ACADEMIE D'AIX-MARSEILLE
UNIVERSITE D'AVIGNON ET DES PAYS DE VAUCLUSE

THESE

présentée pour obtenir le grade de **Docteur en Sciences**
de l'Université d'Avignon et des Pays de Vaucluse

Spécialité : Sciences Agronomiques

Exploration du polymorphisme moléculaire et protéique de la tomate pour l'identification de QTL de qualité du fruit

par Jiaxin XU

soutenue le 31 octobre 2012 devant un jury composé de :

Judith BURSTIN	Directrice de recherches INRA, Dijon	Rapporteur
Patrice THIS	Directeur de recherches INRA, Montpellier	Rapporteur
Christophe ROTHAN	Directeur de recherches INRA, Bordeaux	Examinateur
Félicie LAURI	Maître de Conférence à l'Université d'Avignon	Examinateur
Stéphane MUNOS	Ingénieur INRA, Toulouse	Examinateur
Mathilde CAUSSE	Directrice de recherches INRA, Avignon	Directrice de thèse

Ecole Doctorale :
ED536 Sciences et Agrosciences – UAPV Avignon

Laboratoire d'accueil :
INRA-UR1052 « Génétique et Amélioration des Fruits et Légumes », CS60094 – F84143 Montfavet (France)

Acknowledgements

Formost, I would like to express my deeply-felt thanks to my supervisor Dr. Mathilde Causse for offering me the PhD opportunity in INRA-GAFL and leading me working on exciting PhD project. I thank her for supporting and guiding of my PhD. She was always taking care of me during my difficult time when I was writing manuscripts and this thesis. She taught me not only about knowledge but also how to be a good scientist. I am lucky to be her student and I will remember her for my life.

I would like to express my gratitude to Dr. Stephane Muňos. He was my initial co-supervisor who helped me to develop skills in molecular biotechnology, encouraged me and gave me good advice during the first year of my PhD.

I would like to thank Mireille Faurobert for supervising the proteome part of my PhD. She helped me to write the proteomic dataset manuscript and gave me insightful comments on my thesis.

I would like to express my gratitude to my Chinese supervisor Professor Liang Yan, for her motivation, encouragement and valuable comments on my PhD project. She was always concerning me and came to France to see me even she was quite busy during that period.

I would like to thank the rest of my PhD committee members Laurent Urban, Nadia Bertin, Dominique Brunel for their comments and suggestions on my PhD project.

I am happy to thank Claudie Arliaud, for her patience, enthusiasm, kindness. She helped me to take care of all the administrative things such as visa, apartment which were particularly hard for me as I only knew a little French language. I could not forget that each time when I met her and to ask her for help for something. We had to translate all what we want to say in Yahoo or Google. Now, I hope we could leave the internet and communicate with each other.

I am grateful to Caroline Callot for helping to do the CAPS marker experiment, sugar content measurement and 2-DE experiment. I could not finish my PhD manipulation and got so nice results without her kind help.

I would like to thank Sophie Rolland for assistance with PCR manipulation, RNA extraction, SNPlex data analysis and tomato fruit harvest as well as other technical aspects.

My sincere thanks also goes to Laura Pascual who helped me a lot for data analysis, protein spots classification, and proteomics dataset manuscript preparation. I could not forget she was always giving me a ride home during the last winter. We were working together, discussing the project and had a lot of fun.

I am glad to thank Christopher Sauvage for thoughtful comments, suggestions and revisions on my association mapping manuscript, CAPS marker manuscript and this thesis. We discussed a lot about population genetics, association mapping and statistics. I learned a lot from him. I wish I could meet him earlier.

I would like to thank Guillaume who gave me a lot of help for my PhD project concerning data analysis, manuscript preparation, thesis writing. He helped me in all the time of my staying in France. I am happy to have a friend like him.

Thanks to Jean-paul Bouchet and Nelly Desplat for helping to do bioinformatics analysis.

Thanks to Aurore Desgroux for helping to analyze the metabolome data.

Thanks to Justine Gricourt and Frederique Bitton for assistance with the measurement of Vitamin C.

Thanks to Nicolas Ranc who helped me to get into the project and showed me how to use GeneMapper.

Thanks to Esther Pelpoir for fruit harvesting and sampling.

Thanks to Yolande Carretero for taking care of the plants in the greenhouse.

Thanks to Rebecca Stevens for critically reviewing the proteome manuscript.

Thanks to Elsa Desnoues and Ye Yang, we shared the same office and passed a good time together.

Thanks to all members in INRA-GAFL research units for laboratory guidance and friendships. I am pleased to come to study here and to meet you every day.

Thanks to my friends Christopher Sauvage, Guillaume Bauchet, Laura Pascual, Renaud Duboscq, Noé Gest, Ana Giner, Elsa Desnoues, Laurence Ouibrahim from GAFL Institute, we passed a lot of good time together and had a lot of fun. I am happy to meet you all and thanks for all you have done for me.

Thanks to my friends Kaiyun Tsao, Zhijun Shen, Ye Yang, Rouan Chen, Si Liang, Yiqiao Chen, Yuedong Xu, Donghao Ma for continuous support and encouragement. I was happy to spent most of the week-ends with you together. I will feel lonely without your accompany.

I would like to gratefully acknowledge the China scholarship Council and MAGICTomSNP project for scholarship funding.

I am deeply indepted to my parents and sister for their love, long-term support and encouragement in all aspects of my life.

Thanks to my girlfriend for her love, patience and support. I don’t even have time to talk to her when I was writing this thesis during the past months.

This thesis is dedicated to my parents, my sister and my girlfriend for their unfailing love, and many years of sacrifice.

INDEX

GENERAL INTRODUCTION **1**

CHAPTER I: REVIEW OF LITERATURE **7**
1.1. Tomato **9**
1.1.1. Tomato plant biology, origin, evolution and economic importance. 9
1.1.2. Tomato genetic and genomic resources 15
1.1.3. Tomato fruit quality traits 21
1.2. Genetic control of quality traits **25**
1.2.1. QTL mapping: methodology and application in tomato 25
1.2.2. Mapping of QTLs for fruit quality traits in tomato 29
1.2.3. Positional cloning 37
1.2.4. Association mapping 37
1.2.5. Novel approaches towards the improvement of important traits 45
1.2.5.1. Nested association mapping 45
1.2.5.2. Interest of building MAGIC population 47
1.2.5.3. Systems biology approaches 49
1.3. Objectives and content of this study **61**

CHAPTER II: MATERIALS AND METHODS **65**
2.1. Plant materials **67**
2.2. Methods **69**
2.2.1. Methods of the polymorphism identification and validation 69
2.2.2. Methods of the association analysis 69
2.2.3. Methods of the integration of system biology approaches 73

CHAPTER III: COMBINED LONG RANGE PCR AND NEXT-GENERATION SEQUENCING FOR THE IDENTIFICATION OF POLYMORPHISMS BETWEEN TWO TOMATO LINES **83**
Abstract **87**
Introduction **89**
Material and Methods **93**
Results and Discussion **99**
Conclusion **111**

CHAPTER IV: PHENOTYPIC DIVERSITY AND ASSOCIATION MAPPING FOR FRUIT QUALITY TRAITS IN CULTIVATED TOMATO AND RELATED SPECIES **123**
Abstract **127**
Introduction **129**
Materials and methods **135**
Results **145**
Discussion **159**
Conclusion **171**

CHAPTER V: AN EXTENSIVE PROTEOME MAP OF THE TOMATO (SOLANUM LYCOPERSICUM) FRUIT PERICARP **183**
Abstract **187**
Introduction. **189**
Materials and methods **191**
Results and discussion **193**
Conclusion **199**

CHAPTER VI: GENETIC DIVERSITY AND INHERITANCE OF PROTEOMIC, ENZYMATIC AND METABOLOMIC PROFILES IN TOMATO FRUIT PERICARP **217**
Abstract **221**
Introduction **223**
Materials and methods **229**

Results. **237**
Discussion **265**
Conclusion **279**

CHAPTER VII: DISCUSSION AND PERSPECTIVES **323**
From the re-sequencing of targeted regions in two lines to whole genome re-sequencing of hundreds lines **325**
From mid-throughput SNPlexTM assay to high-throughput SNP-chip for association mapping…. 331
High-throughput technologies opened the way for Systems biology approaches 335
Conclusion. 341

REFERENCES **343**

ANNEXES **371**

GENERAL INTRODUCTION

GENERAL INTRODUCTION

Tomato (*Solanum lycopersicum*, formerly *Lycopersicon esculentum*) belongs to *Lycopersicon* section*, Solanum* genus and the nightshade family Solanaceae. It originated from South America, in the Andes Mountains of Peru, Ecuador and Chile. Mexico is considered as the most probable region of domestication or diversification (Robertson. and Larate 2007). *S. pimpinellifolium* is thought to be the wild ancestor of cultivated tomato. *S. lycopersicum* var. *cerasiforme* accessions produce larger fruits (commonly red and round) than *S. pimpinellifolium*. These accessions are considered as a primitive type of cultivated tomato or as a transitional form between the *S. pimpinellifolium* wild species and the *S. lycopersicum* cultivated one. Recent genetic investigations have shown that the genome of the plants known as 'cerasiforme' (cherry type) are a mixture of wild and cultivated tomatoes, rather than the 'ancestre' of the cultivated tomato (Nesbitt and Tanksley 2002; Ranc et al. 2008).

Tomato is among the most important vegetable in human diet for its health value, due to the amount consumed (2/3 as fresh and 1/3 as processed product in France) and to its original composition of antioxidant compounds (Di Mascio et al. 1989). Including or increasing fresh tomato and tomato products in the diet would thus benefit to human health. Apart from being an economic important plant species, tomato is one of the best characterized plant systems at both genetic and genomic levels. It has abundant genetic and genomic resources such as sequenced genome (Sato et al. 2012), thousands molecular markers and many genetic maps. However, many markers were developed based on polymorphism in wild species, they are not polymorphic when used within the cultivated species (Jimenez-Gomez and Maloof 2009). Therefore, it is still quite important to develop new molecular markers to be used in tomato genetics. Single nucleotide polymorphisms (SNPs) are the most abundant markers in the genome. Next generation sequencing technologies (NGS) offer new capacities for producing high volume of sequence data which can be used to identify SNPs (Barbazuk et al. 2007).

These technologies evolved very fast, with the length and number of reads increasing each month, and it is a real opportunity to test such strategy for SNP discovery in tomato.

Nowadays, the social demand concerns tomato fruit quality (organoleptic and nutritional value). The best way to answer this demand is through breeding improved cultivars. It is therefore important to consider the fruit quality traits in breeding schemes. For this purpose, it is necessary to understand the genetic bases of quality traits. Most tomato fruit quality traits (sweetness, sourness, aroma and texture) are complex because they are controlled by the joint effect of several genes of small effects. Dissection of fruit quality traits in tomato has been firstly achieved by QTL (Quantitative trait loci) mapping. QTL analyses were performed for fruit size, shape and quality traits on bi-parental populations (Paterson et al. 1991; Frary et al. 2000; Eshed and Zamir 1995; Grandillo and Tanksley 1996; van der Knaap and Tanksley 2001; Causse et al. 2002; van der Knaap and Tanksley 2003; Barrero and Tanksley 2004; Causse et al. 2004; Lecomte et al. 2004). A few genes controlling tomato fruit QTL have been cloned like *FW2.2* (Frary et al. 2000) which controls fruit weight; *Lin5* which is responsible for fruit sugar content (Fridman et al. 2000) and *LC* which controls locule number (Munos et al. 2011).

QTL mapping only allows the detect of QTLs with large effects due to restricted allelic variation and modest degree of recombination in bi-parental mapping population (Hall et al. 2010). Association or linkage disequilibrium (LD) mapping represents an alternative approach to identify the genes and genomic regions associated with phenotypic trait variation. It surveys a large range of allelic variation and exploits a large number of historical recombination events in a diverse set of genetic resources. This method was first used in allogamous species or species with wide range of genetic diversity, for several traits such as flowering time and pathogen resistance in Arabidopsis (Aranzana et al. 2005), yield and its components in rice (Agrama et al. 2007), leaf architecture in maize (Tian et al. 2011), iron

deficiency chlorosis in soybean (Mamidi et al. 2011). Very few convincing studies were performed in autogamous crop species. Tomato is a highly autogamous plants and previous studies of genetic diversity in cultivated tomato showed that its diversity is very low. An intermediate level of polymorphism between the wild and the closest cultivated species was identified in accessions of cherry type (*S. lycopersium* var. *cerasiforme*), which may help to overcome the high LD in this autogamous species.

Besides association mapping, novel approaches were also developed for the dissection of quantitative traits. Because DNA sequence variation (SNP or Indel) may not affect the traits directly. There are several intermediate levels between DNA genotypes and the phenotypes. The cascade of effects from DNA variation to phenotype is organized in complicated biological networks (Kliebenstein 2010; Sulpice et al. 2010). Intermediate molecular phenotypes such as transcript and protein abundance also genetically vary in populations and are themselves quantitative traits (Rockman and Kruglyak 2006). New analytical approaches proposed in the context of systems biology consist in combining information from different levels such as metabolome, proteome, transcriptome and genome levels. This should enable us to understand the biology inside the black-box of quantitative genetics relating genotype to phenotype in terms of causal networks of interacting genes. System approaches have been applied in yeast (Ideker et al. 2001), in the model plant Arabidopsis (Hirai et al. 2007) and in tomato (Mounet et al. 2009), at several levels.

In this view, the objective of the study was to characterize tomato genetic diversity at the molecular and proteome levels and to try to identify QTLs, proteins responsible for fruit quality traits in tomato. For this purpose, my PhD was organized in three independent parts**:**

1. We first used Next Generation Sequencing technology (GA2 and 454 platform) to re-sequence about 0.2% of the tomato genome of two contrasted lines in order to identify polymorphisms and develop new markers.

2. Technologies evolve fast and in the same time, SNPlex™ genotyping technology emerged. We used a SNPlex™ array, composed of 192 SNPs selected from resequencing experiment or from public database to genotype 188 tomato accessions characterized at phenotypic level. We conducted association mapping to identify SNP associated with fruit quality traits.

3. Finally, we applied systems biology approaches focused on proteome, metabolome and phenotypic analysis to characterize the fruits of eight contrasted lines as well as four of their hybrids at two stages, cell expansion and orange red stage.

I will present these results in chapters III to VI of this document, following a review of the literature and a rapid summary of the materials and methods used in this study. The results are presented as manuscripts to be submitted to the journals. A general discussion and a few prospects will then conclude this manuscript.

