	201	52,2***	***	TRAD < MOD
a(D65)	448	67,6***	258	42,1***	ns	
b(D65)	822	89,7***	483	59,1***	ns	

Evaluating minimum and maximum values for Firmness and colorimetric parameters for TRAD and MOD, all variables, except for maximum Firmness, were higher in TRAD. This higher range of variation was expected, because the population size of TRAD was greater. The minimum value of 2.7 for a(D65) was due to the green colour of the fruit of VA2290. The VA2290 accession represents a non-ripening mutant, in fact, the same accession showed also the highest value of firmness (Table 2.5).

Table 2.5. Minimum and maximum values of firmness and colorimetric variables.

Variable	TRAD				MOD			
	Min	Acc.	Max	Acc.	Min	Acc.	Max	Acc.
Firmness	23.88	TRVA0590	88.07	TRVA2290	33.02	Marinda	88.24	DZ68
L(D65)	28.64	TRCA0460	60.95	TRVA2150	33.27	Micron	53.41	DZ68
a(D65)	2.7	TRVA2290	35.95	TRPO1590	18.19	DZ 68	34.96	DZ104
b(D65)	13.8	TRCA1210	57.79	TRVI1810	16.47	TOI LS PAN	47.47	DZ68

2.3.2.4 Correlation analyses

Means of variables used for ANOVA were used to perform analysis of correlation. The heatmap of Pearson coefficients were reported in Figure 2.8.

In agreement with the previous results obtained on the Unitus collection, Fruit weight had a negative correlation with Number of fruits and a positive correlation with Ribbing at calyx end and Easiness to detach from the pedicel (bigger fruits are more difficult to detach).

These traits characterized the fasciated accessions classified as flattened or oxheart. Flowering date and Ripening were inversely related to the Fruit load, suggesting that plants with a higher number of fruits flourish and ripen later.

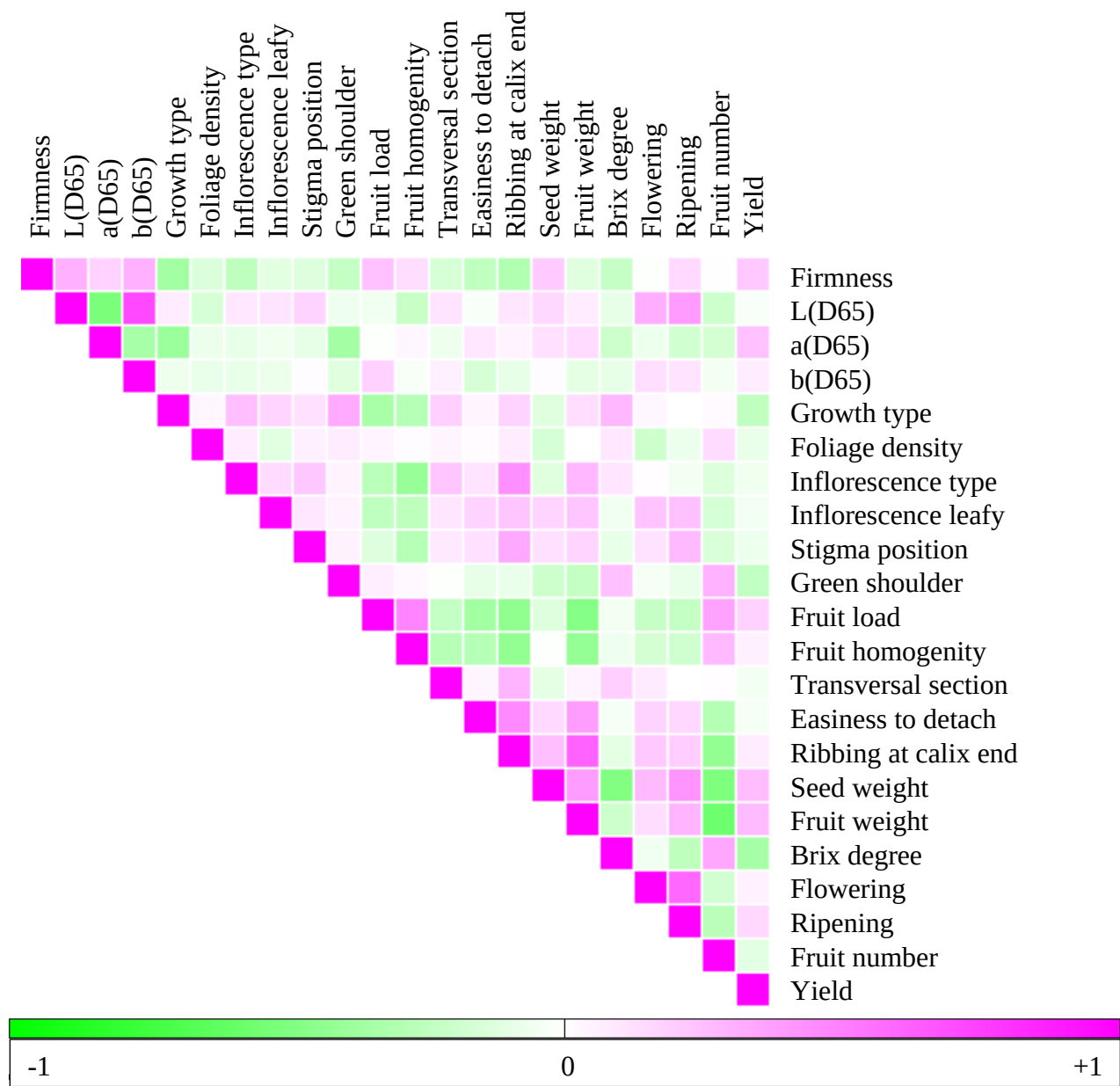


Figure 2.8 Analysis of correlation among firmness, colorimetric and morphological traits of Traditom CC.

2.3.2.5 Multivariate analysis of principal components

Principal component analysis was performed on the Traditom CC, using the 23 variables described and used to study the correlation. The analysis evidenced that the PC1, PC2 and PC3 explained about 44% of the total variability (Table 2.6).

Fruit weight and Ribbing at calix end contributed positively to the first component, while Fruit load, Fruit homogeneity and Number of fruits influenced negatively the PC1.

The second component was positively loaded by Firmness and Seed weight and negatively by Growth habit and Brix degree.

Finally, the third component was positively loaded by the b(D65) and negatively by a(D65) parameter of colorimetric measures.

Table 2.6 Coefficient values of the first three principal components (PC1-PC3) of the PCA. Data are referred to Firmness, colorimetric parameters and 19 morphological variables (Values in bold have an absolute coefficient >0.3).

Variable	PC1	PC2	PC3
Firmness	-0.100	0.364	0.125
LD65	0.132	0.140	0.527
aD65	0.003	0.173	-0.475
bD65	-0.007	0.176	0.450
Growth habit	0.115	-0.341	0.157
Foliage density	-0.034	-0.159	-0.079
Inflorescence type	0.195	-0.233	-0.004
Leafy inflorescence	0.200	-0.043	0.053
Stigma position	0.184	-0.057	0.100
Green shoulder	-0.094	-0.275	0.153
Fruit load	-0.313	0.153	0.059
Fruit homogeneity	-0.301	0.153	-0.064
Transversal section	0.124	-0.182	0.084
Easiness to detach	0.257	-0.061	-0.160
Ribbing at calix end	0.364	-0.093	-0.096
Seed weight	0.217	0.328	-0.038
Fruit weight	0.368	0.053	-0.164
Brix	-0.104	-0.361	0.080
Flowering	0.205	0.112	0.218
Ripening	0.245	0.202	0.228
Number of fruits	-0.311	-0.183	0.089
Yield	0.046	0.292	-0.127
% Variance	0.193	0.137	0.108
Cumulative % Variance	0.193	0.330	0.438

Plotting values of the PC1 and PC2, depicted a more complex clustering compared to that obtained from the Unitus collection accessions (Fig. 2.9). This was not unexpected, since the CC was selected

in order to maximize diversity and avoid clustering. Considering that some accessions were grouped in generic types such as big flat (BF), big round (BR), medium round (MR) etc., more variability along the graph was expected. Generic types were classified mainly considering characteristics of fruit morphology, while the PCA represents all the 23 traits previously described. In fact, categories such as MR or MF have a distribution in all quadrants of the graph.

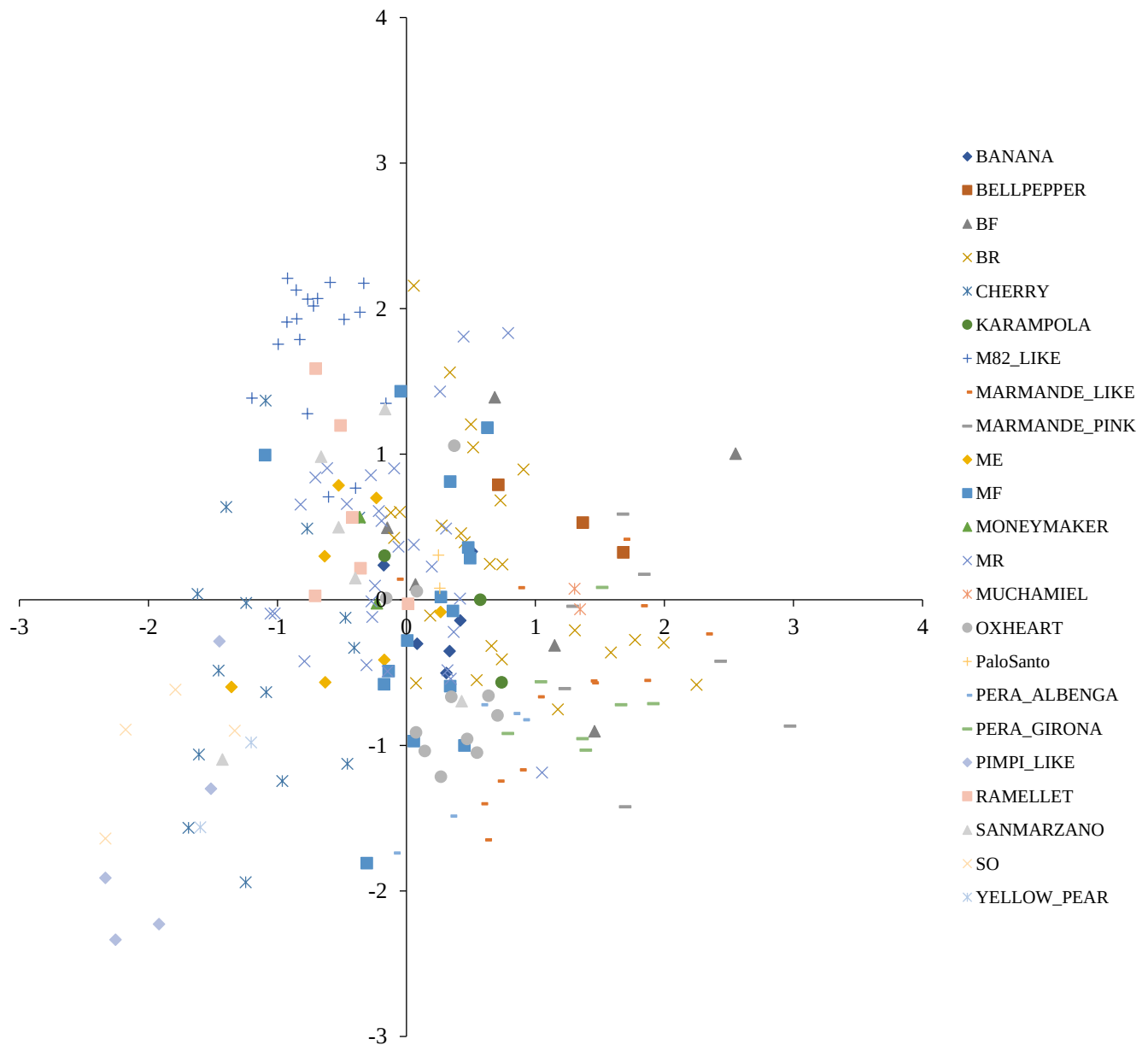


Fig. 2.9 Distribution of the Traditional accessions grouped in types along the X- and Y-axis referred to the first two components (PC1-PC2).

The comparison of TRAD and MOD graphs of PCA revealed an overall reduction of variability in the MOD accessions (Fig 2.9 and 2.10).

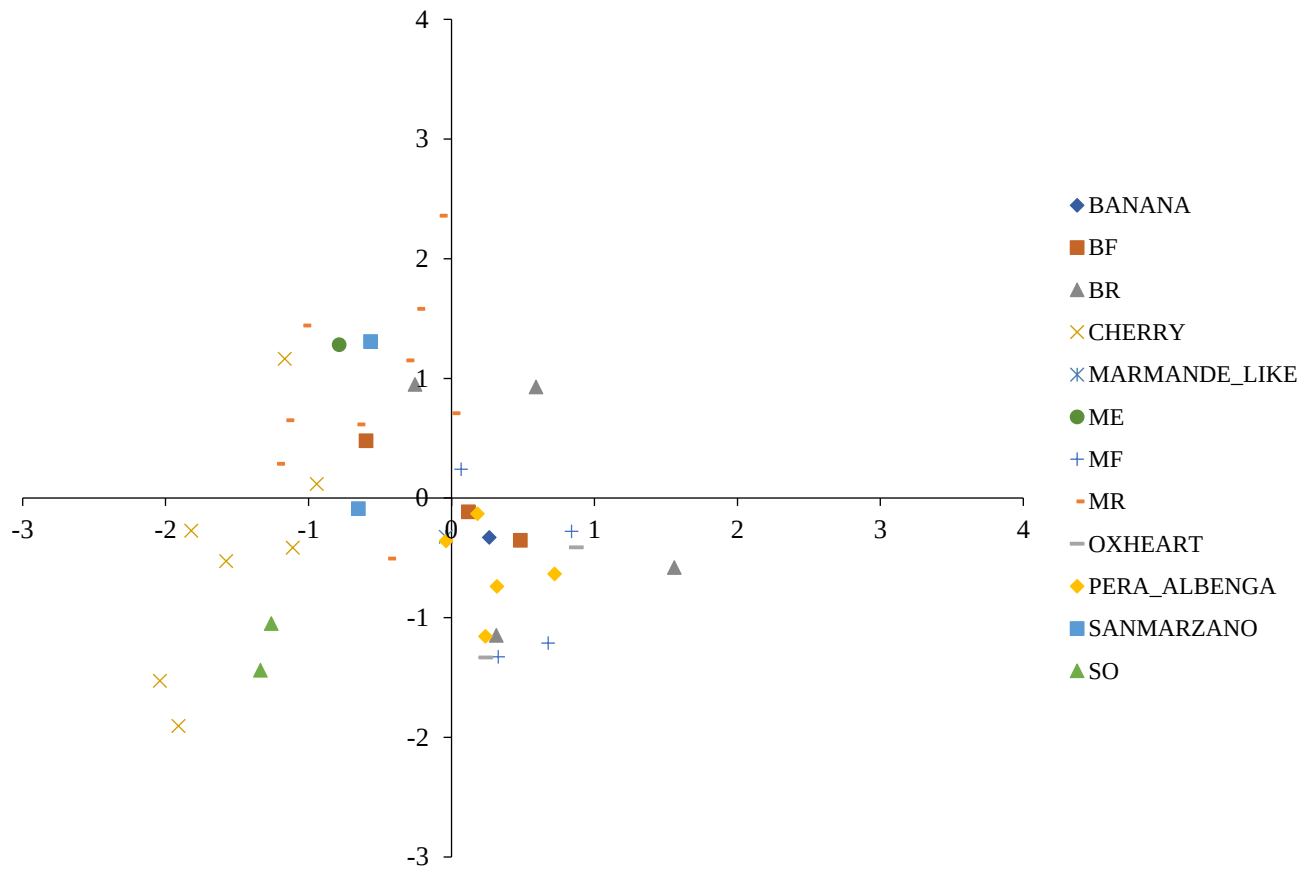


Fig. 2.10 Distribution of the Modern accessions grouped in types along the X- and Y-axis referred to the first two components (PC1-PC2).

2.4 Discussion

Phenotypic investigations on Italian and European CC landraces, reveal a wide variability in the populations, indicating that landraces could improve modern varieties for traits of interest, but also could be improved with modern needs (e.g. resistance genes) to have modern versions of traditional varieties.

As reported in the results, a simultaneous evaluation of a modern counterpart collection was carried out; the phenotyping of traditional and modern varieties gave the opportunity to obtain a comparison for important agronomic traits, in order to get useful information for breeders, farmers and other stakeholders.

The overall comparison between modern and traditional genotypes evidenced the breeding's influence on modern accessions for some agronomic traits.

It is interesting to note that modern varieties show more frequently an inserted stigma, demonstrating that breeding efficiently focused on this trait. A stably inserted stigma, in fact, guarantees higher and stable fruit set and yield.

Another focus of modern breeding was to develop varieties with a determinate growth habit. Determinate plants show more inflorescences, higher fruit set sequence, earliness flowering/ripening. Instead, the higher difficulty to detach the fruit from the pedicel in traditional varieties is probably due to the higher number of fasciated or flattened types (normally harder to detach) in this group, rather than a genetic intervention on this trait.

Seed weight was found higher in traditional varieties; this is surprisingly because seed weight is associated to seed and seedling vigour gained with domestication. It is possible that modern varieties carry introgressions from wild species that contribute to reduce the seed size.

A clear task achieved by modern breeding is the earliest flowering and ripening time, with the aim of developing precocious varieties that fasten the productive cycle.

Analyses of correlation highlighted that increasing of fruit weight was in part obtained by a selection of fruits with a higher number of locules, fasciated and ribbed. These types, mostly flattened in shape, have an indeterminate growth and a lower number of flower and fruits.

Another goal of the modern selection is the increase in soluble solids content, defined as brix degree; from the reported results, in the Italian as well as in the European CC of tomato landraces, there is a lack of correlation between fruit size and brix, denoting the variability of the latter in populations of landraces (Fig. 2.11). Probably, the recent breeding efforts in tomato for increasing the brix degree were more directed towards small than big fruits. It is known that cherry or other small tomato types are tastier and/or sweeter than big flattened or globular types. If we perform a regression analysis for

FW and brix in the landrace population and in the modern counterparts, we can note that in both cases the smallest fruit also present the highest value of brix. However, from the comparison of regression analyses, only the modern accessions show statistical significance ($P \leq 0.01$; Fig. 2.11).

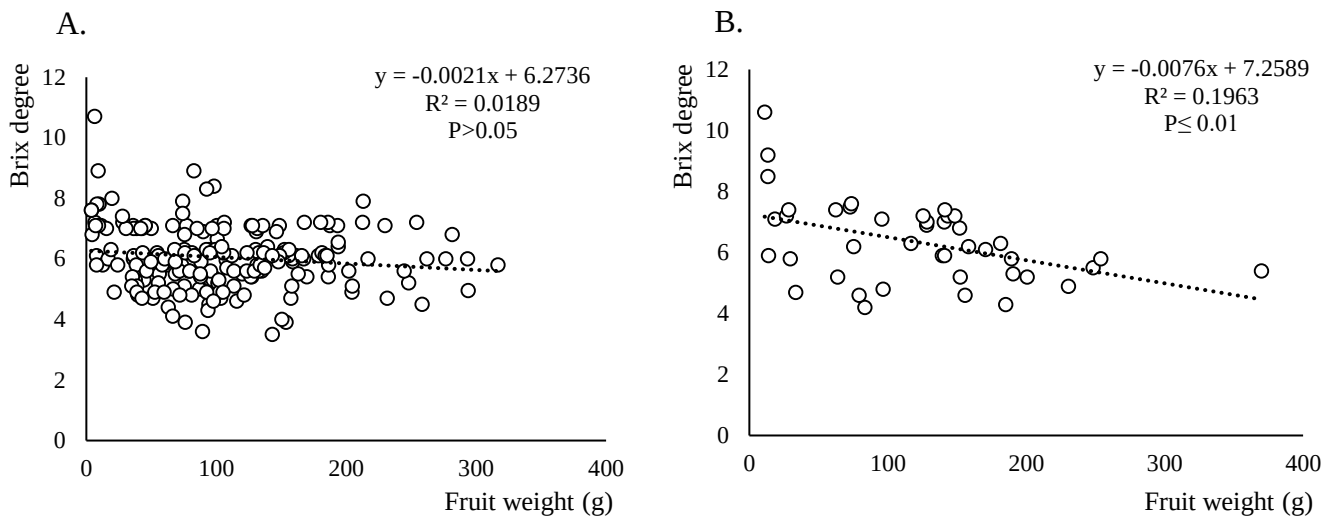


Fig.2.11 Result of regression analysis between Fruit weight and brix degree in (A) European CC landraces and (B) in modern counterparts; regression analysis of modern shows statistical significance for $P \leq 0.01$.

Analysing the overall results, a wide variation exists for phenotypic and agronomically important traits in traditional varieties. Many traditional accessions show adequate and promising characteristics, such as brix degree, fruit load, fruit homogeneity, earliness of flowering and ripening, yield.

In this and other studies it is reported that the yield of some landraces and the modern counterparts are comparable (Andreakis et al. 2004; Traditom consortium, unpublished data)

These variations make it possible the selection of traditional tomato varieties with improved features and qualities as required in the modern market.

References

- Albert, E., Segura, V., Gricourt, J., Bonnefoi, J., Derivot, L., & Causse, M. (2016). Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. *Journal of experimental botany*, 67(22), 6413-6430.
- Andreakis, N., Giordano, I., Pentangelo, A., Fogliano, V., Graziani, G., Monti, L. M., & Rao, R. (2004). DNA fingerprinting and quality traits of Corbarino cherry-like tomato landraces. *Journal of agricultural and food chemistry*, 52(11), 3366-3371.
- Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., ... & Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC genomics*, 16(1), 257.
- Casañas, F., Simó, J., Casals, J., & Prohens, J. (2017). Toward an evolved concept of landrace. *Frontiers in plant science*, 8, 145.
- Causse, M., Friguet, C., Coiret, C., Lépicier, M., Navez, B., Lee, M., ... & Grandillo, S. (2010). Consumer preferences for fresh tomato at the European scale: a common segmentation on taste and firmness. *Journal of Food Science*, 75(9), S531-S541.
- Cocaliadis, M. F., Fernández-Muñoz, R., Pons, C., Orzaez, D., & Granell, A. (2014). Increasing tomato fruit quality by enhancing fruit chloroplast function. A double-edged sword? *Journal of experimental botany*, 65(16), 4589-4598.
- Díez M.J., Nuez F. (2008) Tomato. In: Prohens J., Nuez F. (eds) Vegetables II. Handbook of Plant Breeding, vol 2. Springer, New York, NY
- Ercolano, M. R., Sacco, A., Ferriello, F., D'Alessandro, R., Tononi, P., Traini, A., ... & Delledonne, M. (2014). Patchwork sequencing of tomato San Marzano and Vesuviano varieties highlights genome-wide variations. *BMC genomics*, 15(1), 138.
- García-Martínez, S., Andreani, L., Garcia-Gusano, M., Geuna, F., & Ruiz, J. J. (2006). Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars. *Genome*, 49(6), 648-656.
- Giordano, I., Pentangelo, A., Villari, G., Fasanaro, G., & Castaldo, D. (2000). Caratteristiche bio-agronomiche e idoneità alla trasformazione di pomodori dell'ecotipo "Corbarino". *Industria Conserve*, 75, 317-329.
- Grillo, S., & Leone, A. (Eds.). (2012). *Physical stresses in plants: genes and their products for tolerance*. Springer Science & Business Media.
- Mazzucato, A., Pepponi, F., Chiocchia, G., & Soressi, G. P. (1998). Ricostituzione e valorizzazione della varietà locale di pomodoro 'Scatolone di Bolsena'. *Italus Hortus*, 5(3), 30-35.

- Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., ... & Veronesi, F. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theoretical and Applied Genetics*, 116(5), 657-669.
- Mazzucato, A., Ficcadenti, N., Caioni, M., Mosconi, P., Piccinini, E., Sanampudi, V. R. R., ... & Ferrari, V. (2010). Genetic diversity and distinctiveness in tomato (*Solanum lycopersicum* L.) landraces: the Italian case study of 'A pera Abruzzese'. *Scientia horticultrae*, 125(1), 55-62.
- Monti, L. M., Santangelo, E., Corrado, G., Rao, R., Soressi, G. P., & Scarascia Mugnozza, G. T. (2004). Il "San Marzano": problematiche e prospettive in relazione alla sua salvaguardia e alla necessità di interventi genetici. *Agroindustria*, 3(2), 161-170.
- Perez-Llamas, C., & Lopez-Bigas, N. (2011). Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PloS one*, 6(5), e19541.
- Rao, R., Corrado, G., Bianchi, M., & Di Mauro, A. (2006). (GATA) 4 DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. *Plant Breeding*, 125(2), 173-176.
- R. Rao, M. La Mura, G. Corrado, O. Ambrosino, I. Foroni, E. Perri and G. Pugliano (2009) Molecular diversity and genetic relationships of southern Italian olive cultivars as depicted by AFLP and morphological traits. *The Journal of Horticultural Science & Biotechnology*, 84 (3) 261-266.
- Sacco, Adriana & M, Cammareri & A, Vitiello & Palombieri, Samuela & R, Riccardi & P, Spigno & S, Grandillo. (2017). ITALIAN TRADITIONAL TOMATO VARIETIES: A FOCUS ON THE CAMPANIA REGION. 10.4995/TOMAVAl2017.2017.6526.
- Soressi, G. P. (1969). *Il pomodoro*. Edagricole.

Chapter 3

Evaluation of floral and reproductive traits in the Traditom Core Collection: phenotypic and association mapping analyses

3.1 Introduction

3.2 Materials and methods

3.2.1 Plant material and data collection

3.2.2 Fertility traits phenotyping

3.2.3 Selection of a sub-core collection

3.2.4 Measurement of ovule and ovary size

3.2.5 Stigma position (SP) phenotypic data

3.2.6 Seed weight (SW) phenotypic data

3.2.7 Statistical analyses

3.2.8 Genotypic data and structure analysis

3.2.9 Association mapping and linkage disequilibrium

3.3 Results

3.3.1 Fertility traits phenotyping

3.3.2 Correlation analysis

3.3.3 Dimensional analysis of reproductive organs at anthesis at fruit ripening

3.3.4 Investigations on stigma position (SP)

3.3.4.1 Populations object of the study and phenotyping

3.3.4.2 Estimation of heat sensitivity by stigma position

3.3.4.3 Distribution of stigma position in different tomato types

3.3.4.4 Genotyping and linkage disequilibrium

3.3.4.5 Structure analysis of Traditom CC

3.3.5 GWAS analyses

3.3.5.1 Stigma position (SP) association mapping

3.3.5.2 Seed weight (SW) association mapping

3.4 Discussion

References

Chapter 3

3.1 Introduction

The flower is the reproductive structure of flowering plants (also known as angiosperms or Magnoliophyta), generally carrying both male and female organs (hermaphroditism) or specialized male or female functions (monoecism or dioecism). In the hermaphrodite tomato flower, the male organs, the stamens, enclose the female organ, the pistil; male and female organs are surrounded by petals and sepals (normally in a number of 5-6 each) and inserted in the receptacle.

The stamen consists of a stalk named filament and a bilobed structure at the top of the filament, the anther. The number of stamens is variable (6-8); lateral hairs of adjacent stamens intermingle and form a cone around the pistil. The pistil is formed by two or more fused carpels and consists of a basal part, forming the ovary, that contains the ovules (macrosporangia); at the top of the ovary is present the style and the tip of the style an expanded surface that represents the stigma (Fig. 3.1).

The biological function of a flower is to effect reproduction; the reproductive development begins with the conversion of the shoot apical meristem into an inflorescence meristem, giving rise to floral meristems and finally to the flowers. Successive floral meristems are developed from the inflorescence meristem at the base of each preceding flower bud; this process culminates in the production of a terminal flower (Allen and Sussex, 1996; Park et al. 2012).

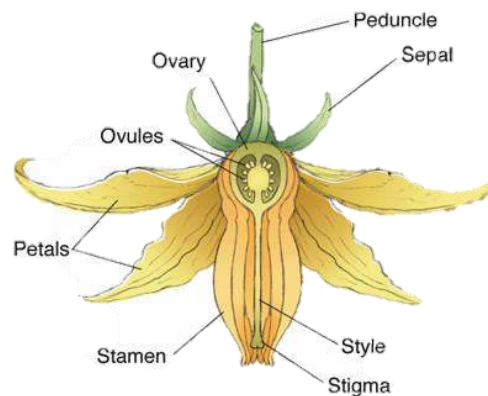


Fig. 3.1 The tomato flower and its organs

The process of pollination, a prerequisite to the fusion of sperm and eggs cells, is different with flowers that may facilitate outcrossing (transfer of pollen from different plants) or that allow selfing (pollination by the same flower or plant). The flower morphology modifications during domestication have determined changes in the reproductive habits.

Wild species relative to tomato, such as *S. pennellii*, *S. peruvianum*, *S. chilense*, *S. habrochaites*, have an exerted stigma (stigma protruding from the anther cone) and an allogamous and self-incompatible mating system (Peralta and Spooner, 2000).

Cultivated tomato (*S. lycopersicum*) is an autogamous crop with self-pollinating flowers. The pollination can occur before the opening of the flower, with a risk of unwanted cross-pollination low but always present at the anthesis.

The “selfing syndrome”, the change from allogamy to autogamy, is generally associated with a set of changes due to the morphology and function of flowers; for example, selfing species often have flowers smaller than their outbreeding relatives (Sicard and Lehnard, 2011; Goodwillie et al. 2010). The major morphological changes of the flower during the domestication are particularly clear in tomato and affected the style length, the number of carpels and locules, the ovary size and shape.

In wild tomato species the style is longer than the anther cone and the stigma exerted, promoting allogamy; differently the style is shorter leaving the stigma at the same level of anthers or inserted in transition- and cultivated-forms of tomato, promoting autogamy (*S. lycopersicum* var. *cerasiforme* and *S. lycopersicum*).

A major QTL controlling stigma position was identified as *se2.1* (Bernacchi and Tanksley, 1997) and later cloned in 2007 by Chen et al. Others few works were published on the study of alleles controlling stigma position and almost all were focused on the study of the trait using interspecific populations. Study of the stigma position trait using a landraces population of *S. lycopersicum* represents a novelty considering that variability in this kind of materials still exists.

Considering the increase in size and changes in shapes of the tomato fruit, all ancestral forms of tomato are characterized by bicarpellary ovaries with two locules, whereas the evolution into domesticated forms increased the number of locules up to ten or more (Munos et al. 2011). These and other major changes of fruit morphology are associated to variation in genes involved in meristem development, such as *WUSCHEL* (*WUS*), *FASCIATED* (*FAS*), *LOCULE NUMBER* (*LC*) and others that increase the size of the ovary and determine the shape of the developing fruit (Lippman and Tanksley, 2001; Barrero et al., 2006; Bauchet et al. 2014; Van der Knaap et al. 2014; Li et al. 2017). Increasing in fruit size was also reflected in other organs, an example is the seed. In fact, the seed of the cultivated tomato are normally larger than those of its wild counterparts (Bai and Lindhout, 2007). The unitary seed weight ranges from about 1 mg in *S. pimpinellifolium* to 4 mg of *S. lycopersicum* (Khan et al. 2012); seeds of *S. l. var. cerasiforme* have an intermediate weight of about 2.5 mg (Fig.3.2; Di Iorio, 2017).



Fig. 3.2 Increase in seed size of three different forms of tomato; the wild *S. pimpinellifolium*, the transition form *S. lycopersicum* var. *cerasiforme* and the cultivated *S. lycopersicum*.

There are two main reasons at the origin of increases of different organs, such as the seed; the first is the domestication process that indirectly selected larger organs, that increase plant vigour and yield (Doganlar et al. 2000). It is thought that larger seeds present higher germination power and seedling vigour. The second reason is that common genetic components coordinately determine the size of seeds and other organs in plants as the fruit (Linkies et al. 2010).

As for the previous examples of genes increasing ovary size during domestication, the *KLUH* gene was described in tomato to promote the growth of ovule and seed in tomato (Adamski et al. 2009). *KLUH* was responsible for a significant portion of the variation in fruit and seed weight in tomato (Chakrabarti et al. 2013). Other QTLs were associated to seed weight in previous studies (Weller et al. 1988; Grandillo and Tanksley, 1996; Doganlar et al. 2000; Orsi and Tanksley, 2009; Hernandez-Bautista et al. 2015). A common result of all studies was the identification of *Sw4.1*, that accounted for a large portion of the genetic variation for seed size. Subsequently, *Sw4.1* was associated to an ABC transporter gene on Chr 4 (Orsi and Tanksley, 2009).

In tomato, environmental conditions are important for the reproductive success and finally for fruit set. In fact, the reproductive processes are adversely affected by high temperature, resulting in a poor fertilization and fruit set, with consequences on fruit quality and yield.

Temperatures higher than 32°C or lower 15°C can have negative effects on the reproductive process, with flower abortion or with reduced pollen production and viability (Lukyanenko, 1991; Mazzucato et al. 1998; Osei et al. 2018).

In the context of climate change and increasing episodes of weather extremes for plants, the study of reproductive aspects in tomato is of crucial importance because the environmental factors have primary roles in the final yield.

Traits related to reproductive biology and its evolution, such as floral organ size, may be difficult to investigate because they require the study of continuous variation regulated by polygenes, commonly

defined as QTL. The quantitative traits do not show the discrete segregation typical of monogenic or digenic systems; in fact, these are regarded as having a more complex genetic control.

Association mapping uses natural genome-wide distribution of various alleles causing phenotypic variation together with other detectable loci (markers) in predicting marker-trait associations (Gupta et al. 2014).

Advances in high-throughput genomic platforms with the interest of studying quantitative traits, make association mapping a common and affordable approach to identify genes involved in complex traits (Georgiadis et al. 2002).

Genome wide association studies (GWAS) are therefore a powerful tool to study quantitative traits in a mixed population differing in genotypes and phenotypes of interest. The GWAS approach consists in evaluating the association of each genotyped marker with a phenotype scored among a variable number of individuals (Korte and Farlow, 2013). GWAS relies on linkage disequilibrium (LD), the non-random co-occurrence of two or more alleles. LD naturally occurs between loci in proximity and is broken down by recombination. In cultivated tomato, the extent of LD is relatively high and could be estimated up to several Mb. Because of this extent of LD, it is possible to conduct GWAS using fewer markers than with many other allogamous species having low LD mean values (Mazzucato et al. 2008; Ranc et al. 2012; Xu et al. 2013).