CHAPTER I: REVIEW OF LITERATURE

CHAPTER I: REVIEW OF LITERATURE

1.1 Tomato

1.1.1 Tomato plant biology, origin, evolution and economic importance

Tomato (*Solanum lycopersicum* formerly *Lycopersicon esculentum*) belongs to *Lycopersicon* section, *Solanum* genus and the nightshade family Solanaceae. This family consists of about 90 genera and 3000–4000 species. Almost half of them are in the large and diverse genus *Solanum* (Knapp et al. 2004). The genus *Solanum* includes commonly cultivated plants such as potato (*Solanum tuberosum*), pepper (*Capsicum annuum*), tobacco (*Nicotiana tabacum*) and tomato, and ornamental plants such as false jasmine nightshade (*Solanum jasminoides*). Some species of the Solanaceae family are known to be toxic as black nightshade (*Solanum nigrum*) or bittersweet (*Solanum dulcamara*). Other species are known for their psychoactive properties as the Mandrake (*Mandragora officinarum*) or belladonna (*Atropa belladonna*). The cultivated tomato was originally named *Solanum lycopersicum* by Linnaeus (1753). It was then re-classified into the newly designed *Lycopersicon* genus named as *esculentum* by Miller (1754) because of its distinct characteristics of anthers and leaves. Recent taxonomic study of the Solanaceae has re-integrated *Lycopersicon* species into the genus *Solanum* with a new nomenclature (Peralta and Spooner 2001; Spooner et al. 2005; Peralta et al. 2008). Now, the classification of tomato to the genus *Solanum* has been widely accepted by most taxonomists (Fridman et al. 2004; Schauer et al. 2005; Mueller et al. 2009). It was supported by several phylogenetic studies on molecular and morphological characters (Peralta et al. 2005; Spooner et al. 2005).

The *Solanum* genus, lycopersicon section consists of cultivated tomato (*Solanum lycopersicum*), which includes the domesticated tomato and wild or weedy forms of the cherry type tomato (*S. lycopersicum* var. *cerasiforme*) and the 12 wild species: *Solanum arcanum*, *S.cheesmaniae*, *S. chilense*, *S. chmielewshii*, *S.corneliomuelleri*, *S. galapagense*, *S. habrochaites*, *S. huaylasense*, *S. neorickii*, *S. pennellii*, *S. peruvianum*, *S. pimpinellifolium* (Peralta et al. 2005; Spooner et al. 2005). Species from this section are diploid (2n=24) and closely related, and are to some degree inter-crossable (Taylor 1986). The principal ecological, botanical, and reproductive characteristics of the wild tomatoes are presented in **Table 1-1**.Tomato and its wild relatives originated from South America, in the Andes Mountains of Peru, Ecuador, and Chile. However, the original site of domestication and the

Table 1-1. Principal ecological, botanical, and reproductive features of the wild tomatoes (*Solanum* sect. *Lycopersicon*). Reviewed by Grandillo et al. (2011)

Species	Geographic distribution	Habitat	Mating system[a]	Crossability to tomato[b]	Distinguishing morphological features[c]
S. lycopersicum "cerasiforme"	Adventive worldwide in the tropics and subtropics (near sea level – 2,400 m); perhaps native in Andean region	Usually mesic sites, often feral or weedy	SC-autogamous	BC	Plants semi-erect to sprawling; fruits red, 1.5–2.5 cm
S. cheesmaniae	Endemic to Galápagos Islands (sea level – 1,500 m)	Arid, rocky slopes, prefers shaded, cooler sites	SC-autogamous	BC	Plants semi-erect to sprawling, flowers very small, pale; fruit purple, greenish-yellow, or orange, 0.5–1.5 cm
S. galapagense	Endemic to Galápagos Islands (sea level – 650 m)	Arid, rocky outcrops and slopes, sometimes near shoreline	SC-autogamous	BC	Plants erect; leaves highly subdivided; internodes short; flowers small, pale, fruit orange (0.5–1 cm)
S. pimpinellifolium	Lowland Ecuador and coastal Peru (sea level – 500 m)	Arid, sandy places, often near sources of water or on the edges of farm fields	SC-facultative	BC	Plants semi-erect to sprawling, flower small-large; fruit red (0.5–1 cm)
S. chmielewskii	Inter-Andean valleys of central and southern Peru (1,600–3,100 m)	Rather moist, well-drained, rocky slopes	SC-facultative	UI	Plant sprawling or trailing; flowers small, pale; fruit green (1–1.5 cm)
S. neorickii	Inter-Andean valleys from Cusco to central Ecuador (1,500–2,500 m)	Rather moist, well-drained, rocky slopes	SC-autogamous	UI	Plants sprawling or trailing; flowers tiny, pale; fruit green; seeds tiny
S. arcanum	Northern Peru, coastal and inter-Andean valleys, middle watershed of Marañón (500–3,000 m)	Varied, but generally dry, rocky slopes	Mostly SI, rarely SC-facultative	UI	EL Plants erect to prostrate, reduced leaflet no.; flowers mostly straight anther tubes and undivided inflorescences; fruit whitish-green with dark stripe
S. chilense	Southern Peru, northern Chile (50–3,500 m)	Very arid and sometimes saline, rocky slopes or washes	SI	UI, EL	Plants erect; leaves finely pubescent;anthers straight; inflorescences compound; peduncles long; Fruit purplish-green
S. peruvianum	Mostly coastal central/southern Peru and northern Chile (sea level – 2,500 m)	Arid, sandy, or rocky dry washes, sometimes near agricultural fields	Mostly SI, rarely SC-facultative	UI, EL	Plants procumbent; anthers bent; inflorescence simple; fruit purplish-green

Table 1-1 Continued

Species	Geographic distribution	Habitat	Mating system[a]	Crossability to tomato[b]	Distinguishing morphological features[c]
S. arcanum	Northern Peru, coastal and inter-Andean valleys, middle watershed of Marañón (500–3,000 m)	Varied, but generally dry, rocky slopes	Mostly SI, rarely SC-facultative	UI	EL Plants erect to prostrate, reduced leaflet no.; flowers mostly straight anther tubes and undivided inflorescences; fruit whitish-green with dark stripe
S. chilense	Southern Peru, northern Chile (50–3,500 m)	Very arid and sometimes saline, rocky slopes or washes	SI	UI, EL	Plants erect; leaves finely pubescent; anthers straight; inflorescences compound; peduncles long; fruit purplish-green
S. peruvianum	Mostly coastal central/southern Peru and northern Chile (sea level – 2,500 m)	Arid, sandy, or rocky dry washes, sometimes near agricultural fields	Mostly SI, rarely SC-facultative	UI, EL	Plants procumbent; anthers bent; inflorescence simple; fruit purplish-green
S. corneliomulleri	Western Andes of central/southern Peru (1,000–3,000 m)	Rocky or sandy slopes and dry washes	SI	UI, EL	Erect to decumbent; leaves glandular pubescent; fruit purplish-green
S. huaylasense	Limited to Callejon de Huaylas, and Río Fortaleza, Peru (1,000–2,900 m)	Rocky slopes and waste places	SI	UI, EL	Spreading, anthers straight, inflorescence compound; fruit purplish-green
S. habrochaites	Northwestern and western central Peru, western and southern Ecuador (40–3,300 m)	Varied, but generally mesic slopes or stream banks	Mostly SI, some SC-facultative	UI	Sprawling shrub or vine; densely pubescent; flowers large; anthers straight; fruit green with dark stripe, hairy
S. pennellii	Coastal valleys of central to southern Peru (near sea level to 1,920 m)	Very arid, sandy or rocky slopes, or dry washes	Mostly SI, some SC-facultative	UI	Spreading shrub; 2 leaves per sympodium[d]; leaflets broad, round; foliage sticky; anthers poricidal; pedicel usually articulated at base

[a]SC=self-compatible; SI= Self-incompatible; autogamous= self-pollinating; allogaous= outcrossing; facultative= may self-pollinate or outcross

[b]BC=bilaterally compatible (i.e., no barrier in either direction); UI= unilateral incompatibility (cross succeed only cultivated tomato is used as the female parent); EL= embryo lethality (can usually be overcome by embryo culture);

[c]Except as noted, all spp, are indeterminate, herbaceous shrubs, with 3 leaves per sympodium; flowers have the standard "lycopersicon" morphology- petals yellow; anthers yellow and fused, with a sterile anther appendage, and lateral pollen dehiscence – and lack floral scent

[d]Values based on Charles Rick's notes at the time of collection, or observation made during regeneration by the TGRC

early events of domestication are not so clear (Peralta and Spooner 2007). Two hypotheses have been proposed for the original place of tomato domestication, Peruvian versus Mexican. Although the precise time and place of domestication are still to be proven, Mexico has been considered to be the most probable original place of domestication or diversification, with Peru as the center of diversity for wild relatives (Robertson. and Larate 2007). Spanish "explorers" first introduced tomato into Europe in the early 16th century. It was first grown for its ornamental value then taken as food by Italians and Spaniards in the 17th century and later by people all over the world (**Figure 1-1**). The earliest citation of the tomato in European literature appeared in an herbal written in 1544 by Pietro Andrea Mattioli, an Italian physician and botanist, who named it "pomo d'oro", or "golden apple". During and following its domestication, tomato has undergone intensive selection particularly for fruit size, shape and color from its closest wild relatives, S. pimpinellifolium. This species produces red, round, and small fruit weighing only a few grams. By contrast, the fruit from modern tomato varieties may weigh up to 1kg, a nearly 1000-fold increase in weight (Tanksley 2004) (**Figure 1-2**). S. pimpinellifolium is thought to be the wild ancestor of cultivated tomato. Nucleotide divergence comparison showed that this wild species has a similar genome sequence to that of its cultivated relative, with 0.6% divergence, which supports this theory (Sato et al. 2012). S. lycopersicum var. cerasiforme produces larger fruits (commonly red and round) than S. pimpinellifolium. These accessions are considered as a primitive type of cultivated tomato or as a transitional form between the S. pimpinellifolium wild species and the S. lycopersicum cultivated one. It also appears that many accessions are original feral accessions (Rick and Holle 1990; Peralta and Spooner 2007). Recent genetic investigations have shown that the accessions classified as 'ceraciforme' are, at the genome level, a mixture of wild and cultivated accessions (Nesbitt and Tanksley 2002; Ranc et al. 2008). The cherry tomato type is also morphologically intermediate between S. pimpinellifolium and S. lycopersicum. Therefore, domestication took probably place in the following sense: from S. pimpinellifolium to S. lycopersicum var cerasiforme then to S. lycopersicum, followed by multiple intercrosses between the three groups.

From the economic point of view, tomato is one of the most commonly consumed vegetables (2/3 as fresh product and 1/3 processed). It is grown in almost every countries of the world. Both worldwide production and harvested area of fresh market and processing tomato increased during the last fifty years and reached 146 million tons and 4.34 million ha in 2010 (http://faostat.fao.org). China is the largest producer of tomato, accounting for 28% of the global production, followed by USA (9%), India (8%), Turkey (7%) and Egypt (6%) etc.

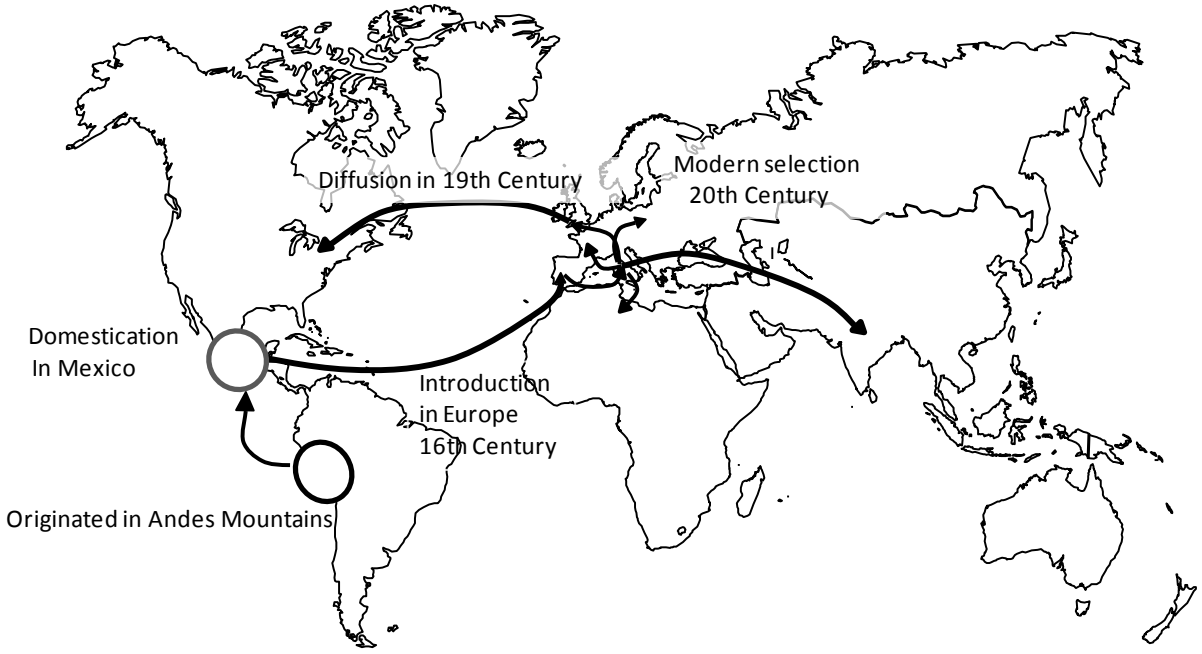


Figure 1-1. Map showing the origin, domestication center and hypothetical diffusion of tomato across the world (Cox, 2000).

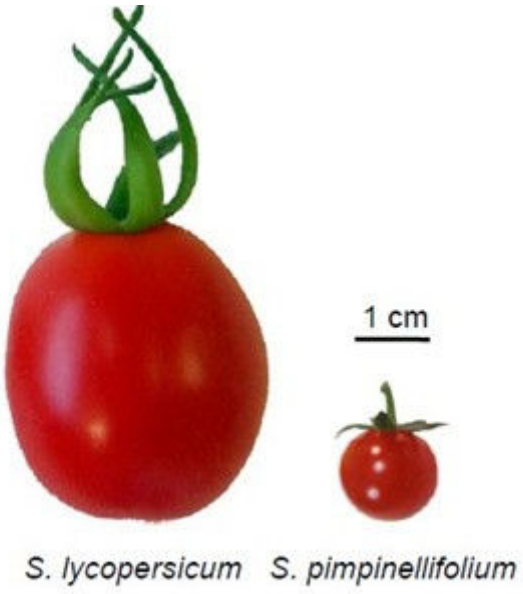


Figure 1-2. Illustration of ripe (58 days after post anthesis) fruits of *S. lycopersicum and S. pimpinellifolium* (From Sato et al. 2012)

Tomatoes are also rich source of essential minerals, vitamins and the dietary antioxidant lycopene, which protects cells from oxidants, and that have been linked to protection against several cancers (Giovannucci 1999; Amarjeet et al. 2011; Kelkel et al. 2011). In addition, tomatoes also contribute to provide fiber to the diet (Davies and Hobson 1981). Including or increasing fresh tomato and tomato products in the diet would thus benefit to health.

1.1.2 **Tomato genetic and genomic resources**

Tomato is among the most important plant species not only because of its economic importance and nutritional value but also because it is one of the best characterized plant systems both at the genetic and genomic levels. Abundant genetic and genomic resources have been developed in tomato. These resources are detailed in the following section.

Genetic resources

More than 75,000 tomato accessions are conserved in gene banks around the world. Traditional tomato genetic resources include the twelve wild and related species, and their collections are publicly available worldwide, for instance, in the Asian Vegetable Research and Development Center (AVRDC), now referred to as the world vegetable center (Taiwan), in the Tomato Genetics Resource Center (Davis, CA, USA), in the Centre for Genetic Resources (Wageningen, Netherlands), in the USDA (Geneva, NY, USA), the N.I. Vavilov Research Institute of Plant Industry (St Petersburg, Russia), and in the research unit of genetics and improvement of fruit and vegetables (Avignon, France). Wild species have been used as sources of genetic variation for tomato improvement. (Rick 1995). Cultivated tomato has undergone a narrowing of the germplasm basis due to genetic bottlenecks and selection (Robertson. and Larate 2007).

Other genetic collections such as mutants also available. This kind of collections are interesting for the investigation of gene function. The biggest collection is maintained by the TGRC (http://tgrc.ucdavis.edu/), including spontaneous and induced mutations affecting many aspects of plant development, disease resistance genes, protein marker stocks and other traits of economically importance (http://solgenomics.net/). Besides, an isogenic mutant library is also available in Solanaceae Genome Network (SGN). It includes a total of 13,000 M2 families produced from tomato cv. M82 by EMS and fast-neutron mutagenesis. These families were phenotypically characterized and categorized into more than 3147 mutations (Menda et al. 2004). More recently the generation of a T-DNA insertion mutagenesis collection has been reported in wild tomato species (Atares et al. 2011)

Table 1-2. Comparision of the most widely used DNA markers in plants.

	RFLP	RAPD	AFLP	SSR	CAPS	SNP
Genomic abundance	High	Very high	Very high	Medium	High	High
Amount of DNA required	High	Low	Medium	Low	Low	Low
Type of polymorphism	Single base changes, insertion, deletion	Single base changes, insertion, deletion	Single base changes, insertion, deletion	Changes in length of repeats	Single base changes, Insertion, Deletion	Single base changes, Insertion, Deletion
Level of polymorphism	Medium	High	High	High	High	High
Inheritance	Co-dominant	Dominant	Dominant	Co-dominant	Co-dominant	Co-dominant
Ease of use	Labor intensive	Easy	Difficult initially	Easy	Easy	Easy
Reproducibility	High	Intermediate	High	High	high	high
Cost	High	Low	Medium	High	Low	Low

RFLPs: Restriction fragment length polymorphism

RAPD: random amplification of polymorphic DNA

RFLPs: restriction fragment length polymorphisms

SSRs: simple sequence repeats

CAPS: cleaved amplified polymorphic sequences

SNPs: single nucleotide polymorphisms

Other TILLING mutant populations are obtained or under construction in several countries and could be used as sources of new phenotypes. Indeed, tomato TILLING populations have been provided for both cultivated varieties (Gady et al. 2009; Rigola et al. 2009; Minoia et al. 2010). Tomato transformation is relatively easy and transgenic lines for delayed ripening (Klee et al. 1991), environmental stress tolerance (Lemaux 2008), pest resistance (Fischhoff et al. 1987), improved nutrition (Roemer et al. 2000), improved taste (Davidovich-Rikanati et al. 2007) have been developed. The first commercially available genetically modified food was a tomato engineered to have a longer shelf life (the "Flavr Savr" variety sold in 1994 in the USA). Currently there are no genetically modified tomatoes available commercially, but scientists are developing tomatoes with new traits like increased resistance to pests or environmental stresses. Furthermore, there are also several other genetic collections such as stress tolerant, cytogenetic, cytoplasmic variants, and collections containing combinations of morphological markers.

Genomic resources available in the tomato

Maps, markers and high-throughput tools for marker identification

Apart from the genetic resources, genomic resources are also rich in tomato. Tomato has been a model organism for fruit bearing plants for long (Giovannoni 2001). The tomato genome is well defined by genetic maps based on molecular markers. During the last 20 years, genetic maps were improved to obtain a current high density map (http://solgenomics.net/). Various types of markers were also developed to characterize genetic resources, construct genetic maps and to associate with phenotypes. The characteristics of different types of markers, restriction fragment length polymorphism (RFLPs), random amplification of polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), cleaved amplified polymorphic sequences (CAPS) and single nucleotide polymorphisms (SNPs), are presented in **Table 1-2**. However, many markers were developed based on polymorphism in wild species and are not polymorphic when used within cultivated germplasm (Jimenez-Gomez and Maloof 2009). Therefore, development of markers continues to be important in tomato. SNPs are the most abundant variation in the genome, most of them are non-gel based markers, they are initially expensive markers but with the development of high-sequencing genotyping technologies, they are quite cost-effective and less time consuming markers (Rafalski 2002). Next generation sequencing (NGS) platforms now help

Table 1-3. Comparison of next-generation sequencing technologies (Deschamps and Campbell 2010).

Sequencing platform	Run time	Read length (bp)	Reads per run (million)	Throughput per run (Gbp)
Roche 454 FLX	10 h	400-500	~1	0.4-0.5
Illumina GAIIx	55 days	100	160	16
ABI SOLID	6-7 days	50	500	25
Helicos HeliScope	8 days	25-55	600-800	21-28
Polonator	80 h	28	300-400	9

to produce an abundance of low-cost, high volume sequence data. They allow millions of bases to be sequenced in one round, at a fraction of the cost relative to traditional Sanger sequencing (Egan et al. 2012). The main commercially available NGS technologies are Roche/454, Solexa/Illumina and AB SOLiD system. Recently, Helicos Biosciences has introduced its version of single-molecule sequencing (tSMS). These technologies present several ways to sequence DNA, prepare template, immobilization, nucleic acid chemistries, synthesis, and detection of nucleotide type and order. Comparison of the next generation sequencing platforms is summarized in **Table 1-3**. These technologies demonstrated the potential to overcome the limitations of Sanger sequencing. For example, sequencing can be multiplexed to a much greater extent by many parallel reactions at a greatly reduced cost. One of the first applications of NGS in plants identified over 36,000 putative maize SNPs using 260,000 and 280,000 ESTs, sequenced using Roche 454 (Barbazuk et al. 2007). Roche 454 technology was used to sequence and assemble 148 Mbp of EST sequences for *Eucalyptus grandis* (Novaes et al. 2008). These technologies evolved very fast, with the length of reads and precision increasing each month, and it is a real opportunity to test such strategy for SNP discovery in tomato. Parallel strategies to develop markers using high-throughput methods include *in silico* mining of SNPs from EST databases (Yang et al. 2004; Labate and Baldo 2005), oligo-based microarray hybridization (Sim et al. 2009), and sequencing introns of conserved orthologous set (COS) genes (Van Deynze et al. 2007; Labate et al. 2009). Recently, whole transcriptome sequencing of six tomato accessions including four representatives of large-fruited cultivated tomato, a cherry tomato and a closely related wild relative led to the identification of 62,576 non-redundant putative SNPs. The utility of these SNPs for assessing genetic variation within cultivated and wild populations was demonstrated (Hamilton et al. 2012). Subsequently, the first large scale SNP genotyping array was also developed for tomato using 8,784 SNPs from this study. The array was optimized for polymorphic SNP markers within the cultivated lineages, allele frequency and genome coverage (Sim et al. 2012).

Furthermore, next generation sequencing technologies 454/Roche GS FLX, SOLID^{TM} sequencer (Applied Biosystems, Foster City, CA), Illumina sequencing combined with Sanger sequencing technology have been used to produce the first high quality tomato whole genome sequence. The size of the genome is estimated at 950 Mb and contains around 35,000 genes which have been annotated (Sato et al. 2012). The genome sequence of domesticated tomato, a draft sequence of its wild relative, *S. pimpinellifolium* are now available to the scientific community. They were compared to each other, and also to potato genome. The two tomato

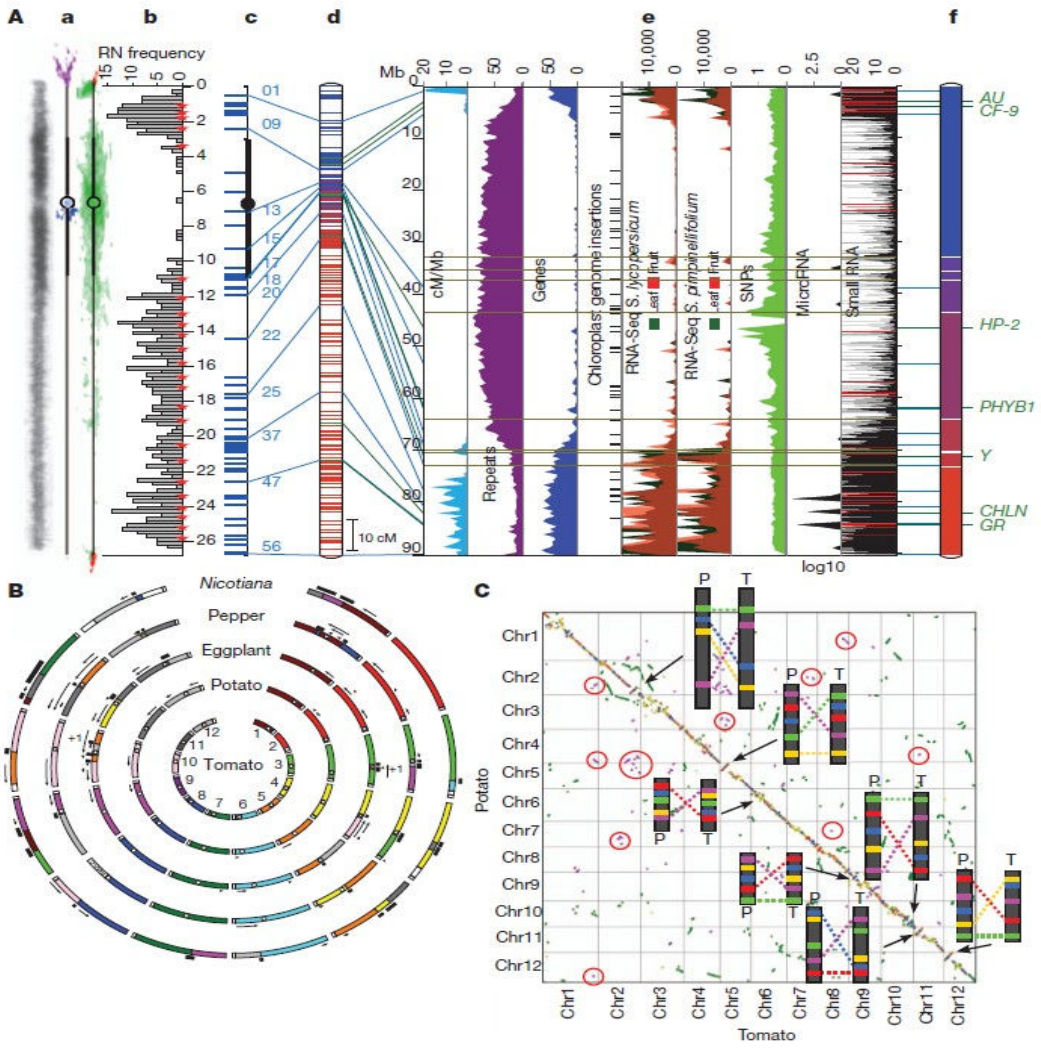


Figure 1-3 Tomato genome topography and synteny. **A**, Multi-dimensional topography of tomato chromosome 1. a, Left: contrast-reversed, 49,6-diamidino-2-phenylindole (DAPI)-stained pachytene chromosome; centre and right: FISH signals for repeat sequences on diagrammatic pachytene chromosomes (purple, TGR1; blue, TGR4; red, telomere repeat; green, Cot 100 DNA (including most repeats). b, Frequency distribution of recombination nodules (RNs) representing crossovers on 249 chromosomes. Red stars mark 5cM intervals starting from the end of the short arm (top). Scale is in micrometers. c, FISH based locations of selected BACs (horizontal blue lines on left). d, Kazusa F2-2000 linkage map. Blue lines to the left connect linkage map markers on the BAC-FISH map (c), and to the right to heat maps (e) and the DNA pseudo molecule (f). e, From left to right: linkage map distance (cM/Mb, turquoise), repeated sequences (% nucleotides per 500 kb, purple), genes (%nucleotides per 500 kb, blue), chloroplast insertions; RNA-Seq reads from leaves and breaker fruits of *S. lycopersicum* and *S. pimpinellifolium* (number of reads per 500 kb, green and red, respectively), microRNA genes (transcripts per million per 500 kb, black), small RNAs (thin horizontal black and red lines, sum of hits-normalized abundances). Horizontal grey lines represent gaps in the pseudo molecule (f). f, DNA pseudo molecule consisting of nine scaffolds. Un-sequenced gaps (approximately 9.8Mb) are indicated by white horizontal lines. Tomato genes identified by map-based cloning are indicated on the right. **B.** Syntenic relationships in the Solanaceae. COSII-based comparative maps of potato, aubergine (eggplant), pepper and Nicotiana with respect to tomato genome. Each tomato chromosome is assigned a different color and orthologous chromosome segment(s) in other species are shown in the same color. White dots indicate approximate centromere locations. Each black arrow indicates an inversion relative to tomato and '11' indicates a minimum of one inversion. Each black bar beside a chromosome indicates translocation breakpoints relative to tomato. Chromosome lengths are not to scale, but segments within chromosomes are. **C**,Tomato–potato syntenic relationships dot plot of tomato (T) and potato (P) genomic sequences based on collinear blocks. Red and blue dots represent gene pairs with statistically significant high and low (Ka/Ks) in collinear blocks, which average Ks<=0.5, respectively. Green and magenta dots represent genes in collinear blocks which average 0.5<Ks<=1.5 and Ks>1.5, respectively. Yellow dots represent all other gene pairs. Blocks circled in red are examples of pan-eudicot triplication. Inserts represent schematic drawings of BAC-FISH patterns of cytologically demonstrated chromosome inversions (Sato et al. 2012).

genomes show only 0.6% nucleotide divergence and signs of recent admixture, but they show more than 8% divergence from potato, with nine large and several smaller inversions (**Figure 1-3**).