Large and mixed populations are the most suitable genetic material to conduct GWAS in tomato (Hall et al. 2010; Chen, 2013). In this Chapter, the natural variation of a number of traits related to flower morphology and function (including stigma position (SP), fruit-set (FS), seeds per fruit (SF), number of fruits (NF), seed weight (SW), fruit weight (FW), flowering time (FLOW), ripening time (RIP), yield (Y), amount of soluble solids content (°B)) has been studied in the Traditom CC. GWAS studies using as main collection 186 accessions of tomato local varieties were undertaken using GBS data of Traditom Consortium. The phenotypic values for SW and SP exhibited by different traditional varieties, collected over different field and years, were used to detect QTLs involved in style and seed development by association mapping approaches. In add, phenotypic and genotypic data obtained from a MAGIC population and a *S.l.* var. *cerasiforme* population, developed at INRA of Avignon (France) were used to confirm and extend the power of the analysis on stigma position (Pascual et al. 2015; Albert et al. 2016).

The revision of existing bibliographical material with respect to the study of stigma and seed weight on tomato produced a limited number of results for scientific literature respect to fruit weight or other morphological and reproductive traits, suggesting a limited attention on these topics in tomato.

Despite this, the discovery of new QTLs on SW could be useful in tomato but also extended to search for orthologous SW genes in other species such as rice, wheat, maize and beans.

At the same time, SP represents a challenging matter for the production of hybrid seed and the increasing interest related to heat stress studies on susceptible crops.

3.2 Materials and Methods

3.2.1 Plant material and data collections

The main collection used in the present experiments was the Traditom Core Collection composed in the frame of the European project TRADITOM. The Core Collection accessions are the most representative based on the phenotypic and genotypic characteristics of the whole Traditom collection of about 1,472 accessions.

The Traditom CC included 218 accessions of *S. lycopersicum* landraces. For data regarding fertility traits, including seed weight, 202 accessions were processed. For association analysis, data relative to 186 accessions were included.

Four plants per each accession were transplanted. Standard agronomic tomato management practices were applied.

3.2.2 Fertility traits phenotyping

Fertility traits investigated in the Traditom CC were: stigma position (SP), fruit-set (FS), seeds per fruit (SF), number of fruits per plant (NF), seed weight (SW), fruit weight (FW), flowering time (FLOW), ripening time (RIP), yield (Y) and the amount of soluble solids content (°B).

All traits were taken on a single plant basis. The screening of SP and SW is described in detail respectively in the 3.2.5 and 3.2.6 paragraphs. FS was evaluated as a score (1, scarce; 9, very abundant) and referred to the homogeneity of fruit set across trusses. SF, expressed as the mean number of seeds per fruit, was calculated weighting all seeds coming from a known number (from 5 to 15) of fruits. NF was the total number of fruits per plant. FW was expressed as the mean weight of all fruits from truss 1 to 4, expressed in g. FLOW was reported as the number of days from sowing until 50% of plants had at least one open flower. RIP was calculated as the number of days from sowing until 50% of plants had at least one ripe fruit. Y, expressed in Kilograms of total fruits produced per plant, was recorded as the production of the total fruits per plant in the 1st to 4th trusses. °B was determined by a digital refractometer (Milwaukee MA871) on a representative sample of fruits.

3.2.3 Selection of a sub-core collection based on fertility-related traits

Considering the fertility-related traits data recorded on 202 accessions, a selection was made in order to preserve the maximum variability in a well-defined sub-collection of 48 accessions.

To select accessions for the sub-core collection, Log_{10} transformed data were used to perform a principal component analysis (PCA) with SAS University Edition software to reduce the dimensionality of dataset.

Subsequently, a hierarchical clustering analysis was performed using the Ward's criterion on the selected principal components. Considering the cluster analysis, a dendrogram was generated and an initial partitioning was performed by cutting the hierarchical tree at the threshold value of 0.6 (Euclidean distance). At least one accession belonging to each branch of the dendrogram was selected. Other parameters were adopted to complete the representative list of 48 accessions. For this purpose, the control accessions coming from the different countries used in the Traditom project and the accessions with extreme values were included. Finally, accessions of the sub-core collection were selected in such a way that all different types were represented.

3.2.4 Measurement of ovule and ovary size

For the 48 accessions of the sub-collection, five flowers per plant at the anthesis stage were collected and immediately stored in an FAA fixative solution (125 ml Ethanol 95%, 12.5 ml Glacial acetic acid 99.7%, 25 ml Formaldehyde 37%, 87.5 ml distilled water). After 3 days in FAA solution, flowers were transferred in 70% Ethanol and stored at 4°C.

Subsequently, three flowers per accession were dissected to obtain three ovaries and a section of placenta with at least 20 ovules. A digital camera linked to a stereomicroscope allowed to take pictures of ovaries and ovules. Images obtained were analysed with the software ImageJ (Schneider et al. 2012) to measure both diameters of each ovary (polar and equatorial diameter) and ovule (major and minor diameter).

Mean data of three ovaries per accession and 10 ovules per accession, were used to carry out a regression analysis between ovule and seed size and between ovary and fruit size.

3.2.5 Stigma position (SP) phenotypic data

Phenotypic data of 186 accessions of *S. lycopersicum*, belonging to the Traditom Core Collection (CC), were obtained from three different locations: Italy (Viterbo - Unitus) Spain (Valencia - UPV)

and France (Avignon - INRA). SP data were collected during three different observations of plant development: early (first-second trusses), medium (third-fourth trusses) and late (fifth-sixth trusses). Phenotypic data of a MAGIC population developed at INRA of Avignon (255 lines derived from four *S. lycopersicum* and four *S. lycopersicum* var. *cerasiforme* parents) were collected in 2012 and 2017 field trials.

Phenotypic data on a Cerasiforme population (136 accessions belonging to *S. lycopersicum* var. *cerasiforme*) were collected at the INRA of Avignon in 2018.

To evaluate SP, a score from 1 to 4 was adopted to describe a stigma inserted (1), at the same level of the anther cone (2), slightly exserted (3), or highly exserted (4) (Fig. 3.3).



Fig.3.3 Description of the score adopted to evaluate the position of the stigma. Numbers defined an inserted stigma (1), stigma at the same level of the anther cone (2), stigma slightly exserted (3), stigma highly exserted (4)

3.2.6 Seed weight (SW) phenotypic data

SW data of 202 accessions of *S. lycopersicum* Traditom CC were collected.

Data regarding SW derived from two different evaluations fields carried out in Spain (2015) and in Italy (2016).

For each seed stock, the processing method to estimate seed size consisted in counting and weighting two replicate samples of 50 seeds each. SW data were reported as single seed weight (mg).

3.2.7 Statistical analyses

The variables described as fertility traits were analysed by correlation analysis procedure (PROC CORR) of the SAS University Edition software and represented by Heat-map using the “Gitools V.2.2.2” software (Perez-Llamas and Lopez-Bigas, 2011).

The relationship between ovule size and seed weight, and ovary size and fruit weight were studied by linear regression analysis.

Differences among types and observations for the SP trait were analysed by Analysis of Variance (ANOVA) according to a factorial design using the PROC GLM (General Linear Model) procedure of the SAS University Edition software.

The main factors were “Type” and “Measure”. Type levels were 23: “Banana”, “Big-Flat” (BF), “Big-Round” (BR), “Bell Pepper”, “Cherry”, “Karambola”, “M82-like”, “Medium-Elliptic” (ME), “Medium-Flat” (MF), “Medium-Round” (MR), “Marmande Pink”, “Marmande-like”, “MoneyMaker”, “Muchamiel”, “Oxheart”, “PaloSanto”, “Pera Albenga”, “Pera Girona”, “Pimpinellifolium-like”, “Ramellet”, “Small-Oval” (SO), “San Marzano” and “Yellow Pear”.

Measure levels were three: “Early”, “Medium” and “Late”. Their interaction (Type*Measure) was taken into consideration. Significant differences among levels of main factors were estimated by Duncan multiple range test, applied to separate means.

3.2.8 Genotypic data and structure analysis

The whole Traditom population was genotyped by the Traditom consortium using the Illumina HiSeq 2000 platform at LGC Genomics, Berlin (Germany). Genotypic data of accessions belonging to the CC were obtained. A minor allele frequency threshold of 0.05 was applied and minor SNP states were removed. After filtering, the set of markers was constituted of 2678 SNPs. Tomato genome version SL2.50 was used to assign the positions of the SNPs in the analysis.

The structure of the Traditom CC population (used to perform SW association mapping) was assessed by STRUCTURE 2.3.3 software (Pritchard et al. 2000), with no a priori information regarding population origin. The degree of admixture was estimated by setting a value of 100.000 for each run, for both burning period and Markov Chain Monte Carlo iterations. Twelve independent runs across a range of K values (K = 1–12) were made. The best number of clusters (K) was estimated with STRUCTURE HARVESTER program (Earl and Von Holdt, 2012) based on the method of Evanno et al. (2005).

Genotypic data of the MAGIC population and of the *S.l.* var. *cerasiforme* population were previously characterized at the INRA of Avignon (France) (Pascual et al. 2015; Albert et al. 2016).

Analysis of principle components of genotypic data used in SP association mapping analyses was performed with SAS University Edition software.

3.2.9 Association mapping and linkage disequilibrium

Association studies were performed using as main population the set of 186 Traditom CC accessions. Statistical models used were the simple general linear model (GLM) and the mixed linear model (MLM). The GLM model was combined with the structured association model (Q model) taking into account the structure of the collection. The MLM model was applied taking into account the kinship (K) and the structure (Q or PCA) as described by Yu et al. 2006. The kinship matrix (K) implemented in the mixed linear model (MLM), was estimated using TASSEL v. 5.2.52.

The association analyses between traits and markers was estimated with TASSEL v.5.2.52. (Bradbury et al. 2007).

Linkage disequilibrium of all SNP pairs on each chromosome were detected using the Haploview software (Barrett et al. 2004) with the following parameters: MAF > 0.05; HardyWeinberg P-value cut-off = 0; percentage of genotyped lines > 0.75. Haplotypes and tag SNPs were predicted based on the estimated LD blocks according to the definition of Gabriel et al. (2002).

Pairwise r^2 between markers was calculated for each chromosome using TASSEL. Values of r^2 were plotted against physical distance and LD decay was inferred via locally weighted scatterplot smoothing (LOWESS) using StatsDirect v.3 and testing smoothing parameters fixed to 0.2 (Breseghello and Sorrels, 2006; Sim et al. 2012).

Values of LD decay was calculated using a R v.3.5.3 script.

3.3 Results

3.3.1 Fertility traits phenotyping

The CC was investigated for ten fertility related traits, including traits directly involved in reproductive process, such as SP, FS, SF, NF. Other fertility-related traits considered were SW, FW, FLOW, RIP and Y. The amount of soluble solids content (°B) was also considered in the analysis. A summary of these phenotypic data is showed in Table 3.1.

Table 3.1 Summary of phenotypic data of fertility traits of the CC. List of abbreviations: stigma position (SP), fruit-set (FS), seeds per fruit (SF), number of fruits per plant (NF), seed weight (SW), fruit weight (FW), flowering time (FLOW), ripening time (RIP), yield (Y) and the amount of soluble solids content (°B).

	SP	FS	SF	NF	SW (mg)	FW (g)	FLOW	RIP	Y (kg)	°B
Min	1.00	1.00	7.00	6.33	1.44	3.68	50.00	92.00	0.76	3.50
Max	4.00	9.00	249.78	555.00	4.19	316.53	79.00	144.00	5.62	10.70
Mean	1.71	5.55	87.97	51.35	2.91	105.13	60.79	108.95	2.91	6.05
St. Dev.	0.72	2.22	47.42	67.88	0.55	66.61	5.10	8.21	0.99	1.03

SP ranged in 4 categories, from the minimum of 1 (inserted stigma) to 4 (highly exerted stigma). Most accessions were classified with a value of 2 (stigma at the level of the anther cone).

FS, ranging from 1 to 9 and recorded in 5 classes, was considerably high; in fact, 86 accessions were scored in the upper classes of 7 and 9.

The maximum number of SF was scored in the Greek accessions TRTH0710 Karambola (about 250 seeds, not shown). Other Greek accessions of the same type of Karambola have a high number of seeds, TRTH1070 and TRTH1100 showed 122.64 and 220.48 seeds, respectively.

The small oval Israeli controls TRIS0010 and TRIS0020 showed the higher amounts of fruits with 555 and 463 fruits per plant on average. SW ranged from 1.44 mg in TRIS0010 to 4.19 mg of TRVA1100 (Marmande type from the Valencia collection).

Another Marmande had the higher FW, with 317 g of the fruit. On the other hand, PIMPI-LIKE types showed the minimum FW.

Early FLOW is a characteristic of some determinate types (as M82-LIKE) and small accessions (PIMPI-LIKE and CHERRY). At the same time, some PIMPI-LIKE, showed the faster RIP.

Considering the Y (kg of fruit per plant), 102 accessions ranged from 3 to 4 kg per plant. Twenty-six ranged from 4.5 until the maximum value of 5.62 kg.

Finally, °B ranged from 3.5 to 10.7; 50 accessions have a °B equal to 7 or higher, with the maximum recorded in TRIS0010.

3.3.2 Correlation analysis

Phenotypic data of the CC (n = 202) were used to perform a correlation analysis for fertility related traits. Values of Pearson correlation coefficients (r) and significance of trait correlation are shown in Fig. 3.4 (A and B). Results showed positive and significant correlation of FW with SW, SF, RIP and Y. FLOW was positively correlated with RIP and SW. Other positive correlations were found for RIP with SW, SF and SP. NF was positively correlated with °B and FS.

Negative and significant correlations were found for °B with SW, RIP and Y. NF was negatively correlated with SW, SF, FW and RIP. Other negative correlations were found for FS with FW, FLOW and RIP.

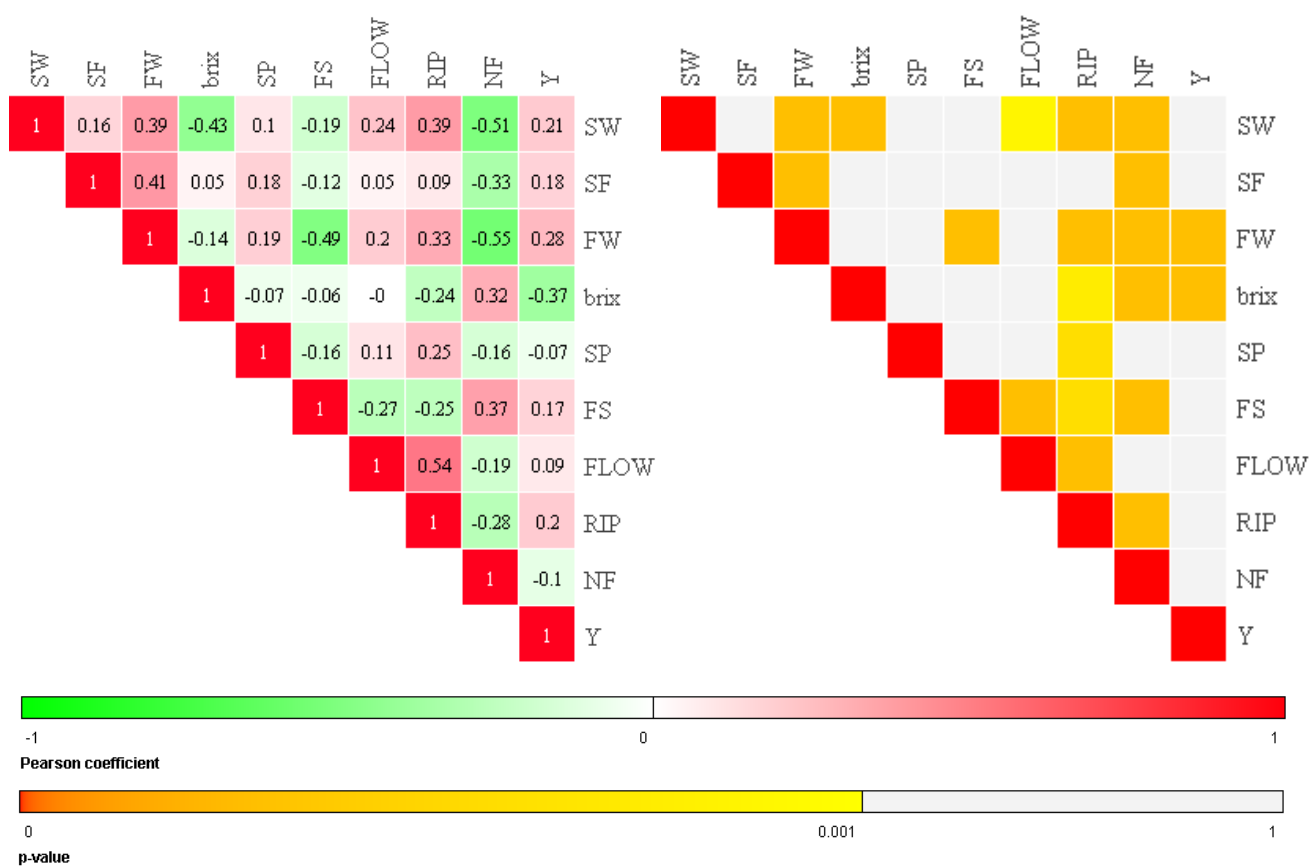


Fig.3.4 Results of correlation analysis of fertility traits of Traditom CC. A- In the boxes are reported the values of Pearson coefficient. Red or green show a positive or negative correlation between traits. B- Coloured boxes indicates correlations where the coefficient has a statistical significance of $p < 0.001$.

3.3.3 Dimensional analysis of reproductive organs at anthesis at fruit ripening

From the CC, a sub-selection of 48 accessions was explored for ovary and ovule size, in order to identifying the relationship between ovary and fruit and ovule and seed size.

Phenotypic dimensional data related to ovary and ovule of a sub-selection of 48 CC accessions were used to investigate how ovule and ovary were related in size with seed and fruit, respectively (Fig. 3.5)

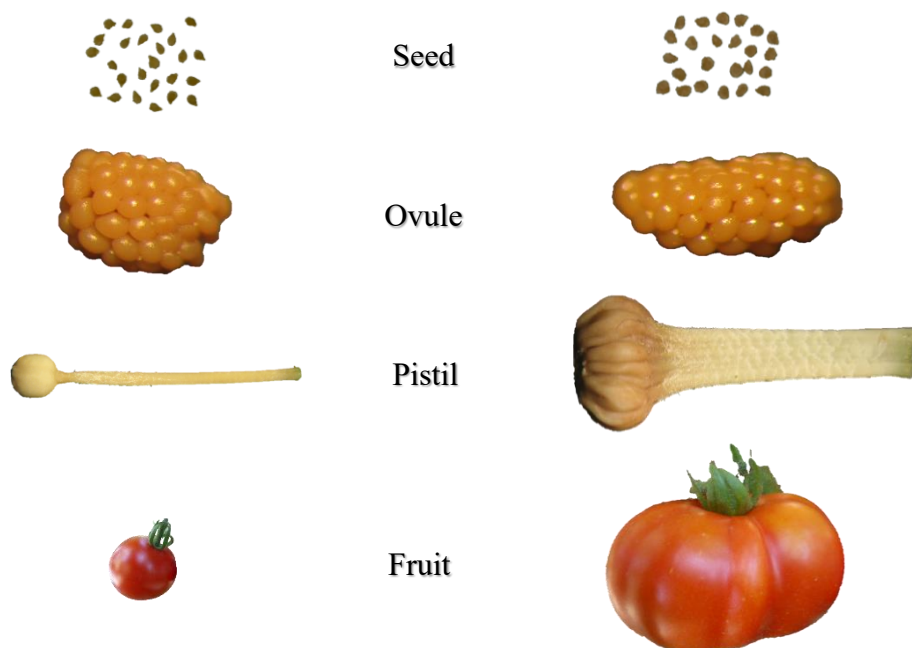


Fig. 3.5 Dimensional comparison of organs in small and big fruited tomatoes

Results of linear regression analysis showed absence of linear relationship between ovule size at anthesis and final seed weight (Fig.3.6A).

The second linear regression (Fig. 3.6B) showed a highly significant ($p < 0.001$) relationship between ovary at anthesis stage and final fruit size, explaining that fruit size is almost fully established before fertilization. The genotype that diverged more from the linear regression was the big flat TRMO0430, showing a high fruit weight compared to a medium ovary size.

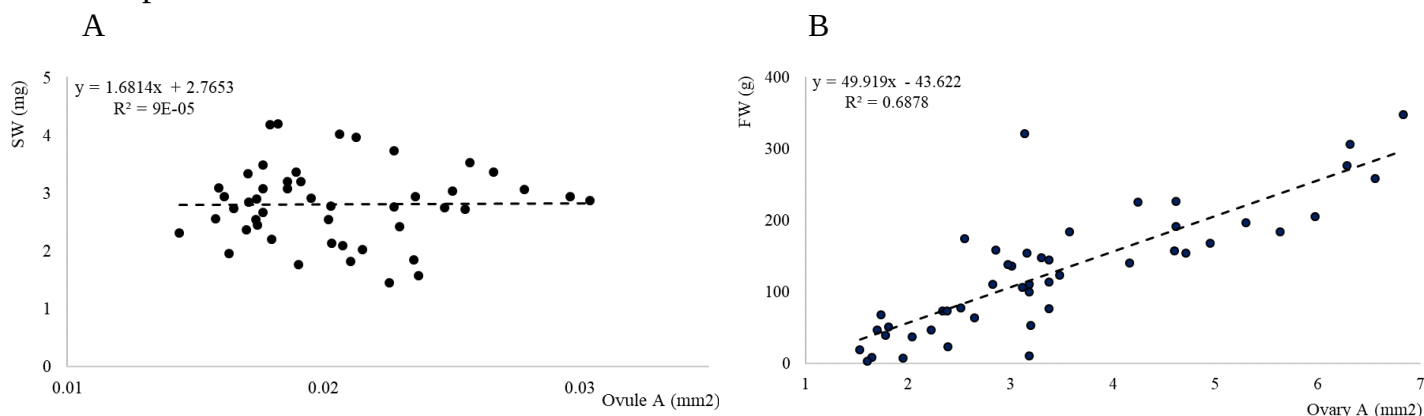


Fig. 3.6 (A) Absence of relationship between ovule and size (B) Highly significant relationship between ovary and fruit size. Flowers were examined at the anthesis stage.

3.3.4 Investigations on stigma position (SP)

3.3.4.1 Populations object of the study and phenotyping

The main population used in the study was the Traditom CC, evaluated for two years and three different field trials in Europe (Italy, France, Spain).

In addition to the Traditom CC, to perform the GWAS analysis on SP phenotypic and genotypic data of two other populations developed at the INRA of Avignon have been used; a MAGIC population (obtained from *S. lycopersicum* and *S. lycopersicum* var. *cerasiforme* parental genotypes) and a CER population (composed exclusively of *S. lycopersicum* var. *cerasiforme* accessions).

The choice to include different populations was made to extend the possibility to detect QTLs correlated with the trait of interest. Furthermore, identification of markers in more than one population, could confirm the reliability and potential interest of emerging QTLs.

The accessions of the three populations showed large phenotypic variation in SP (Table 3.2).

Table 3.2. Data of stigma position used in the GWAS

Population	Traditom CC	Traditom CC	Traditom CC	MAGIC	MAGIC	CER
Field trial	Italy (Unitus Viterbo)	France (INRA Avignon)	Spain (UPV Valencia)	France (INRA Avignon)	France (INRA Avignon)	France (INRA Avignon)
Year	2016	2017	2016	2012	2017	2018
No. of accessions	186	186	186	255	255	136
SP mean	1.73	1.66	1.57	1.55	2.19	1.94
value $\pm$ SD	± 0.75	± 0.52	± 0.77	± 0.98	± 0.92	± 0.88

3.3.4.2 Estimation of heat sensitivity by stigma position

In all field trials, accessions with a stable inserted or exerted stigma or accessions showing a SP modified in response to heat stress were observed.

Comparing data of CER and Traditom CC during normal and heat stress phenotyping (early and late detection, according with increasing temperatures in the season), most of the accessions were not affected by temperatures. On the other hand, when fluctuations were observed, the number of accessions that showed a more exerted stigma (+1, +2 in the score value) was bigger than that of accessions having a decrease in SP (-1, -2). These results highlight a reaction to extend the style beyond the anther cone in response to heat only for some genotypes (Table 3.3). The number of genotypes with fluctuations in SP was higher in the CER population, this could be explained because *S. lycopersicum* var. *cerasiforme* is a transition form between the wild species with exerted stigma and the domesticated forms of *S. lycopersicum*.

Table 3.3 Fluctuations observed during different observations of the same accessions during the growing season. CER is referred to S.lycopersicum var.cerasiforme population. TR. CC. describes the Traditom CC revealed in three experimental fields.

	<i>Insertion</i>		<i>Stationary</i>		<i>Exsertion</i>	
Population obs.	-2	-1	0	+1	+2	Total
CER	1	27	126	61	31	246
TR. CC. FRA	1	8	135	41	1	186
TR. CC. ITA	0	14	133	35	4	186
TR. CC. SPA	1	21	131	25	8	186

These results have shown that in some accessions SP is affected by environmental conditions such as higher temperatures.

In fact, there are accessions with an inserted, or exerted stigma in every condition, and others that present the exerted stigma with increasing temperatures.

3.3.4.3 Distribution of stigma position in different tomato types

To better describe the results of SP phenotyping in the Traditom CC, accessions were grouped in known (or homogeneous) types (see Chapter 2 par. 2.3.2.1) and statistical analyses were performed in three locations (Table 3.4).

Table 3.4 Results of ANOVA for type, measure in the three field trials.

	Spain	France	Italy
Type	***	***	***
Measure	Ns	*	*
Duncan's test	-	E=M<L	E<M=L
T*M	Ns	Ns	Ns

SP was monitored during the growing season with three different observations, early (E), medium (M) and late (L). A summary of ANOVA is showed in Table 3.4. No significant interaction between Type and Measure (T*M) was found. For the main factor Measure, a significant difference was found ($p<0.05$) in France and Italy, but not in Spain. In both cases, the late screening gave the highest values of SP. Significant differences among Types ($p<0.001$) were identified in all the three locations. Duncan mean separation of levels of the Type factor showed a high variability of the trait; five types (Pera Girona, Marmande-like, Pera Albenga, Muchamiel and Bell pepper) showed a mean exerted SP position (Fig. 3.7)

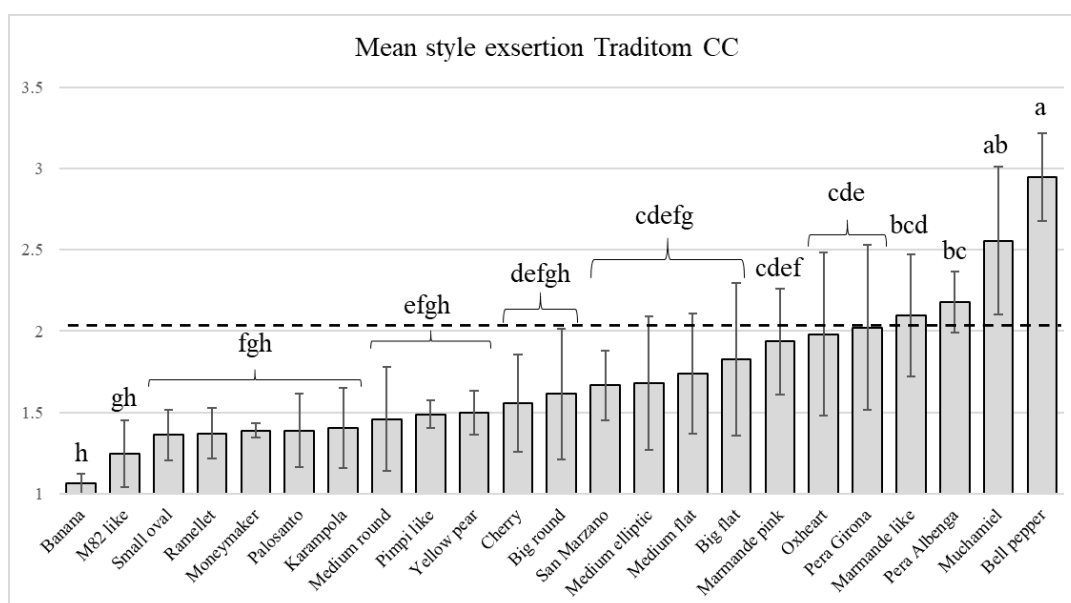


Fig.3.7 Results of phenotyping of SP of Traditom CC considering three observations during the growing season. Accessions were grouped in Types. The dotted lines indicate the value when stigma was at the level of the anther cone. Means with the same letter are not significantly different for Duncan's grouping test.

Considering these results, subsequent GWAS analysis was conducted using separated data for location and the mean of the three annotations for SP.

3.3.4.4 Genotyping and linkage disequilibrium

Genotyping of Traditom CC was performed using the Illumina HiSeq2000 platform at LGC Genomics, Berlin (Germany) yielding a total of 23231 SNPs.

The mean number of SNPs per chromosome was of 1837. After the filtering step (MAF > 0.05, remove minor SNP states) the number of markers obtained was of 2678 with an average of 208 SNPs per chromosome.

The maximum number of SNPs was found on Chr 5 and the minimum on Chr 10.

An investigation on LD decay and LD blocks across each chromosome was also performed.

Considering parameters used, the mean extent of LD blocks along the genome was of 213.04 Kb, with the maximum value on Chr 7 (485 Kb) and a minimum on Chr 10 (0 Kb) (Supplementary figures 3.1 and 3.2).

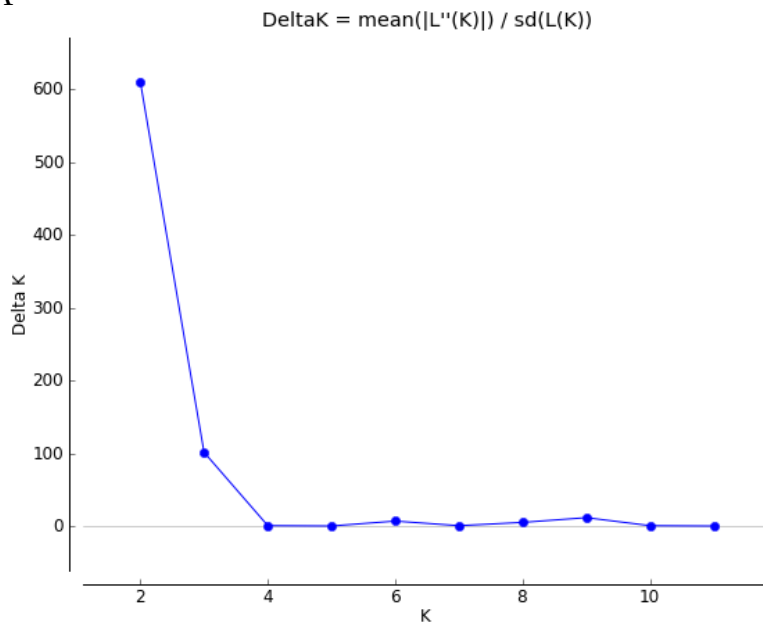
Using the 2678 SNPs, the r^2 values were plotted against the genetic distance for each chromosome. According to Sim et al. 2012, a fixed baseline of 0.2 was used to predict the LD decay. Genetic distance was estimated as the point where the LOESS curve (Cleveland, 1979) crosses the baseline r^2 value.

The distance of LD decay ranged from 0.74 Mb on Chr 3 to 17.38 Mb on Chr 12, with an average value of 4.10 Mb (Supplementary Table 3.1).

3.3.4.5 Structure analysis of the Traditom CC

The STRUCTURE analysis resulted in prediction of the best K as K=2 for the CC population (Fig. 3.5A). Out of 200 genotypes, 153 belonged to the main cluster (>75% kinship). Twenty-nine had identity for the second cluster (>75%) where 18 genotypes were essentially admixed (Fig. 3.8B). The genotypes included in the second cluster, included 13 M82-like types (processing types), four small oval/cherry genotypes plus other types including seven medium/big round fruited accessions.

A



B

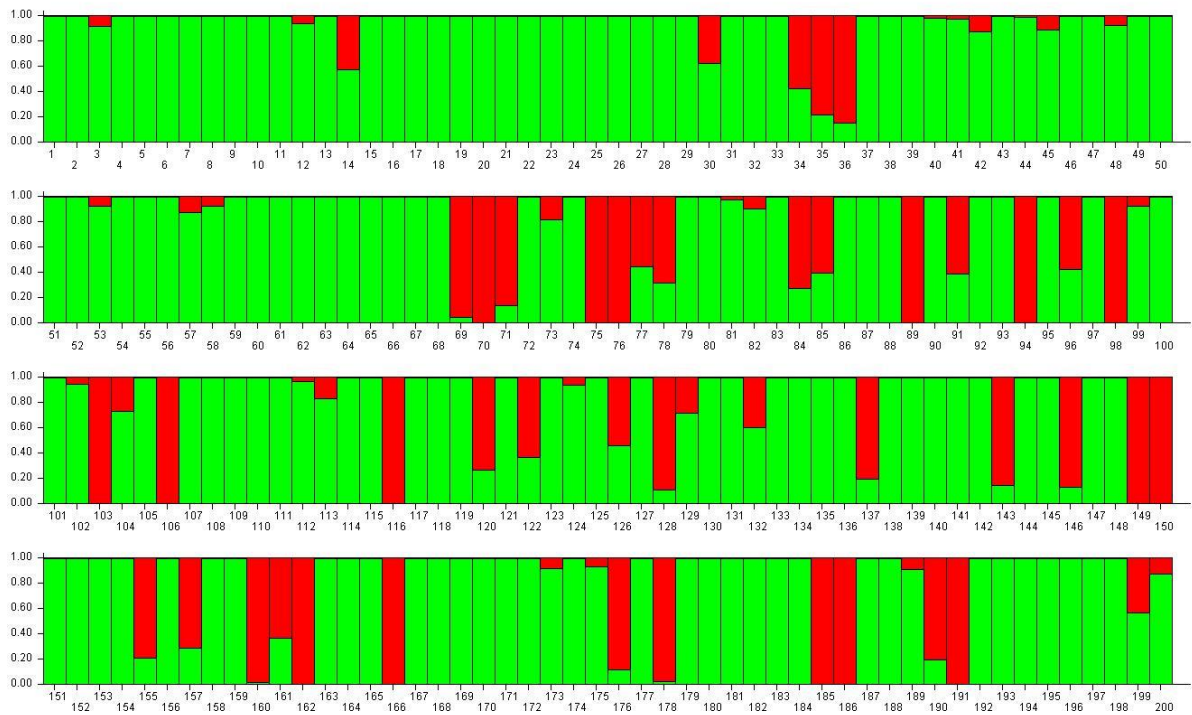


Fig 3.8 Genetic structure of the tomato collection. A) Estimation of the best number of clusters (K) by the Evanno ΔK method. B) Population stratification inferred for K = 2. Each bar stands for a genotype, which is partitioned into color segments that represent the estimated membership fraction in the K cluster.