During the sequencing of the genome, various genomic resources including Bacterial Artificial Chromosome (BAC), physical maps, a large set of molecular markers and a number of computational pipelines for sequence analysis and genome annotation were developed (Mueller et al. 2009). Meanwhile, numerous large-scale functional genomic resources have been developed over the past years. A collection of more than 11,000 tomato cDNA sequences has been released (Aoki et al. 2010). A large number of Expressed Sequence Tags (ESTs) that currently represent approximately 40,000 unigenes derived from more than 300,000 ESTs have been generated (http://www.sgn.cornell.edu). They have been used in several microarray platforms to study the tomato fruit at transcriptome level (Lemaire-Chamley et al. 2005). At the proteomic level, tomato proteome profiles during precise stages of fruit growth and ripening have been established by Faurobert et al. (2007). Proteomics technical platforms were also developed in order to characterize the differences in cell wall structure and composition that occur during tomato fruit development and ripening (Rose et al. 2004a; Rose et al. 2004b). The complete tomato genome sequence (Sato et al. 2012) will increase the efficiency of protein identification after their separation and subsequent mass spectrometry analysis. Fruit metabolite profiles have also been characterised and integrated with transcriptomic and phenotypic data to study the metabolic networks in order to improve tomato fruit quality (Carrari et al. 2006; Osorio et al. 2011). Furthermore, many other web resources that gather data obtained from different tomato "omics" approaches are publicly available and was reviewed by Barone et al. (2008).

1.1.3 **Tomato fruit quality traits**

The objective of tomato breeding initially focused on yield and ancillary traits (adaptation, disease resistance, earliness) as reviewed by Causse et al. (2006). Nowadays, the social demand concerns tomato fruit quality (organoleptic and nutritional value), but yield and ancillary traits remain necessary targets. Tomato fruit is mainly composed of water (95%), and 5 % of dry matter, which comprising around 50% sugars, 10% organic acids, 8% minerals, 7% pectin and less proportion of carotenoids and other secondary metabolites (Davies and Hobson 1981) (see **Figure 1-4**).

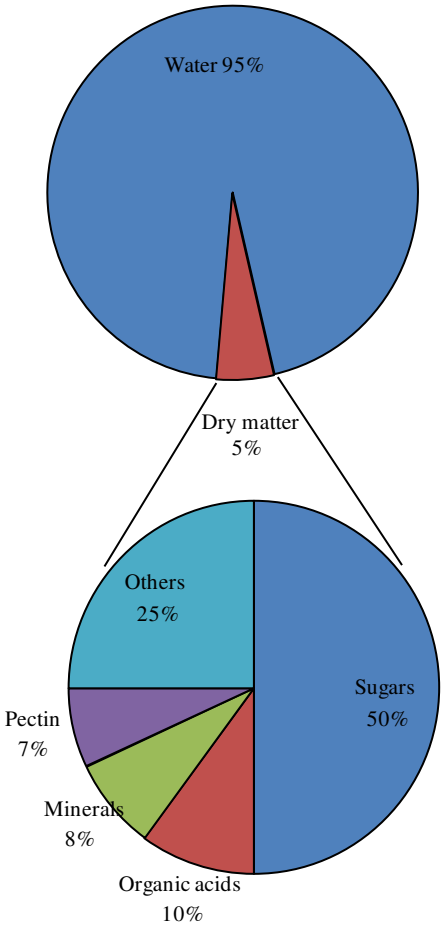


Figure 1-4. Composition of tomato fruit (Adapted from Davies and Hobson 1981)

Fruit quality for fresh market tomato is determined by both external (size, color, and firmness) and internal (flavor, aroma, texture) characteristics. The relationships between fruit characteristics and tomato taste have been widely studied (Causse et al. 2010). Sugar content, acids and their ratio play an important role in determining fruit flavor (Stevens et al. 1977; Stevens et al. 1979; Bucheli et al. 1999). Sugar and acid contents are also related to sweetness and sourness (Stevens et al. 1977) and contribute to sweetness and the overall aroma intensity (Baldwin et al. 1998). Texture traits are more difficult to relate to instrumental measurements, although firmness perceived when eating is partly related to compression tests (Causse et al. 2001). Recently, Tieman et al. (2012) used targeted metabolomics and natural variation in flavor-associated sugars, acids, and aroma volatiles to evaluate the chemistry of tomato fruits, creating a predictive and testable model of liking. This approach provides novel insights into flavor chemistry, the interactions between taste and retronasal olfaction, and paradigm for enhancing liking of natural products. Moreover, consumer preference maps were also constructed in different countries (Netherlands, France, and Italy) and identified the most important characteristics (Causse et al. 2010). The study also showed that preferences were homogeneous across countries. Consumers from different countries have a large range of preference. In addition, consumers appreciated both firm and soft tomatoes. Cherry tomatoes, with fruits rich in acids and sugars, are usually preferred (Hobson and Bedford 1989). In contrast, long shelf cultivars have been described as less tasty than traditional ones (Jones 1986), with lower volatile content (Baldwin et al. 1991). The carotenoid pigments determine the color of tomato fruit. Most tomatoes produce red fruits, but a number of cultivars with yellow, orange, pink, purple, green, black, or white fruit are also available. Ripe fruit contain high levels of lycopene, the pigment that gives tomato its red color (Darrigues et al. 2008).

Both Davis and Hobson (1981) and Stevens (1986) reviewed the literature on the genetic variability for quality traits in natural condition while Dorais et al. (2001) reviewed the impact of environmental conditions in greenhouse production. A large range of genetic variation has been identified for every quality components such as fruit weight, fruit firmness and sugar content (Davies and Hobson 1981; Langlois et al. 1996; Causse et al. 2003; Tikunov et al. 2005). The analysis of trait inheritance shows a polygenic control for most of the quality traits (Stevens 1986; Causse et al. 2003). The genetic dissection of the molecular bases of quality traits is necessary for fruit quality improvement. Up to now, the most widely used methods for

this purpose are based on the identification of Quantitative Traits Locus (QTL) mapping, and more recently on association mapping. In addition, high-throughput sequencing/genotyping technologies are also opening a broad range of new prospects towards the application of systems biology approaches for better understanding the molecular and genetic basis of complex traits.

1.2 **Genetic control of quality traits**

Most quality traits of interest are quantitatively inherited. They are under the control of many genes. Thanks to molecular markers, genetic maps can be constructed and numerous quantitative trait loci (QTLs) were identified in tomato (Paterson et al. 1991; Saliba-Colombani et al. 2001; Wang et al. 2006; Szalma et al. 2007; Orsini et al. 2012). QTL mapping approach allowed screening mapping populations usually derived from the cross between a wild species and a cultivated accession. This approach has several advantages: (i) no population structure in the mapping population; (ii) segregating alleles are at balanced frequency; (iii) it allows the detection of rare alleles and epistasis. However, the linkage mapping approach has several limitations: (i) restricted allelic variation in bi-parental mapping population; (ii) low precision due to limited recombination within the population (Hall et al. 2010). In the following section, we will detail the methodology of QTL mapping and its application in tomato mainly focused on quality traits.

1.2.1 QTL mapping: methodology and application in tomato

There are four main steps in QTL mapping: 1) obtaining a mapping population derived from two contrasted inbred lines for the phenotype of interest; 2) phenotypic characterization of a relatively large number of individuals from a segregating population; 3) build genetic linkage map using molecular markers and based on the recombination rates in the progeny; 4) statistic analysis to identify the loci underlying the genetic architecture of the traits of interest. Such mapping studies are performed to detect the tight linkage of a molecular marker to a gene controlling the variation of the phenotype of interest. Developing a population for QTL mapping involves selecting contrasted parents, crossing them with each other, then advancing the progeny in an appropriate manner to obtain a set of individual plants of lines segregating for traits of interest. Most commonly, a QTL mapping population is derived from the cross of two parental lines that show marked differences for the trait of interest. A typical QTL population consists of 100 to 300 lines or individual plants, each of which is characterized at both genotypic and phenotypic level. Three common types of QTL mapping population are

generally used: F2, recombinant inbred line, and backcross populations. The F2 population can be quickly obtained by selfing the F1 hybrid between two parents. The recombinant inbred line (RIL) population is built through single seed descent from the F2 generation. Backcross populations are developed by crossing an F1 individual to one of its parents. Introgression lines are constructed by an advanced backcross program. Molecular markers for construction of linkage maps in tomato have been described in genomic resources section. Briefly, maps initially mainly relied on the development of RFLP markers (Tanksley et al. 1992). Then PCR-based markers such as SSR, AFLP or CAPs markers were used. During the past two decades, several genetic maps of tomato genome have been reported. More than 2000 loci detected by RFLP, amplified fragment length polymorphism (AFLP), cleaved amplified polymorphic sequence (CAPS), and SSR markers were mapped on populations derived from crosses between a tomato accession and related wild species. In order to increase the average density of initial maps, SNP markers now replace these markers. Recently, a total of 1137 markers, including 793 SNPs and 344 SSR were mapped onto two mapping populations derived from crosses between 'Micro-Tom' and either 'Ailsa Craig' or 'M82' (Shirasawa). SNP array which contains 7720 SNPs was used to generate high-density maps for three interspecific F_2 populations: EXPEN2000 (*Solanum lycopersicum* LA0925 × *S. pennellii* LA0716, 79 individuals), EXPEN2012 (*S.lycopersicum* Moneymaker × *S. pennelli* LA0716, 160 individuals), and EXPIM2012 (*S. lycopersicum* Moneymaker × *S. pimpinellifolium* LA0121, 183 individuals). The EXPEN 2000-SNP and EXPEN 2012 maps consisted of 3,503 and 3,687 markers representing 1,076 and 1,229 unique map positions (genetic bins), respectively. The EXPEN 2000-SNP map had an average marker bin interval of 1.6 cM, while the EXPEN 2012 map had an average bin interval of 0.9 cM. The EXPIM 2012 map was constructed with 4,491 markers (1,358 bins) and an average bin interval of 0.8 cM (Sim et al. 2012).

Statistical methods for QTL mapping

The simplest method for QTL mapping is ANOVA that assesses the relationship of a phenotype with a marker genotype, and thus indicates which markers are associated with the quantitative trait of interest (Soller et al. 1976). This method is simple while easily incorporates covariates, and could be extended to more complex models. Disadvantages are that individuals with missing genotype data are excluded, QTL location is not precise in low-density scan and it only considers one QTL at a time. Simple interval mapping (SIM) is another method for QTL mapping. It uses an estimated genetic map as the framework for the location of putative QTL (Lander and Botstein 1989). It statistically tests for a single QTL at

each location incremented along the ordered markers in the genome. The results of the tests are expressed as LOD (logarithm of the odd ratio) scores, which compare the likelihood function under the null hypothesis (no QTL) with the alternative hypothesis (QTL at the testing position) for the purpose of locating probable QTL. The advantages of this method are (i) it takes into account missing data, (ii) it allows higher power in low-density scans and (iii) it improves the precision of QTL location. The disadvantages are greater computational effort, need for specialized software, and it only considers one QTL per chromosome.

The drawback of SIM mapping was overcome by composite interval mapping (CIM) (Zeng 1994) and MQM (multiple-QTL model) developed by (Jansen and Stam 1994). Both methods combine interval mapping for a single QTL in a given interval with multiple regression analysis on markers associated with other QTL. It considers a marker interval plus a few other well-chosen single markers as covariates in each analysis. The advantages of CIM are as follows: mapping of multiple QTLs can be accomplished by the search in one dimension; by using linked markers as cofactors, the test is not affected by QTL outside the region, thereby increasing the precision of QTL mapping; by eliminating much of the genetic variance controlled by other QTL, the residual variance is reduced, thereby increasing the power of detection of QTL. CIM is more powerful than SIM, but is yet to be extensively used in QTL mapping. There are three limitations for CIM mapping: 1) the use of tightly linked markers as cofactors can reduce the statistical power to detect QTL; 2) the test statistic in a marker rich region may not be compared to that in a marker poor region; 3) Estimation of the joint contributions of multiple linked QTL and epistasis is difficult (Zeng et al. 1999).
Multiple interval mapping (MIM) is the extension of interval mapping of multiple QTLs, just as multiple regression extends analysis of variance. MLM allows one to infer the location of QTLs to positions between markers, makes proper allowance for missing genotype data, and allow interactions between QTLs to be tested.

1.2.2 Mapping of QTLs for fruit quality traits in tomato

In tomato, many studies have been carried out to map specially the QTLs controlling fruit quality related traits in progenies derived from inter-specific crosses (Grandillo et al. 2011) (**Table 1-4**). Here are the details for fruit weight, fruit shape, locule number, firmness, color and sugar related traits.

Table 1-4 Summary of QTL mapping studies conducted in Solanum sect. *lycopersicon* for morphological, yield, fruit quality and reproductive-related traits (reviewed by Grandillo et al. 2011)

Wild or donor parent	Main traits analyzed	No. traits evaluated [a]	No. QTL [b]	Mapping population (pop. size)[c]	Marker type[d]	No. Markers	References[e]
S. arcanum LA1708	Yield, fruit quality, horticultural	35 (29)	166	BC3/BC4 (200)	RFLP, PCR, MO	174	Fulton et al. (1997)
S. arcanum LA1708	Biochemical related to flavor	15	103	BC3/BC4 (200)	RFLP, PCR, MO	174	Fulton et al. (2002)
S. chmielewskii LA1028	Fruit weight, brix, pH	3	15	BC1 (237)	RFLP, ISO, MO	70	Paterson et al. (1988, 1990), Frary et al. (2003)
S. chmielewskii LA1028	Brix	1	ns[f]	LA1563 (BC5S5), derived F2	RFLP	60	Osborn et al. (1987)
S. chmielewskii LA1028	Yield, brix, pH	6	ns	ns LA1500-LA1503, LA1563 (BC5S5), derived F2/F3	RFLP, ISO	132	Tanksley and Hewitt (1988)
S. chmielewskii LA1028	Yield, brix, fruit quality	13	ns	LA1500-LA1503, LA1563 (BC5S5), BILs	RFLP	9	Azanza et al. (1994)
S. chmielewskii CH6047	Flowering time	2	8	F2 (149)	AFLP, CAPS/ SCAR/CG, SSR	225	Jimenez-Gomez et al. (2007)
S. chmielewskii LA1840	Fruit weight and composition	16 (14)	103	ILs (20)	COSII, SSR 133	133	Prudent et al. (2009)
S. galapagense LA0483	Frui size, brix, pH	3	29	F2/F3 (350)	RFLP	71	Paterson et al. (1991)
S. galapagense LA0483	Fruit quality	3	73*	F8 RILs (97)	RFLP, MO, ISO	135	Goldman et al. (1995)
S. galapagense LA0483	Morphological	7	41*	* F8 RILs (97)	RFLP, MO, ISO	135	Paran et al. (1997)
S. habrochaites LA1777	Sexual compatibility factors and floral morphology	9	23	BC1 (149)	RFLP	135	Bernacchi and Tanksley (1997), Chen and Tanksley (2004), Chen et al. (2007)
S. habrochaites LA1777	Yield, fruit quality, horticultural	19	121	BC2/BC3 (315/200)	RFLP	122	Bernacchi et al. (1998a)
S. habrochaites LA1777 and S. pimpinellifolium LA1589	Yield, fruit quality, horticultural	12	22	NILs	RFLP	ns[f]	Bernacchi et al. (1998b), Monforte and Tanksley (2000), Monforte et al. (2001), Yates et al. (2004)
S. habrochaites LA1777	Biochemical related to flavor	15	34	BC2/BC3 (315/200)	RFLP	122	Fulton et al. (2002)
S. habrochaites LA1777	Aroma volatiles	40 (27)	30	ILs, BILs (89)	RFLP	95	Mathieu et al. (2009)
S. habrochaites LA1777	Hybrid incompatibility, floral morphology	25	22	ILs, BILs (71)	RFLP	95	Moyle and Graham (2005) Moyle (2007)
S. habrochaites LA0407	Stem vascular morphology	5	1	BILs (BC2S5), F2:3 (64)	RFLP, PCR	67	Coaker et al. (2002)
S. habrochaites LA0407	Fruit color	3	13	BILs (BC2S5)/F3, F4 (64)	RFLP, PCR	63 ;394	Kabelka et al. (2004)
S. habrochaites PI-247087	Ascorbic acid	2	5	BC2/BC2S1 (130/79,68)	AFLP, RFLP, SSR, MO, CGAA	217	Stevens et al. (2007)

Table 1-4 Continued-1

Wild or donor parent	Main traits analyzed	No. traits evaluated [a]	No.QTL [b]	Mapping population (pop. ize)[c]	Marker type[d]	No. Markers	References[e]
S. habrochaites LYC4 (IL5-1 and IL5-2 lines) and S. habrochaites (IVT-line 1)	Parthenocrapy, stigma exsertion	2	5	(two) BC5S1, F2	CAPS, COS, SSR	34	Gorguet et al. (2008)
S. lycopersicum "cerasiforme" Cervil inbred line	Aroma volatiles (18), fruit quality	32 (26)	81	81 F7-RILs (144)	RFLP, AFLP, RAPD, MO	103	Saliba-Colombani et al.(2001), Causse et al (2002, 2007), Lecomteet al. (2004a, b), Chaïb et al. (2007)
S. lycopersicum "cerasiforme" Cervil inbred line	Sensory attributes (12)	12	49	F7-RILs (144)	RFLP, AFLP, RAPD, MO	103	Causse et al. (2001, 2002, 2007), Lecomte et al. (2004a, b), Chaïb et al.(2007)
S. lycopersicum "cerasiforme" Cervil inbred line	Ascorbic acid	2	6	F7-RILs (144)	RFLP, AFLP, RAPD, MO	103	Stevens et al. (2007)
S. neorickii LA2133	Yield, fruit quality, horticultural	30	199	BC2/BC3 (170)	RFLP, PCR, MO	133	Fulton et al. (2000)
S. neorickii LA2133	Biochemical related to flavor	15	52	BC2/BC3 (170)	RFLP, PCR, MO	133	Fulton et al. (2002a)
S. pennellii LA0716	Fruit weight, seed weight, stigma exsertion, leaflet shape	4	21	BC1 (400	ISO 12	12	Tanksley et al. (1982)
S. pennellii LA0716	Morphological (plant, flower, leaf)	11	74	F2 (432)	RFLP	98	deVicente and Tanksley (1991)
S. pennellii LA0716	Yield, fruit quality	6	104	ILs/HILs/ILs _ Tester (49/50/50)	RFLP	375	Eshed and Zamir (1995, 1996), Alpert et al.(1995), Eshed et al.(1996), Gur and Zamir (2004)
S. pennelli LA0716	Fruit shape	2	1	F2 from IL2-5 (60)	RFLP	15	Ku et al. (1999)
S. pennellii LA0716	Sensory attributes , aroma volatiles	ns	1	ILS (4)	RFLP	ns	Tadmor et al. (2002)
S. pennellii LA0716	Leaf morphology and size	8	30	ILS (58)	RFLP	375	Holtan and Hake (2003)
S. pennellii LA0716	Fruit color, carotenoids	6	50	ILs (75)	RFLP, CG	637 (614, 23)	Liu et al. (2003)
S. pennellii LA0716 and S. pimpinellifolium LA1589	Locule number	2	4	Several F2	ns	ns	Barrero and Tanksley (2004)
S. pennellii LA0716	Fruit size and composition	9	81	ILs (70)	RFLP, CG	671 (592,79)	Causse et al. (2004)
S. pennellii LA0716	Leaf and flower morphology	22 (18)	36	F2 (83)	RFLP, SSR, COS	391, (350, 10, 31)	Frary et al. (2004)
S. pennellii LA0716	Fruit quality, Transcriptomic analysis	6	ns	ILs (6)	RFLP	ns	Baxter et al. (2005)

Table 1-4 Continued-2

Wild or donor parent	Main traits analyzed	No. traits evaluated [a]	No.QTL [b]	Mapping population (pop. ize)[c]	Marker type[d]	No. Markers	References[e]
S. pennellii LA0716	Fruit antioxidants	5	20	ILs (76)	RFLP	~600	Rousseaux et al. (2005)
S. pennelli LA0716	Primary metabolites (74), yield related	83	889,326[g]	ILs (76)	RFLP	~600	Schauer et al. (2006)
S. pennelli LA0716	Yield fitness	35	841	ILs, ILHs (76; 76)	RFLP	~600	Semel et al. (2006)
S. pennelli LA0716	Aroma volatiles (23),organic acids	25 (24)	29	ILs (74)	RFLP	~600	Tieman et al. (2006)
S. pennellii LA0716	Ascorbic acid	1	12	ILs (71)	RFLP	~600	Stevens et al. (2007, 2008)
S. pennelli LA0716	Hybrid incompatibility	4	19	ILs (71)	RFLP	~600	Moyle and Nakazato (2008)
S. pennelli LA0716	Primary metabolites (74)	74	332	ILs, ILHs (68;68)	RFLP	~600	Schauer et al. (2008)
S. pennellii LA1657	Yield, fruit quality, horticultural	25	84	BC2/BC2F1 (175)	RFLP	150	Frary et al. (2004)
S. pimpinellifolium CIAS27	Fruit quality, horticultural	18	85	F2 (1,700)	MO, ISO 6	6,4	Weller et al. (1988)
S. pimpinellifolium LA1589	Fruit quality, flower morphology, flowering and ripening time	19	54	BC1 (257)	MO, RAPD, RFLP	120	Grandillo and Tanksley (1996), Alpert et al. (1995), Grandillo et al.(1996), Ku et al. (2000)
S. pimpinellifolium LA1589	Yield, fruit quality, horticultural	21 (18)	87	BC2/BC2F1/BC3 (~170/170)	MO,RAPD,CAPS,RFLP	121	Tanksley et al. (1996)
S. pimpinellifolium LA0722	Fruit quality, lycopene	7	59	BC1/BC1S1 (119)	RFLP	151	Chen et al. (1999)
S. pimpinellifolium LA1589	Fruit shape	2	2	F2 (82)	RFLP	82	Ku et al. (1999)
S. pimpinellifolium LA1589	Fruit size and shape	7	30	F2 (114)	CAPS	90	Lippman and Tanskley (2001)
S. pimpinellifolium LA1589	Fruit and ovary shape	2	1	F2 (100)	RFLP, SNP	108	van der Knaap and Tanksley (2001)
S. pimpinellifolium LA1589	Fruit quality, horticultural	22	71	BC2F6 – BILs (196)	RFLP, MO	127	Doganlar et al. (2002)
S. pimpinellifolium LA1589	Biochemical related to flavor	15	33	BC2/BC2F1/BC3 (~170)	MO, RAPD, CAPS, RFLP	121	Fulton et al. (2002)
S. pimpinellifolium LA1589	Fruit shape	3	4	F2 (85)	RFLP	97	van der Knaap et al. (2002)
S. pimpinellifolium LA1589	Fruit shape and size	10	50	F2 (200)	RFLP	93	van der Knaap and Tanksley (2003)
S. pimpinellifolium LA1589	Fruit shape	15	36, 32, 27[h]	(two) F2, BC1 27(99; 130; 100)[h]	RFLP, PCR	111, 111, 108[h]	Brewer et al. (2007)
S. pimpinellifolium LA1589	Fruit shape	14	20, 23, 20[h]	(three) F2 (130;106; 94)	RFLP, PCR	111, 96, 97	Gonzalo and van der Knaap (2008)
S. lycopersicum IVT KTl (breeding line containing S.pimpinellifolium and S.neorickii introgressions)	Fruit size, flowering and ripening time	6	3	F2, F3 (292)	RFLP	45	Lindhout et al. (1994)
S. pimpinellifolium LA1237(the "selfer") and LA1581 (the "outcrosser")	Flower morphology and number	6 (4)	5	F2 (147)	RFLP	48	Georgiady et al. (2002)

[a]The number of traits for which QTLs were identified is indicated in parenthesis
[b]An "*" indicates the number of significant markers _ traits associations
[c]*ILH* Introgression line hybrid
[d]For markers abbreviations text
[e]Some of the related and/or follow-up studies are also listed
[f]ns not specified
[g]Fruit metabolism and yield-associated traits, respectively
[h]per population

Fruit weight

In tomato, a minimum of 28 loci accounting for fruit mass has been identified (Grandillo et al. 1999; Paran and van der Knaap 2007). Six of them *fw1.1*, *fw2.2*, *fw2.3*, *fw3.1*, *fw3.2*, *fw4.1* and *fw9.1* are responsible for more than 20% of fruit weight variation. However, only fw2.2 has been cloned and studied at the molecular level so far (Tanksley 2004). *fw2.2* encodes a protein controlling fruit growth and mutations at this locus resulted in a major increase in fruit size during tomato domestication (Alpert et al. 1995; Frary et al. 2000). This locus makes the largest contribution to the difference in fruit size between most cultivated tomatoes and their small-fruited wild species counterparts (Alpert et al. 1995). More recently, *fw3.2* was fine mapped by Zhang et al. (2012) to a 51.4kb interval corresponding to a region comprising seven candidate genes. Fruit shape analysis indicated that *fw3.2* primarily played a role in controlling fruit weight, with a minor effect on fruit shape (Zhang et al. 2012).

Fruit shape

Most of the tomato cultivars produce round fruits but many shape variants exist, either flattened or elongated. The major loci that have been identified as contributing to an elongated shape in tomato are *sun* (van der Knaap and Tanksley 2001; van der Knaap et al. 2002; 2004), *ovate* (Ku et al. 1999; Liu et al. 2002; van der Knaap et al. 2002) and *fs8.1* (Grandilio et al. 1996; Ku et al. 2000). *Sun* encodes a protein that is a positive regulator of growth factor. It controls elongated and pointed fruit shape and is hypothesized to alter hormone or secondary metabolite levels (Xiao et al. 2008). The QTL *ovate* was cloned and the gene corresponds to a new class of regulatory protein responsible for the variation of the length / width index (Liu et al. 2002). The QTL *fs8.1* induces a square shape of the fruit. This character was used in breeding cultivars for industrial production because it increased the strength of the fruit during mechanical harvesting (Ku et al. 2000).

Locule number

Locule number of tomato fruit influences fruit shape and size. The locules are directly derived from the carpels in the flower. Two QTLs *fas* and locule number (*lc*, also named lcn2.1), have major effects on the phenotype. Both have been positionally cloned. Mutations in the *fas* gene are described as key factors leading to the increase in fruit size (Cong et al. 2008).

The *lc* QTL is located in a 1,600-bp region 1,080-bp downstream from 3' end of WUSCHEL, which encodes an homeodomain protein that regulates stem cell fate in plants (Munos et al. 2011). In addition to *fas* and *lc*, several QTLs controlling locule number have been mapped, and a candidate gene approach has been used to map genes regulating floral meristem development that might colocalize with known QTLs for locule number (Barrero et al. 2006).

Firmness

Fruit firmness is determined by a number of factors including cell wall structure, turgor (Saladie et al. 2007) and cuticle properties (Chaib et al. 2007) and is therefore likely to be a highly complex trait, involving numerous genes and pathways (Brummell and Harpster 2001). A total of 64 QTLs controlling firmness were mapped using seven different populations (Tanksley et al. 1996; Fulton et al. 1997; Bernacchi et al. 1998a; Causse et al. 2002; Doganlar et al. 2002; Frary and Doganlar 2003). Many candidate genes tested linked to cell wall modification but none validated as QTL. One QTL was fine mapped on chromosome 2 using the *S. pennellii* interspecific introgression lines (IL) and fine mapped in a population consisting of 7500 F2 and F3 lines from IL 2-3 and IL 2-4 (Chapman et al. 2012).

Color

Eight QTL that modify lycopene content in the fruit, including a major QTL accounting for 12% of the total phenotypic variation, were identified in a segregating population involving *S. pimpinellifolium* (Chen et al. 1999). Five QTL that modify fruit color intensity were identified by Tanksley et al. (1996). Saliba-Colombani et al. (2001) identified three QTL for L*, a* and b* color parameters, two QTL for lycopene and three for carotene content. QTLs related to these three components were also mapped on a recombinant inbred line population *S. lycopersicum* × *S. pennellii* (Yong-Sheng et al. 2003). Sixteen QTLs were identified but only three corresponded to previously known mutations (r, B and Del) involved in lycopene synthesis pathway.

Sugar and related traits

Genomic regions carrying QTLs for sugar content or related traits (Brix, fructose, glucose, or sucrose content) were detected on different populations involving several species (Paterson et al. 1988; Paterson et al. 1990; Paterson et al. 1991; Azanza et al. 1994; Eshed and Zamir 1995; Goldman et al. 1995; Grandillo and Tanksley 1996; Tanksley et al. 1996; Fulton et al. 1997; Fulton et al. 2000; Fulton et al. 2002; Bernacchi et al. 1998a; Chen et al. 1999; Saliba-Colombani et al. 2001; Doganlar et al. 2002; Causse et al. 2004) (see **Figure 1-5**).