3.3.5 GWAS analyses

3.3.5.1 Stigma position (SP) association mapping

Association analyses considering the MLM+K+PCA and GLM+PCA approach detected a total of 50 associations for the SP trait.

In the GLM approach, P-values $< -\text{Log}_{10}^6$ were considered, for the MLM approach P-values $< -\text{Log}_{10}^4$ were used.

Among these associations, the most significant were identified on Chr 2 of MAGIC population (SP_year2, with P-value of 1.57E-08. The physical position of this SNP is closely related to the previously identified *se2.1* locus (Bernacchi and Tanksley, 1997; Chen et al. 2007).

Associations were distributed over 10 chromosomes, except for Chr 6 and 10. The percentage of phenotypic variance explained by single QTLs ranged from 4.45% to 17.43%.

For the Traditom CC population, 11 significant SNPs were identified on Chr 1, 5, 8, 11 and 12 (Table 3.6). Associations were found over two fields (SP_ESP, SP_FRA) and SP_Mean. Of these, the SNP on Chr 11 (physical position 55.20 Mb) was the most significant ($P < 3.99\text{E-}07$), with a maximum R^2 of 14.12% and detected over three analyses.

Finally, investigation on LD associated the two pairs of SNPs on Chr 12 in the same LD block, with a final number of nine QTLs were identified in TR population (Table 3.5).

Table 3.5 Single nucleotide polymorphism (SNP) significantly associated with stigma position (SP) in Traditom (TR) population

Chr	Trait	Physical position (Mb) ^{a, b}	P-value	R ² (%)
1	SP_ESP	13.85	1.10E-06	15.7
	SP_FRA	79.61	7.76E-04	6.87
5	SP_Mean	0.17	1.89E-04	9.74
	SP_FRA	63.23	7.22E-04	8.85
8	SP_FRA	37.3	9.87E-06	10.48
11	SP_ESP	54.87	9.94E-04	7.81
	SP_ESP; SP_FRA; SP_Mean	55.2	3.99E-07 ; 3.46E-05 ; 2.66E-06	14.12 ; 11.7 ; 12.15
12	SP_FRA	3.42	3.70E-04	9.05
	SP_FRA	3.44	9.18E-04	7.78
	SP_ESP	39.29	9.94E-04	6.28
	SP_FRA	39.42	2.25E-04	9.46

^a SNPs in the same LD block are evidenced in gray. ^b Physical positions based on the tomato reference genome assembly v. SL2.5 at the Sol Genomics Network (<http://solgenomics.net>).

For the CER population, 28 significant SNP loci were found on nine chromosomes. Results obtained with CER returned the highest number of associations among the three populations used. Among the associations, the most significant was found on Chr 11 ($P < 1.19\text{E-}07$) with an R^2 of 17.43%. Nine SNPs reported on this analysis were enclosed in a 0.65 Mb region of Chr 11. In addition, the most significant SNP identified in the TR population and previously described, was confirmed in this region of Chr 11.

The SNPs on Chr 3, in position 52.04 and 52.05 Mb, belong to the same LD block of 133 kb.

On Chr 11, two couples of markers are in LD; positions 53.93 and 54.96 are located in a 35 kb LD block, SNPs in position 55.06 and 55.07 Mb are in a 12kb LD block. Considering LD, the number of QTLs identified from the CER population was reduced to 25 (Table 3.6).

Table 3.6 Single nucleotide polymorphisms (SNPs) significantly associated with stigma position (SP) in CER population

Chr	Trait	Physical position (Mb) ^a _b	P-value	R ² (%)
2	SP_CER	40.35	4.55E-04	12.83
	SP_CER	46.44	3.92E-04	13.22
3	SP_CER	52.04	6.04E-05	14.32
	SP_CER	52.05	3.53E-05	16.8
	SP_CER	65.88	8.53E-04	11.63
	SP_CER	67.08	5.84E-04	10.62
4	SP_CER	2.92	2.82E-06	14.45
	SP_CER	58.25	2.24E-06	14.14
5	SP_CER	6.17	4.76E-06	13.38
	SP_CER	6.4	1.25E-06	15.89
7	SP_CER	63.76	7.52E-04	10.4
8	SP_CER	0.89	7.48E-06	13.42
	SP_CER	59.19	9.97E-04	11.33
9	SP_CER	0.75	5.18E-04	10.74
	SP_CER	63.35	2.72E-04	11.48
	SP_CER	68.17	3.19E-04	12.37
11	SP_CER	3.2	1.32E-04	14.84
	SP_CER	54.93	1.48E-06	14.26
	SP_CER	54.96	1.00E-07	16.97
	SP_CER	55.06	2.73E-07	16.77
	SP_CER	55.07	4.48E-06	13.3
	SP_CER	55.15	6.18E-07	15.31
	SP_CER	55.2	1.52E-06	14.87
	SP_CER	55.36	1.19E-07	17.43
	SP_CER	55.37	2.54E-07	15.92
	SP_CER	55.58	2.60E-06	14.27
12	SP_CER	4.39	8.85E-06	12.53
	SP_CER	5.07	2.76E-07	15.86

^a SNPs in the same LD block are evidenced in grey. ^b Physical positions based on the tomato reference genome assembly v. SL2.5 at the Sol Genomics Network (<http://solgenomics.net>).

Twelve significant associations were found on seven chromosomes when using SP data from the MAGIC population.

The most significant was the SNP on Chr 2 and was the only one confirmed over two years. As for the other populations studied, it was localized a QTL in the long arm of Chr 11, the two markers in the physical position of 53.43 and 54.68 Mb (Table 3.7).

Table 3.7 Single nucleotide polymorphism (SNP) significantly associated with stigma position (SP) in MAGIC population

Chr	Trait	Physical position (Mb) ^a	P-value	R ² (%)
1	SP_year2	1.05	5.52E-05	6.94
	SP_year2	84.13	6.88E-04	4.86
2	SP_year1; SP_year2	50.31	3.69E-06 ; 1.57E-08	5.96 ; 12.82
3	SP_year2	31.26	3.18E-05	7.37
	SP_year2	49.33	2.25E-05	7.64
	SP_year1	64.52	9.41E-06	5.48
7	SP_year1	6.36	6.75E-06	5.65
8	SP_year2	60.17	9.19E-04	4.61
	SP_year1	63.9	5.26E-08	8.04
11	SP_year1	53.42	5.04E-05	4.61
	SP_year2	54.68	5.27E-04	4.82
12	SP_year1	0.26	1.27E-06	6.5

^a Physical positions based on the tomato reference genome assembly v. SL2.5 at the Sol Genomics Network (<http://solgenomics.net>).

After merging results from the three populations under study, SNPs likely associated with the same QTL were found on Chr 3, identified respectively in CER (65.88 Mb) and MAGIC (64.52 Mb) populations. Another QTL confirmed in both the CER (59.19 Mb) and the MAGIC (60.17 Mb) populations was identified on Chr 8.

The QTL on Chr 11 was the most interesting because it was confirmed in all populations investigated. In fact, a total of 13 SNPs was found in a region comprised between 54.68 Mb and 55.58 Mb, probably describing the same QTL of Chr 11.

Finally, two pairs of SNPs were identified in the distal region of the short arm of Chr 12 and identified in TR and CER populations.

The overall results of GWAS on SP showed a total of 46 QTLs associated with the trait.

3.3.5.2 Seed weight (SW) association mapping

Phenotypic data of SW were resumed in Fig. 3.9 . The histogram showed that the 24 types spread over a SW scale ranging from 1.5 to 3.6 mg. In particular, the types Small elliptic, *Pimpinellifolium*-like and Small oval represent the types with smallest seed; in this case, the seed weight was in the range of 1.5 to 2 mg. On the other hand, types such as Muchamiel, Marmande pink, and Medium round were in the range of the biggest seeds, 3.1 to 3.6 mg.

On a single accession basis, the sample with smallest seeds was TRVA0590 Pimpi-like with a mean value of 1.494 mg, whereas that with biggest seeds was TRVA1300 Big round with a value of 3.995 mg (data not shown).

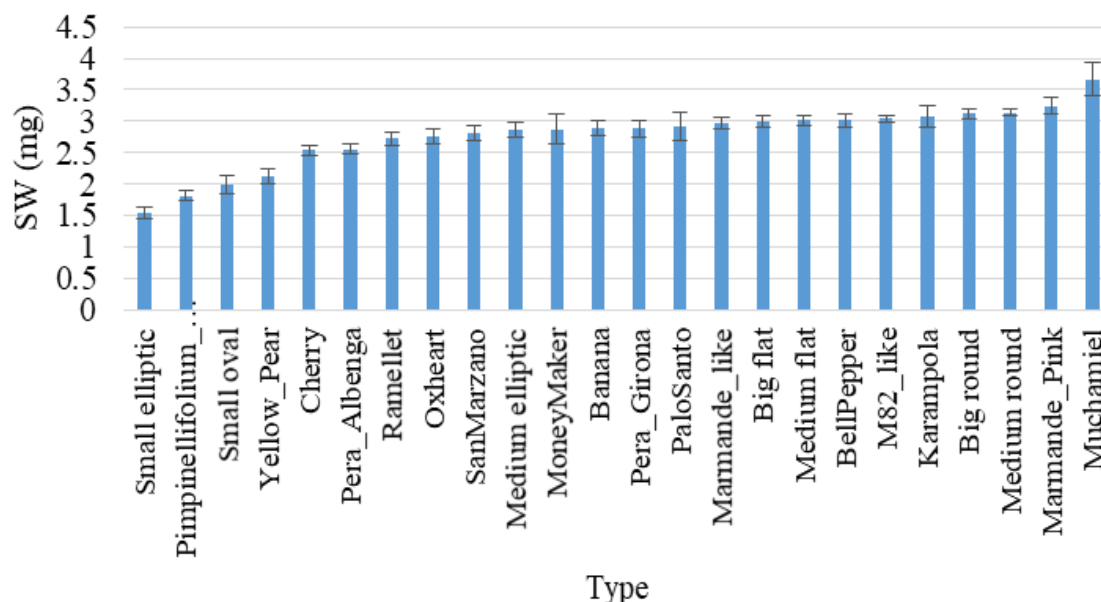


Fig.3.9 SW distribution according to known types of Traditom CC

The GLM+Q mapping approach identified a total of 32 associations with SW considering a P-value $< -\text{Log}_{10}^6$. The percentage of total phenotypic variance explained (R^2) ranged from 10.41% to 28.15%. These SNPs were distributed over 9 chromosomes; only Chr 6, 7 and 12 showed no SNP (Table 3.5).

The most significant association was observed for mSW in Chr 11, with a P-value of $1.22\text{E-}14$ and phenotypic variation explained of 28.15%. The same association was also found in SW2015 and SW2016 with P-values, respectively, of $7.49\text{E-}11$ ($R^2 = 20.97$) and $1.76\text{E-}10$ ($R^2 = 19.85$). The highest number of associations was found considering the mSW value (32 SNPs), whereas six SNPs were found for SW2015 and 15 SNPs were identified for SW2016. Analyzing the data for genome position, Chr5 showed nine associations mapped; of these, 5 fell in a small region comprised from 62.48 to 63.24 Mb. Investigation on LD showed that SNPs in position 63.1 Mb and 63.24 Mb were located in the same LD block of 146 Kb.

Considering associations confirmed by different analyses, 5 were found over the all three analyses, 12 were found in two analyses. If associations in LD were considered, a final number of 31 QTLs were identified for SW. The complete list of associations found is shown in Table 3.8.

Table 3.8 Single nucleotide polymorphisms (SNPs) significantly associated with seed weight (SW)

Chr	Trait	Physical position (Mb) ^{a, b}	P-value	R ² (%)
1	mSW, SW2015, SW2016	9.35	7.31E-08 ; 5.32E-06 ; 9.33E-06	16.98 ; 12.86 ; 12.06
	mSW	13.85	4.96E-06	15.31
	mSW, SW2015	29.09	1.04E-07 ; 7.54E-06	16.15 ; 12.13
	mSW	30.83	2.13E-07	15.82
	mSW	55.86	1.55E-07	17.47
	mSW, SW2015	70.42	1.60E-08 ; 6.91E-07	17.86 ; 14.4
	mSW	83.45	1.33E-06	12.66
2	mSW, SW2015, SW2016	50.43	1.43E-11 ; 6.48E-07 ; 2.25E-10	24.74 ; 15 ; 21.92
3	mSW	2.9	4.62E-06	12.79
	mSW	69.73	6.02E-06	12.54
4	mSW, SW2016	2.78	4.03E-07 ; 1.60E-06	14.9 ; 13.35
	mSW, SW2016	53.09	4.60E-08 ; 5.57E-07	19.3 ; 16.37
	mSW, SW2016	63.42	1.06E-06 ; 5.46E-06	13.08 ; 11.31
5	mSW	2.39	4.21E-06	14.93
	mSW, SW2016	2.78	3.45E-06 ; 6.20E-06	12.94 ; 10.67
	mSW, SW2016	6.44	2.30E-06 ; 7.00E-06	13.08 ; 11.66
	mSW	62.45	2.81E-06	13.61
	mSW, SW2016	62.48	2.27E-07 ; 2.67E-06	14.92 ; 12.16
	mSW	62.98	4.32E-06	11.59
	mSW	62.99	8.78E-06	11.9
	mSW	63.1	1.91E-06	14.92
	mSW	63.24	8.05E-06	13.55
8	mSW	0.28	2.84E-06	13.72
	mSW, SW2016	0.35	7.25E-07 ; 1.95E-06	14.42 ; 13.23
	mSW	64.94	8.87E-06	10.41
9	mSW, SW2016	6.44	3.27E-06 ; 6.44E-06	14.1 ; 13.11
	mSW, SW2015, SW2016	67.11	4.19E-08 ; 8.93E-06 ; 5.27E-07	19.85 ; 14.07 ; 16.83
10	mSW	0.2	4.78E-07	15.3
11	mSW, SW2016	17.97	6.14E-06 ; 9.94E-06	11.71 ; 11
	mSW, SW2016	28.87	2.72E-06 ; 2.12E-06	13.32 ; 13.29
	mSW, SW2016	32.83	4.33E-06 ; 1.26E-06	11.97 ; 13.06
	mSW, SW2015, SW2016	55.42	1.22E-14 ; 7.49E-11 ; 1.76E-10	28.15 ; 20.97 ; 19.85

^a SNPs in the same LD block are evidenced in gray. ^b Physical positions based on the tomato reference genome assembly v. SL2.5 at the Sol Genomics Network (<http://solgenomics.net>).

3.4 Discussion

Evaluation on fertility-related traits in the tomato Traditom CC revealed high variability for these traits. All these characteristics could result very useful for breeding activities, considering that the population of landraces showed an average elevated FS, earliness of FLOW and RIP, medium-high Y and a °B higher than 7 in a quarter of the population. Other traits, such as SF, may be considered of interest in the selection of genotypes with a high or low number of seeds (e.g following the ideotype of seed producers or consumer preferences).

Investigation on dimensional aspects of the ovary evidenced that, at the anthesis stage, the size of resulting fruit can be predicted by the size of the ovary, as showed by the high correlation between ovary and fruit dimensions.

On the other hand, the lack of correlation between ovule and seed size, suggest that size of seed is determined in a later, post-fertilization stage of development.

At the same time, a high positive correlation was showed between seed and fruit weight.

As reported in previous studies, 28 tomato fruit weight QTLs have been identified (Grandillo and Tanksley, 1999). Some of these QTLs have been cloned and characterized, including *fw2.2* (Frary et al. 2000), *locule number (lc)* (Rodriguez et al. 2011), *fw3.2* (Chakrabarti et al. 2013), *fw7.3* (Li et al. 2018), *fasciated (fas)* (Barrero and Tanksley 2004; Cong et al. 2008) and *fw11.3* (Mu et al. 2017). All these genes are expressed in the developing ovary and fruit determining the final fruit size.

The sub-CC was phenotypically characterized for ovule, seed, ovary and fruit size, confirming and adding information on the involvement of these genes in the final size of reproductive organs. The main result of this analysis was that ovule and seed size are not correlated as ovary and fruit size. This indicated that not all gene controlling fruit and seed size are working in parallel ways in the two organs.

As described by Chakrabarti et al. (2013), *fw3.2* is a gene involved in determining both fruit and seed size; however, considering that at anthesis the ovule size is the same for small and big seed, the expression of this gene before the anthesis influences the ovary and not the ovules size. Novel SW QTLs detected in this study, that do not overlap with FW weight QTLs may be interesting candidates to dissect the regulation of seed development from that of fruit development

Environmental stresses, such as heat stress, strongly affect the reproductive development of tomato plants (Sato et al. 2000; Giorno et al. 2009; Hatfield et al. 2015). One of the issues of growing tomato at relatively high temperatures is the induction of stigma exsertion (Pan et al. 2017; defined in this work as stigma position, SP), that may affect self-pollination and fruit set.

Even if the present work was not specifically focused on establish heat stress effects, the SP was monitored during the growing season in three different periods that usually represent temperature increases: early, medium and late development.

Investigation observing this trend evidenced that several accessions showed an elongation of the style during the season. For this reason, it is clear an involvement of heat stress in stigma exertion mechanisms, that draws attention on genes related to abiotic stresses.

For example, heat shock transcription factors serve as the terminal components of signal transduction, mediating the expression of heat shock proteins and other heat shock induced transcripts (Kotak et al. 2007; Jung et al. 2012). Genes involved in the metabolic pathway of abscisic acid (ABA), salicylic acid (SA), ethylene, oxidative burst, and calmodulin are known to play significant roles in the heat stress response (Larkindale et al. 2005; Zhang et al. 2009; Firon et al. 2012).

Other hypotheses of genes involved in SP could be related with development or elongation of the stylar tissue. The gene involved in the *se2.1* mutation was defined by Chen et al. (2007) as a positive regulator of cell elongation, because greater accumulation of the transcript was associated with exerted stigma. *se2.1* was identified in interspecific studies and was considered the major determinant for SP in the evolution from allogamy to autogamy. The *se2.1* mutation is essentially fixed in the cultivated tomato (Chen et al. 2007). Despite this, a high variability for SP in landraces of cultivated tomato still exist and therefore other QTLs, different from *se2.1* remains to be elucidated.

Fig. 3.10 combines the results of this work to previous studies on SP, with the identification of novel QTLs for SP.

In addition to the previously cited *se2.1*, are reported the two QTLs reported by Xu et al. (2017), *qSP1* and *qSP3*. A confirmation of *se2.1* was also reported as *qSL2* by these authors. *se5.1* was reported by Gorguet et al. (2008) on Chr 5. All these previously reported QTLs were confirmed in this work.

Firstly, *se2.1* was confirmed by the MAGIC population with data of two different years (SP_year1 and 2). Xu et al. (2017) found a style length QTL very close to that position, while a SP_CER marker was found about 3.5 Mb distant. No positions of TR population were found on Chr2, suggesting a different control of the SP in a cultivated germplasm of landraces where actually the *se2.1* mutation is fixed (Chen et al. 2004).

Considering the QTL identified by Xu et al. (2017) on Chr1, it is coincident with a marker found in the MAGIC population GWAS.

Analysing the region in the distal part of Chr 1 (between 79.61 and 84.13 Mb), we found two markers in the Traditom CC and in the MAGIC population. These positions coincide with the overlapping

region of the introgression lines IL1-1 and IL1-2, described as stigma exerted lines (Liu and Zamir, 1999).

About 2.5 Mb separates the Xu et al. (2017) result of Chr3 and a position identified here with the MAGIC population.

In the long arm of Chr5 *se5.1* was 2.5 Mb distant from a marker found in the Traditom population (FRA), suggesting the confirmation of this QTL.

Several novel QTLs for SP were revealed by this study; among them, the positions represented in more than one population deserve more attention. For example, the same marker of Chr12 (position 39.29 Mb) was identified in the FRA and ESP Traditom field trials, close to four HSP genes (Solyc12g042830, Solyc12g043110, Solyc12g043120, Solyc12g045560).

However, the most interesting multi-represented novel position was on Chr11 (55.20 Mb). This marker was found with phenotypic data of the Traditom field trials of ESP and FRA and for the Mean of the three locations (ITA, FRA, ESP). In addition, several position from the CER data were found near the position of 55.20 Mb, and both evaluation years of MAGIC were represented on this location of Chr11.

Two hypotheses of candidate genes inside the region of interest of Chr11 are Solyc11g071830 and Solyc11g071840. The first is the heat shock protein 40 (HSP40), involved in the response to different abiotic stresses (heat and cold) and plant hormones (salicylic acid, jasmonic acid, abscisic acid). In addition, this protein is pistil and stamen specific, with a regulation during the reproductive phases (Li et al. 2012). The association of this protein with heat stress and the pistil specific expression makes it a candidate for controlling SP in tomato.

The second hypothesis of gene is annotated as a Calmodulin binding protein and called SUN-like protein 31 (SISUN31). *SISUN31* is one of the 34 members of the SUN-like gene family and is expressed during flower and fruit development, showing an involvement in fruit growth and polar elongation (Wu et al. 2011; Sacco et al. 2015).

In this region of Chr11 we also found the *FASCIATED* gene (*FAS*; Solyc11g071810). *FAS* is a well characterized gene involved in the control of locule number; if mutated, it confers fruit fasciation and generally flat shape (Rodriguez et al. 2011). The *fas* mutation is present in the Traditom population and could actually be associated with SP because this trait is more frequent in flattened and fasciated types. Nevertheless, we found also markers in this region when using the MAGIC and CER data, where the *fas* mutation is not present, thus rejecting the hypothesis that the *fas* mutation is directly responsible of the association with SP.

Other positions found in this GWAS could be considered of interest if confirmed in future works or if it will be identified a connection with some putative genes.

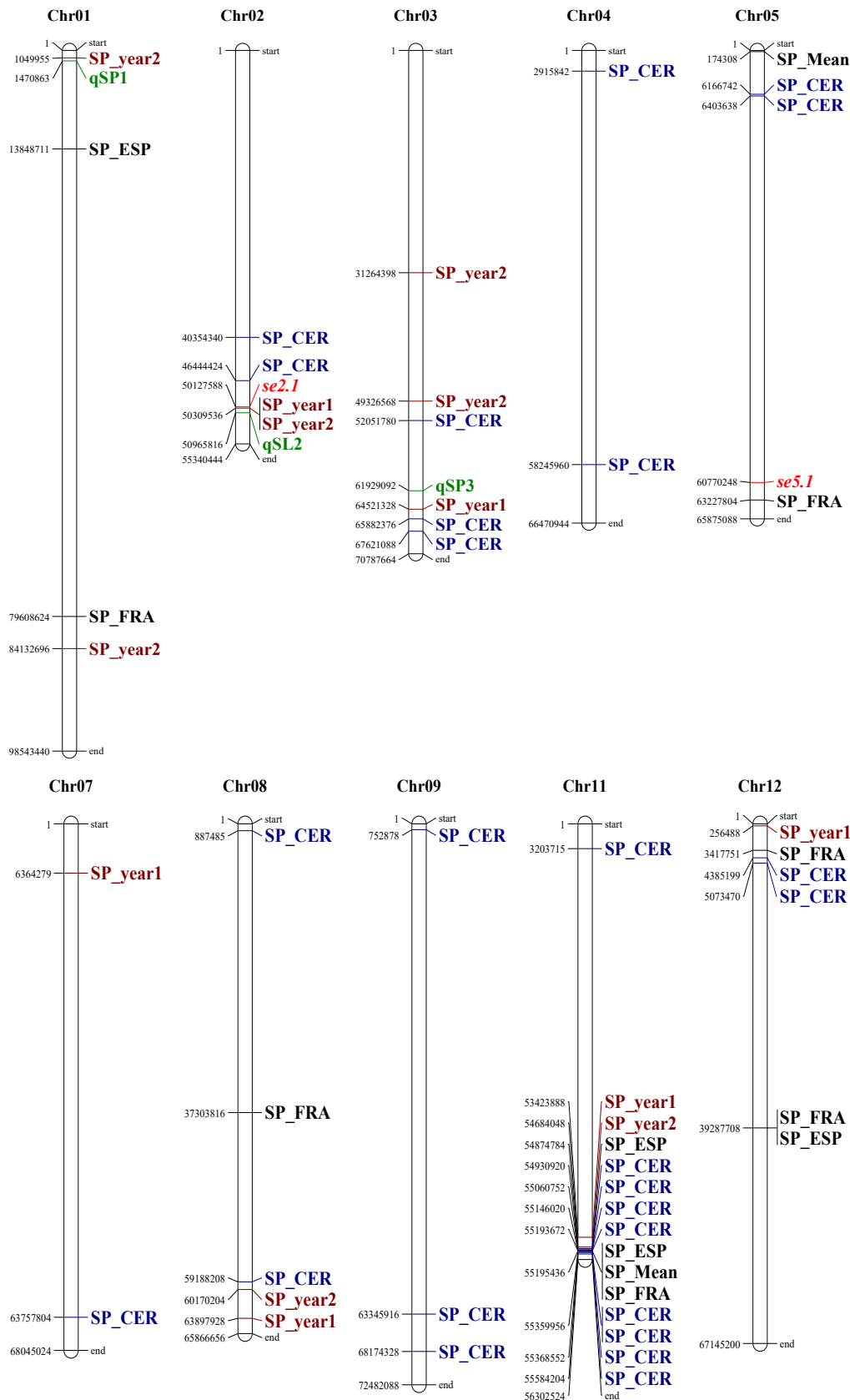


Fig. 3.10 Mapping of markers identified by GWAS and main previous results on SP. Physical map of the tomato genome showing the position of the associated SNP markers (left). The colours represent the different populations used in the GWAS; Traditom results are reported in black, blue was used for the CER population, dark red for the MAGIC population. Previous works were reported as the QTLs names in red (Chen et al. 2007;Gorguet et al. 2008) and green (Xu et al. 2017).

Concerning the GWAS analysis on SW, we found interesting results that confirmed the outcome of previous screenings and added new QTLs for this trait in tomato.

In Figure 3.11, are showed the physical positions of the markers identified in this study and previous QTLs that co-localize with them considering at least two different studies (Tanksley et al. 1982; Weller et al. 1988; Goldman et al. 1995; Grandillo and Tanksley 1996; Doganlar et al. 2000; Khan et al. 2012).

Among these, *sw4.1* is considered one of the most significant QTLs underlying the evolution of seed size in tomato and was characterized as an ABC transporter gene (Solyc04g055120; Orsi and Tanksley, 2009).

Other four QTLs found in this study co-localize with QTLs previously reported by Doganlar et al. (2000).

Further confirms were found on Chr1 (pos. 70.42 Mb) and Chr9 (pos. 67.11 Mb) with the corresponding results of Khan et al. (2012).

Many QTLs for SW have also been identified in other crops, where the seed size is considered more important being the final product or the edible part (e.g. soybean, rice, wheat).

SW in tomato increased with domestication, as it is a trait important for seed germination and seedling vigour; larger seeds have better seedling survival rate (Moles and Westoby, 2004).

As other dimensional variables, SW is a complex trait controlled by many genetic loci. Furthermore, the environmental interaction may affect the identification of this loci, as shown with the different numbers and positions of the QTLs discovered under two different seasons.

For this reason, marker positions of QTLs confirmed in more than one year result more reliable and worthy of attention. However, also the other positions found only in one analysis may be of greater interest if further confirmed.

According to the results of this work, the most significant association was located on Chr11, validated in the three analyses and with an average phenotypic variation explained of 22.99%.

The marker position fall in an intronic region of the locus Solyc11g071710, a DnaJ-like zinc-finger protein.

This and other positions found in this work deserve to be deeply investigated in order to find the underlying genes.

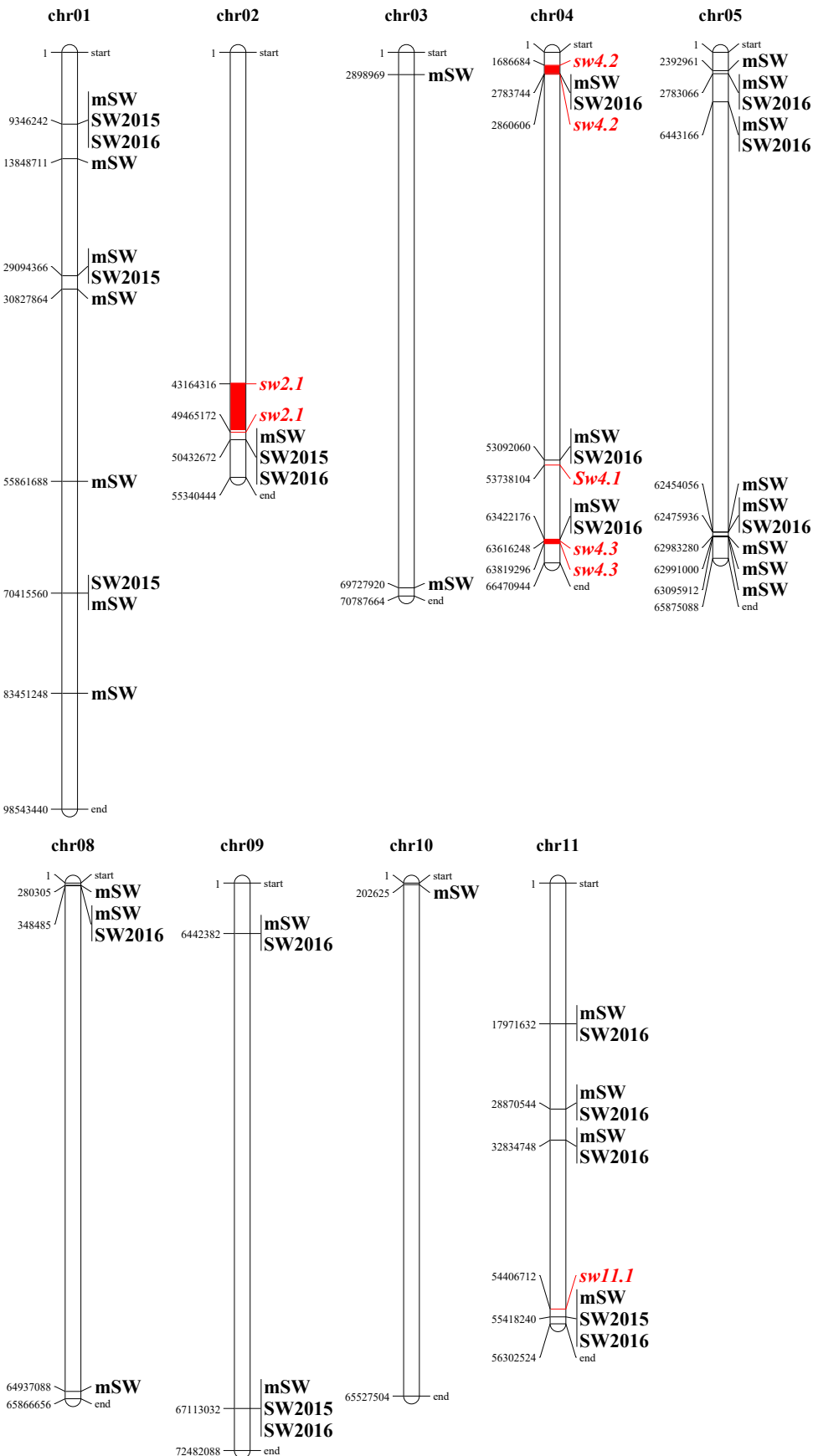


Fig.3.11 Mapping of markers identified by GWAS and main previous results on SW. Physical map of the tomato genome showing the position of the associated SNP markers. The red coloured position represent the previous QTLs reported for SW (Tanksley et al. 1982; Goldman et al. 1995;Grandillo and Tanksley, 1996;Doganlar et al. 2000).

References

- Adamski N. M., Anastasiou E., Eriksson S., O'Neill C. M., Lenhard M. (2009). Local maternal control of seed size by KLUH/CYP78A5-dependent growth signalling. *Proceedings of the National Academy of Sciences*, 106(47), 20115-20120.
- Albert, E., Segura, V., Gricourt, J., Bonnefoi, J., Derivot, L., & Causse, M. (2016). Association mapping reveals the genetic architecture of tomato response to water deficit: focus on major fruit quality traits. *Journal of experimental botany*, 67(22), 6413-6430.
- Allen, K. D., & Sussex, I. M. (1996). Falsiflora and anantha control early stages of floral meristem development in tomato (*Lycopersicon esculentum* Mill.). *Planta*, 200(2), 254-264.
- Bai, Y., & Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of botany*, 100(5), 1085-1094.
- Barrero, L. S., & Tanksley, S. D. (2004). Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theoretical and Applied Genetics*, 109(3), 669-679.
- Barrero, L. S., Cong, B., Wu, F., & Tanksley, S. D. (2006). Developmental characterization of the fasciated locus and mapping of Arabidopsis candidate genes involved in the control of floral meristem size and carpel number in tomato. *Genome*, 49(8), 991-1006.
- Barrett, J. C., Fry, B., Maller, J. D. M. J., & Daly, M. J. (2004). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263-265.
- Bauchet, G., Munos, S., Sauvage, C., Bonnet, J., Grivet, L., & Causse, M. (2014). Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection. *BMC plant biology*, 14(1), 279.
- Bernacchi, D., & Tanksley, S. D. (1997). An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. *Genetics*, 147(2), 861-877.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633-2635.
- Breseghele, F., & Sorrells, M. E. (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*, 172(2), 1165-1177.
- Chakrabarti, M., Zhang, N. A., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., ... & Causse, M. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences*, 110(42), 17125-17130.
- Chen, K. Y., & Tanksley, S. D. (2004). High-resolution mapping and functional analysis of se2. 1: a major stigma exertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. *Genetics*, 168(3), 1563-1573.
- 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(5850), 643-645.