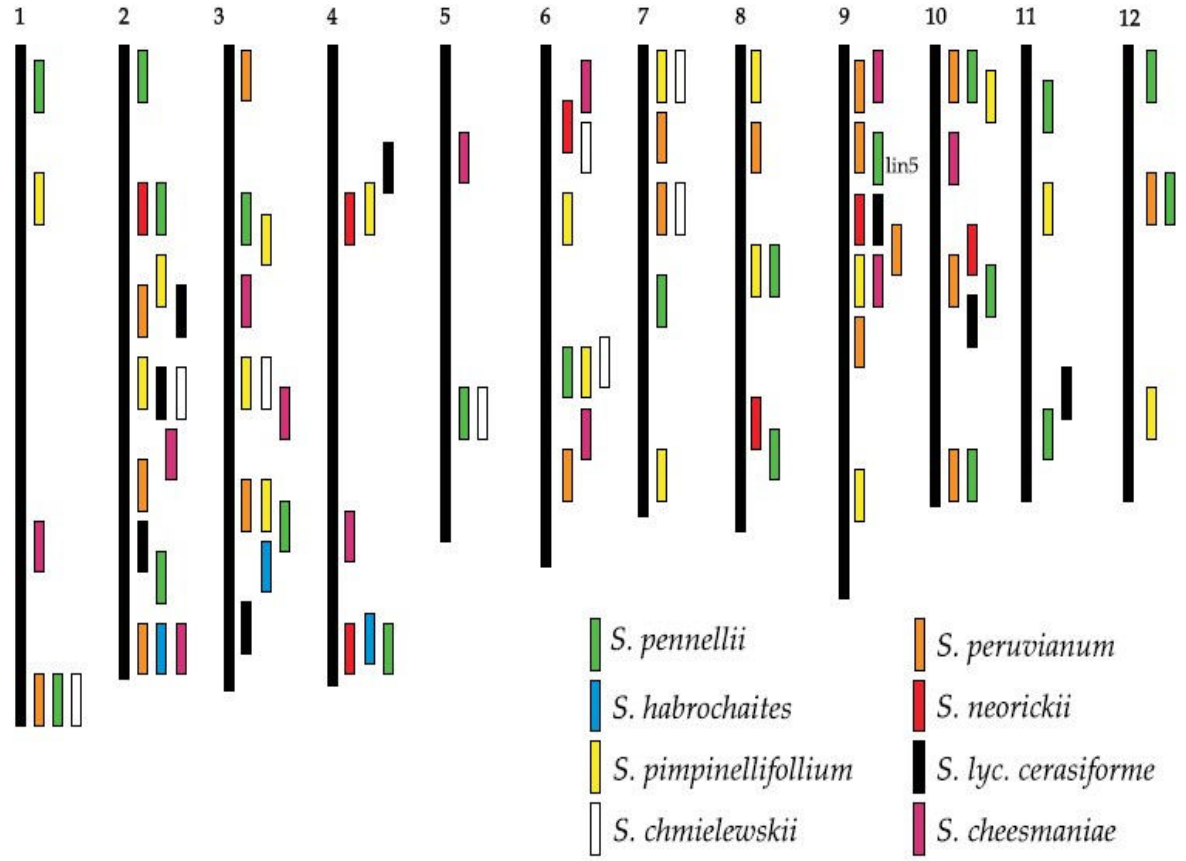


Figure 1-5 Summary of QTL for sugar content or related traits (Brix° or hexose content) in one of the following progeny: *S. lycopersicum* × *S. cheesmaniae* F2 population (Paterson et al. 1991); *S. lycopersicum* ×*S. cheesmaniae* recombinant inbred population (Goldman et al. 1995); *S. lycopersicum* × *S. chmielewskii* F2 and advanced backcross lines (Paterson et al. 1988, 1990; Azanza et al. 1994);*S. lycopersicum* × *S. habrochaites* advanced backcross population (Bernacchi et al. 1998); *S. lycopersicum* × *S. neorickii* advanced backcross population (Fulton et al. 2000); *S. lycopersicum* × *S. pimpinellifolium* advanced backcross population (Tanksley et al.1996; Doganlar et al. 2002); *S. lycopersicum* × *S. pimpinellifolium* backcross populations (Grandillo and Tanksley 1996; Chen et al. 1999); *S. lycopersicum* × *S. pennellii* introgression lines (Eshed and Zamir 1995; Causse et al. 2004); *S. lycopersicum* × *S. pennellii* advanced backcross population (Frary et al. 2004); *S. lycopersicum* × *S. peruvianum* advanced backcross population (Fulton et al. 1997); *S. lycopersicum cv cerasiforme*×*S. lycopersicum*recombinant inbred line population (Saliba-Colombani et al.2001). The data concerning the advanced backcross involving *S. pimpinellifolium*, *S. peruvianum*, *S. neorickii* and *S. habrochaites* were summarized by Fulton et al. (2002). The QTLs were positioned on the tomato reference map (Tanksley et al. 1992), based on the nearest marker (Adapted from Labate et al. 2007).

Some QTLs controlling sugar related traits were identified in the same chromosomic region. Brix9-2-5, a major QTL that increase sugar content of tomato was delimited through fine mapping to 484 bp of the apoplastic invertase (LIN5), which operates in sugar transport to the developing fruit (Fridman et al. 2000). For soluble solid content, 23 QTLs controlling the content in soluble solids have been mapped (Fridman et al. 2000). Causse et al. (2004) mapped 63 genes encoding enzymes involved in the Calvin cycle, glycolysis, the Krebs cycle, sugars and starch metabolism, transport and a few other functions, and a few co-localizations between candidate genes and QTL were detected. Lin5 (Fridman and Zamir 2003; Fridman et al. 2004), together with the absence of a gene regulating the ADPGppase on chromosome 9 (Schaffer et al. 2000), supports the hypothesis that high starch accumulation in L9 was stimulated by the higher sucrose unloading, and hexose transport within the cells (Ho 1996; Schaffer et al. 1999), as observed in another study (Li et al. 2002).

1.2.3 Positional cloning

QTL mapping often detect QTLs within large marker intervals (>= 10cM). The next step after QTL detection is to localize the QTL to a precise genomic region. It is also critical to determine the number of QTLs segregating within the region. Fine-mapping experiments have been carried out to identify QTLs where a single QTL was detected. Two genes involved in tomato fruit size have been isolated using a positional cloning strategy, *fw.2.2*, the first case in which a quantitative trait locus (QTL) was resolved in the corresponding gene (Frary et al. 2000) and fasciated, which affects tomato fruit size through its effect on locule number (Cong et al. 2008). More recently, Munos (2011) reported the map based cloning for *lc*, a QTL responsible for locule number, and identified two SNPs downstream of WUSHEL that controls the trait using association mapping strategy.

1.2.4 Association mapping

Although QTL mapping will continue to be an important tool for QTL detection in crops, it is now a classical approach. However, the development of a mapping population is time consuming and very costly. The main pitfall of QTL mapping is the detection of large effect QTLs due to the restricted allelic variation and modest degree of recombination in bi-parental mapping population (Hall et al. 2010); whereas in association mapping historical recombination and natural genetic variation were employed for high resolution mapping (Zhu et al. 2008) (Figure 1-6)

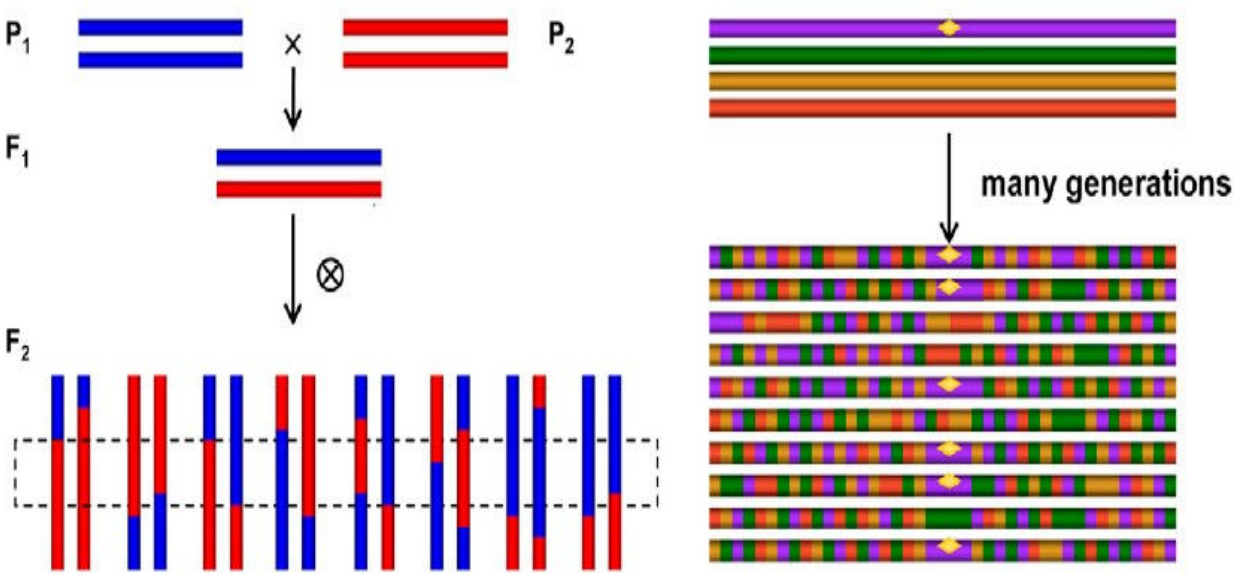


Figure 1-6. Schematic comparison of linkage analysis with designed mapping populations and association mapping with diverse collections (Adapted from Zhu et al. 2008)

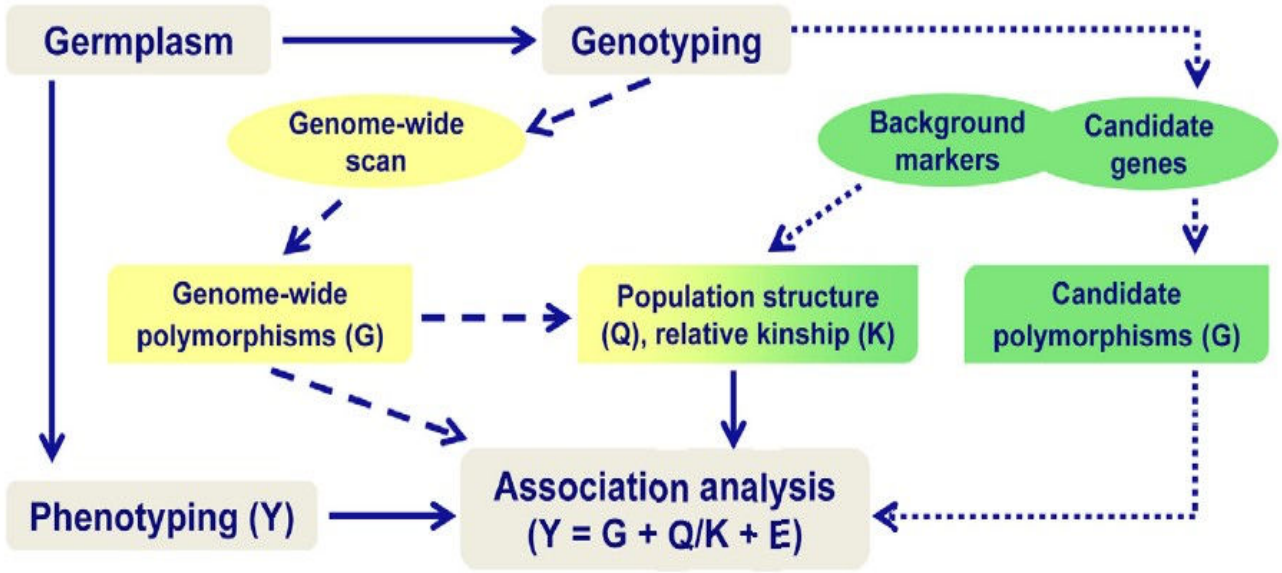


Figure 1-7. The scheme of association mapping for tagging a gene of interest using germplasm accessions (Adapted from Zhu et al, 2008)

Association mapping also called linkage disequilibrium (LD) mapping refers to the analysis of statistical association between a polymorphic site and the variation of a trait in a set of unrelated individuals. The main steps of the procedure are illustrated in **Figure 1-7**. LD refers to non-random association between two markers (alleles at different loci), two genes or QTLs, a gene/QTL and a marker locus (Gupta et al. 2005). Generally, the strength of the correlation between two markers is a function of the distance between them: the closer two markers are, the stronger the LD. The resolution with which a QTL can be mapped is a function of how quickly LD decays over distance (Myles et al. 2009). Thus, to identify the causal polymorphism responsible of a trait variation, it is essential to know the pattern of LD at the whole genome level, and to know how LD is decreasing with the physical distance and genetic distance between loci. When a large LD exists at the whole genome level, one can hope to capture a large diversity for useful genes in sampling maximum allelic diversity at neutral loci (McKhann et al. 2004), but association mapping will not be much more precise than QTL mapping. On the contrary a limited LD for closely linked sites (few kb) will allow the fine mapping of the polymorphism of the candidate genes responsible of a trait variation (Remington et al. 2001), but hundreds of thousands SNPs will be necessary for genome wide association. Association mapping employs occurring variation in genetic collections thus overcome the limited variation that characterizes QTL mapping population. Moreover, many recombination events that have occurred over evolutionary history also means that linkage blocks are smaller in an association mapping population than in a QTL mapping population, hence association mapping results in much more fine-scale mapping (Nordborg et al. 2002). Association mapping was first implemented in human and then adapted to other organisms such as plants. In plants, many association studies have been published to date for several traits, such as flowering time and pathogen resistance in Arabidopsis (Aranzana et al. 2005), yield and its components in rice (Agrama et al. 2007), leaf architecture in maize (Tian et al. 2011), iron deficiency chlorosis in soybean (Mamidi et al. 2011) as illustrated in **Table 1-5**. However, a potentially serious obstacle to association mapping is the population structure. Population structure may cause false positive results. The real structure of the association mapping population is not totally known; it has thus to be inferred with various statistical methods using molecular markers. Several statistic methods were developed to deal this problem. Pritchard et al. (2000) introduced the so-called “structured association mapping” for reducing confounding effect due to population structure.

Table 1-5. Some successful reports of Association mapping in plants. Adapated from Al-Maskri et al. (2012).

Species	Mapped traits	References
Self-pollinated		
Arabidopsis	Growth response, flowering time, branching architecture and pathogen resistance	Caicedo et al.(2004); Oslen et al. (2004); Ehrenreich et al. (2007); Zhao et al. (2007)
Rice	Plant height, flag leaf length and width, tiller number, stem diameter, stigma characteristics, flowering date, panicle length, grain length and width, grain thickness, 1000-grain weight.	Zhang et al. (2005); Agrama et al. (2007); Yan et al. (2009)
Barley	Plant height, heading date, flowering date, rachila length, yield stability, yield, mildew and leaf rust resistance.	Igartua et al. (1999); Ivandic et al. (2003); Kraakman et al. (2004, 2006)
Wheat	Plant height, milling quality, High molecular weight glutenin, 1000-kernal weight, protein contents, drought tolerance, sedimentation value, test weight, starch concentration, insect and disease resistance.	Breseghello and Sorrells (2006); Ravel et al. (2006); Roy et al. (2006); Crossa et al. (2007); Peng et al. (2007); Reif et al. (2011); Adhikari et al. (2011); Zhang et al. (2011).
Potato	Resistance to wilt disease, phytophthora, bacterial blight, tuber shape, flesh color, under water weight and maturity etc.	Gebhardt et al. (2004); Simko et al. (2004); Malosetti et al. (2007); D'hoop et al. (2008)
Soybean	Seed protein contents	Jun et al. (2008)
Tomato	Fruit quality traits fruit weight, soluble solid content, locule number and morphological traits	Nesbitt and Tanksley (2002), Munos et al. (2011); Ranc et al. (2012) ; Mazzucato et al. (2008)
Cross-pollinated		
Maize	Plant height, endosperm color, starch production, flowering time, maysin and chlorogenic accumulation forage quality, cell wall digestibility and oleic acid concentration.	Remington et al. (2001); Thornsberry et al. (2001); Guillet-Glaude et al. (2004); Wilson et al. (2004); Anderson et al. (2007)
Forage grasses	Cold tolerance, forage quality, flowering time and carbohydrate contents	Dobrowolski and Forster (2007); Skøt et al. (2007)
Forest trees	Early-wood micro fibril angle, wood growth rate, and wood density	Thumma et al. (2005); Wicox et al. (2007)

The approach is based on first assigning individuals to subpopulations using a model-based Bayesian clustering algorithm, STRUCTURE, and then carrying out all analyses conditional on the inferred assignments. Yu et al. (2006) recently introduced a mixed linear model (MLM) approach to control for population structure. The key of their approach is to use a random effect to estimate the fraction of the phenotypic variation that can be explained by genome-wide correlations. The individual random deviations from the population mean are constrained by assuming that the (phenotypic) covariance between individuals is proportional to their relative relatedness (or kinship), which is estimated using genome-wide marker data. In addition to this random effect, Yu et al. (2006) used the population assignments produced by the STRUCTURE algorithm (the Q matrix) as a fixed effect in the model, as in structured association (Pritchard et al. 2000; Thornsberry et al. 2001; Yu and Buckler 2006). Recent studies demonstrated that MLM models successfully corrected for the genetic relatedness in association mapping in maize and Arabidopsis panel data sets. However, the current available mixed-model suffers from computational difficulty. A new method, efficient mixed model association (EMMA), corrects for population structure and genetic relatedness in model organism association mapping (Kang et al. 2008). It allows to substantially increasing the computational speed and reliability of the results. Multiple corrections for multiple testing have been used to control significant associations. Failure to correct for multiple tests may produce false positive associations. Several correction approaches have been proposed to control the false positive results. Bonferroni correction and False Discovery Rate (FDR) are the two most commonly used approaches for multiple comparisons. Bonferroni correction is quite stringent but offers the most conservative approach to control false positives. FDR correction controls the expected proportion of incorrectly reject null hypotheses (type I errors) (Benjamini and Hochberg 2000). It is less conservative compared to Bonferroni correction. Permutation tests are simple and commonly used in linkage mapping to control genome wide error. The advantage of permutation testing is that it controls false positives due to multiple testing. The major drawback of this approach is that it is computationally expensive (Pattin et al. 2009).

In general, LD is expected to be higher in autogamous species than in allogamous species because less efficient recombination events occurred in autogamous species and their individuals are more likely to be homozygous at a given locus than in allogamous species (Flint-Garcia et al. 2003). Thus, the resolution for performing association mapping in autogamous species is expected to be relatively lower than in allogamous species.

Therefore, association mapping was first used in allogamous species or species with wide range of genetic diversity. Several studies were carried out to assess the LD and population structure in tomato using different type of markers. Van Berloo et al. (2008) investigated population structure and LD decay along chromosomes within a diverse set of 94 cultivated tomato cultivars, representing a wide range of phenotypic diversity of fruit quality traits using 882 AFLP markers. The results showed a clear substructure consistent with a grouping based on fruit size. LD was up to 20cM for cherry type tomato. Large fruited lines showed the similar LD pattern. Strong LD is limited to certain hotspots. Strong LD between markers on different chromosomes is rare. Ranc et al. (2008) studied the population structure in 340 accessions consistinf of 130 *S. lycopersicum*, 144 *S. l. cerasiforme* and 66 *S. pimpinellifolium* accessions with 20 SSR markers. They showed that the 144 *S. l. cerasiforme* accessions were structure into two groups; one close to the domesticated group and one resulting from the admixture of the *S. lycopersicum* and *S. pimpinellifolium* genomes. Robbins et al. (2011) used 340 PCR-based markers including SNPs to analyze the LD in 102 tomato lines representing wild species, landraces, vintage cultivars, and contemporary (fresh market and processing) varieties. The genome wide analysis showed that LD decays over 6-8cM when taking into account all cultivated tomatoes, including vintage and contemporary tomatoes. It was 6-14 cM and 3-16 cM for contemporary processing varieties and fresh market varieties, respectively.

A few association studies have been carried out on tomato. Nesbitt and Tanksley (2002) evaluated the polymorphism around the fw2.2 locus, a fruit weight QTL positionally cloned, in a small sample of tomato cultivars and cherry tomato accessions. The authors failed to identify any association but showed the admixture structure of cherry tomato accessions. Mazzucato et al. (2008) studied associations between 29 SSR markers and 15 morpho – physiological traits in 50 tomato landraces. The markers were selected to include a group of loci in regions harboring identified quantitative trait loci (QTLs) that affect fruit size and/or shape (Q-SSRs) and a group of markers that have not been mapped or shown to have a priori known linkage (NQ-SSRs). They showed that the proportion of significant associations is higher between the Q-SSR subset of markers and the subsets of traits related to fruit size and shape than for all of the other combinations. Munos et al. (2011) used association analysis to identify two SNP located in a small region of chromosome 2 involved in the control of locule number in tomato fruit. They used a combination of map-based cloning to identify the locus region and association mapping to refine its molecular characterization. The map based cloning step identified a 1,608-bp region.

For the association mapping step, this region was sequenced in a core collection of 88 accessions chosen from a large genetic collections to maximize the genetic diversity. This step identified two SNPs that were strongly associated with the locule number phenotype. The two SNPs were then validated in the study of Ranc et al. (2012) who conducted a pilot association study using Mixed linear model focused on chromosome 2 and characterized 90 tomato accessions for fruit weight, soluble solid content and locule number and the 341 SNPs. They identified respectively 37, 13 and 3 associations for fruit weight, soluble solid content and locule number with a model taking into account structure and kinship based on 20SSR. However, lower numbers of associations (11, 3 and 3 for the three traits, respectively) were detected when taking into account the structure and kinship based on 341 SNPs. Their results illustrated the importance of the correction for population structure. They showed the efficiency of genome admixture in cherry tomato to overcome the low resolution of association mapping for inbred crop (Annexe1). Thanks to the high-throughput sequencing technologies and full sequence of the tomato genome, it is now possible to extend the analysis to the whole genome level using thousands of SNP.

1.2.5 Novel approaches towards the improvement of important traits

Apart from the two most commonly used approaches, linkage analysis and association mapping, novel approaches were also developed for the dissection of complex traits. Two types of novel approaches can be underlined: (1) building genetic materials such as Nested association mapping (NAM) and multi-parent advanced generation intercross (MAGIC) population and (2) new analytical approaches referred as 'system biology' which consists in combining information from different levels such as metabolomic, proteomic, transcriptomic and genomic levels. The details of these approaches are described in the following paragraphs.

1.2.5.1 Nested association mapping

Nested association mapping (NAM) approach has been proposed to combine the advantages of QTL mapping and association mapping approaches for identifying quantitative loci responsible for phenotypes of interest. It has high mapping resolution and low marker density requirement compared to association mapping.

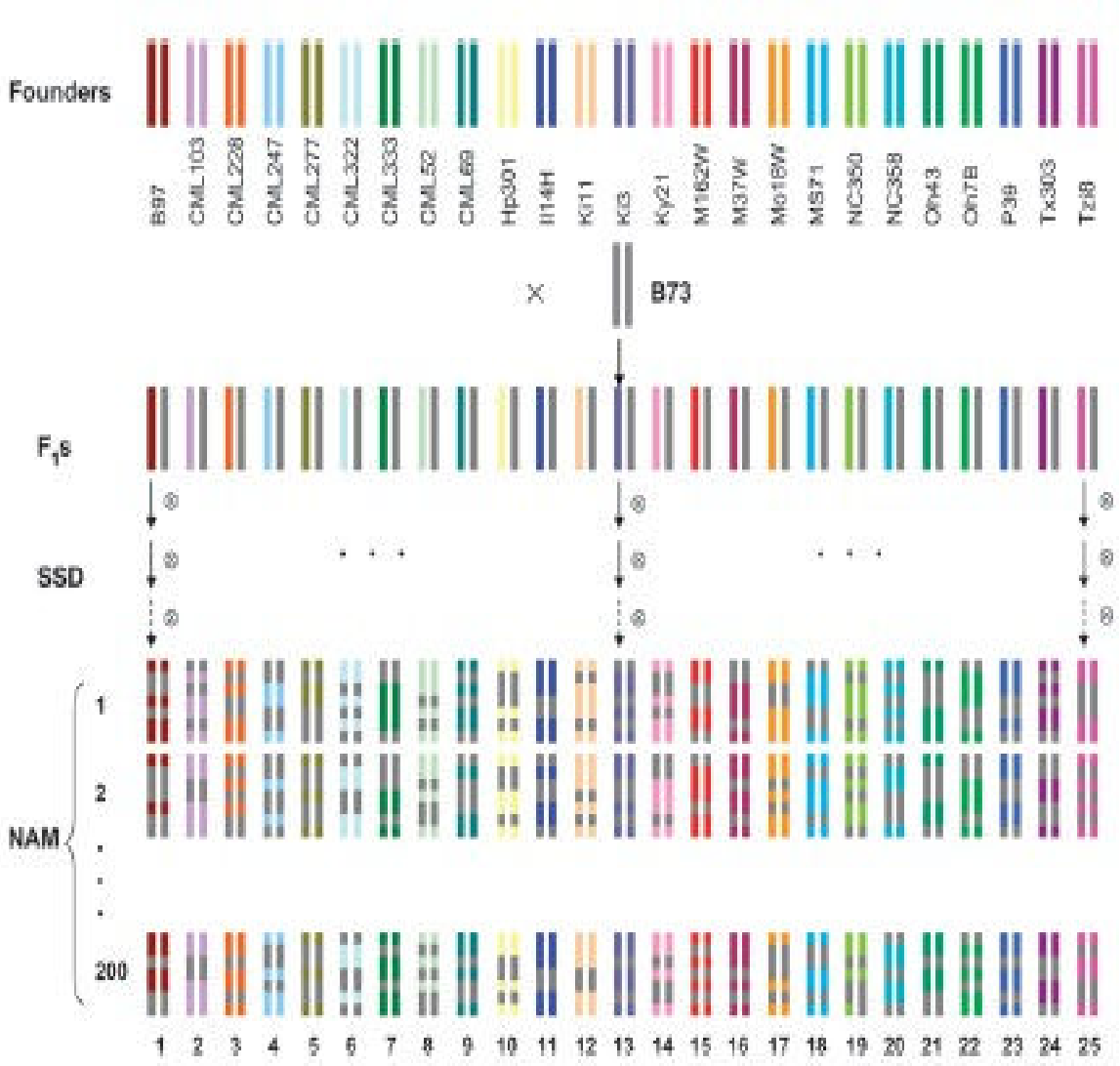


Figure 1-8. Construction of a NAM population in maize. Diagram of genome reshuffling between 25 diverse founders and the common parent and the resulting 5000 immortal genotypes. Due to diminishing chances of recombination over short genetic distance and a given number of generations, the genomes of these recombinant inbred lines (RILs) are essentially mosaics of the founder genomes. SSD, single-seed descent. Modified from Yu et al. (2008).

Both historic and recombination events were employed in this approach (Yu et al. 2008). NAM population has been established in maize (Yu et al. 2008) and Arabidopsis (Brachi et al. 2010). The populations derives from the cross of a central parent with a panel of other diverse parents (Huang et al. 2011). **Figure 1-8** showed an example of the construction of NAM population in maize. A total of 5000 NAM lines were generated from 25 families, each developed by crossing one of the 25 diverse maize lines to one central parent, B73 (Yu et al. 2008; Buckler et al. 2009; McMullen et al. 2009).

The first application of NAM to identify QTLs was reported by Buckler et al. (2009) on the genetic architecture of maize flowering time. This study identified 39 QTLs explaining 89% of the variance in days to silking and days to anthesis and 29 QTLs explaining 64% of the variance in the silking-anthesis interval. It showed that this approach was useful for the characterization of agronomic traits in maize as well as other species. However, interactions of QTL with genetic background may not be detected, as one parent is common to all subpopulations.

1.2.5.2 Interest of building MAGIC population

The challenge for plant geneticists is to build appropriate populations so that the calculation progress, analysis and profiling will be fast. The large number of parental accessions increases the allelic and phenotypic diversity over traditional RILs, potentially increasing the number of QTL that segregate in the cross. Advanced intercross (AIC) was first used in mice employing multiple parents to map a QTL explaining 10% of the phenotypic variation for anxiety in mice to a 4.8 Mb region (Yalcin et al. 2005). The complex trait consortium was then built to analyze complex traits in mice using eight-way funnel breeding scheme (Churchill et al. 2004). Cavanagh et al. (2008) proposed to use this scheme in crops which was named multi-parent advanced generation intercross (MAGIC) population. This approach is ideal: very diverse, no population structure, and suitable for both fine and coarse mapping (Cavanagh et al. 2008). **Table 1-6** presents the comparison of classical genetic mapping strategies, association mapping and the use of MAGIC population. MAGIC requires two simple extensions to the more traditional method of QTL mapping in which both parents are crossed and associations are screened in the segregating progeny. Firstly, rather than simply crossing two lines, a population is determined by crossing several lines (8, 16 or more).