- Chen, Y. (2013). Dissection of agronomic traits in crops by association mapping. In *Diagnostics in Plant Breeding* (pp. 119-142). Springer, Dordrecht.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368), 829-836.
- Cong, B., Barrero, L. S., & Tanksley, S. D. (2008). Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nature genetics*, 40(6), 800. Data. Genetics. 2000; 155(2): 945–959. PMID: 10835412
- Di Iorio (2017). Phenotypic and genetic studies on seed size in traditional tomato (*Solanum lycopersicum* L.) varieties. Master thesis. Dipartimento di Scienze Agrarie e forestali, Università degli Studi della Tuscia.
- Doganlar, S., Frary, A., & Tanksley, S. D. (2000). The genetic basis of seed-weight variation: tomato as a model system. *Theoretical and Applied Genetics*, 100(8), 1267-1273.
- Earl DA, Von Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 2012; 4(2): 359–361.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals STRUCTURE: a simulation study. *Molecular ecology*, 14(8), 2611-2620.
- Firon, N., Pressman, E., Meir, S., Khoury, R., & Altahan, L. (2012). Ethylene is involved in maintaining tomato (*Solanum lycopersicum*) pollen quality under heat-stress conditions. *AoB plants*, 2012.
- Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Van Der Knaap, E., Cong, B., ... & Tanksley, S. D. (2000). Fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, 289(5476), 85-88.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., ... & Liu-Cordero, S. N. (2002). The structure of haplotype blocks in the human genome. *Science*, 296(5576), 2225-2229.
- Georgiady, M. S., Whitkus, R. W., & Lord, E. M. (2002). Genetic analysis of traits distinguishing outcrossing and self-pollinating forms of currant tomato, *Lycopersicon pimpinellifolium* (Jusl.) Mill. *Genetics*, 161(1), 333-344.
- Giorno, F., Wolters-Arts, M., Grillo, S., Scharf, K. D., Vriezen, W. H., & Mariani, C. (2009). Developmental and heat stress-regulated expression of HsfA2 and small heat shock proteins in tomato anthers. *Journal of Experimental Botany*, 61(2), 453-462.
- Goldman, I. L., Paran, I., & Zamir, D. (1995). Quantitative trait locus analysis of a recombinant inbred line population derived from a *Lycopersicon esculentum* x *Lycopersicon cheesmanii* cross. *Theoretical and Applied Genetics*, 90(7-8), 925-932.
- Goodwillie, C., Sargent, R. D., Eckert, C. G., Elle, E., Geber, M. A., Johnston, M. O., ... & Winn, A. A. (2010). Correlated evolution of mating system and floral display traits in flowering plants and its implications for the distribution of mating system variation. *New Phytologist*, 185(1), 311-321.

- Gorguet, B., Eggink, P. M., Ocana, J., Tiwari, A., Schipper, D., Finkers, R., ... & van Heusden, A. W. (2008). Mapping and characterization of novel parthenocarp QTLs in tomato. *Theoretical and Applied Genetics*, 116(6), 755-767.
- Grandillo, S., & Tanksley, S. D. (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theoretical and Applied Genetics*, 92(8), 935-951.
- Grandillo, S., Ku, H. M., & Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics*, 99(6), 978-987.
- Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2014). Association mapping in crop plants: opportunities and challenges. In *Advances in genetics* (Vol. 85, pp. 109-147). Academic Press.
- Hall D, Tegstrom C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Brief Funct Genomics* 9:157–165.
- Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and climate extremes*, 10, 4-10.
- Hernández-Bautista, A., Lobato-Ortiz, R., Cruz-Izquierdo, S., García-Zavala, J. J., Chávez-Servia, J. L., Hernández-Leal, E., & Bonilla-Barrientos, O. (2015). Fruit size QTLs affect in a major proportion the yield in tomato. *Chilean journal of agricultural research*, 75(4), 402-409.
- Hernández-Bautista, A., Lobato-Ortiz, R., García-Zavala, J. J., López-Fortoso, F., Cruz-Izquierdo, S., Chávez-Servia, J. L., & Cadeza-Espinosa, M. (2016). Quantitative trait locus mapping associated with earliness and fruit weight in tomato. *Scientia Agricola*, 73(5), 478-486.
- Jung, K. H., Ko, H. J., Nguyen, M. X., Kim, S. R., Ronald, P., & An, G. (2012). Genome-wide identification and analysis of early heat stress responsive genes in rice. *Journal of Plant Biology*, 55(6), 458-468.
- Khan, N., Kazmi, R. H., Willems, L. A., Van Heusden, A. W., Ligterink, W., & Hilhorst, H. W. (2012). Exploring the natural variation for seedling traits and their link with seed dimensions in tomato. *PloS One*, 7(8), e43991.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9(1), 29.
- Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., & Scharf, K. D. (2007). Complexity of the heat stress response in plants. *Current opinion in plant biology*, 10(3), 310-316.
- Li, H., Qi, M., Sun, M., Liu, Y., Liu, Y., Xu, T., ... & Li, T. (2017). Tomato transcription factor SIWUS plays an important role in tomato flower and locule development. *Frontiers in plant science*, 8, 457.
- Li, J., Zhang, H., Hu, J., Liu, J., & Liu, K. (2012). A heat shock protein gene, CsHsp45. 9, involved in the response to diverse stresses in cucumber. *Biochemical genetics*, 50(7-8), 565-578.
- Li, M., Wang, X., Li, C., Li, H., Zhang, J., & Ye, Z. (2018). Silencing GRAS2 reduces fruit weight in tomato. *Journal of integrative plant biology*.

Linkies A., Graeber K., Knight C., Leubner-Metzger G. (2010). The evolution of seeds. *New Phytologist*, 186(4), 817-831.

Lippman, Z., & Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. *Genetics*, 158(1), 413-422.

Lukyanenko, A. N. (1991). Parthenocarpy in tomato. In *Genetic improvement of tomato* (pp. 167-177). Springer, Berlin, Heidelberg.

Mazzucato, A., Papa, R., Bitocchi, E., Mosconi, P., Nanni, L., Negri, V., ... & Veronesi, F. (2008). Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theoretical and Applied Genetics*, 116(5), 657-669.

Moles, A. T., & Westoby, M. (2004). Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology*, 92(3), 372-383.

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., ... & van der Knaap, E. (2017). Fruit weight is controlled by cell size regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLoS genetics*, 13(8), e1006930.

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., ... & Brunel, D. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiology*, 156(4), 2244-2254.

Orsi, C. H., & Tanksley, S. D. (2009). Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. *PLoS genetics*, 5(1), e1000347.

Osei, M. K., Annor, B., Adjepong-Danquah, J., Danquah, A., Danquah, E. Y., Blay, E., & Adu-Dapaah, H. (2018). Genotype× Environment Interaction: A Prerequisite for Tomato Variety Development. In *Recent Advances in Tomato Breeding and Production*. IntechOpen.

Pan, C., Ye, L., Zheng, Y., Wang, Y., Yang, D., Liu, X., ... & Lu, G. (2017). Identification and expression profiling of microRNAs involved in the stigma exertion under high-temperature stress in tomato. *BMC genomics*, 18(1), 843.

Park, S. J., Jiang, K., Schatz, M. C., & Lippman, Z. B. (2012). Rate of meristem maturation determines inflorescence architecture in tomato. *Proceedings of the National Academy of Sciences*, 109(2), 639-644.

Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J. P., ... & Causse, M. (2015). Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant biotechnology journal*, 13(4), 565-577.

Peralta, I. E., & Spooner, D. M. (2000). Classification of wild tomatoes: a review. *Kurtziana*, 28(1), 45-54.

Perez-Llamas, C., & Lopez-Bigas, N. (2011). Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PloS one*, 6(5), e19541.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype

- Ranc, N., Muños, S., Xu, J., Le Paslier, M. C., Chauveau, A., Bounon, R., ... & Causse, M. (2012). Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3: Genes, Genomes, Genetics*, 2(8), 853-864.
- Rodríguez, G. R., Muños, S., Anderson, C., Sim, S. C., Michel, A., Causse, M., ... & van der Knaap, E. (2011). Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. *Plant physiology*, 156(1), 275-285.
- Sacco, A., Ruggieri, V., Parisi, M., Festa, G., Rigano, M. M., Picarella, M. E., ... & Barone, A. (2015). Exploring a tomato landraces collection for fruit-related traits by the aid of a high-throughput genomic platform. *PLoS One*, 10(9), e0137139.
- Sato, S., Peet, M. M., & Thomas, J. F. (2000). Physiological factors limit fruit set of tomato (*Lycopersicon esculentum* Mill.) under chronic, mild heat stress. *Plant, Cell & Environment*, 23(7), 719-726.
- Sicard, A., & Lenhard, M. (2011). The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, 107(9), 1433-1443.
- Sim, S. C., Van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganai, M. W., ... & Panthee, D. R. (2012). High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PloS one*, 7(9), e45520.
- Tanksley, S. D., Medina-Filho, H., & Rick, C. M. (1982). Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity*, 49(1), 11.
- van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., ... & Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Frontiers in plant science*, 5, 227.
- Weller, J. I., Soller, M., & Brody, T. (1988). Linkage analysis of quantitative traits in an interspecific cross of tomato (*lycopersicon esculentum* x *lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics*, 118(2), 329-339.
- Wu, S., Xiao, H., Cabrera, A., Meulia, T., & van der Knaap, E. (2011). SUN regulates vegetative and reproductive organ shape by changing cell division patterns. *Plant Physiology*, 157(3), 1175-1186.
- Xu, J., Ranc, N., Muños, S., Rolland, S., Bouchet, J. P., Desplat, N., ... & Causse, M. (2013). Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theoretical and Applied Genetics*, 126(3), 567-581.
- Xu, J., Driedonks, N., Rutten, M. J., Vriezen, W. H., de Boer, G. J., & Rieu, I. (2017). Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (*Solanum lycopersicum*). *Molecular breeding*, 37(5), 58.
- Zhang, W., Zhou, R. G., Gao, Y. J., Zheng, S. Z., Xu, P., Zhang, S. Q., & Sun, D. Y. (2009). Molecular and genetic evidence for the key role of AtCaM3 in heat-shock signal transduction in *Arabidopsis*. *Plant Physiology*, 149(4), 1773-1784.

Chapter 4

Bulk RNA-seq analysis to dissect stigma position regulation in tomato

4.1 Introduction

4.2 Materials and methods

4.2.1 Plant material and data collection

4.2.2 Material collection and RNA extraction

4.2.3 RNA-sequencing, filtering and mapping

4.2.4 Enrichment analysis

4.2.5 Tissue-specific genes

4.2.6 RT-PCR and RT-qPCR

4.2.7 Statistical analysis of RT-qPCR expression data on single genotypes

4.2.8 Identification of structural variants from RNA-seq Data

4.3 Results

4.3.1 The “stigma position” population

4.3.2 Dynamic of stamen and pistil growth in flower buds

4.3.3 Bulk RNA-seq analysis

4.3.4 Sequence reads filtering and mapping

4.3.5 Analysis of differentially expressed genes (DEGs)

4.3.6 Validation of RNA-seq data by qRT-PCR

4.3.7 GO Enrichment and network analyses

4.3.8 Transcription factors listed among DEGs involved in stigma position

4.3.9 Analysis on tissue specificity

4.3.10 RT-PCR and RT-qPCR analysis of the selected DEGs on single genotypes

4.3.11 Identification of structural variants from RNA-seq Data

4.4 Discussion

4.4.1 Phenotyping of SP in tomato landraces and selection of contrasting genotypes

4.4.2 RNA-seq analysis, specificity and enrichment

4.4.3 Validation of DEGs on single genotypes

4.4.4 Identification of structural variants from RNA-Seq Data

References

4.1 Introduction

Tomato (*Solanum lycopersicum* L.) is characterized by hermaphrodite flowers that have five or more anthers fused laterally to form a staminal cone that encloses the pistil. The pistil is usually shorter than the anthers and an inserted stigma represents a crucial point for the transition from wild to cultivated species of tomato.

The position of the stigma with respect to the anther cone has been previously studied in tomato (Rick and Dempsey, 1969; Bernacchi and Tanksley, 1997) and other species (Yan et al. 2009; Takano-Kai et al. 2011; Fourquin et al. 2014) because this aspect of floral morphology is linked with important aspects of tomato breeding and cultivation.

First, the position of the stigma is related to the mating system. As described by Kalisz et al. (1999), changes in floral morphology are normally required to promote either autogamy or allogamy. Stigma exertion above the anther cone promotes outcrossing, while an inserted stigma promotes self-fertilization (Rick et al. 1977; Motten and Antonovics 1992; Karron et al. 1997; Richards 1997; Motten and Stone 2000).

Among tomato relatives, wild allogamous species, (*S. pennellii*, *S. peruvianum*, *S. chilense*, *S. habrochaites*, *S. chmielekii*) exhibit highly exerted stigmas (Rick and Lamm, 1955; Rick et al. 1976; Rick 1982). Other species varies greatly with respect to stigma exertion, such as *S. pimpinellifolium* (Rick et al. 1977), while the cultivated *S. lycopersicum*, together with *S. parviflorum* and *S. cheesmaniae*, are definitely autogamous with an inserted style (Rick 1963; 1995).

A major stigma exertion QTL, associated with the evolution from wild to cultivated forms and therefore from allogamy to autogamy, was previously described as *style2.1* (*se2.1*), identified in a backcross population of *S. lycopersicum* X *S. habrochaites* (Bernacchi et al. 1997, Chen and Tanksley, 2004).

se 2.1 was subsequently cloned resulting in a helix-loop-helix (HLH) transcription factor promoting cell elongation; the association with the length of style was due to a change in the promoter of the gene. Therefore, the greater accumulation of the transcript in the allogamous species with respect to the autogamous cultivated tomato explains part of the transition from long to short styles (Chen et al. 2007). Few other QTLs for stigma position have been later identified.

Gorguet et al. 2008 identified a QTL in the long arm of Chr 5, named *se5.1*. *se5.1* was detected by interval mapping in a population carrying the complete introgression of Chr5 from *S. habrochaites* (stigma exerted) in a *S. lycopersicum* cv. Moneymaker genetic background (stigma inserted).

Recently, Xu et al. (2017), phenotyping an F₂ population derived after crossing two heat tolerant contrasting parents of *S. lycopersicum*, have identified two new QTLs on Chr 1 and 3 and confirmed the previously mapped *se2.1*.

Nevertheless, regulation of the stigma position continues to fascinate because variability for SP in cultivated tomato still exists and the genetic determinants of this variability remain to be elucidated.

Another issue related to stigma is heat stress. Records of global average temperature show long-term warming trends since the end of the 19th century, which have been most rapid since the 1970s. The year 2017 was one of the world's three warmest years on record (together with the years 2016 and 2015) with temperatures between 1.0°C and 1.1°C above pre-industrial levels. According to different observational records of global average annual near-surface (land and ocean) temperature, the last decade (2008–2017) was 0.89°C to 0.93°C warmer than the pre-industrial average, which makes it the warmest decade on record.

Europe has experienced seven extreme heat waves since 2000. In southern Europe, under a high emissions scenario, extreme heat waves are projected to be particularly strong (Global and European temperature, report 2018 of European Environment Agency (EEA)). Therefore, high temperature is becoming an increasingly serious factor in limiting crop production with global climate change (Liu et al. 2015; Pan et al. 2019; Hussain, 2019).

Under high-temperature stress, both the extrusion of stigma position above the anther cone and male fertility are factors limiting self-fertilization, fruit productivity, fruit quality and seed set (Pan et al. 2019; Akthar et al. 2012; Dane et al. 1991; Levy et al. 1978). Pan et al. (2019) reported that stigma exsertion induced by heat stress in cultivated tomato is caused by more seriously shortened stamens than longer pistils. These results differ from others (Chen and Tanksley 2004; Xu et al. 2017) affirming that stigma exsertion is due largely to style length and to a lesser degree to stamen length.

However, the physiological and genetic mechanisms inducing stigma exsertion during heat stress remain largely unknown.

Other physiological processes of flower development are influenced by high temperatures: male and female gametophyte development, stigma receptivity, pollen tube growth, embryo development (Giorno et al. 2013). Although the heat stress could induce style elongation in tomato, Saeed et al. (2007) reported that a considerable proportion of the variation in the stigma exsertion was genetically determined.

Elucidating the molecular and physiological mechanisms of stigma exsertion could thus be useful in breeding to develop more thermotolerant accessions with an inserted style, in order to obtain a high and stable yield.

On the other hand, genetic control of stigma position could be used for hybrid seed production. The commercial exploitation of hybrid vigor is the most important application of genetics, which plays a key role in ensuring global food security and reducing poverty (Cheng et al. 2007).

Hybrid tomato varieties have considerable advantages; hybrids produce higher yields, generally mature earlier and more uniformly, quality fruit is superior and resistance to diseases is improved. With all these advantages, many farmers prefer to sow hybrid seeds despite the higher seed costs.

Generally, hybrid seed production requires the removal of male organs from the seed parent plants before anthesis (emasculature), to avoid self-fertilization. Other ways were developed to produce hybrid seed in tomato, such as use of male sterility. Male sterility is a powerful system in hybrid seed production in

monoecious and hermaphrodite crop species because it eliminates the tedious and laborious hand emasculation of the flowers and reduces the cost associated with seed production (Sawhney, 1997).

However, the use of male sterile lines must be coupled with a style elongation above the anther cone, in order to have an exposed stigma, further decreasing the possibility of self-fertilization and facilitating pollination. For this reason, stigma exertion could be an interesting and useful feature to facilitate the pollination techniques, bypassing the emasculation phase, which is very laborious and expensive for tomato hybrid seed production.

In order to dissect genes involved in style length regulation, in this experiment, a bulk RNA sequencing (RNA-seq) approach was performed. We selected two groups of 13 genotypes each, contrasting for having stigma insertion or exertion. We adopted the strategy of pooling for two reasons: a) pooling facilitates the sampling of adequate tissue for the analysis and b) it should allow to evidence common mechanisms regulating style exertion differently from the study of a single pair of genotypes. Comparative results from the study of a pair of specific genotypes would have been uncertain in lack of near-isogenicity between them.

To understand the patterns of genotypic contribution to the discovery of DEGs, a number of DEGs identified in the bulk RNA-seq experiment, have been subsequently validated in all the single accessions composing the bulk.

4.2 Materials and methods

4.2.1 Plant material and data collection

A total of 26 accessions of tomato presenting stigma inserted (13) and stigma exerted (13) were selected from the Traditom Core Collection (CC) composed of 218 tomato landraces and 45 commercial hybrid genotypes. The selected accessions included 24 landraces and two hybrids.

Phenotypic data were collected for two years (2017, 2018) in open field growth conditions at University of Tuscia's experimental farm, at Viterbo, Italy (42°260'N, 12°040'E). Standard agronomic tomato management practices were applied. Temperatures were monitored during all observations and were reported in Table S4.1. Stigma position (SP), fruit weight (FW), fruit shape index (SI), locule number (LN), seed weight (SW), fruit set (FS) and yield (Y) were considered to describe the accessions.

For a deeper evaluation of SP, the first year data were collected in three times (early, mid, late growth season, corresponding to the 1-2nd, 4-5th and 7-8th truss respectively). A score from 1 to 4 was adopted to describe a stigma inserted (1), at the same level of the anther cone (2), slightly exerted (3), or highly exerted (4). The selected accessions were grown to confirm the score of the stigma position and to collect pistils for subsequent RNA-seq analysis. The second year, the 26 accessions were submitted to a fine phenotyping to study the stigma score during the whole flowering season. In addition, from the third truss of each plant (4 plants per accession) flowers were collected and studied by measuring with a stereomicroscope stamen and pistil length at different developmental stages (2E, 2M, 2L, 3, 4). Flower developmental stages were based on previous description (Mazzucato et al. 1998), with Stage 2 further divided in early (2E), mid (2M) and late (2L), Stage 3 being the opening flower and Stage 4 the flower at anthesis (Fig.4.1).

Stage	Description	Representative image
2E	Floral bud, 75% of final size	
2M	Floral bud, final size, petals not visible	
2L	Floral bud, final size, petals visible	
3	Opening flower	
4	Open flower, anthesis	

Figure 4.1 Description of the flower developmental stages adopted to study the dynamic of stamen and pistil growth in tomato flowers

FW was evaluated as the mean weight in g of all mature fruit on trusses 1-4. Fruit SI was calculated as the ratio between polar diameter and equatorial diameter measured on 8 fruits and LN was evaluated as the number of locules of the same fruit sample. SW, reported as single seed weight (mg), was estimated by counting and weighting two replicates of 50 seeds. FS was evaluated with a score using 5 categories (1, very poor/ 3, poor/ 5, intermediate/ 7, good/ 9, very good). The Y variable (Kg/plant) was calculated multiplying the mean fruit weight by the total number of fruits.

4.2.2 Material collection and RNA extraction

For RNA sequencing experiments, a pool of 12 pistils was collected from 4 plants per each accession at pre-anthesis stage (including stages 2L and 3 as previously described). Materials were collected in three biological replicates.

After sampling, tissues were immediately frozen in liquid nitrogen and stored at -80°C until RNA isolation. Total RNA was extracted using Trizol reagent (Invitrogen, CA, USA) following the manufacturer’s protocol and treated with DNase I (Amersham Pharmacia Biotech). The yield and quality of the RNA was measured using a spectrophotometer and agarose gel analysis.

4.2.3 RNA sequencing, filtering and mapping

Qubit® 2.0 Fluorometer (Thermo Fisher Scientific Inc.) was used to quantify 100 ng of RNA for accession used to compose each bulk of 13 accessions. Three biological replicates were composed for each bulk. The cDNA libraries were constructed using the TruSeq™ RNA Sample Preparation Kit (Illumina, Inc.) according to the manufacturer's instructions.

Next generation sequencing experiments were performed by Genomix4life S.R.L. (Baronissi, Salerno, Italy). Libraries were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies) and pooled such that each index-tagged sample was present in equimolar amounts, with final concentration of the pooled samples of 2 nM. The pooled samples were subject to cluster generation and sequencing using an Illumina HiSeq 2500 System (Illumina) in a 2x100 paired-end (RNA-seq) format.

RNA-seq raw sequencing data were submitted to a quality check, removing low quality portions while preserving the longest high-quality part of NGS reads. The minimum length was set to 35 bp and the quality score to 15. The software BBDuk (sourceforge.net/projects/bbmap/) was used for this scope. FeatureCounts (version 1.5.1) was used to calculate gene expression values as raw fragment counts (annotation ITAG3.1). The following step has been the removal of the not expressed genes and the ones showing too much variability. The HTSFilter package (Rau et al. 2013) was chosen for this scope, which implements a filtering procedure for replicated transcriptome sequencing data based on a Jaccard similarity index. The Trimmed Means of M-values (TMM) normalization strategy was used. Processing of raw sequencing data was performed by Sequentia (Sequentia Biotech SL).

Differential gene expression was considered as Log₂FC (Fold Change): fold change being relative to the Inserted bulk, taken as reference.

$$\text{Log}_2\text{FC} = \frac{\text{Exserted}}{\text{Inserted}}$$

A positive or negative Log₂FC indicates a gene up- or down-regulated in the Exserted bulk.

The identification of differentially expressed genes (DEGs) was performed with the packages edgeR (Robinson et al. 2010) and NOISeq (Tarazona et al. 2015).

However, due to the very low number of significant differential genes identified with edge R, the final analysis was performed with the package NOISeq . The software NOISeq is particularly suited for experiments where high variability is observed, but it must be pointed out that NOISeq shows higher false positives with respect to edgeR.

In order to control the false discovery rate (FDR), multiple testing correction was performed using the Benjamini-Hochberg method. The total number of DEGs were identified at an FDR of 5%.

4.2.4 Enrichment analysis

The Gene Ontology (GO) analysis was based on hypergeometric test of proportion of genes in each GO category against all the GO annotated genes in the genome.

DEGs with an $FDR \leq 0.05$ based on Benjamini-Hochberg correction and categories containing at least 3 DEGs were maintained.

Additionally, the enriched GO categories were detected and visualized with the Cytoscape plugin ClueGO (version 3.7.1; Bindea et al. 2009; 2013; Shannon et al. 2003). Functional enrichment analysis including GO and Kyoto Encyclopaedia of Genes and Genomes (KEGG) terms in comparison with the whole-transcriptome background was performed and only pathways with $p < 0.05$ were showed.

The GO categories searched included biological process, cellular component, molecular function and KEGG pathways.

The intervals of each GO level were set at a min level = 3 and a max level = 8, and a minimum of two genes per category as selected. Analysis in ClueGO were conducted using a right-sided hypergeometric test for enrichment based on the *S. lycopersicum* reference genome.

MapMan (version 3.6.0; Thimm et al. 2004) was used to generate a list of transcription factor (TF) genes resulting from DEGs using *S. lycopersicum* ITAG2.3 mapping.

4.2.5 Tissue-specific genes

DEG expression specificity has been investigated using data on specificity from the CoNekT database (Proost and Mutwil, 2018). The cut-off specificity measure (SPM) used (Kryuchkova-Mostacci et al. 2017), as recommended, was set at 0.85 (higher values are more specific, lower ones include more broadly expressed genes).

4.2.6 RT-PCR and RT-qPCR

Twenty selected DEGs were validated by semi-quantitative and quantitative RT-PCR. Extracted RNA was used to perform a reverse transcription PCR to synthesize cDNA using the Maxima reverse transcriptase (Thermo Fisher Scientific, USA) and following the manufacturer's instructions.

Quantitative RT-PCR (qRT-PCR) analyses were performed in order to validate RNA-seq data on bulks of 20 DEGs selected for high significance ($FDR < 1e^{-10-5}$), high Log_2FC (> 0.8), including information of enrichment analysis and information of previous works (auxin-related, style specific and genes involved in elongation of tissues).

The RT-qPCR experiments were carried out using the GoTaq® qPCR Master Mix (Promega, USA) according to the manufacturer's instruction in the Bio-Rad CFX96 Manager system (Bio-Rad, USA). Three independent biological replicates of each samples were carried out. The *SlCAC* gene was used as internal standard and the relative expression ($2^{-\Delta\Delta C_t}$) was calculated using CFX Manager software v.1.6 (Bio-Rad Laboratories). Table 4.1 show the list of primers, number of cycles of RT-PCR and amplicon size.

The expression of the same 20 DEGs was evaluated by semi-quantitative RT-PCR on single genotypes. The number of cycles used in the RT-PCR experiments was specific for each gene and *SlCAC* (Soly08g006960; Exposito-Rodriguez et al. 2008) was used as reference gene. Amplified products were visualized on a 1% (w/v) agarose gel stained with ethidium bromide. The software Image J v.1.8 was used to perform the gel analysis and calculate the amount of expression of each gene using two biological replicates.

For six selected DEGs, the analysis on single genotypes was repeated by qRT-PCR using the same procedure described above.

Table 4.1. List of primers used in RT-PCR and qRT-PCR. ^a Number of cycles are referred only to the RT-PCR analysis.

Name ^a	Primers		Product size (bp)	N° of cycles
BHLH135	F-CAAGAGGAAATGGGAAGTGC	R-GCGTCCATCTCCTTTCTGAG	168	28
XYLO630	F-TTGATCCAACCAAGGGCTAC	R-TTTCTCAAGCCCACCTCTTG	200	27
PRP	F-CACAAACCCAAGCCTAAACC	R-TTCACTATTGGTGGCAGTGG	200	30
ARF8	F-CTGCTCAAACCCAAATGCTGTC	R-GGTAAGTGTGTTGGTGAGCCTG	96	27
EXT100	F-CCATCAAAACCACCAGCTAAG	R-TGGTGTGTTGTCACACTAGC	169	20
COLDRP	F-ATCAACATGAGCAGCGGAAG	R-CCATCAGCATCCTTATCTCCTC	128	27
IAA9	F-TAGATGCTTTACCTGATTACGACA	R-TGCAGACAAACTCCAATATCAAAC	119	23
GLU6PHOS	F-CTGGCTGCTGGCTCTCTTAT	R-GCGGGCTCTCCACTCTTAAT	194	28
UBI140	F-TGATGATGACGGTGTGCGTA	R-TCTGCGATAAGCGGATCACT	150	35
DEF710	F-AATATGGGGAACCTCCCTTCG	R-CCTTTGAAATGGTGGCTCTG	125	30
HXXXD380	F-GTCAAATGCATGTCCCTTGC	R-CCTGATGCTGGATAAACATGAC	130	33
V-TYPE	F-ATGTTTGGAGACTGGGGACA	R-AGGATCGCGACAACCATAAG	237	35
S-ADEN	F-GGAGCTTTTATCCATGGCTCT	R-GCTCTCATGCACTTTGACACA	243	24
GLU-S-TRA	F-TTCCCGCTCTTATCAGGATG	R-CAAACGGTTCCACAATCAGC	181	28
FBOX000	F-AGATCTGCCGGAAGGAAAAC	R-TTACATGGCATCCCTTCACC	158	35
SEN510	F-TCAGGTAAAACGGCATCTCC	R-CAATCAGGGACATTACAGG	110	23
CCCH	F-TGAAACAAGACCTCGGAAC	R-CCCTGTGTGGAGAAACACCT	197	35
SKP1	F-AGGTAGATAAGGCGGTGGCAT	R-TCCTCTTCAGCAGTCTGATCC	178	35
TAGL11	F-GGGAACCTGAAGCAATTGGAG	R-TCCCTCTTGTGCAAATCCTC	106	25
EF1G1	F-TTCCAGATGGGTGTGTCAA	R-GCATCAACTTCAGTGGCAGA	222	26
CAC	F-GATGTCCTTATCAACCGTCTCTAC	R-ACAAGAAAGAACAGCCTCCAATCT	188	27

^aName of genes adopted in this study. The Soly code of the genes selected are reported in Table 4.4

4.2.7 Statistical analysis of RT-qPCR expression data on single genotypes

For each gene, normalized expression data were subjected to hierarchical analysis of variance considering “Bulk” as main effect and “Genotype within bulk” as nested effect. Homogeneity of variances within each bulk was assessed by Bartlett’s test and the homogeneity of error variances between bulks by a bilateral F test. In the 2-ways Anova, differences between bulks were tested using GEN(BULK) variance as error term. When Bartlett’s test was significant, the analysis was repeated excluding the second biological replicate (replicate B) that was omitted also in the RNA-seq analysis (Figure 4.7). For *SKP1* and *CCCH*, data were transformed into Log_{10} to meet the homoscedasticity assumption.

When differences among genotypes within bulk were detected, genotype means were separated by Duncan multiple range test. All analyses were carried out using SAS software package (SAS® University Edition).

4.2.8 Identification of structural variants from RNA-seq Data

Mapped reads of each bulk were filtered with Opossum (Oikkonen and Lise, 2017) software to remove artefacts derived from RNA-seq data. The software Platypus (Rimmer et al. 2014) was used to perform the variant calling analysis.

High quality filtered variants were maintained and considering positions where almost two of the three biological replicates were represented. For a total of 7137 variants, was calculated the mean allelic frequency of each bulk and the difference (Delta) between bulks (bulk exserted – bulk inserted). The allelic frequency is defined as the ratio between the reads with a variation and the total number of reads mapped for each site.

4.3 Results

4.3.1 The “stigma position” population

Two groups of 13 accessions each were selected within the Traditom Core Collection established using European tomato landraces considering the extreme values of stigma position (inserted or exerted). As stigma position in some genotypes is a heat-dependent trait, the selection was done considering those accession that showed a stable inserted or exerted stigma.

In addition, the two groups were composed considering the inclusion of similar types to maximize the common features between the two groups except for the trait of interest (Table 4.2).

Table 4.2. List of accessions selected to compose the stigma inserted (INS) and exerted (EXE) bulks to perform the RNA-seq analysis with a phenotypic characterization for the stigma position (SP) carried out at early (E), mid (M) and late (L) growth stages, fruit weight (FW), fruit shape index (SI), locule number (LN), locule number (LN), seed weight (SW), fruit set (FS) and plant yield (Y).

N°	Bulk	Accession Traditom code	Name	Fruit shape	SP (score)			FW (g)	SI	LN	SW (mg)	FS (score)	Y (kg)
					E	M	L						
1		TRVI0500	Valentino	Elongate	1.1	1	1	125.8	1.86	4.5	2.96	3	2.14
2		TRTH0900	Kouniadi	Round	1.3	1.2	1	208.6	0.78	6	2.66	5	2.54
3		TRVA1940	Tomate valenciano	Round	1.3	1	1	131.5	0.96	9.1	2.29	5	2.52
5		TRVI0870	Scatolone di Bolsena	Flattened	1	1	1	137.2	0.67	8	2.73	5	3.89
6		TRBA1980	Processing Tomato	Elliptic	1.1	1.4	1.1	70.5	1.6	2	3.08	5	3.09
7		TRMO0170	Bargemont	Elliptic	1	1	1	172.7	1.51	6	2.38	5	2.83
8	INS	TRPA0140	Patataro	Flattened	1.3	1.3	1	53	0.72	5	2.79	9	3.02
9		TRCA1500	Tomaquet rosa	Medium round	1	1.1	1	144	0.96	4	3.2	3	2.42
10		TRVI1900	Oval Ukrainian	Round	1.8	1.2	1	69.4	1.21	2	3.01	7	2.98
11		TRVA2130	Muchamiel	Flattened	1.3	1.6	1.3	102	0.75	6.1	3.36	1	1.44
12		TRBA2000	Cor de bou	Oxheart	1.3	1.2	1	105.4	0.92	5.5	3.66	5	3.17
13		TRHY0050	Borsalina	Pear (Albenga)	1	1	1	178.8	1.09	7	2.8	7	4.77
14		TRCA0010	Pera Girona	Pear (Girona)	1	1	1	240.5	0.97	5	1.77	3	4.6
16		TRTH2190	Roz Exohis	Round	3.8	2.6	2.7	229.6	0.72	10	2.02	3	3.64
17		TRVI1830	Bell pepper	Bell pepper	2.8	2.8	3.4	147.4	0.87	3.3	2.75	5	1.82
18		TRVA1300	Tomate Alicante	Round	2.9	2.8	3.5	220	0.82	5.2	4.19	5	2.34
20		TRMO0510	Marmande ancienne	Flattened	3.2	2.9	3.6	188.8	0.59	12.3	3.73	5	2.5
21		TRPO1340	Pumecetto	Elliptic	2	2.7	3.2	84.9	1.38	2.5	2.51	5	2.84
22		TRPO2520	Friulano	Elliptic	2.2	2.3	2.3	16.9	1.35	2.5	2.2	9	2.75
23	EXE	TRMO0280	Comète	Flattened	2.6	2.4	3.2	58.5	0.81	3.8	3.6	7	2.25
24		TRPO2000	Laviano	Round	1.9	1.6	2	56.3	0.92	2.5	2.45	5	3.39
25		TRVA1770	Pimiento	Round	2.3	2.4	2.1	51.5	0.9	3.7	3.32	9	2.78
26		TRVA0010	Muchamiel	Flattened	2.4	2.3	2.3	158	0.79	9	3.96	5	2.49
27		TRBA1930	Cor de bou	Oxheart	3.1	2.7	2.5	84	1.02	5.3	2.72	5	3.09
28		TRHY0470	Voluptuoso	Pear (Albenga)	2.8	2.5	2.9	216.5	0.96	5	2.65	5	2.1
29		TRCA0910	Montserrat	Pear (Girona)	3	2.3	2.6	100.6	0.94	4.4	2.67	3	2.18

| Chapter 4

Scoring of stigma position (Fig.4.4) in the 26 accessions largely confirmed the selection criteria; all accessions within the INS group showed a stably inserted stigma (score <2), whereas those within the EXE group scored values >2 with only an exception of genotype 24 (Fig. 4.2 and Fig. 4.3).

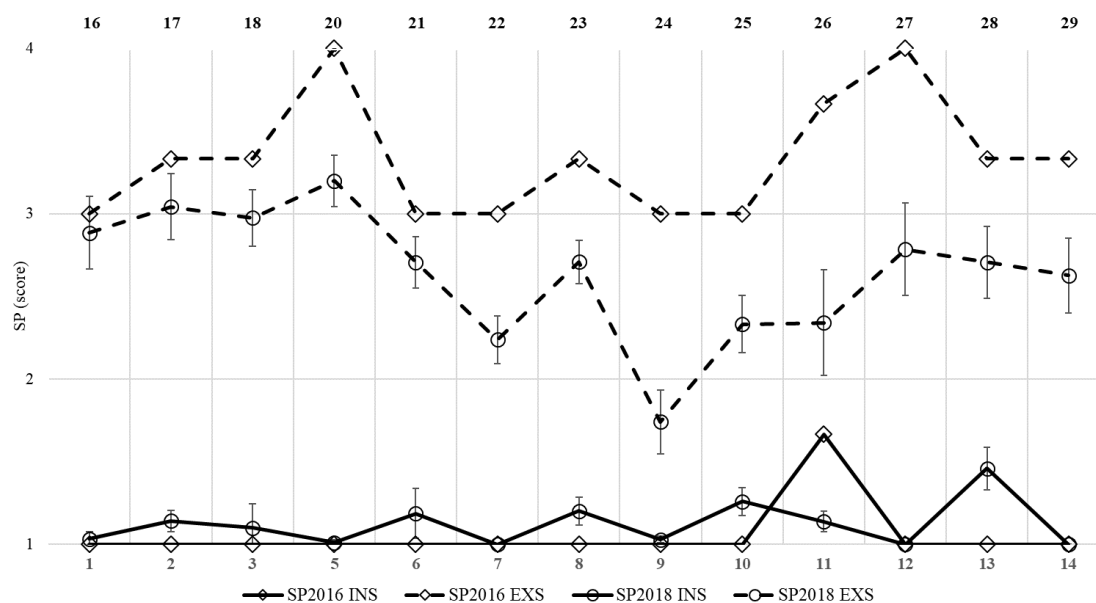


Fig 4.2 Comparison of two years and two methods to assign SP score. Continuous lines represent data of the INS group collected in the Traditom screening (-◇-) and in 2018 (-○-). Dashed lines represent the same screenings for the EXE group.

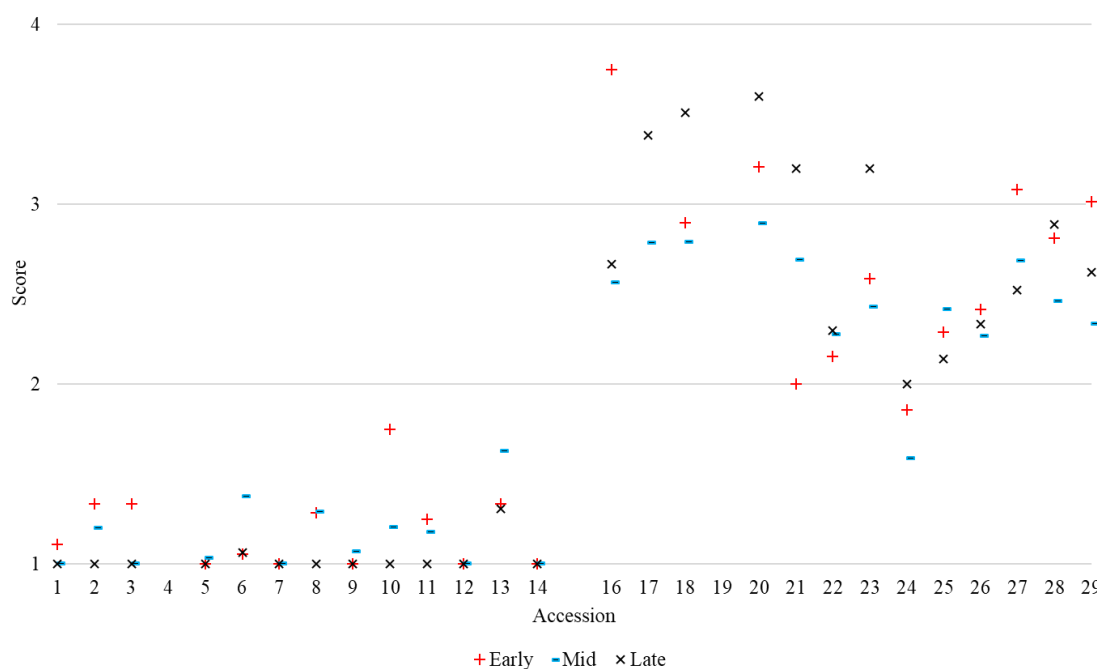


Fig. 4.3 Score of SP evaluated during the year 2108. Early, Mid and Late stage were represented with a plus (red), minus (light blue) and multiplication (black) sign. Genotypes belonging to the INS group are on the right (1-14); those belonging to the EXE group are on the left (16-29).

Considering data of two years of evaluations, value of SP was evaluated in 2016 assigning the score of opening flowers in three times (early, mid, late observations); these data were confirmed in 2017 with the same method (data not shown); in 2018, phenotyping has been considered for the entire flowering season (30-May to 6-August) with a number of observations (1572) considerably higher. This explains why Fig.4.2 shows lower values of SP in 2018, this is particularly noticeable for EXE accessions. In add, lack of interaction between years in the EXE group, confirmed the absence of strong environmental influence in these genotypes, and that the difference between observations could be mostly due to the method of investigation. However, the trend of SP in 2016 and 2018 is the same and EXE showing significantly higher values respect to the INS group. Notwithstanding the temperature tended to increase during the growing season (Supplementary Fig.4.1), the stigma position in the selected accession remained stable (Table 4.2).



Fig. 4.4 Description of the score adopted to evaluate the position of the stigma. Numbers defined an inserted stigma (1), stigma at the same level of the anther (2), stigma slightly exserted (3), stigma highly exserted (4)

The collection was very variable for fruit traits, FW ranging from about 50 to 250 g, LN from 2 to above 12 and SI from 0.67 to 1.86 (Table 4.2).

SW was also very variable, ranging from less than 2 to more than 4 g per 1000 seeds (Table 4.2). Finally, the selected accessions were highly differentiated for the scored fruit set and the estimated yield per plant. All traits, except SP, were not significantly different between groups after ANOVA (Table 4.3).

Table 4.3. Means of phenotypic traits in the two groups and ANOVA analysis. ns, not significant; ***, significant for $P \leq 0.001$.

Mean values	SP	FW (g)	SI	LN	SW (mg)	FS	Y (kg)
INS	1.12	133.8	1.1	5.3	2.8	4.8	3
EXE	2.64	142	0.9	5.4	3	5.5	2.6
ANOVA (F=0.05)	***	ns	ns	ns	ns	ns	ns

A subsequent principal component analysis (PCA) was carried out considering the dataset including (4.5A) or excluding (Fig.4.5B) the variable SP. A clear separation (Y-axis) between the two bulks can be observed if stigma data are included in the analysis. Otherwise, when stigma was excluded, the other components were describing the population without a clear specific clustering (Fig. 4.5B).

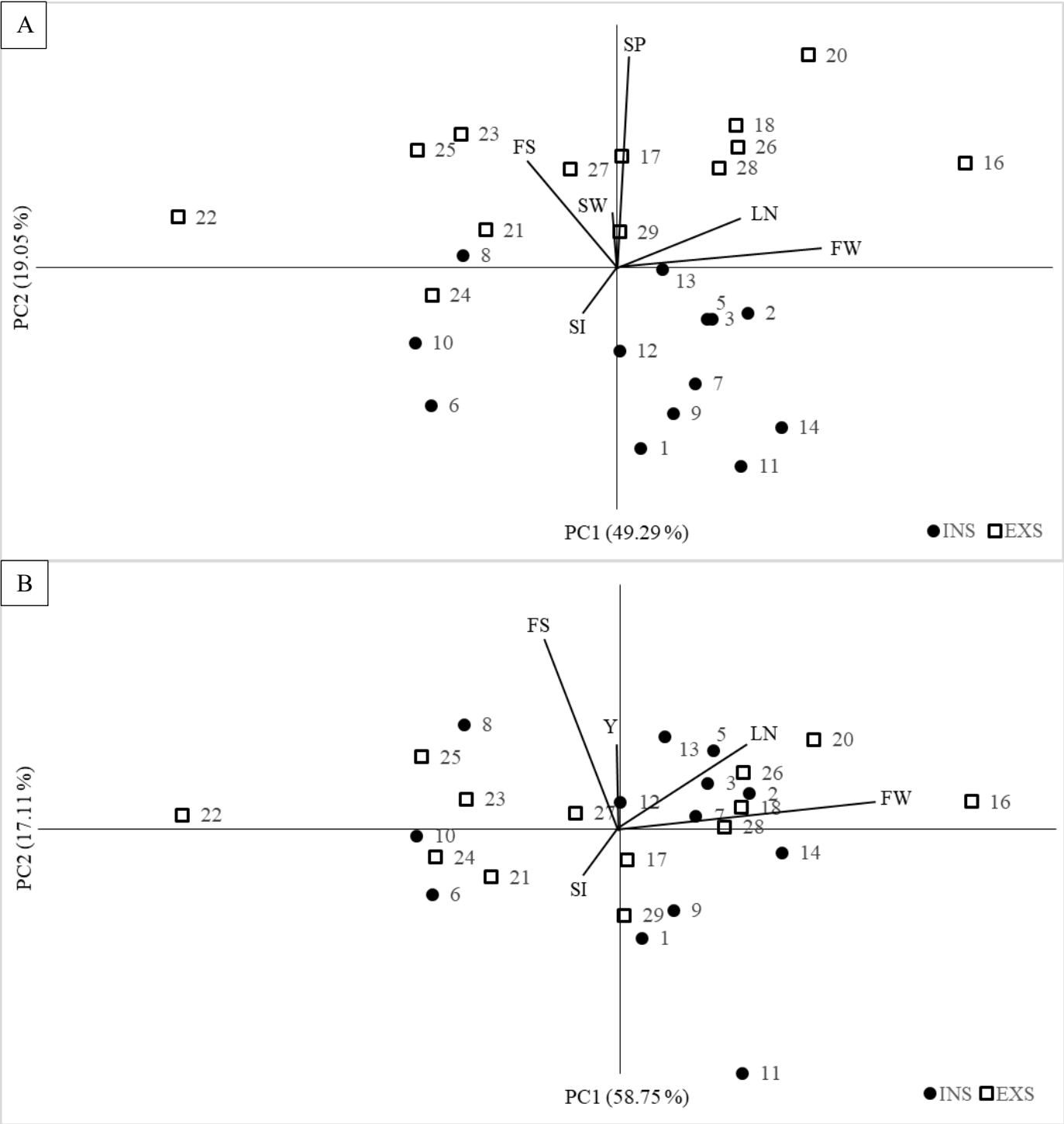


Figure 4.5A. PCA including all phenotypic data of Table1 representing the first two components. Fig. 4.5B. PCA including all data except stigma position. Numbers refer to single genotypes as described in Table 4.2

4.3.2 Dynamic of stamen and pistil growth in flower buds

In order to get information on the right stage of flower development to sample for the subsequent experiments of RNA-seq, the growth of flower organs was monitored for five stages of flower bud development. The Stages considered spanned the occurrence of male and female gametogenesis up to open flower.

Results of observations for all accessions suggest that between stage 2m and 2l was recorded the higher elongation of style in comparison to stamens growth. This was appreciable in the EXE group (Fig. 4.7), while INS accessions showed a parallel growth of pistil and stamens (Fig. 4.6).

In consequence, the present investigation on the dynamics of growth of pistils and stamens indicated the best stage to perform RNA extraction for RNA-seq analysis.

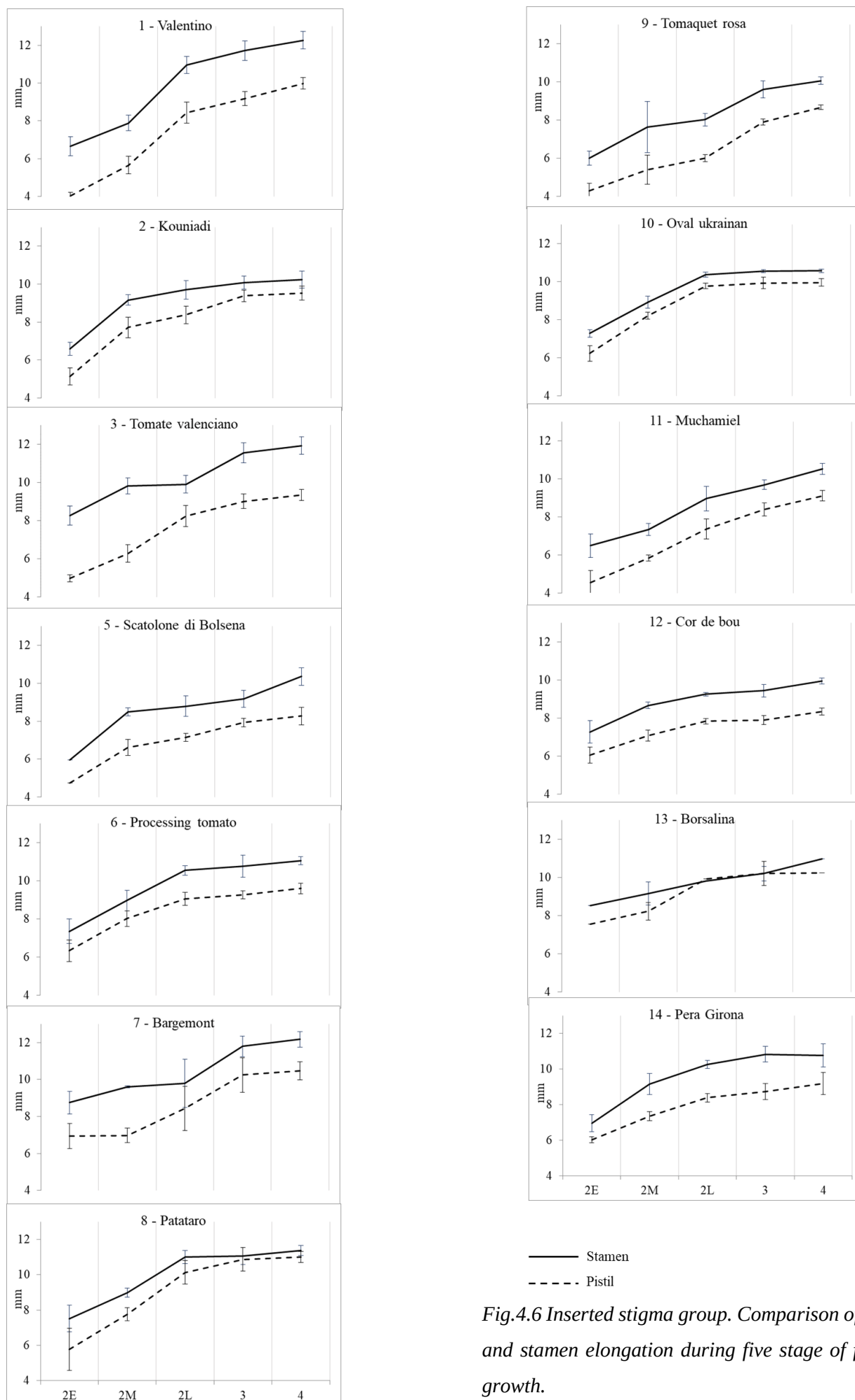


Fig.4.6 Inserted stigma group. Comparison of pistil and stamen elongation during five stage of flower growth.

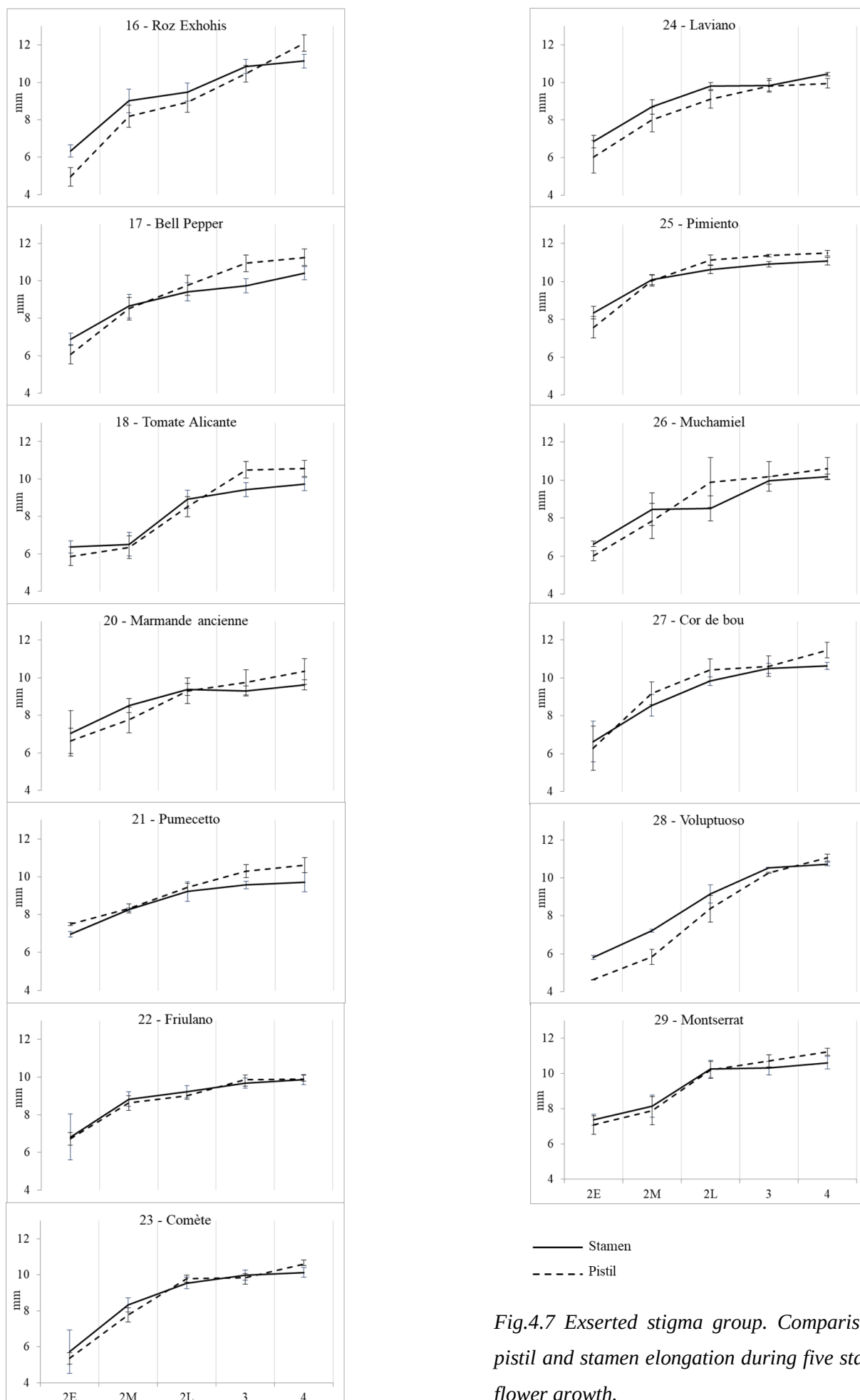


Fig.4.7 Exserted stigma group. Comparison of pistil and stamen elongation during five stage of flower growth.

4.3.3 Bulk RNA-seq analysis

Genes differentially expressed between accession with inserted and exserted stigma were identified through the comparison of three biological replicates of two contrasting bulks of 13 accessions.

RNA-seq data analysis followed three steps: read processing, statistical analysis of the data, functional annotation of the identified loci. The first step included the alignment of the reads against the reference genome. The second step involved the analysis of the read counts to each annotated locus in order to identify differential expression among samples. The third step included the attachment of biological information to the differentially expressed genes, allowing a better understanding about the relationship between an organism's genome and its phenotype.

4.3.4 Sequence reads filtering and mapping

The high-quality reads were aligned against the *S. lycopersicum* reference genome sequence (SL3.0) with STAR aligner (version 2.5.2b). Alignment statistics are shown in Table 4.3. As it can be observed, about 80% of the reads could be mapped uniquely to the genome.

Table 4.3 Alignment statistics of RNA-seq analysis showing the three replicates for each bulk analysed, number of reads before and after the quality check, percent of reads that aligned uniquely, multiple or unmapped in the reference genome.

Sample	Phenotype	Number Before QC	Number After QC	% Uniquely Mapped Reads	% Multiple Mapped Reads	% Unmapped Reads
Bulk1_A	Style Inserted	29'539'434	29'387'220	82.5%	1.5%	15.9%
Bulk1_B	Style Inserted	36'825'606	36'656'424	82.4%	1.5%	15.9%
Bulk1_C	Style Inserted	29'019'866	28'769'018	79.1%	1.5%	19.2%
Bulk2_A	Style exserted	43'618'390	43'453'684	82.5%	1.5%	15.9%
Bulk2_B	Style exserted	43'177'780	42'965'718	82.9%	1.4%	15.6%
Bulk2_C	Style exserted	37'693'658	37'472'620	82.8%	1.5%	15.6%

The overall quality of the experiment was evaluated, on the basis of the similarity between replicates, by a PCA using the normalized gene expression values as input (Figure 4.8).

Using this approach, it can be detected if any of the samples does not behave as expected.

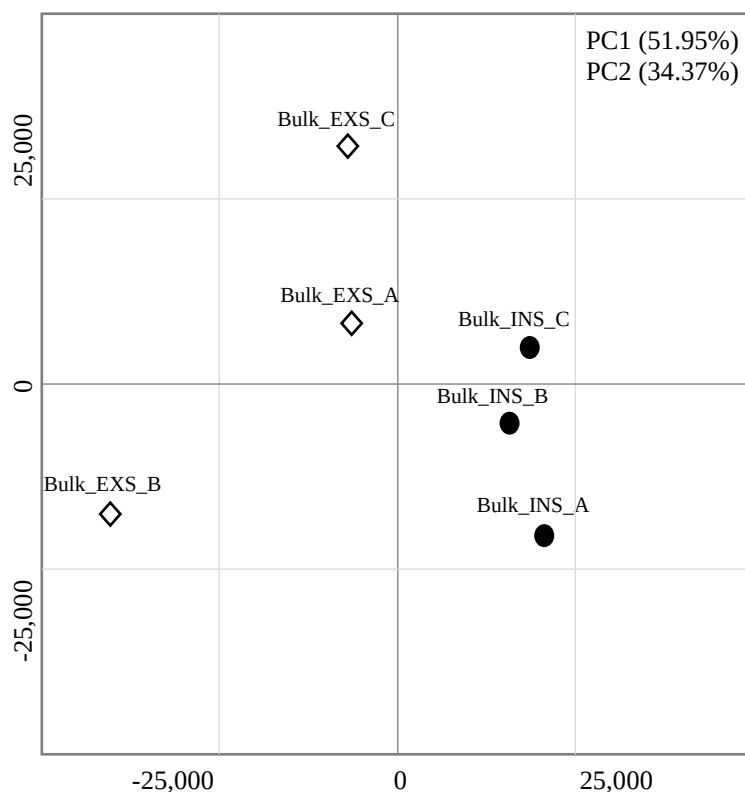


Figure 4.8 Principal Component Analysis (PCA) conducted on the normalized gene expression values of the samples. X- and Y-axes show the PC1 and PC2, respectively, with the amount of variance explained by each component is reported in parenthesis. Each point in the plot represents a sample, symbols of the same colors are replicates of a same experimental group.

Figure 4.8 shows that the two bulks are separated on the X-axis, however the sample Bulk_EXE_Rep. B seems quite different from the other two replicates of the same group. It must be considered that each sample is actually a bulk of different plants, so a certain degree of variability could be expected. However, after comparing the two bulks it appeared that most of the DEGs identified were due to a different expression specific for the Bulk_EXE_Rep.B sample. As a consequence, it was decided to adopt to conservative strategy to exclude the Bulk_EXE_Rep.B sample from the final analysis .

4.3.5 Analysis of differentially expressed genes (DEGs)

Out of 35,768 annotated coding sequences, 26,701 showed expression in at least one replicate sample in the RNA-seq analysis (not shown).

The histogram in Fig. 4.8 shows a summary of the differentially expressed genes which were found.

An MA plot (reporting log-intensity ratios M-values and log-intensity averages A-values) and a Volcano plot were also generated (Supplementary fig. 4.3 and 4.4).

RNA-seq analyses produced 17552 genes that passed the low expression filter. Among them, 801 resulted differentially expressed genes (DEGs), divided in 566 upregulated and 235 downregulated genes (Figure 4.9).

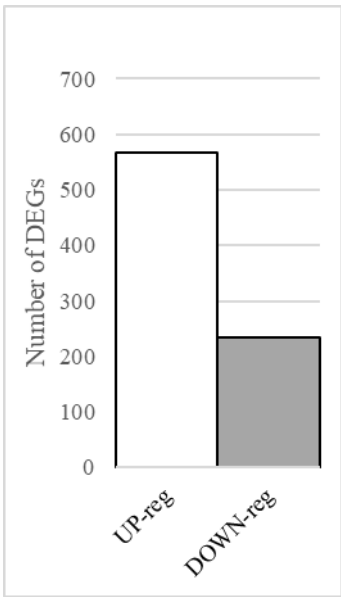


Figure 4.9 Differentially expressed genes (DEGs) identified in the RNA-seq platform as divided in upregulated (UP-reg) and downregulated (DOWN-reg) in the EXE bulk.

4.3.6 Validation of RNA-seq data by RT-qPCR

In order to validate bulk RNA-seq data, a quantitative RT-PCR was performed on the same set of bulked RNA samples used for RNA-seq analysis.

The validation was done on 20 DEGs (13 upregulated and 7 downregulated) selected for their high significance ($FDR < 1e10^{-5}$), high Log_2FC (> 0.8), including information of enrichment analysis and information of previous works. Among these, the choice was focused on DEGs auxin related (Tian et al. 2004; Wang et al. 2005; Zhang et al. 2007; De Jong et al. 2009; Li et al. 2016), style specific and involved in elongation of tissues (Hyodo et al. 2003; Leseberg et al. 2008; Lee et al. 2010; Hara et al. 2013; Huang et al. 2017).

The list of 20 genes and criteria of selection were reported in Tab. 4.4.

The correlation between the two methods was measured by regression analysis of log_2 fold changes (log_2FC) calculated after RNA-seq and after RT-qPCR. A very high correlation was observed ($P < 0.001$) between the results from RT-qPCR and RNA-seq, with an R^2 value of 0.93 (Figure 4.10).

A comparison of gene expression estimation of validated genes by RT-qPCR is also given in Fig. 4.11. Concordance of directionality of fold change values between the two experiments confirms the RNA-seq differential gene expression results. Further details on the real interest of this group of genes, could be sought from an investigation on singles accessions used to compose each bulk.

Table 4.4 List of 20 DEGs selected for validation on RT-qPCR.

Up/Down regulation	Gene ID	Gene Description	Name ^a	Log ₂ FC	FDR	Reasons for selection
Up-	Solyc01g010135.1	Basic helix-loop-helix DNA-binding superfamily	BHLH135	0.63	1.27E-14	High FDR
Up-	Solyc01g099630.3	Xyloglucan endotransglucosylase-hydrolase 1	XYLO630	0.87	6.86E-07	Growing tissues related (Hyodo et al. 2003; Hara et al. 2013; Lee et al. 2010)
Up-	Solyc01g111145.1	Proline-rich protein	PRP	0.79	1.89E-15	High FDR
Up-	Solyc02g037530.3	Auxin response factor 8B	ARF8	0.11	4.80E-02	Auxin metabolism (De Jong et al. 2009; Li et al. 2016; Tian et al. 2004)
Up-	Solyc02g078100.3	Pollen Ole e 1 allergen/extensin	EXT100	0.55	0.00E+00	High FDR/pistil extension-like protein
Up-	Solyc03g026000.3	Cold regulated protein 27	COLDRP	0.57	5.00E-02	Abiotic stress related/Style predominant
Up-	Solyc04g076850.3	Entire	IAA9	0.03	4.88E-02	Auxin metabolism (Wang et al. 2005; Zhang et al. 2007)
Up-	Solyc05g045670.3	Glucose-6-phosphate/phosphate-translocator	GLU6PHOS	1.64	0.00E+00	High FDR/High Log ₂ FC
Up-	Solyc06g051140.3	Ubiquitin-conjugating enzyme E2	UBI140	3.97	2.03E-14	High FDR/High Log ₂ FC
Up-	Solyc07g007710.3	Defensin protein	DEF710	1.39	4.94E-02	High Log ₂ FC/Style predominant
Up-	Solyc07g008380.2	HXXXD-type acyl-transferase family protein	HXXXD380	0.71	0.00E+00	High FDR/pistil specific
Up-	Solyc07g032080.3	V-type proton ATPase subunit a	V-TYPE	1.68	1.56E-10	High FDR/High Log ₂ FC
Up-	Solyc09g091550.3	S-adenosyl-L-methionine: salicylic acid carboxyl	S-ADEN	1.17	5.33E-15	High FDR/High Log ₂ FC
Down-	Solyc01g081250.3	Glutathione s-transferase, putative	GLU-S-TRA	-3.34	7.06E-14	High FDR/High Log ₂ FC
Down-	Solyc01g100000.3	F-box protein	FBOX000	-0.53	4.96E-02	GOEA -multi represented
Down-	Solyc02g080510.1	Senescence regulator	SEN510	-0.46	2.84E-07	High FDR/Pistil predominant/Gibberellin metabolism
Down-	Solyc04g064763.1	Zinc finger CCCH domain-containing protein 38	CCCH	-4.7	1.98E-14	High FDR/High Log ₂ FC
Down-	Solyc10g055570.2	SKP1-like protein	SKP1	-4.23	8.10E-15	High FDR/High Log ₂ FC
Down-	Solyc11g028020.2	Tomato AGAMOUS like 11	TAGL11	-0.57	4.28E-02	Tissue differentiation (Leseberg et al. 2008; Huang et al. 2017)
Down-	Solyc11g028100.1	Elongation factor 1-gamma 1	EF1G1	-0.65	1.21E-14	High FDR

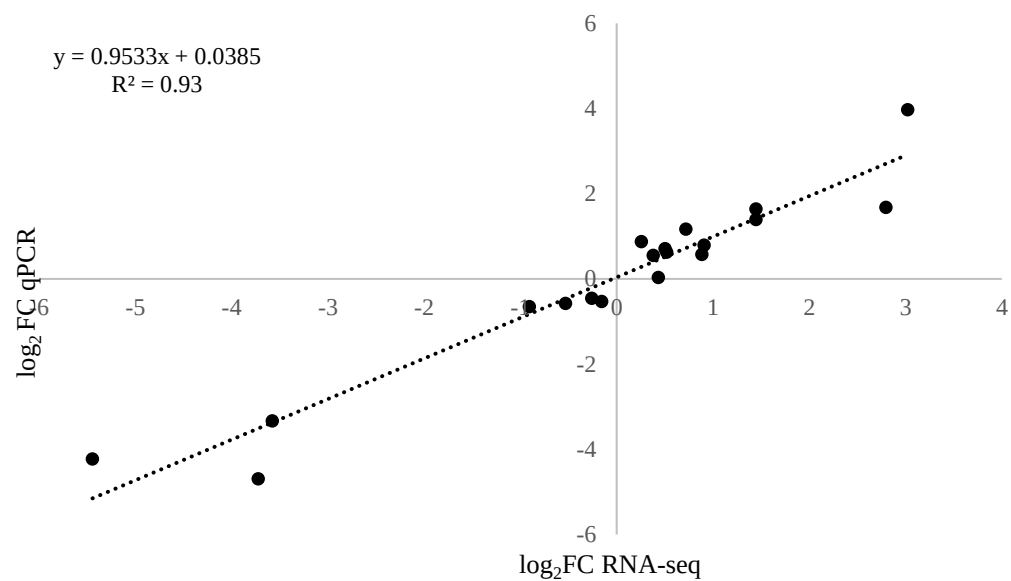


Figure 4.10 Validation of RNA-seq selected genes using RT-qPCR. Regression plot indicating the relationship between $\log_2FC$ RT-qPCR results (Y-axis) of 19 selected genes and the corresponding data from $\log_2FC$ RNA-seq analysis (X-axis).

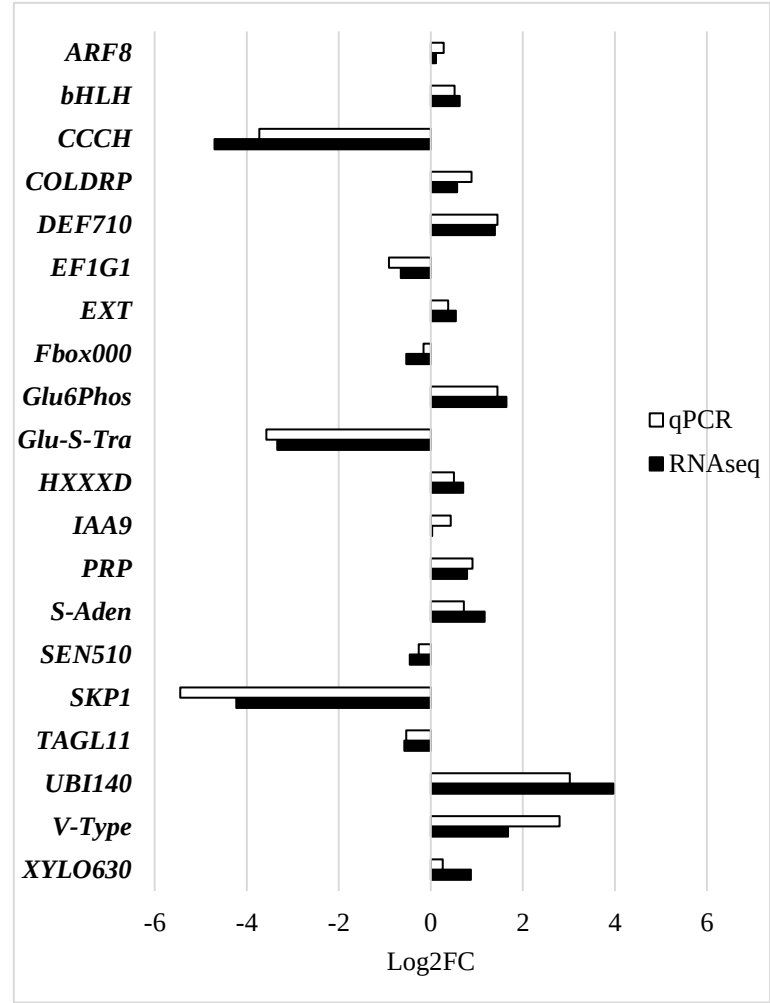


Figure 4.11 Validation of RNA-seq differential gene expression results. $\log_2$ transformed fold change values comparisons were obtained from RNA-seq platform data (black bars) and from quantitative RT-PCR (white bars).

4.3.7 GO Enrichment and network analysis

For the significantly differentially expressed genes a Gene Ontology Enrichment Analysis (GOEA) was performed to identify the most enriched categories across the down- and up-regulated genes (Fig. 4.12). GO terms that occurred frequently in the list of DEGs are said to be over-represented or enriched (Chen, 2016).

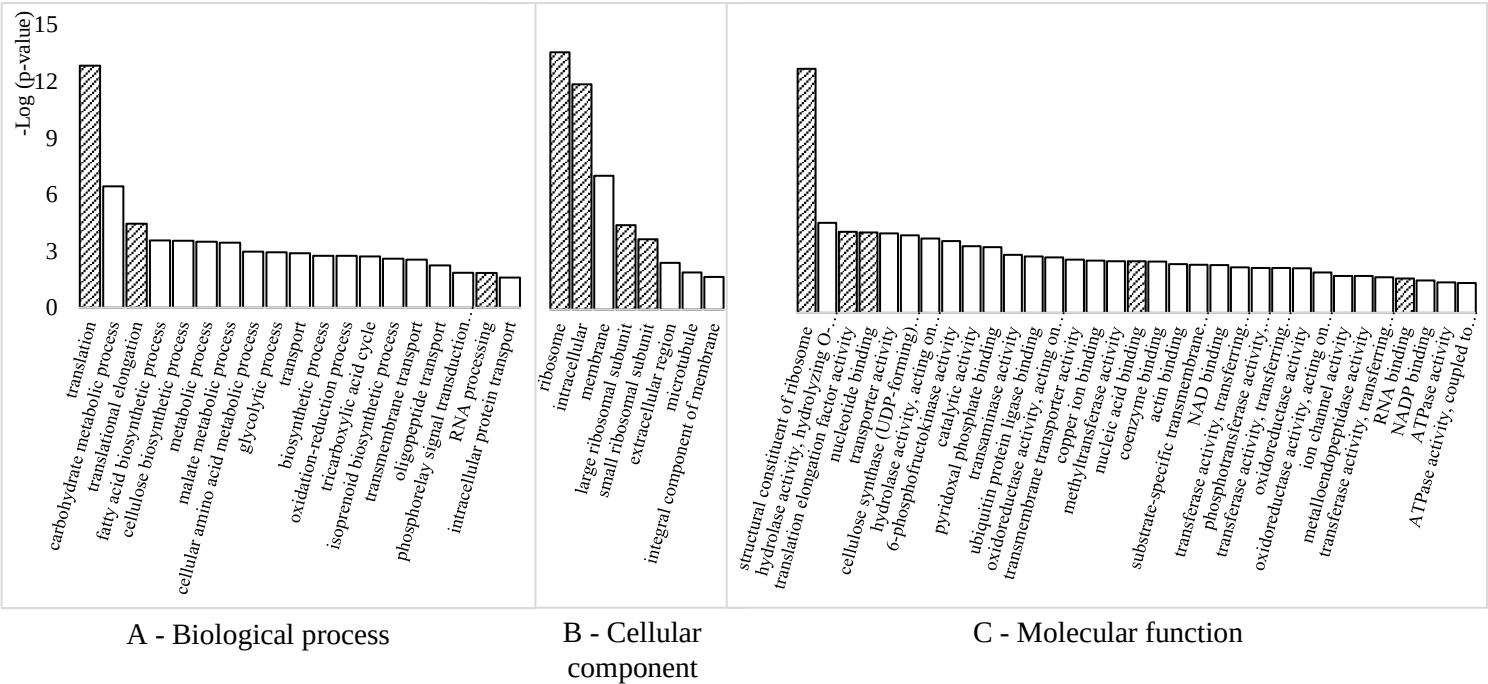


Figure 4.12 Go analysis of the DEGs up-regulated (white bars) and down-regulated (striped bars). The X-axis represents various GO terms in the three categories of Biological process, Cellular component, and Molecular function. The Y-axis indicates the Log10 p-value of a specific GO term in the corresponding category.

Enriched Biological process categories of up-regulated genes were represented by metabolism-related processes, such as carbohydrates, lipid and amino acid metabolism (Fig. 4.12A). The most enriched term for these genes in Cellular component was Membrane (Fig. 4.12B) and in Molecular function were those related to hydrolytic and biosynthetic activity (Fig. 4.12C).

Downregulated genes were enriched in terms related to mechanism of translation and translational elongation, expressed in ribosomes and in the intracellular matrix. Terms referred to nucleotide and nucleic acid binding were also enriched in down-regulated DEGs (Fig. 4.12C).

Successively, DEGs were submitted to further analysis, in order to integrate GO terms with KEGG pathways, to create functionally organized GO category networks (Bindea et al 2009; 2013). Only significant ($p \leq 0.05$) terms belonging to GO and KEGG ontologies were considered for the network analysis.

Up-regulated genes were represented by several terms related to the “Cell wall” GO term group. Biological processes involved in this category were carbohydrate, glucan, sucrose, polysaccharide metabolic processes. The complete list of genes and terms description related to cell wall was represented in Table 4.5 .

Here were found four members on Chr 1; one of these, XTH1 (in this study XYLO630) , was selected and validated in RT-qPCR analysis. Other pathways over-represented in Up-regulated genes, were Lipid transport, Auxin metabolism and Amino acid metabolism (Fig. 4.13).

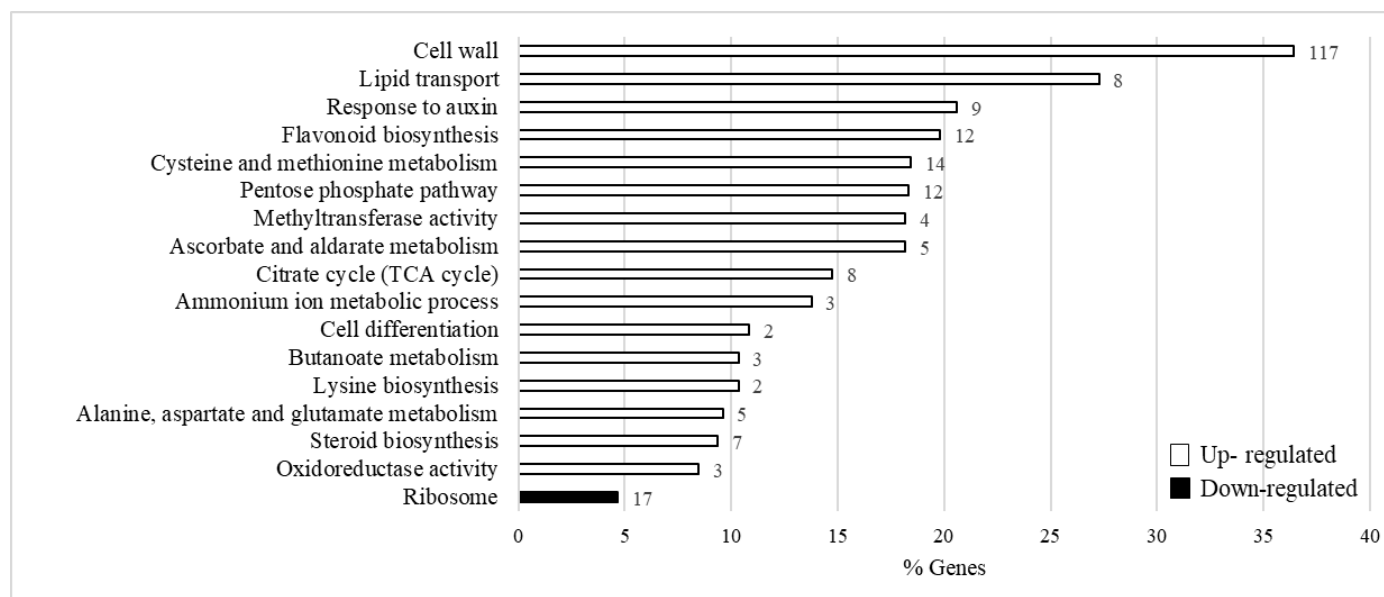


Fig.4.13 Percent of DEGs for each group of GO term. White and black bars represent respectively Up- and Down-regulated genes.

Down-regulated DEGs were only enriched in terms of ribosome pathway, suggesting that protein synthesis was suppressed in the EXE (exserted stigma) genotypes.

Table 4.5. List of DEGs involved in cell wall metabolism and pathway. Genes were reported with the ratio of expression between the EXE and the INS bulk (logFC) and FDR from RNA-seq analysis

Solyc code	Gene name		Description	logFC	FDR
Solyc01g067660.3	1,4-alpha-glucan-maltohydrolase	bmy3	polysaccharide metabolic process; hydrolase activity	-0.18	4.96E-02
Solyc01g099630.3	Xyloglucan endotransglucosylase-hydrolase 1	XTH1	cell wall polysaccharide metabolic process; hemicellulose metabolic process	0.87	6.86E-07
Solyc01g104950.3	LEXYL2	LOC543515	hydrolase activity; hydrolyzing O-glycosyl compounds	0.10	4.97E-02
Solyc01g109790.3	ADP-glucose pyrophosphorylase large subunit 1	AgpL1	cellular glucan metabolic process; glucan metabolic process	0.31	4.57E-02
Solyc02g084720.3	Beta-galactosidase 6	TBG6	galactosidase activity; cell wall	0.25	4.98E-02
Solyc02g088100.3	Expansin precursor 5	EXP5	plant-type cell wall organization; plant-type cell wall organization or biogenesis	0.47	3.99E-02
Solyc02g091920.3	Xyloglucan endotransglucosylase-hydrolase 7	XTH7	cell wall polysaccharide metabolic process; hemicellulose metabolic process	0.96	4.69E-02
Solyc03g019790.3	Alpha-galactosidase	AGAL	galactosidase activity; hydrolase activity	0.79	4.75E-02
Solyc04g007690.3	SIPIN3	PIN3	auxin-activated signaling pathway; cellular response to auxin stimulus	0.42	4.19E-02
Solyc05g005550.3	Polygalacturonase non-catalytic subunit AroGP2	AROGP2	cell wall; external encapsulating structure	0.19	4.78E-02
Solyc05g007830.3	Expansin12	exp12	plant-type cell wall organization; plant-type cell wall organization or biogenesis	-0.19	4.97E-02
Solyc05g053010.1	Lectin receptor kinase	LOC101247920	cell periphery;	0.26	4.78E-02
Solyc06g071400.3	SWEET	std1	organic substance transport; cell periphery	0.51	4.08E-02
Solyc07g007790.3	Sucrose-phosphate synthase	sps	sucrose synthase activity; UDP-glucosyltransferase activity	0.27	4.00E-02
Solyc07g043490.1	UDP-Gal:Tomatidine Galactosyltransferase	GAME1	transferase activity; transferring hexosyl groups	0.53	2.22E-16
Solyc09g010860.3	Expansin precursor 4	EXPA4	plant-type cell wall organization; plant-type cell wall organization or biogenesis	0.48	3.32E-02
Solyc12g009300.2	SuSy	LOC543961	sucrose synthase activity; UDP-glucosyltransferase activity	0.16	4.26E-02

4.3.8 Transcription factors listed among DEGs involved in stigma position

Considering the result of enrichment of Down-regulated genes, a general reduction of transcription could be expected. Thus, an investigation on Transcription Factors (TFs) has been also reported. Out of 801 DEGs, 69 genes belonging to 25 family classes were annotated as TFs in the tomato genome. Overall, 37 TFs were up-regulated and 33 down-regulated in the EXE bulk (Table 4.6). The TFs involved in auxin response, such as ARF, Aux/IAA and ARR TFs, were all up-regulated in the EXE bulk.

Table 4.6 Up- and down-regulated DEGs annotated as transcription factors.

TF family	Up-regulated DEGs			Down-regulated DEGs		
	N	Code	Name	N	Code	Name
ARF	3	Solyc02g037530; Solyc11g069190; Solyc12g006340	ARF4, 6, 8B	0		
ARR-B	2	Solyc05g054390; Solyc06g048600	ARR1, ARR17	0		
AT-HOOK	0			1	Solyc04g076220	
Aux/IAA	2	Solyc03g120390; Solyc04g076850	IAA9, IAA15	0		
B3	1	Solyc08g006260		0		
BES1	1	Solyc02g063010		0		
bHLH	4	Solyc01g010135; Solyc04g007300; Solyc10g006430; Solyc11g005780		2	Solyc07g005400; Solyc07g052670	
bZIP	1	Solyc06g009640		2	Solyc04g071510; Solyc04g078840	
CAMTA	1	Solyc01g105230		0		
DOF	1	Solyc11g010940		1	Solyc09g010680	
EIL	0			1	Solyc01g009170	
GATA	1	Solyc08g074865		0		
GRAS	2	Solyc06g082530; Solyc11g011260	GAI	2	Solyc01g090950; Solyc04g014830	
HD-ZIP	1	Solyc02g080260	Woolly/PDF 2	0		
MICK-MADS	1	Solyc07g055920	TAGL1	1	Solyc11g028020	TAGL1 1
MYB	0			1	Solyc06g009710	MYB11 1
MYB-rel	3	Solyc02g086683; Solyc02g086687; Solyc06g082510		0		
NAC	0			3	Solyc06g060230; Solyc06g073050; Solyc12g013620	LEJA2
SRS	1	Solyc11g064800		0		
SWI	2	Solyc05g054560; Solyc06g050510	CHR11, CHR42	1	Solyc06g060120	CHB4
TAZ	0			1	Solyc12g038420	
TRIHILIX	0			1	Solyc01g096470	
WD40	2	Solyc06g051380; Solyc07g025140		5	Solyc01g087670; Solyc01g111590; Solyc04g078780; Solyc07g008860; Solyc07g066410	
ZF	5	Solyc07g047940; Solyc09g008430; Solyc02g087970; Solyc08g077060; Solyc08g082960	ZF49	6	Solyc10g083460; Solyc01g057780; Solyc01g107170; Solyc02g068990; Solyc04g064763; Solyc05g052550	ZF38
Unclassified	3	Solyc05g051240; Solyc02g081860; Solyc01g005430		5	Solyc03g116680; Solyc10g009410; Solyc08g078670; Solyc09g011830; Solyc04g072500	

Figure 4.14 describes how the percent of TF in Down-regulated group is higher, despite the similar number of TFs in the two groups. This result confirms that the transcription and translational activities in the exerted group is lower respect to the insert group of genotypes.

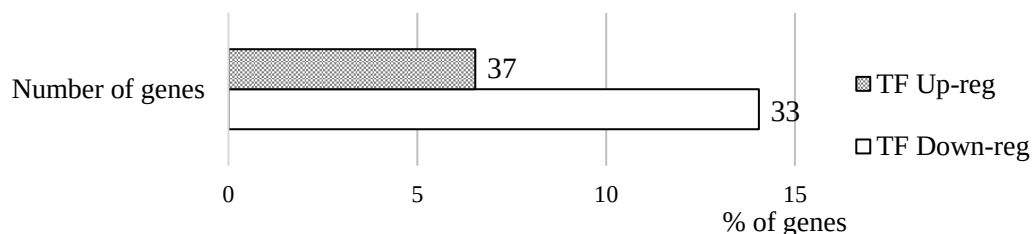


Fig. 4.14 Percent and total number of TF genes . X-axis represent the percent of TF in Up- and Down-regulated genes calculated on the total number of DEGs. Total number of TF genes is reported beside each bar.

4.3.9 Analysis on tissue specificity

Nearly 20% of total DEGs were found with a tissue specificity expression according to the CoNekT annotation. The most represented tissue was “Style”, with 62 (7.74% of total DEGs) organ specific DEGs found.

Other organs of reproductive and vegetative part were represented; the second organ with a high number of specific DEGs was “Pollen” where about 97% of the genes were up-regulated. This indicated that in the EXE bulk events of pollination had probably taken place. Other genes specific of the female reproductive organs were found, with a predominant presence of up-regulated genes. In total, the 85.5% of specific expression genes were upregulated. Only genes specific of “Ovule” were mainly (87.5%) down-regulated. Also, genes reported as specific of vegetative organs were found, as root, seedling and hypocotyl specific genes; no DEG with specificity for leaf tissue was found (Table 4.7).

Table 4.7 . Result of investigation in categories and sub-categories with the number and percent of genes specific for each tissue. The table include that tissue-specific genes used in RT-qPCR validation

Category	Sub-category	Tissue	DEGs				Validated
			No. of DEGs	Percent on total DEGs (%)	Up	Down	
Reproductive	Flower development		9	1.12	8	1	
	Female organs development	<i>Style</i>	62	7.74	59	3	<i>COLDRP, S-ADEN</i>
		<i>Ovule</i>	8	1.00	1	7	
		<i>Fruit</i>	1	0.12	1	0	
		<i>Pericarp</i>	14	1.75	11	3	
		<i>Septum</i>	4	0.50	3	1	
		<i>Seed</i>	3	0.37	2	1	<i>SKP1</i>
	Male organs development	<i>Pollen</i>	35	4.37	34	1	
Vegetative		<i>Root</i>	14	1.75	11	3	<i>CCCH</i>
		<i>Seedling</i>	4	0.50	3	1	
		<i>Hypocotyl</i>	5	0.62	4	1	
		<i>Leaf</i>	0	-	-	-	
Total specific			159	19.85	136	23	
Total non-specific			642	80.15	430	212	
Total			801	100	566	235	

Among these genes, were found *COLDRP* and *S-ADEN* genes, included in a subsequent validation. Other genes validated, as *SKP1*, has been found with a seed specific expression, while *CCCH* has been found specific of root tissue.

4.3.10 RT-PCR and RT-qPCR analysis of the selected DEGs on single genotypes

From the list of 20 genes, a semi-quantitative RT-PCR analysis (RT-PCR) was performed for all single genotypes of the two bulks (not shown). This analysis has been useful to observe how the expression of selected genes was distributed on the genotypes of each bulk.

RT-PCR on single genotypes revealed that bulk's results not necessarily indicate a higher or lower expression of a gene for all genotypes. In fact, in many cases, differences between bulks were due to the expression pattern of one or few single genotypes. In other cases, there was not a trend for genotypes to describe a clear pattern.

Results of RT-PCR were used to select six genes to be studied on single genotypes by RT-qPCR. The six genes selected were *BHLH135*, *CCCH*, *COLDRP*, *S-ADEN*, *SKP1*, *XYLO630*. All genes, except for *SKP1* were selected for a general up- or down-regulation in most of the genotypes of the group.

SKP1 was selected because it showed the highest significant value as down-regulated gene.

Analysis with RT-qPCR was performed for all the single genotypes belonging to each bulk for the six selected genes (Fig. 4.15).

To meet the assumption of homoscedasticity, for three DEGs (*COLDRP*, *S-Aden*, *CCCH*) it was necessary to exclude the B replicate (that was excluded also in the RNA-seq analysis) and for *CCCH* and *SKP1* to adopt a logarithmic transformation. After two-ways hierarchical Anova, four genes showed significant differences between bulks and four among genotypes within bulk (Table 4.8).

Table 4.8. Results of hierarchical analysis of variance of the six genes selected for RT-qPCR.

Name ^a	Log ₂ FC ^b	FDR ^b	Bartlett Rep A B C		Bartlett Rep A C		ANOVA1 (Gen)		Mean Squared Error		F (P=0.05) Crit. value df(12)=2.69	ANOVA2	
			INS	EXE	INS	EXE	INS	EXE	INS	EXE		Bulk	Gen (Bulk)
<i>BHLH135</i>	0.63	1.27E-14	*	ns			***	ns	0.05	0.15	3 - *	***	*
<i>XYLO630</i>	0.87	6.86E-07	ns	ns			ns	ns	0.22	0.33	1.5 - ns	**	ns
<i>COLDRP</i>	0.57	5.00E-02	ns	***	ns	ns	ns	ns	0.25	0.26	1.04 - ns	*	ns
<i>S-ADEN</i>	1.17	5.33E-15	**	ns	ns	ns	***	ns	0.05	0.09	1.8 - ns	ns	***
<i>CCCH</i>	-4.7	1.98E-14	ns	**	ns	*	**	ns	0.24	0.17	1.41- ns	*	**
<i>SKP1</i>	-4.23	8.10E-15	ns	ns			***	ns	0.6	0.4	1.50- ns	ns	***

^bData from RNA-seq analysis.

BHLH135 showed both differences between bulks and among genotypes within bulk. One-way anova indicated that significant differences within bulk were only found for the INS bulk. Considering INS genotypes, Duncan's test evidenced that five genotypes were significantly different from the others (Figure 4.15-A). These five accessions have an expression of the gene similar to the genotypes of exerted group. Difference between bulks was significant for $p < 0.001$.

Results for *XYLO630* underlined no differences among genotypes in each group, with a difference between bulks for $p < 0.01$. The gene is overall expressed at higher levels in the EXE group.

Results for *COLDRP* confirm a higher expression in the EXE group ($p < 0.05$). Differences among genotypes were not detected (Table 4.9; Figure 4.15C).

Results obtained for *S-ADEN* differ somehow from what was expected from the RNA-seq platform, validation on bulks and semi-quantitative preliminary screening (Table 4.9).

As it can be observed from the figure 4.15D, Genotypes 11 and 13 in the INS group have higher expression than all the others of the two bulks. In fact, statistical analysis confirms differences among genotypes of the INS group while no significant difference between bulks were observed.

CCCH revealed significant differences both between bulks and among genotypes. In the INS group, the highest expression was in the Genotype 3 (Tomate Valenciano), with values 85 times greater than the average expression of other genotypes of the same bulk. However, the significant difference ($p < 0.05$) observed between the bulks is also due to other genotypes of the INS group.

SKP1 was selected for its highly significant down-regulation and because, as observed in the semi-quantitative preliminary analysis, the expression of the gene was relevant in the accession 5 (Scatolone di Bolsena), an unusual Marmande-like tomato with an inserted style.

Duncan's test evidenced the difference between this accession and another of the INS group (Genotype 6); the expression of all the others was comparable to zero.

Table 4.9 shows the comparison among RNA-seq platform and the other validation experiments. In all cases the up- or down-regulation of each gene was respected. Nevertheless, when result of RNA-seq platform and validation were compared with the grouping of single genotypes analysis, for *S-ADEN* it was not found a corresponding Log₂FC association.

Table 4.9 Comparison of Log₂FC values in the six selected genes. Results of RNA-seq Log₂FC were compared with the validation on bulks RT-qPCR and the analysis considering single genotypes of each group.

Gene	bulk RNA-seq	bulk RT-qPCR	Singles RT-qPCR
<i>BHLH135</i>	0.67	0.63	0.80
<i>XYLO630</i>	0.86	0.75	0.87
<i>SKP1</i>	-4.23	-5.29	-8.19
<i>COLD RP</i>	0.57	0.61	0.39
<i>S-ADEN</i>	1.16	0.71	0.16
<i>CCCH</i>	-4.69	-3.12	-4.51

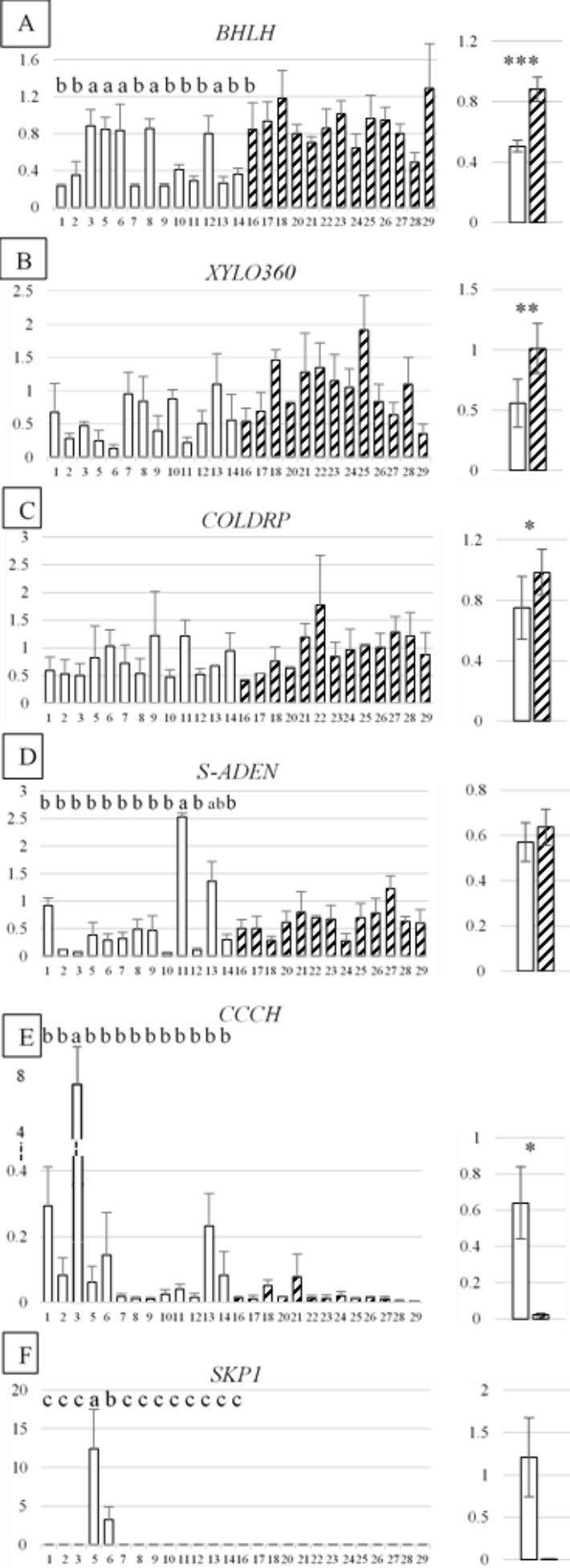


Figure 4.15 Results of RT-qPCR on single genotypes for the six selected DEGs; *bHLH135* (A), *XYLO630* (B), *COLD RP* (C), *S-ADEN* (D), *CCCH* (E), *SKP1* (F). Mean values are reported for single genotypes (left) and for the whole bulks (right). Differences among genotypes within bulk are reported after Duncan's multiple range test ($P \leq 0.05$) using different lowercase letters; differences between bulks are reported with symbols *, ** and *** indicating a significance for $P \leq 0.05$, 0.01 and 0.001, respectively.

4.3.11 Identification of structural variants from RNA-seq Data

Reads of RNA-seq have been used to search nucleotide variants occurring with different frequency between the two bulks in order to assess if a structural polymorphism could be associated to the phenotypic trait of interest.

A strong positive or negative difference between the allelic frequencies of the two bulks shows a possible relationship between the genotype and phenotype. For each variant identified, was applied a z-test to evaluate the significance of difference between bulks. After a multiple testing correction, eight variants were identified with an $FDR \leq 0.01$ (Fig. 4.16).

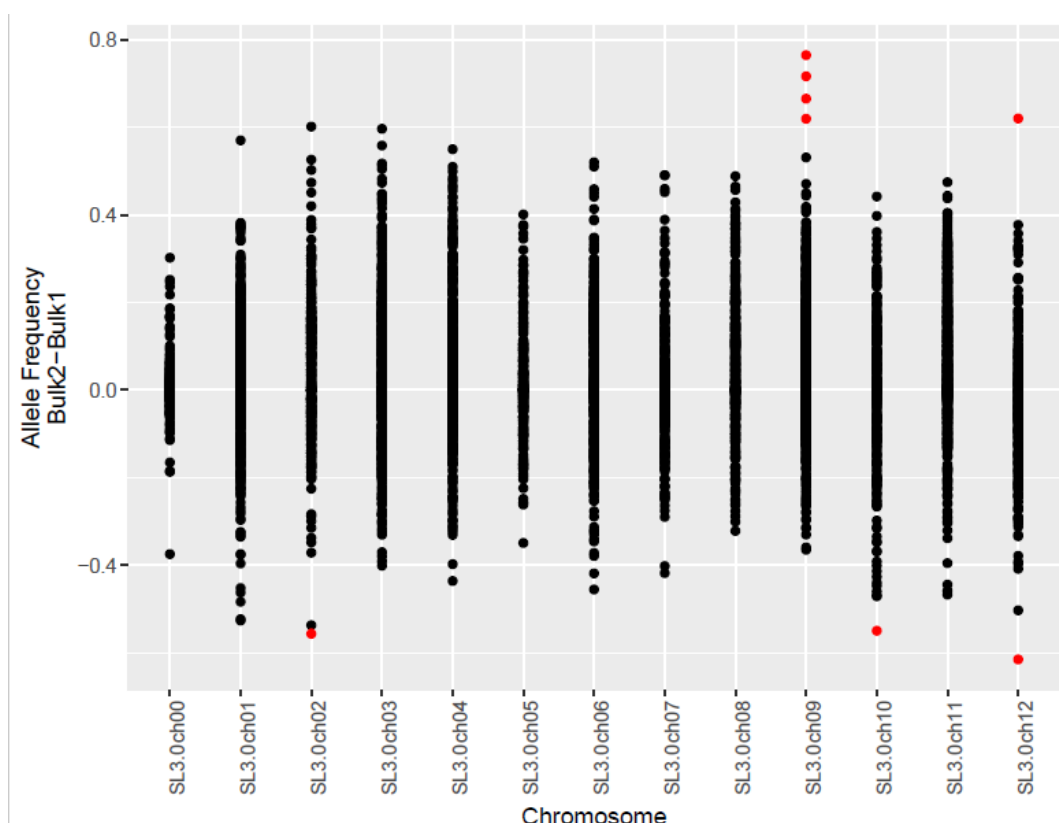


Figure 4.16 Result of variants calling performed on RNA-seq data in the comparison of the two groups. Points reported in red evidence the eight variants with a $FDR \leq 0.01$

Nevertheless, all variants with a delta value higher than 0.5 were considered in the analysis, for a total of 28 variants. Using the tomato genome SL3.0 as reference, has been studied the possible impact of the variant on the amino acid sequence.

From the analysis, 10 variants fall in UTR regions, 9 variants generated a synonym change, 7 variants appeared non-synonymous and 2 were not found, probably due to an error of annotation in the SL3.0 version or small contamination from genomic DNA. The non-synonymous variants were analysed with PROVEAN (Protein Variation Effect Analyzer) tool to predicts whether the amino acid substitution would have an impact on the

biological function of a protein. PROVEAN predicted four variants as neutral substitutions and two as deleterious. Finally, one variant represented a non-sense mutation.

The two deleterious and the non-sense variants were considered interesting for further analysis (Table 4.10).

Two of these variants were located with a distance of 94,971 kb each other, appearing in strong linkage disequilibrium (LD). In fact, the allelic frequency found in these variants are similar, with a greater frequency in the EXE bulk.

Table 4.10 List of more interesting variants identified in RNA-seq data.

Variant	Bulk INS	Bulk EXE	Delta Bulk	p-value	Gene description	Effect
A/G	0.27	0.86	0.60	2.17E-05	C2H2 Zinc finger family protein	STOP-Gln
C/G	0.28	0.80	0.52	2.23E-04	DHHA1 domain protein	Ala-Gly Predicted deleterious
C/T	0.23	0.83	0.60	1.85E-05	Heavy metal ATPase	Ala-Thr Predicted deleterious

4.4 Discussion

4.4.1 Phenotyping of SP in tomato landraces and selection of contrasting genotypes

The choice of accessions to compose the groups used in this experiment was a basic step for the transcriptomic analysis described in this chapter.

The wide variability and deep phenotyping carried out on the Traditom CC allowed the selection of genotypes with stably inserted and exerted stigma mainly belonging to European landraces.

The composition of two extreme groups allowed to compare the expression data and investigate molecular determinants of the stigma position in cultivated tomato.

A number of problems, that could affect the phenotypic characterization of the accessions, have been dealt with in this work.

The degree of SP is frequently influenced by environmental conditions, mostly by heat stress. In this case, if the plants are grown in open field, the season temperatures could influence the degree of SP of susceptible accessions. To overcome this possibility, we addressed the selection not only to the extreme tails of the CC distribution, but rather to the accessions that in multiple observations (i.e. at different temperatures) showed a stably inserted or exerted stigma.

The accession number 24, showed an average SP during the 2018 characterization with the lowest value for all EXE accessions. This accession is a round tomato type from Laviano, province of Salerno, Italy (TRPO2000) and was the only that showed two SP values (E and M) below 2 (considered as the stigma at the level of stamen cone). The rest of the selected accessions showed high stability in their degree of SP.

Investigation on pistil development of the accessions used in this work, showed that genetic determinants of style elongation in EXE varieties are expressed generally at the end of stage 2 of flower bud development. This indicated the best sampling stage for the RNA extraction.

4.4.2 RNA-seq analysis, specificity and enrichment

Analysis of DEGs showed that the differences between the INS and the EXE group mainly consisted in genes up-regulated in the EXE group.

Validation of 20 genes by qRT-PCR showed a very high degree of correspondence and thus reliability of RNA-seq analysis.

A high number of the DEGs were annotated as tissue-specific genes and most of them represented style-specific genes. This validated the reliability of the adopted experimental scheme as it could highlight genes that are specifically involved in the development of the style. Also, several pollen specific DEGs were found

in the EXE bulk probably indicating that in these samples pollen and pollen tubes could contaminate the pistil tissue sampled. Exposure of the style actually favours cross pollination in tomato that may occur before the flower is fully open (Chen et al. 2007). Notably, most of the ovule specific genes found among DEGs were down-regulated in EXE genotypes.

Enrichment analysis revealed that cell wall related genes were highly represented in upregulated DEGs. High cell wall synthesis may be related to high cell expansion and development of EXE genotypes styler tissues. Also, the only gene characterized in this context, *se2.1*, is described by Chen et al. (2007) as a positive regulator of cell elongation. Another characteristic that was noted in the cultivated germplasm, is the connection of SP with the heat stress. A congruous candidate gene for SP, could be refer to a gene of elongation of tissue, activated by abiotic stimuli (such as heat stress).

Genes encoding transcription factors were also highly represented among DEGs. This category is represented by groups of genes involved in growth and development. Family members of ARF, bHLH, GRAS and IAA deserve attention.

An involvement of genes related to auxin function in the EXE group is evident. Three ARF genes are present among up-regulated DEGs (ARF4, ARF6, ARF8B), together with an Aux/IAA (Soly03g120390) and an auxin transport protein (Soly04g076850).

The role of auxin is involved in the apical-basal patterning development of the *Arabidopsis thaliana* gynoecium; the inhibition of polar auxin transport elongates the style and stigma in the apical part of the gynoecium (Ståldal and Sundberg 2009).

Among genes down regulated in EXE are those involved in the ribosomal function, indicating that the transcriptional mechanism is partially inhibited in the EXE group. As reported by Ståldal and Sundberg (2009), many different transcription factors regulate gynoecium development and are required for carpel fusion and correct style and stigma development. However, the response of different mutants to inhibition of polar auxin transport suggests that generally auxin promotes style and stigma development and can act independently of some transcription factors involved in the gynoecium development.

In *Arabidopsis*, ARF3 plays important roles in floral meristem determinacy and floral organ patterning, leaf polarity, developmental timing and patterning, and gynoecium development (Sessions et al. 1997 ; Fahlgren et al. 2006).

4.4.3 Validation of DEGs on single genotypes

Investigation on single genotypes of 20 DEGs (RT-PCR) was useful to select the 6 most interesting genes for an evaluation in RT-qPCR.

The overall results showed that difference of expression was overall between bulks (INS/EXE), while in some cases there were strong differences among genotypes of the same bulk.

According to the statistical analysis *BHLH135* and *XYLO630* are the most interesting results, for the difference of expression between bulks.

BHLH135 is one of the six bHLH members present in the DEGs list. Is interesting to note that *se2.1*, the only characterized gene for style growth in tomato (Chen et al. 2007), is a bHLH TF member, too. Expression evidences on CoNekT database (<http://conekt.mpimp-golm.mpg.de/pub/>), show that the four up-regulated bHLH factors found in this work and *se2.1* were remarkably style and flower predominant, whereas the two down-regulated genes were mostly expressed in different organs.

Another bHLH member of this group showed orthology to *ETTIN* (*ETT/ARF3*) and *HECATE3*, two genes involved in the auxin-regulated style differentiation and growth in Arabidopsis (Sessions et al. 1997)

Other examples of interaction of ARF genes and bHLH interactions and a role in tissue elongation are reported. For example, *BIGPETALp* (*BPEp*) interacts with *ARF8* to regulate petal growth (Varaud et al. 2011); down-regulation of *ARF3* altered the expression of six bHLH in tomato (Zhang et al. 2015).

bHLH TFs have been reported to regulate many diverse biological processes; among these, they are involved in the regulation of abiotic stress-related signal transduction network (Castilhos et al. 2014; Filiz et al. 2017). Plotting the phenotypic and *BHLH135* expression data obtained in this work, regression analysis reveals high significance ($P \leq 0.001$; Fig. 4.17).

The genotype number 24 (Laviano), showed the lowest score of SP and the second lowest value of expression in the EXE group. The group of five genotypes of the INS bulk that showed higher expression values of *BHLH135* is clearly recognisable on the scatterplot.

The results of this work and the previous information on this TF family, increase the possibility that *BHLH135* is a novel HLH gene involved in the style elongation process.

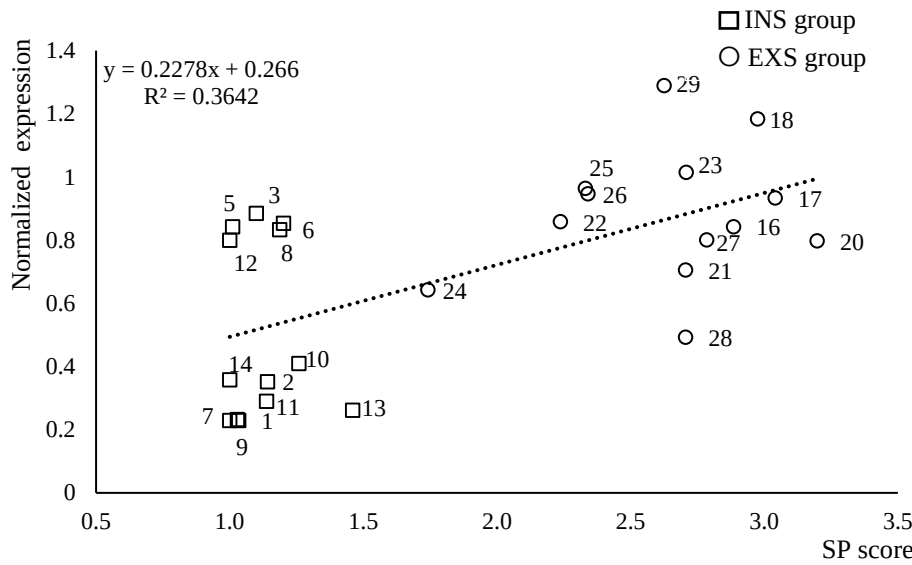


Fig.4.17 Regression analysis of phenotypic score of SP and molecular data for the BHLH135 gene. Value of statistical test: $P \leq 0.001$

XYLO630 belong to the xyloglucan endotransglucosylase/hydrolases (XTHs). Members of this family are enzymes that mediate the construction and restructuring of xyloglucan cross-links, thereby controlling the extensibility or mechanical properties of the cell wall in a wide variety of plant tissues (Kurasawa et al. 2009). The major function of XTHs in plants is to act at the microfibril–matrix interface during cell elongation (Vissenberg et al. 2005). The present work gives evidence that XTH genes are involved in style growth in tomato.

A xyloglucan endo-transglycosylase gene (*XTH7*), homologous of *XYLO630*, was described as an *ETT* target in style differentiation (Simonini et al. 2018). *ETT*, as previously described, is an orthologous of a bHLH member found in this work.

Two up-regulated DEGs in our system were homologous to *XTH7*. The one with highest FDR, Solyc01g099630, when validated on single genotypes, confirmed a lower expression in most of the INS genotypes (Fig. 4.18).

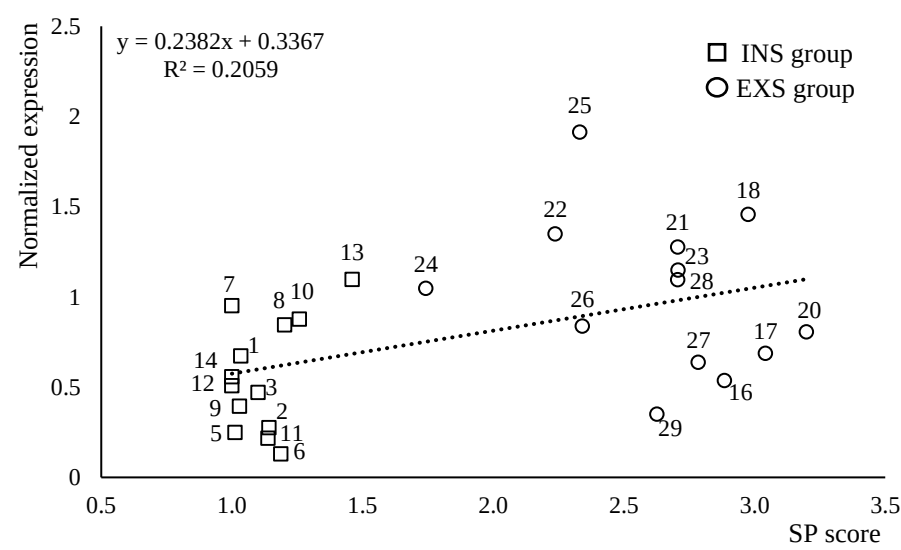


Fig. 4.18 Regression analysis of phenotypic SP scores and molecular data for the XYLO630 gene. Statistical test show significance for $P \leq 0.05$.

Table 4.11 shows the results of the regression analyses for the genes selected in RT-qPCR, considering phenotypic and molecular data of the two groups of accessions. As described, *BHLH135* and *XYLO630* show significance for $P \leq 0.05$. Regression results for the other genes are not significant, except for *CCCH* that is near to the level of significance of 0.05.

Table 4.11. Results of regression analyses between phenotypic SP score and normalized expression of each gene validated in RT-qPCR of 26 accessions.

	<i>BHLH135</i>	<i>XYLO630</i>	<i>COLDRP</i>	<i>S-ADEN</i>	<i>CCCH</i>	<i>SKP1</i>
Significance	0.001	0.012	0.40	0.60	0.057	0.12
F	***	*	ns	ns	ns	ns

Although there is a lack of correlation between molecular and phenotypic data for *COLDRP*, *S-ADEN*, *CCCH* and *SKP1*, interesting information could be obtained from these genes. *COLDRP* was found related to abiotic stresses (Wang et al. 2017) and style predominant. Not enough information was found in tomato for *COLDRP*; in Arabidosis it is involved in the regulation of circadian rhythm, response to cold, transition from vegetative to reproductive phase of meristem (Gilmour et al. 2000). In this work, RT-qPCR confirmed the result of RNA-seq, with a significant high expression in the EXE group, but the FC found between the bulks was lower respect to RNA-seq experiment.

The *CCCH* family encode zinc finger proteins that play important roles in plant growth and development, and in response to biotic and abiotic stresses. These genes are involved in the response to mannitol, salicylic acid, ethylene and approximately 82 % of all tomato *CCCH* genes have heat stress relative elements in their promoter regions (Xu et al. 2014).

S-ADEN was associated in previous works in the response to abiotic stress (salinity, water, mannitol) and hormonal treatments (abscisic acid and auxin; Espartero et al. 1994; Van Breusegem et al. 1994; Sánchez-Aguayo et al. 2004). The up- regulation of the transcripts indicate a decrease in Salicylic acid accumulation (Iberkleid et al. 2015) and in silico investigation of this work reveals a style predominant expression.

The different results obtained for this gene between qRT-PCR and RNA-seq were due to the higher expression of *S-ADEN* in the genotype 11 (Muchamiel) and 13 (Borsalina).

The greater accumulation of the transcripts in these genotypes influences the significance of the validation result.

SKP1 (*S-Phase Kinase Associated Protein 1*) is a small protein that serves as an adaptor that links the Cullin1 and F-box proteins to form the SCF complex and mediate protein degradation; it is involved in different developmental processes (Liu et al. 2004).

This gene plays important roles in regulating hormone signalling pathways (auxin, jasmonic acid and gibberellic acid; Zhang et al. 2015). Overexpression of this gene evidenced an involvement in abiotic tolerance and developmental processes in tomato (Chen et al. 2018).

Expression analyses found in this work are coincident with previous results of Zhang et al. 2015, with low expression in pistil tissues of the SPK1 family members. The expression of this genes is higher in seed and ripening stages of the fruit.

Interestingly, the expression of the gene in the INS group, is specific of two genotypes (Scatolone di Bolsena and Processing tomato). The first, is a landrace of the center of Italy, flattened and classified as Marmande-like in this work. Considering SP data of different Scatolone di Bolsena belong to the Unitus collection, they have all an inserted stigma, differing from the majority of the others Marmande-like of the Core Collection. In fact, the other 17 accessions classified as Marmande-like have an average SP score of 2.31.

4.4.4 Identification of structural variants from RNA-Seq Data

The analysis of SNPs of the RNA-seq reveal us three interesting mutation to better investigate their roles.

One of these is a non-sense mutation that blocking the transcription of a C₂H₂ Zinc finger protein, as the consequence this protein is unable to bind DNA or RNA segments and control transcription of target genes.

Further investigations are necessary to confirm if these SNP are involved in determining the stigma position in tomato; the first hypothesis could be the screening of these interesting SNPs in a wide sample of accessions, using phenotypic data available for the Traditom or other characterized populations.

Successively, it would be necessary to do a functional validation of the allele, for example with the aim to knock out the gene simulating the effect of the STOP codon.

References

- Akhtar, S., Ansary, S. H., Dutta, A. K., Karak, C., & Hazra, P. (2012). Crucial reproductive characters as screening indices for tomato (*Solanum lycopersicum*) under high temperature stress. *J Crop Weed*, 8, 114-117.
- Bernacchi, D., & Tanksley, S. D. (1997). An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. *Genetics*, 147(2), 861-877.
- Bindea G, Galon J, Mlecnik B, CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics*. 2013, 29(5):661-3.
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009, 25(8):1091-3.
- Castilhos, G., Lazzarotto, F., Spagnolo-Fonini, L., Bodanese-Zanettini, M. H., & Margis-Pinheiro, M. (2014). Possible roles of basic helix-loop-helix transcription factors in adaptation to drought. *Plant Science*, 223, 1-7.
- Chen, K. Y., & Tanksley, S. D. (2004). High-resolution mapping and functional analysis of se2. 1: a major stigma exertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus *Lycopersicon*. *Genetics*, 168(3), 1563-1573.
- 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(5850), 643-645.
- Chen, Y., Chi, Y., Meng, Q., Wang, X., & Yu, D. (2018). GmSK1, an SKP1 homologue in soybean, is involved in the tolerance to salt and drought. *Plant physiology and biochemistry*, 127, 25-31.
- Chen, Y., Lun, A. T., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, 5.
- Cheng, S. H., Zhuang, J. Y., Fan, Y. Y., Du, J. H., & Cao, L. Y. (2007). Progress in research and development on hybrid rice: a super-domesticated in China. *Annals of botany*, 100(5), 959-966.
- Dane F, Hunter AG, Chambliss OL (1991) Fruit set, pollen fertility, and combining ability of selected tomato genotypes under high-temperature field conditions. *J Am* 906–910 *Soc Hortic Sci* 116:
- De Jong, M., Mariani, C., & Vriezen, W. H. (2009). The role of auxin and gibberellin in tomato fruit set. *Journal of experimental botany*, 60(5), 1523-1532.
- Espartero, J., Pintor-Toro, J. A., & Pardo, J. M. (1994). Differential accumulation of S-adenosylmethionine synthetase transcripts in response to salt stress. *Plant molecular biology*, 25(2), 217-227.
- Expósito-Rodríguez, M., Borges, A. A., Borges-Pérez, A., & Pérez, J. A. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC plant biology*, 8(1), 131.

- Fahlgren, N., Montgomery, T. A., Howell, M. D., Allen, E., Dvorak, S. K., Alexander, A. L., & Carrington, J. C. (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. *Current biology*, 16(9), 939-944.
- Filiz, E., Vatansever, R., & Ozyigit, I. I. (2017). Dissecting a co-expression network of basic helix-loop-helix (bHLH) genes from phosphate (Pi)-starved soybean (*Glycine max*). *Plant Gene*, 9, 19-25.
- Fourquin, C., & Ferrándiz, C. (2014). The essential role of NGATHA genes in style and stigma specification is widely conserved across eudicots. *New Phytologist*, 202(3), 1001-1013.
- Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., & Thomashow, M. F. (2000). Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. *Plant physiology*, 124(4), 1854-1865.
- Giorno, F., Wolters-Arts, M., Mariani, C., & Rieu, I. (2013). Ensuring reproduction at high temperatures: the heat stress response during anther and pollen development. *Plants*, 2(3), 489-506.
- Gorguet, B., Eggink, P. M., Ocana, J., Tiwari, A., Schipper, D., Finkers, R., ... & van Heusden, A. W. (2008). Mapping and characterization of novel parthenocarp QTLs in tomato. *Theoretical and Applied Genetics*, 116(6), 755-767.
- Hara, Y., Yokoyama, R., Osakabe, K., Toki, S., & Nishitani, K. (2013). Function of xyloglucan endotransglucosylase/hydrolases in rice. *Annals of botany*, 114(6), 1309-1318.
- Huang, W., Peng, S., Xian, Z., Lin, D., Hu, G., Yang, L., ... & Li, Z. (2017). Overexpression of a tomato miR171 target gene Sl GRAS 24 impacts multiple agronomical traits via regulating gibberellin and auxin homeostasis. *Plant biotechnology journal*, 15(4), 472-488.
- Hussain, S., Khaliq, A., Ali, B., Hussain, H. A., Qadir, T., & Hussain, S. (2019). Temperature Extremes: Impact on Rice Growth and Development. In *Plant Abiotic Stress Tolerance* (pp. 153-171). Springer, Cham.
- Hyodo, H., Yamakawa, S., Takeda, Y., Tsuduki, M., Yokota, A., Nishitani, K., & Kohchi, T. (2003). Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in Arabidopsis thaliana. *Plant molecular biology*, 52(2), 473-482.
- Iberkleid, I., Sela, N., & Miyara, S. B. (2015). Meloidogyne javanica fatty acid-and retinol-binding protein (Mj-FAR-1) regulates expression of lipid-, cell wall-, stress-and phenylpropanoid-related genes during nematode infection of tomato. *BMC genomics*, 16(1), 272.
- Kalisz, S., Vogler, D., Fails, B., Finer, M., Shepard, E., Herman, T., & Gonzales, R. (1999). The mechanism of delayed selfing in Collinsia verna (Scrophulariaceae). *American Journal of Botany*, 86(9), 1239-1247.
- Karron, J. D., Jackson, R. T., Thumser, N. N., & Schlicht, S. L. (1997). Outcrossing rates of individual Mimulus ringens genets are correlated with anther–stigma separation. *Heredity*, 79(4), 365.
- Kryuchkova-Mostacci, N. and Robinson-Rechavi, M. (2017) A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinform.*, 18, 205–214.

- Kurasawa, K., Matsui, A., Yokoyama, R., Kuriyama, T., Yoshizumi, T., Matsui, M., ... & Nishitani, K. (2009). The AtXTH28 gene, a xyloglucan endotransglucosylase/hydrolase, is involved in automatic self-pollination in *Arabidopsis thaliana*. *Plant and cell physiology*, 50(2), 413-422.
- Lee, J., Burns, T. H., Light, G., Sun, Y., Fokar, M., Kasukabe, Y., ... & Allen, R. D. (2010). Xyloglucan endotransglycosylase/hydrolase genes in cotton and their role in fiber elongation. *Planta*, 232(5), 1191-1205.
- Leseberg, C. H., Eissler, C. L., Wang, X., Johns, M. A., Duvall, M. R., & Mao, L. (2008). Interaction study of MADS-domain proteins in tomato. *Journal of experimental botany*, 59(8), 2253-2265.
- Levy A, Rabinowitch HD, Kedar N (1978) Morphological and physiological characters affecting flower drop and fruit set of tomatoes at high temperatures. *Euphytica* doi:10.1007/BF00039137 27:211–218.
- Li, S. B., Xie, Z. Z., Hu, C. G., & Zhang, J. Z. (2016). A review of auxin response factors (ARFs) in plants. *Frontiers in plant science*, 7, 47.
- Liu, F., Ni, W., Griffith, M. E., Huang, Z., Chang, C., Peng, W., ... & Xie, D. (2004). The ASK1 and ASK2 genes are essential for *Arabidopsis* early development. *The Plant Cell*, 16(1), 5-20.
- Liu, J., Feng, L., Li, J., & He, Z. (2015). Genetic and epigenetic control of plant heat responses. *Frontiers in plant science*, 6, 267.
- Motten, A. F., & Antonovics, J. (1992). Determinants of outcrossing rate in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). *American Journal of Botany*, 79(4), 419-427.
- Motten, A. F., & Stone, J. L. (2000). Heritability of stigma position and the effect of stigma-anther separation on outcrossing in a predominantly self-fertilizing weed, *Datura stramonium* (Solanaceae). *American Journal of Botany*, 87(3), 339-347.
- Oikkonen, L., & Lise, S. (2017). Making the most of RNA-seq: pre-processing sequencing data with opossum for reliable SNP variant detection. *Wellcome open research*, 2.
- Pan, C., Yang, D., Zhao, X., Jiao, C., Yan, Y., Lamin-Samu, A. T., ... & Lu, G. (2019). Tomato stigma exsertion induced by high temperature is associated with the jasmonate signalling pathway. *Plant, cell & environment*, 42(4), 1205-1221.
- Proost, S., & Mutwil, M. (2018). CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic acids research*, 46(W1), W133-W140.
- Rau, A., Gallopin, M., Celeux, G., & Jaffrézic, F. (2013). HTSFilter: Independent data-based filtering for replicated transcriptome sequencing experiments.
- Richards, A. J. (1997). *Plant breeding systems*. Garland Science.
- Rick, C. M. (1963). Barriers to interbreeding in *Lycopersicon peruvianum*. *Evolution*, 216-232.
- Rick, C. M. (1976). Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. *Genetica Agraria (Italy)*.

- Rick, C. M. (1982). Genetic-relationships between self-incompatibility and floral traits in the tomato species. *Biologisches Zentralblatt*, 101(2), 185-198.
- Rick, C. M., & Chetelat, R. T. (1995, March). Utilization of related wild species for tomato improvement. In *I International Symposium on Solanacea for Fresh Market* 412 (pp. 21-38).
- Rick, C. M., & Dempsey, W. H. (1969). Position of the stigma in relation to fruit setting of the tomato. *Botanical Gazette*, 130(3), 180-186.
- Rick, C. M., & Lamm, R. (1955). Biosystematic studies on the status of *Lycopersicon chilense*. *American Journal of Botany*, 663-675.
- Rick, C. M., Fobes, J. F., & Holle, M. (1977). Genetic variation in *Lycopersicon pimpinellifolium*: Evidence of evolutionary change in mating systems. *Plant Systematics and Evolution*, 127(2-3), 139-170.
- Rick, C. M., Fobes, J. F., & Holle, M. (1977). Genetic variation in *Lycopersicon pimpinellifolium*: Evidence of evolutionary change in mating systems. *Plant Systematics and Evolution*, 127(2-3), 139-170.
- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R., Wilkie, A. O., ... & WGS500 Consortium. (2014). Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics*, 46(8), 912.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140.
- Saeed, A., Bhatti, K. H., & Khan, A. (2007). Heat tolerance studies in tomato (*Lycopersicon esculentum* Mill.). *International Journal of Agriculture and Biology*, 4, 649–652.
- Sánchez-Aguayo, I., Rodríguez-Galán, J. M., García, R., Torreblanca, J., & Pardo, J. M. (2004). Salt stress enhances xylem development and expression of S-adenosyl-L-methionine synthase in lignifying tissues of tomato plants. *Planta*, 220(2), 278-285.
- Sawhney, V. K. (2004). Photoperiod-sensitive male-sterile mutant in tomato and its potential use in hybrid seed production. *The Journal of Horticultural Science and Biotechnology*, 79(1), 138-141.
- Sessions, A., Nemhauser, J. L., McColl, A., Roe, J. L., Feldmann, K. A., & Zambryski, P. C. (1997). ETTIN patterns the Arabidopsis floral meristem and reproductive organs. *Development*, 124(22), 4481-4491.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–2504.
- Simonini, S., Stephenson, P., & Østergaard, L. (2018). A molecular framework controlling style morphology in Brassicaceae. *Development*, 145(5), dev158105.
- Ståldal, V., & Sundberg, E. (2009). The role of auxin in style development and apical-basal patterning of the Arabidopsis thaliana gynoecium. *Plant signaling & behavior*, 4(2), 83-85.

- Takano-Kai, N., Doi, K., & Yoshimura, A. (2011). GS3 participates in stigma exertion as well as seed length in rice. *Breeding science*, 61(3), 244-250.
- Tarazona, S., Furió-Tarí, P., Turra, D., Pietro, A. D., Nueda, M. J., Ferrer, A., & Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic acids research*, 43(21), e140-e140.
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., ... & Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37(6), 914-939.
- Tian, C. E., Muto, H., Higuchi, K., Matamura, T., Tatematsu, K., Koshiba, T., & Yamamoto, K. T. (2004). Disruption and overexpression of auxin response factor 8 gene of Arabidopsis affect hypocotyl elongation and root growth habit, indicating its possible involvement in auxin homeostasis in light condition. *The Plant Journal*, 40(3), 333-343.
- Van Breusegem, F., Dekeyser, R., Gielen, J., Van Montagu, M., & Caplan, A. (1994). Characterization of a S-adenosylmethionine synthetase gene in rice. *Plant Physiology*, 105(4), 1463.
- Varaud, E., Brioudes, F., Szécsi, J., Leroux, J., Brown, S., Perrot-Rechenmann, C., & Bendahmane, M. (2011). AUXIN RESPONSE FACTOR8 regulates Arabidopsis petal growth by interacting with the bHLH transcription factor BIGPETALp. *The Plant Cell*, 23(3), 973-983.
- Vissenberg, K., Fry, S. C., Pauly, M., Höfte, H., & Verbelen, J. P. (2005). XTH acts at the microfibril-matrix interface during cell elongation. *Journal of Experimental Botany*, 56(412), 673-683.
- Wang, J. W., Wang, L. J., Mao, Y. B., Cai, W. J., Xue, H. W., & Chen, X. Y. (2005). Control of root cap formation by microRNA-targeted auxin response factors in Arabidopsis. *The Plant Cell*, 17(8), 2204-2216.
- Wang, P., Cui, X., Zhao, C., Shi, L., Zhang, G., Sun, F., ... & Xu, X. (2017). COR27 and COR28 encode nighttime repressors integrating Arabidopsis circadian clock and cold response. *Journal of integrative plant biology*, 59(2), 78-85.
- Xu, J., Driedonks, N., Rutten, M. J., Vriezen, W. H., de Boer, G. J., & Rieu, I. (2017). Mapping quantitative trait loci for heat tolerance of reproductive traits in tomato (*Solanum lycopersicum*). *Molecular Breeding*, 37(5), 58.
- Xu, R. (2014). Genome-wide analysis and identification of stress-responsive genes of the CCCH zinc finger family in *Solanum lycopersicum*. *Molecular genetics and genomics*, 289(5), 965-979.
- Yan, W. G., Li, Y., Agrama, H. A., Luo, D., Gao, F., Lu, X., & Ren, G. (2009). Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Molecular breeding*, 24(3), 277-292.
- Yan, W. G., Li, Y., Agrama, H. A., Luo, D., Gao, F., Lu, X., & Ren, G. (2009). Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Molecular breeding*, 24(3), 277-292.
- Zhang, H., Kim, M. S., Krishnamachari, V., Payton, P., Sun, Y., Grimson, M., ... & Paré, P. W. (2007). Rhizobacterial volatile emissions regulate auxin homeostasis and cell expansion in Arabidopsis. *Planta*, 226(4), 839.

Zhang, X., Yan, F., Tang, Y., Yuan, Y., Deng, W., & Li, Z. (2015). Auxin response gene SlARF3 plays multiple roles in tomato development and is involved in the formation of epidermal cells and trichomes. *Plant and Cell Physiology*, 56(11), 2110-2124.

Conclusions

Phenotypic diversity of the Unitus and the Core Collection yielded many information on the European germplasm of tomato. A lot of features hidden in traditional varieties are now available to rediscover these materials and to develop hybrids and/or varieties improved for modern versions of them.

Tomato is the second most consumed vegetable in Europe and the conservation and increase of the competitiveness in local and global market was one of the major aims of the Traditom project.

The overall results improve the knowledge on traditional varieties, involving also the social, cultural and environmental aspects.

First, data collected permit to growers and other stakeholders to invest in the potential of traditional varieties; consumers acquire knowledge on biodiversity, sites of origin and qualitative aspects of these varieties.

In addition, we can consider a romantic and cultural context that could be an advantage of traditional varieties; in fact, consumers could have the idea to eat the same tomato that their ancestors developed some centuries ago.

Markets could be enriched of traditional improved varieties, where fantasy names of modern hybrid could be replaced with the original name of the landrace, conserving all the features of distinctiveness, quality, taste and connection to local areas.

A specific investigation of this thesis was done on fertility related traits.

Results obtained improve the knowledge on aspects that control fertility-related organs and processes, confirming and adding new information of QTLs involved in seed size and stigma position in tomato.

Previous works were focused on the study of interspecific populations to find QTLs regulating these traits, while in this thesis it has been studied a large population of landraces to identify QTLs considering the variability still existing in traditional varieties.

GWAS on seed size identified 31 QTLs. Of them, 5 known QTLs were confirmed.

Combining GWAS of three population for a total of seven datasets, a total of 46 QTLs for stigma position have been identified. Among these, all known QTLs were confirmed and one of the novel (Chr 11) was found in all population tested.

Finally, the transcriptomic analysis described in Chapter 4 identifies a number of candidate genes for a functional analysis, in order to confirm the involvement of them in the stigma position regulation process. 801 DEGs were revealed by bulk RNA-seq analysis, with several TF, auxin-related and elongation -related process genes.

| Conclusions

In this last part of conclusions, are reported other example of genes deserving attention, obtained by crossing the results of GWAS (Chapter 3) and RNA-seq (Chapter 4); the aim is to obtain further candidates in order to improve the set of putative regulators of stigma position in the population studied.

In Chr 1, close to a marker identified in the Magic population (0,35 Mb), we found a *GRAS* TF (Solyc01g090950) that is down- regulated in the EXE genotypes.

The *GRAS* genes are transcriptional regulators which play a significant role in plant growth and development; the functions of several *GRAS* genes have been associated to phytohormones (gibberellins) signal transduction, light signalling, and abiotic/biotic stress response (Niu et al. 2017).

This particular gene, in tomato, was found regulated by the miRNA171 (Huang et al. 2015).

Furthermore, many members of *tomato GRAS* identified were described with a sharply increase or decrease upon pollination/fertilization either or both in stamen and ovary, indicating their potential active roles during ovary and anther development.

Other interesting candidate genes were found in Chr 2, at 0.53 Mb from two positions that co-localize in the MAGIC population. The candidates are a *WAT1* like (Solyc02g087060) and two *MYB* (Solyc02g086683, Solyc02g086687) members, all up-regulated in EXE genotypes.

WAT1 (Walls are thin 1), is a member of the plant drug/metabolite exporter family. *WAT1* is preferentially expressed in vascular tissues and *WAT1* mutants show defects in cell elongation (Ranocha et al. 2010). Another interesting activity of *WAT1* is the involvement as auxin transport facilitator (Ranocha et al. 2013)

At the same time, *MYB* genes play roles in many developmental processes. The overexpression of one *MYB* gene, *MYBH*, significantly increased hypocotyl elongation in *Arabidopsis thaliana*. Moreover, *MYBH* expression enhanced auxin biosynthesis (Kwon et al. 2013).

The *MYBH* gene found in *Arabidopsis* is the ortholog on another gene of tomato chromosome 8, despite this, the evidences of the involvement of this family gene in elongation and auxin biosynthesis make our candidates as interesting genes deserving further investigations (Kwon et al.2013).

Following these examples, we can consider of interest that significant genes revealed by RNA-seq co-localize with markers positions of GWAS.

In conclusion, phenotypic, molecular, association analysis and transcriptomic approaches are different strategies that singularly or together may shed light on fertility traits of tomato.

This work identifies wide perspectives to improve the knowledge on molecular and genetic control of stigma position in tomato and therefore for improving the trait by breeding.

References

- Huang, W., Xian, Z., Kang, X., Tang, N., & Li, Z. (2015). Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC plant biology*, 15(1), 209.
- Kwon, Y., Kim, J. H., Nguyen, H. N., Jikumaru, Y., Kamiya, Y., Hong, S. W., & Lee, H. (2013). A novel Arabidopsis MYB-like transcription factor, MYBH, regulates hypocotyl elongation by enhancing auxin accumulation. *Journal of experimental botany*, 64(12), 3911-3922.
- Niu, Y., Zhao, T., Xu, X., & Li, J. (2017). Genome-wide identification and characterization of GRAS transcription factors in tomato (*Solanum lycopersicum*). *PeerJ*, 5, e3955.
- Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratgé-Faillie, C., Novák, O., ... & Jin, X. (2013). Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nature Communications*, 4, 2625.
- Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., ... & Boerjan, W. (2010). Walls are thin 1 (WAT1), an Arabidopsis homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant Journal*, 63(3), 469-483.

Supplementary materials

Supplementary Table 2.1. Phenotyping kit 1. Variables of plant and fruit evaluated in the Unitus collection in 2015 with a brief description.

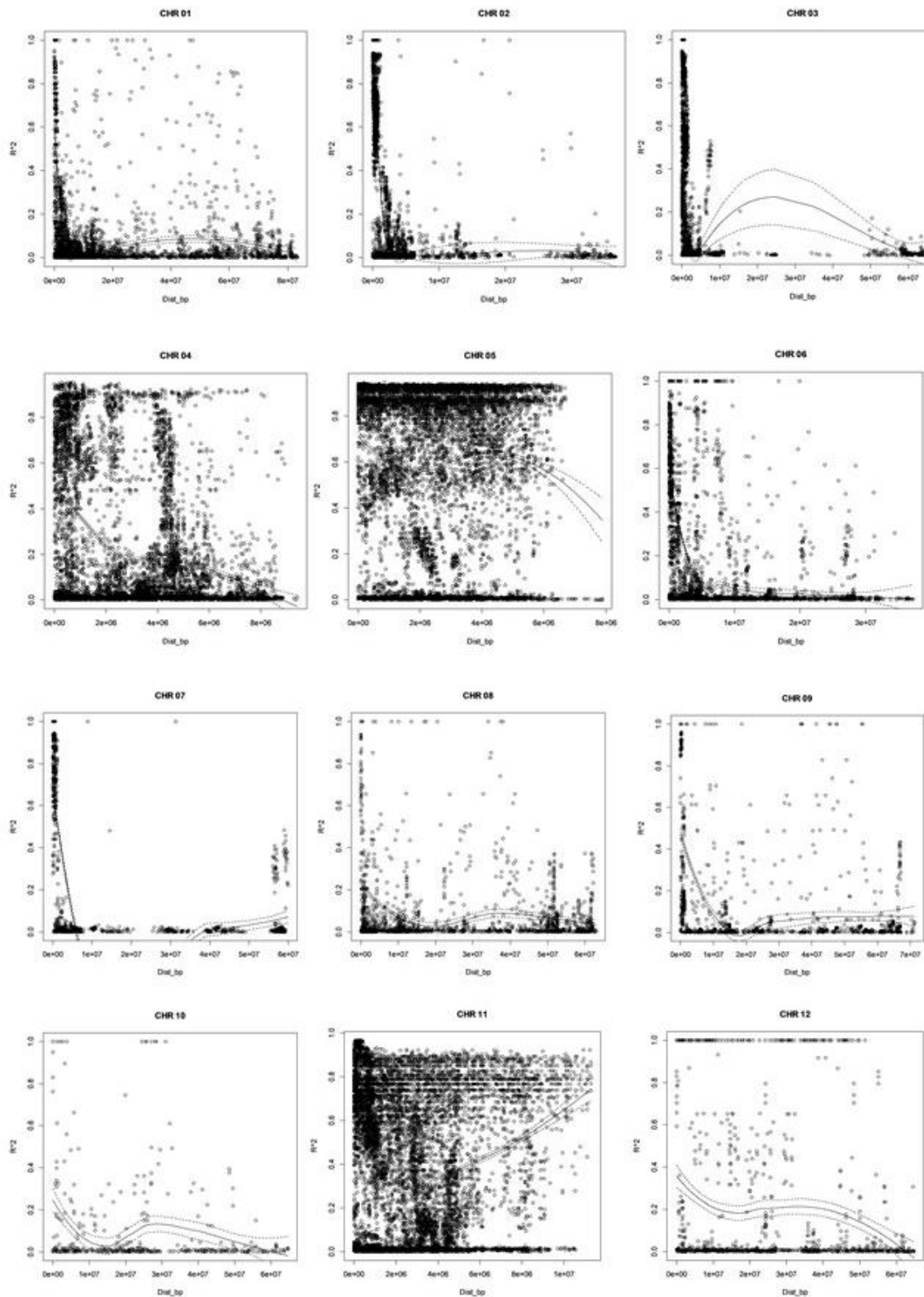
Variable name	Measure unit	Description
Brix degree	%	Content of total soluble solids
External fruit color	Score: 1, yellow; 2, orange; 3, pink; 4, red; 5, purple; 6, brown; 7, green	Colour of the fruit at ripening
Fruit fasciation	Score: 1, absent; 3, slight; 5, intermediate; 7, severe	Degree of fasciation of the flower, degree of cresting of the fruit
Fruit Load	Score: 1, low 3, medium-low 5. medium 7. medium-high 9. high	Charge of fruit of the plant
Fruit shape	Score: 1, flat; 2, rectangular; 3, ellipsoid; 4, obovoid; 5, round; 6, oxheart; 7, long; 8, heart	Predominant fruit shape recorded after the fruits turn colour
Fruit weight	g	Average fruit weight of a sample of representative fruits
Green shoulder	Score: 1, uniform; 2, light green; 3, medium green; 4, dark green	Presence and intensity of green shoulder on the upper part of the fruit around the calyx
Growth habit	Score: 1, determinate; 2, semideterminate; 3, indeterminate	The continuous alternation between vegetative and reproductive phases until plant end growth
Height 1st inflorescence	cm	Height of the 1st inflorescence measured at sixty days from the transplant
Height last inflorescence	cm	Height of the last inflorescence measured at sixty days from the transplant
Inflorescence distance	cm	Mean distance among inflorescences measured at sixty days from the transplant
Inflorescence leafy	Score: 1, not leafy; 2, leafy; 3, leafy with shoots	Referred to an inflorescence that exhibits a reversion of inflorescence meristems to vegetative growth after forming flowers
Inflorescence number	n	Number of inflorescence measured at sixty days from the transplant
Inflorescence type	Score: 1, uniparous; 2, fishbone; 3, forked; 4, irregular; 5, compound	Referred to flowers and their organization into branching systems
Jointless	Score: 0, absence; 1, presence	Presence or absence of jointless pedicel
Leaf attitude	Score: 1, semierect; 2, horizontal; 3, horizontal dropping; 4, dropping	Orientation of the leaf
Leaf border	Score: 1, entire; 2, undulate; 3, serrated; 4, strong serrated	Shape of the leaf border
Locule number	n	Number of chambers within the fruit, measured cutting the fruit along the transversal section
Puffiness	Score: 1, absent; 3, slight; 5, intermediate; 7, severe	Presence of cavity between pericarp and seeds
Ribbing at calix end	Score: 1, very weak; 2, weak; 3, intermediate; 4, strong	Presence of lobes at the peduncular part of fruit
Shape Index	-	Ratio between polar and equatorial diameter
Shape of scar	Score: 1, dot; 2, stellate; 3, linear; 4, irregular	Predominant shape of pistil scar
Shoulder shape	Score: 1, flat; 2, slightly depressed; 3, moderately depressed; 4, strongly depressed	Fruit depression degree at the proximal end
Skin color	Score: 1, colourless; 2, yellow	Colour of the epidermis of the fruit
Stigma position	Score: 1, inserted; 2, stigma at the same level; 3, slightly exerted; 4 highly exerted	Relative position of the stigma compared with the stamens in early, medium and late trusses
Total Height	cm	Height of the plant measured at sixty days from the transplant

Supplementary Table 2.2 . Phenotyping kit 2. Variables of plant and fruit evaluated in the Core collection and a brief description.

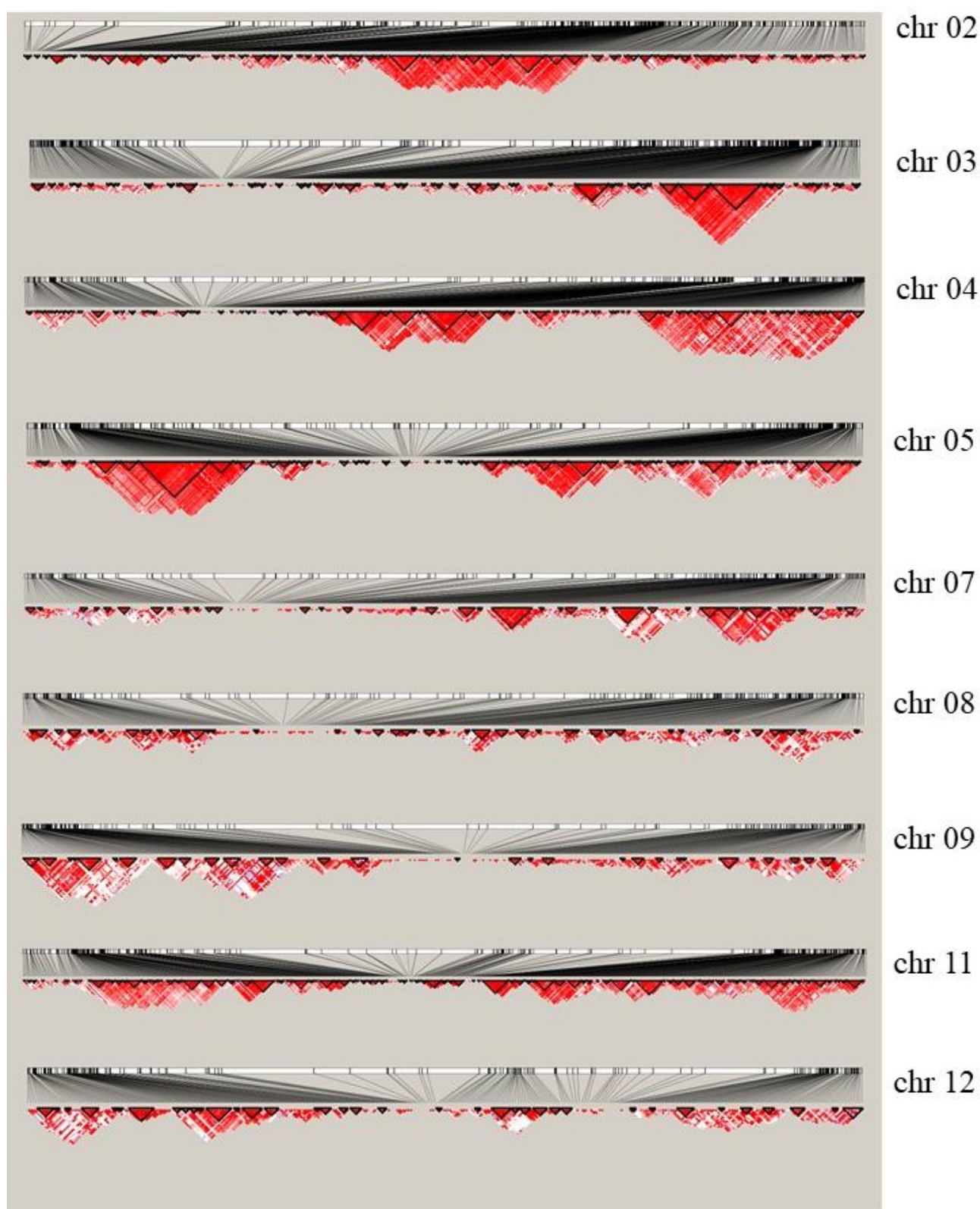
Name	Values	Description
Brix degree	Percentage	Content of total soluble solids
External fruit color	Score: 1, yellow; 2, orange; 3, pink; 4, red; 5, purple; 6, brown; 7, green	Colour of the fruit at ripening
Fruit Load	Score: 1, low 3, medium-low 5. medium 7. medium-high 9. high	Charge of fruit of the plant
Fruit shape	Score: 1, flat; 2, rectangular; 3, ellipsoid; 4, obovoid; 5, round; 6, oxheart; 7, long; 8, heart	Predominant fruit shape recorded after the fruits turn colour
Fruit weight	g	Average fruit weight of a sample of representative fruits
Green shoulder	Score: 1, uniform; 2, light green; 3, medium green; 4, dark green	The presence of green(shoulder) trip on the upper part of the fruit, around calyx and the green intensity
Growth habit	Score: 1, determinate; 2, semideterminate; 3, indeterminate	The continuous alternation between vegetative and reproductive phases until plant end growth
Inflorescence leafy	Score: 1, not leafy; 2, leafy; 3, leafy with shoots	Referred to an inflorescence that exhibits a reversion of inflorescence meristems to vegetative growth after forming flowers
Inflorescence number	n	Number of inflorescences considering the 1st-5th trusses
Puffiness	Score: 1, absent; 3, slight; 5, intermediate; 7, severe	Presence of cavity between pericarp and seeds
Ribbing at calix end	Score: 1, very weak; 2, weak; 3, intermediate; 4, strong	Presence of lobes at the peduncular part of fruit
Stigma position	Score: 1, inserted; 2, stigma at the same level; 3, slightly exerted; 4 highly exerted	Relative position of the stigma compared with the stamens in early, medium and late trusses
Foliage density	Score: 3, sparse; 5, intermediate; 7, dense	Visual evaluation of foliage density
Fruit set sequence	Score: 3, poor; 5, intermediate; 7, good; 9, very good	Referred to homogeneity of fruit set across trusses
Fruit size	Score: 1, very small (<3cm); 2, small (3-5 cm); 3, intermediate (5,1-8cm); 4, large (8,1-10cm); 5, very large (>10cm)	Average size of the fruit of a sample of representative fruits
Fruit size homogeneity	Score: 3, low; 5, intermediate; 7, high	Visual evaluation of fruit size homogeneity across all plant of the same accession
Irregular transversal section	Score: 1, round; 2, angular; 3, irregular; 4, highly	Referred to fruit cross sectional shape
Easiness of fruit to detach from the pedicel	Score: 3, easy; 5, intermediate; 7, difficult	Easiness of fruit to detach from the pedicel while harvesting
Blossom end scar condition	Score: 1, open; 2, closed; 3, both	Shape of the abscission zone where the pistils detaches
Macrocalix	Score: 1, absence; 2, presence	Referred to the presence of long size sepals
Fruit firmness	unit	Measured with a penetrometer
Puffiness appearance	Score: 3, slight; 5, intermediate; 7, severe	Presence of cavity between pericarp and seeds
Flowering earliness	days	Number of days to from sowing until 50% of plants have at least one open flower in a uniform growing environment.
Ripening earliness	days	Number of days from sowing until 50% of plants with at least one ripe fruit
Presence and incidence of concentric cracking	Score: 1, absent; 3, scarced; 5, intermediate; 7, abundant	Referred to the splitting of the epidermis in circular patterns around the stem scar.

Supplementary Table 3.1 LD decay at fixed baseline of 0.2 for each chromosome (Chr). Distance of LD decay was expressed in Mbp.

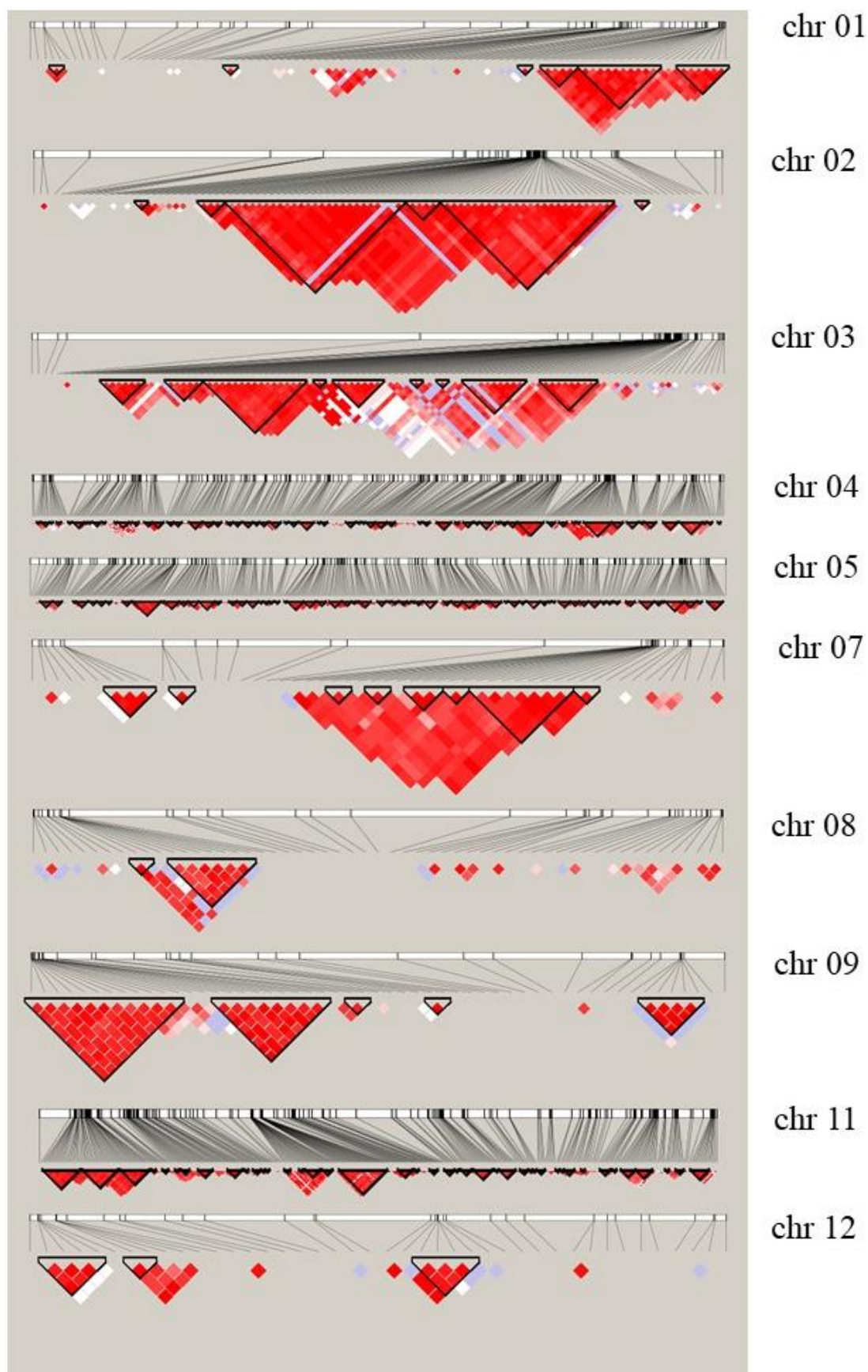
Chr	$r^2=0.2$
Chr01	3.07
Chr02	1.53
Chr03	0.74
Chr04	2.45
Chr05	-
Chr06	2.78
Chr07	3.87
Chr08	1.33
Chr09	6.15
Chr10	1.66
Chr11	-
Chr12	17.38
Average	4.10



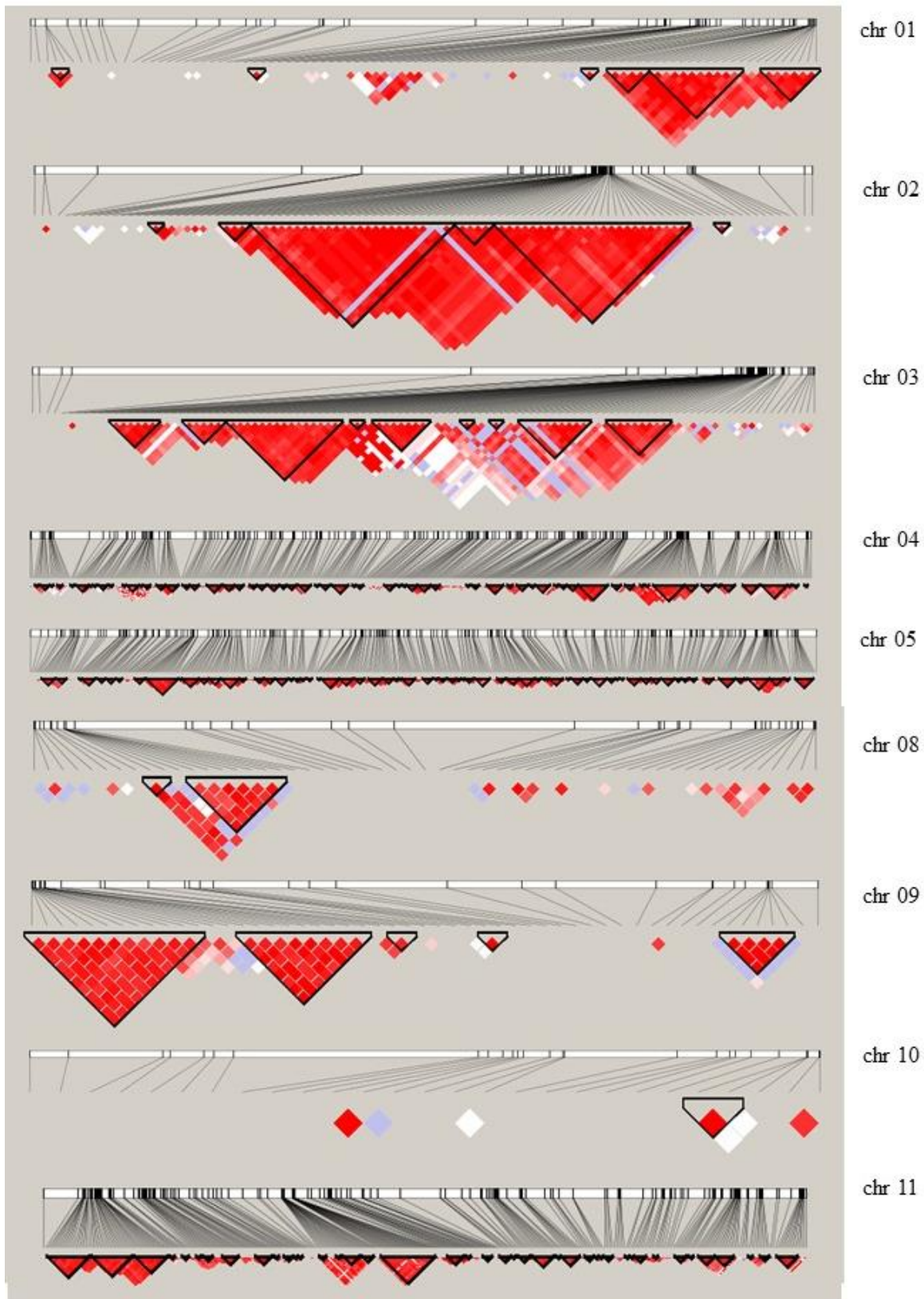
Supplementary Fig. 3.1. LD decay estimated using the locally weighted scatterplot smoothing LOESS algorithm. Values of r^2 were plotted against physical distance (bp).



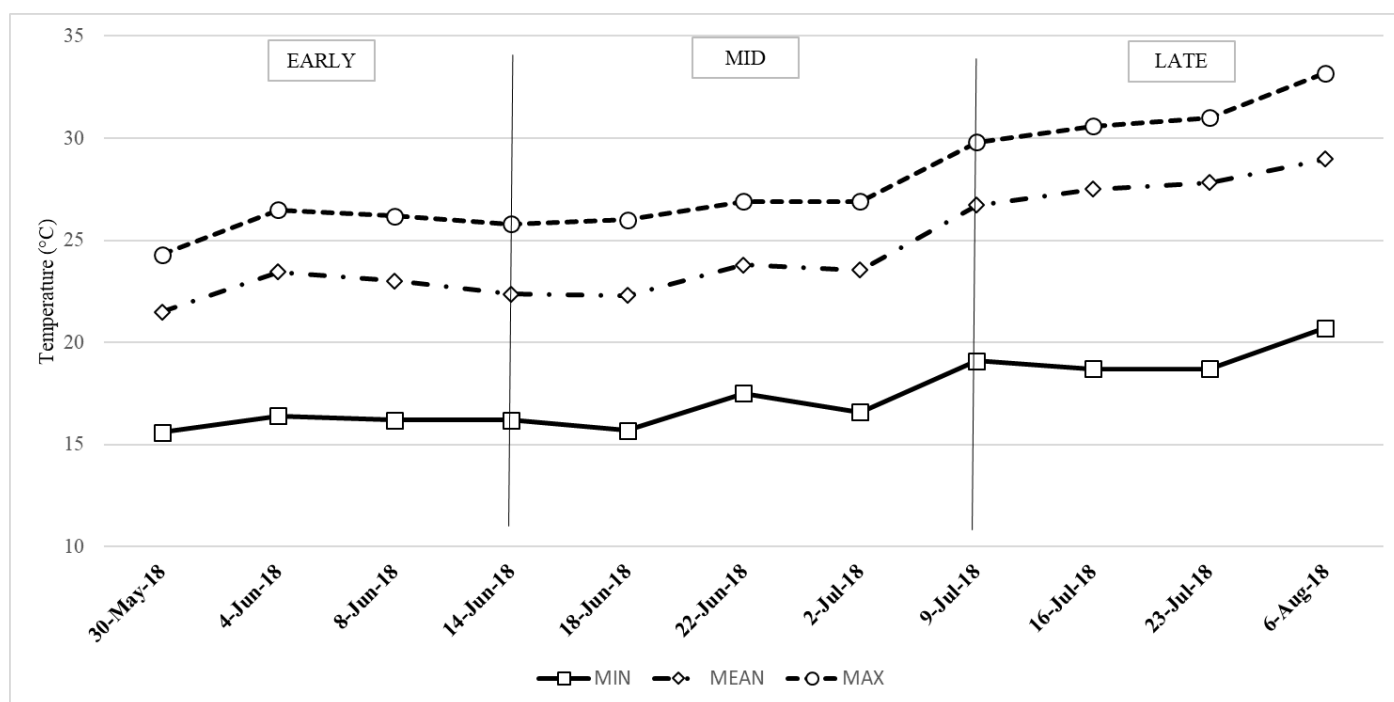
Supplementary Fig. 3.2. LD blocks found in the CER population for chromosomes where associated markers for SP were localized. Blocks of markers that are in strong LD using confidence intervals algorithm in Haploview software (black triangle).



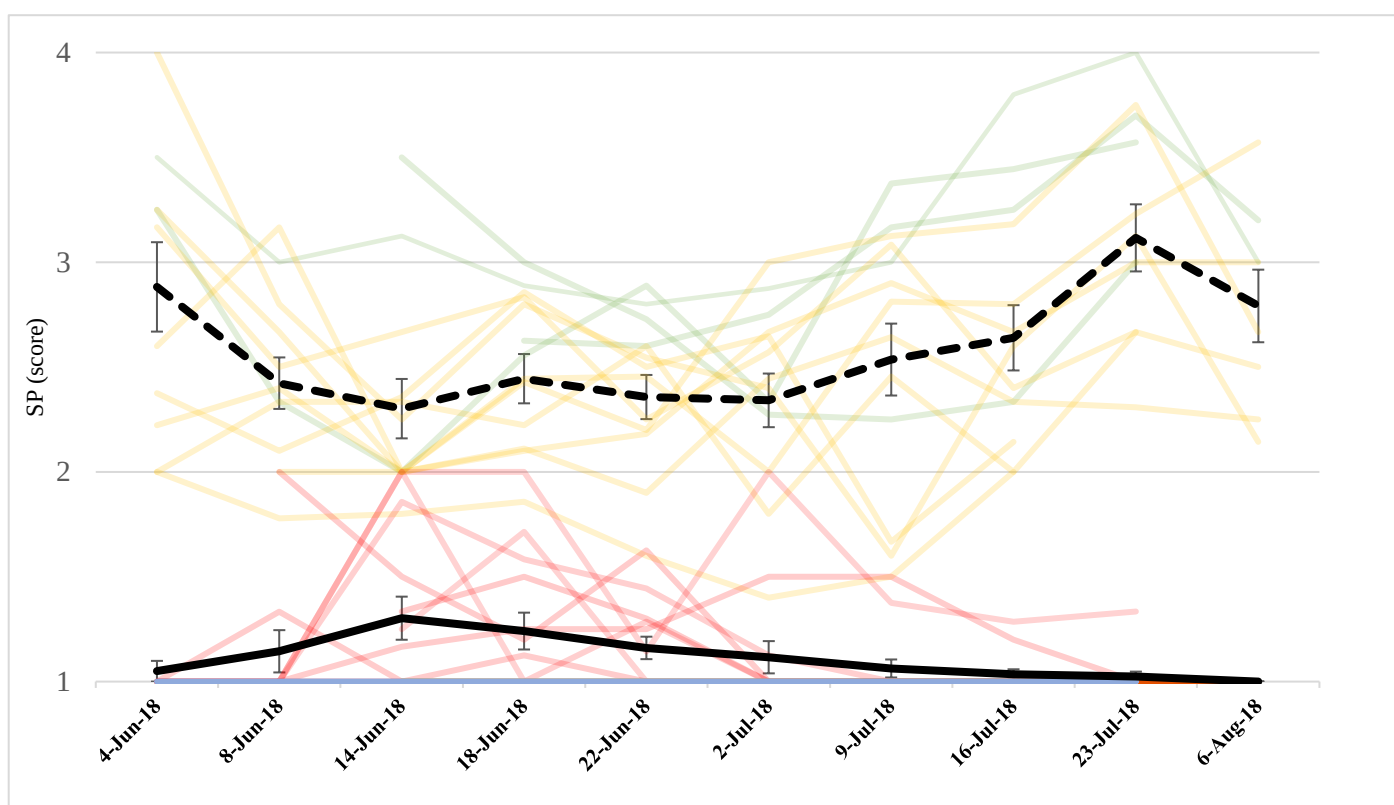
Supplementary Fig. 3.3. LD blocks found in Traditum CC population for chromosomes where associated markers for SP were localized. Blocks of markers that are in strong LD using confidence intervals algorithm in Haploview software (black triangle).



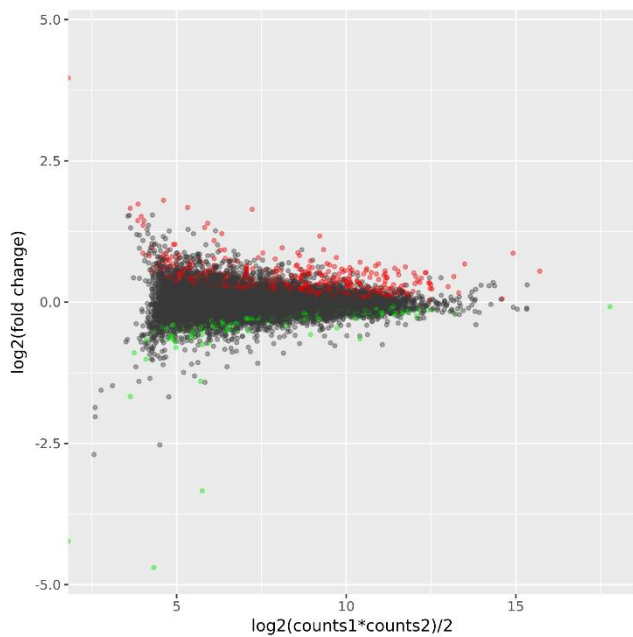
Supplementary Fig. 3.4. LD blocks found in Traditom population for chromosomes where associated markers for SW were localized. Blocks of markers that are in strong LD using confidence intervals algorithm in Haploview software (black triangle).



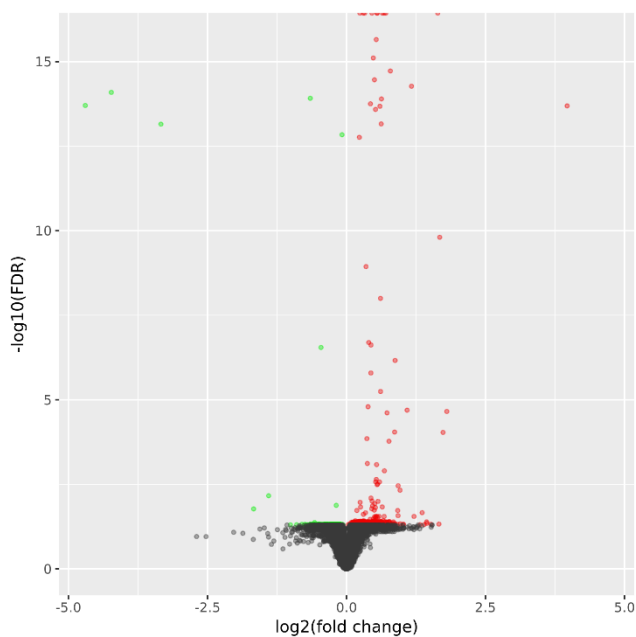
Supplementary Fig. 4.1 Temperatures recorded during the phenotyping season in summer 2018. Early, Mid and Late are referred to the three observations of SP.



Supplementary Fig. 4.2 Mean of SP in the inserted and exerted accessions during the phenotyping season in summer 2018.



Supplementary Fig. 4.3 The MA plot shows the relationship between the average expression value (on the X-axis) and the fold change (Y-axis) for each gene in the genome. The distribution of the dots in the MA-plot can be useful to check if the differentially expressed genes are uniformly distributed across the different ranges of expression values and the relationship with the fold-change.



Supplementary Fig. 4.4 The Volcano plot shows the relationship between the fold-change (on the X-axis) and the significance of the differential expression test (Y-axis) for each gene in the genome. The distribution of the dots in the Volcano plot can be useful to check if a range of fold-changes is associated with a stronger or a weaker significance of differential expression.

Supplementary file 4.1. List of DEGs of bulk RNA-seq experiment.xls

Aknowledgements

This work of Ph.D. thesis is part of the TRADITOM project, performed between the 2015 - 2018 time frame. It has received funding from the Horizon 2020 European Union funding for Research and Innovation.

I want to thanks all partners and colleagues contributing to the outcomes of the project and especially the project coordinator, Professor Antonio Granell of Consejo Superior de Investigaciones Científicas (Spain). The project gave me the opportunity to spend time abroad; I would like to thanks Professor Mathilde Causse and Professor Dani Zamir that gave me the chance to work in their laboratories at INRA of Avignon and The Robert H. Smith Institute in Rehovot in Israel, respectively.

The work was conducted mainly at the Laboratory of “Biotechnology of Vegetable Crops”, Dept. Agricultural and Forest Sciences (DAFNE) of the University of Tuscia of Viterbo. All results were achieved only thanks to my supervisor Professor Andrea Mazzucato. I want to express him my special gratitude for his support, aid and presence at all times during my activities. During the field and laboratory activities, I have met a wonderful team: Dr. Maurizio Enea Picarella and Dr. Gabriella Dono. Every day passed with Maurizio is a treasure for his advices and competence in the work and in the life context. Gabriella, is a brilliant colleague and has become a great friend of mine.

Surely, I have to thanks Dr. Mario Sampalmieri, Dr. Giulio Paolini, Dr. Andrea Marrucci, Dr. Luca Sersali, very valid trainee students. In this period, I had the possibility to collaborate as a thesis advisor at the Master thesis of Dr. Chiara Di Iorio and Dr. Fabrizio De Angelis, which are asserting themselves in the working life; it was a pleasure for me share a part of the activities with you, *good luck guys!* I want to thanks all other technicians, friends and colleagues of other laboratories of University of Tuscia, with a special thanks to Eleonora, Luca, Francesco and Elisa. I wish to thanks the coordinators of the Ph.D. in Plant and Animal Science, Professor Stefania Masci and Professor Roberta Bernini.

Last year of my Ph.D. was particularly significant for me, with the waiting and the born of my son, Raffaele, and I would like to do a special mention to my future bride, Ashanti, and my family, supporting me every day and every moment. Grazie

*“Tutto ciò non è mio,
ma di ogni persona che ha creduto in me...”*