Table 1-6 Relative strengths and weaknesses of three methods for the identification of QTL in crops, bi-parental linkage analysis (linkage), association mapping (association) and Multi-parent Advanced Generation Inter-crosses (MAGIC) (Cavanagh et al. 2008)

Application	Linkage	Association	MAGIC
Suitability for coarse mapping	+	-	+
Suitability for fine mapping	-	+	+
Low genotyping requirement	+	-	-
Resistant to population substructure	+	-	+
Relevance to breeders	-	+	+
Relevance over time	-	+	+
Time to establish	-	-	+

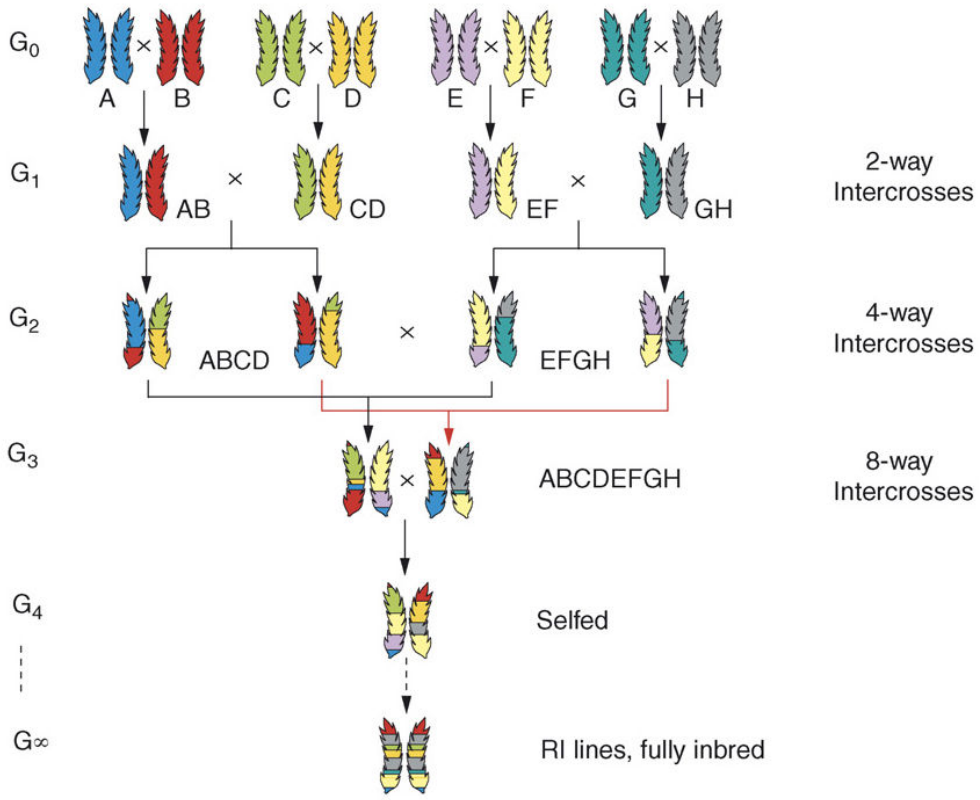


Figure 1-9. Creation of a Multi-parent Advanced Generation Inter-Cross (MAGIC) population with 8 parental lines (Modified from Cavanagh et al. 2008)

Secondly, rather than looking for associations immediately after crossing, the population is primarily recycled through several more generations of crosses (**Figure 1-9**) or selfing, increasing the number of efficient recycle. MAGIC population has been used in *Arabidopsis thaliana* to fine map development related traits (Kover et al. 2009). In this study, a set of 527 MAGIC lines was derived from 19 diverse founders. These lines were genotyped with 1,260 SNP. QTL responsible for 10% of the phenotypic variation could be mapped in most cases with an average of 300 kb confidence interval, and if the number of lines were doubled, the interval would be under 200 kb (Kover et al. 2009). Moreover, this approach was also used to study candidate genes for flowering time in 275 Arabidopsis MAGIC lines (Ehrenreich et al. 2009). Several other MAGIC populations were produced in wheat (http://www.niab.com/pages/id/93/MAGIC_Populations_in_Wheat), rice (http://www.intl-pag.org/19/abstracts/W80_PAGXIX_500.html) and also in tomato (presently created in INRA, GAFL). Statistical package like HAPPY has been developed to analyze heterogeneous stocks including MAGIC (http://spud.well.ox.ac.uk/arabidopsis). Huang and George (2011) have recently developed a R/mpMap package to solve the complex statistical need for MAGIC population. It has interface with earlier mapping platforms, R/qtl (Broman et al. 2003) and R/happy (Mott et al. 2000). With the availability of tomato genome, next generation sequencing/genotyping and advanced statistical methods, MAGIC will be ideal to generate high density maps and to fine map QTLs in tomato.

1.2.5.3 Systems biology approaches

High throughput approaches have opened new and exciting prospects for analyzing biological systems and their complex functions at different levels including metabolomic, proteomic, transcriptomic and genomic (**Figure 1-10**). The systems biology approaches integrating 'omics' resources and technologies are revolutionizing biology and offer new strategies for a better understanding of the molecular and genetic bases of complex traits.

Metabolome analysis and enzymatic activities

Metabolome is the distribution and concentration of metabolites in the cell. Few metabolomic studies have been performed in tomato fruit. Comparative analysis of the metabolite composition in leaves and fruits from six tomato species was studied by Schauer et al. (2005). In this study, a wide range of variation was observed for sugar content, amino acid composition and for secondary metabolite levels in both leaves and fruits of the wild species.

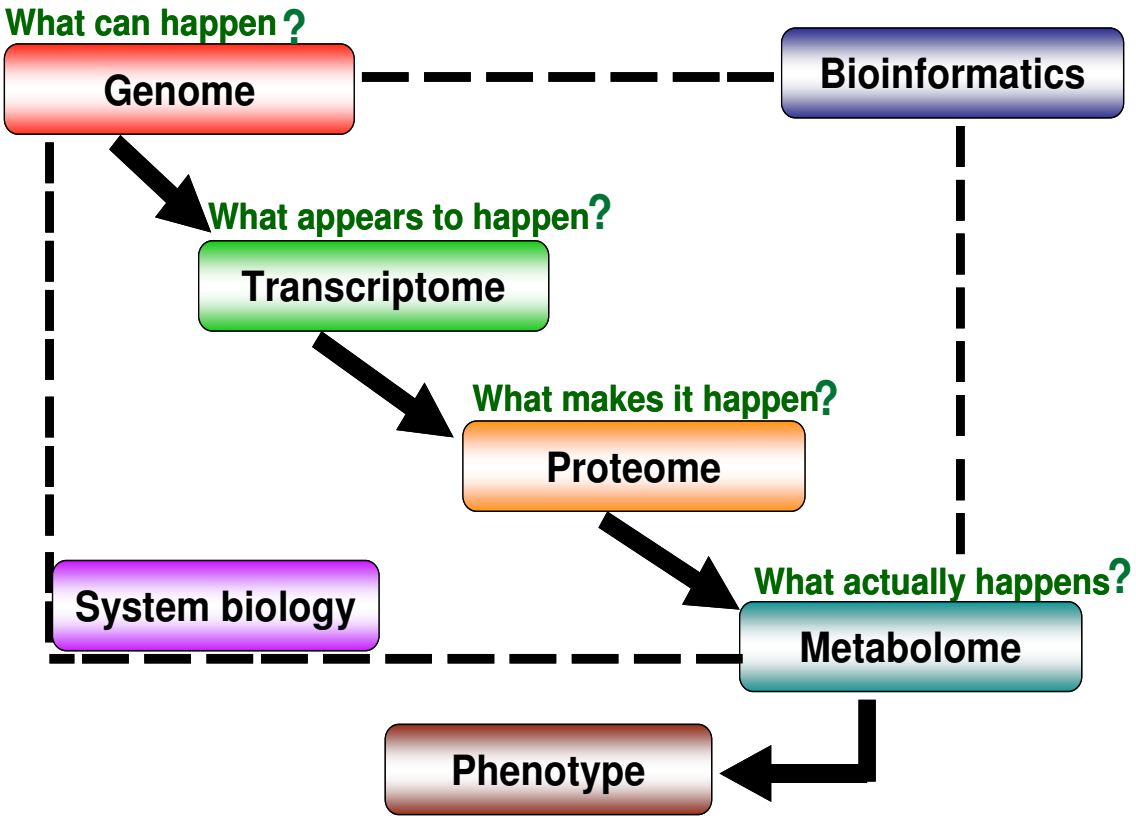


Figure 1-10. Systems biology approaches and the integration of different levels of genome expression

Carrari et al. (2006) reviewed the metabolic regulation underlying tomato fruit development and focused on primary metabolites and secondary metabolites that are important for tomato fruit quality. Moreover, Roessner-Tunali et al. (2003) identified 70 small molecular weight metabolites and catalogued the metabolite composition of developing fruit in three tomato transgenic lines. Fraser et al. (2007) studied the intermediary metabolism variation according to phytoene levels in tomato. Moco et al. (2007) studied the secondary metabolism in the fruit of commercial tomato cultivars using a combined approach of liquid chromatography (LC) and photo-diode array (PDA) detection, fluorescence detection (FD), and mass spectrometry (MS). Related metabolite profiles of peel and flesh were found between several commercial tomato cultivars indicating similar metabolite trends despite the genetic background. They also investigated metabolite profiles of different fruit tissues (vascular attachment region, columella and placenta, epidermis, pericarp, and jelly parenchyma) at five fruit developmental stages in a single tomato cultivar. Unrelated to the chemical nature of the metabolites, behavioral patterns could be assigned to specific ripening stages or tissues. Metabolite profiling has also been used to phenotype introgression lines to identify metabolites QTLs (Schauer et al. 2006). Causse et al. (2002) conducted a complementary approach on broad genetic crosses, and identified QTL for organoleptic properties of tomatoes. Do et al. (2010) mapped QTL for metabolites introgression lines derived from the cross between *Solanum lycopersicum* and the wild species *Solanum chmielewskii* in response to to carbon availability.

Very little information is available for enzymatic activities in tomato fruit. Steinhauser et al. (2010) investigated the activity of 22 enzymes from central metabolism during fruit development in modern tomato cultivar *Solanum lycopersicum* 'M82' and its wild relative *Solanum pennellii* (LA0716). Enzymes showed different development response for the two species. Most enzyme activities decreased during fruit development in *S. lycopersicum* cultivars, they remained high or even increased in *S. pennellii*, particularly enzymes involved in organic acid synthesis. They pursued the analysis by mapping QTL for enzyme activities in the Introgression Lines derived from the cross of the two accessions (Steinhauser et al. 2011).

Studies on Enzymes activitiy have been measured on other species. Gibon et al. (2004) set up a platform for rapid analysis and measured the activity of 23 enzymes that are involved in central carbon and nitrogen metabolism in *Arabidopsis thaliana*. The largest diurnal changes in activity were found for AGPase and nitrate reductase mainly because of posttranslational regulation. The changes of enzyme activity are strongly delayed, with the delay varying from

enzyme to enzyme. It is proposed that enzyme activities provide a quasi-stable integration of regulation at several levels and provide useful data for the characterization and diagnosis of different physiological states.

Response of these enzymes to the carbon and nutrient status and to temperature (Osuna et al. 2007; Usadel et al. 2008; Mounet et al. 2009) and to map enzyme activity QTLs in a Cape Verde Islands × Landsberg erecta *Arabidopsis* recombinant inbred line population (Keurentjes et al. 2008) has also been investigated.

Proteome analysis

Proteomics is a valuable approach for studying the biology of living organisms and their interaction with the environment in post genomic area (Maghuly et al. 2011). The protein level integrates post-transcriptional and translational modifications that modulate the quantity, the localization and the efficiency of the final gene product within the cell. The plasticity of a phenotype is driven by these altered levels of proteins and metabolites (Weckwerth 2008). Recently protein metabolism and especially protein stability was suggested to play a major role in plant growth, yield, and heterosis (Goff 2011). Thus, dissection of complex traits should take into account the analysis of the proteome. This has been hampered by the difficulty and low throughput of this approach. Thanks to the advances in the identification of proteins by mass spectrometry following their separation in two-dimensional electrophoresis and in methods for large-scale analysis of proteome variations, proteomics is becoming an essential methodology in various fields of plant biology.

Knowledge on the fruit proteome is a challenging area of research, as reviewed by Palma et al (2011). Proteome in tomato fruit is poorly documented, partly due to technical issues. While custom-made or commercial arrays are available for transcript profiling and widely used techniques like gas chromatography-mass spectrometry and liquid chromatography-mass spectrometry are available for metabolite profiling, it is still a technical challenge to obtain quantitative information about large numbers of proteins (Rose et al. 2004a; Baerenfaller et al. 2008). Efforts have begun to establish proteomics technical platforms in order to characterize the differences in wall structure and composition that occur during tomato fruit development and ripening (Rose et al. 2004a; Rose et al. 2004b). Samples extraction methods for tomato were also compared by Saravanan and Rose (2004). Faurobert et al. (in press) recently reviewed the tomato proteomic studies. They include: response of the plant to several environmental stresses (Iwahashi and Hosoda 2000; Page et al. 2010; Manaa et al. 2011; Marjanovic et al. 2012) and proteome variation along seed development and germination

(Sheoran et al. 2005) and along fruit development and ripening (Rocco et al. 2006; Faurobert et al. 2007). The improvement of the protein identification technologies and the availability of the complete high quality tomato genome sequence (Sato et al. 2012) increase the efficiency of protein identification after their separation and subsequent mass spectrometry analysis.

Transcriptome analysis

Several studies have been carried out on the transcriptome dynamics during tomato fruit development. Transcriptome profiling via cDNA microarray analysis identified 869 genes that are differentially expressed in developing tomato pericap (Alba et al. 2005). Transcriptional profiles at cell expansion stage of tomato fruit development were also studied by Lemaire-Chamley et al. (2005). They performed an initial sequencing of around 1,200 sequence tags related to cell expansion stage. Results showed that approximately up to 35% of the expressed sequence tags showed no homology with available tomato expressed sequence tags and up to 21% with any known gene. It provides a basis for tissue-specific analyses of gene function in growing tomato fruit. Carrari et al. (2006) studied the transcriptome changes during tomato fruit development. It was apparent that transcript abundance was less strictly coordinated by functional group than metabolite abundance. Transcriptome analysis of ovaries was also performed by Vriezen et al. (2008) using two complementary approaches: cDNA-amplified fragment length polymorphism (CDNA-AFLP) and microarray analysis. Gene expression profiles suggested that, in addition to auxin and gibberellins, ethylene and abscisic acid (ABA) are involved in regulating fruit set.

Recent tomato genome study allowed to cluster the protein coding genes of tomato, potato, Arabidopsis, rice and grape into 23,208 gene groups (+2 members), around which 8,615 are common to all five genomes, 1,727 are confined to eudicots (tomato, potato, grape and Arabidopsis), and 727 are confined to plants with fleshy fruits (tomato, potato and grape). Relative expression of all tomato genes was determined by replicated strand-specific Illumina RNA-Seq of root, leaf, flower (two stages) and fruit (six stages) in *S. Lycopersicon,* in addition to leaf and fruit (three stages) of *S. pimpinellifolium.* Chromosomal organization of genes, transcripts, repeats and small RNAs (sRNAs) were very similar in the two species (Sato et al. 2012).

Analysis of tomato genome structure

Zamir and Tanksley (1988) carried out the first analysis aimed to make deductions about the organization and evolution of the tomato genome. In this study, fifty random clones (350-2300 bp) were obtained from sheared DNA using Southern analysis. This work along with

others (Ganal et al. 1988), provided the first general sketch of the tomato genome at the molecular level and showed that it is comprised largely of single copy sequences (70%) and these sequences, together with repetitive sequences, are evolving at a rate faster than the coding portion of the genome. Peterson et al. (1998) characterized the tomato genome using *in vitro* and *in situ* DNA reassociation. Fluorescence in situ hybridization (FISH) analysis demonstrated that tomato heterochromatin contains substantial quantities of single-copy and middle repetitive DNA, while tomato euchromatin contains little if any highly repetitive sequences. Recently, the inbred tomato cultivar 'Heinz 1706' was sequenced and assembled using a combination of Sanger and next generation sequencing technologies. The genome of *S. pimpinellifolium* LA1589 was also sequenced and assembled *de novo* using Illumina shorts reads, yielding a 739 Mb draft genome. Estimated divergence between the wild and domesticated genome is 0.6% (5.4 million SNP distributed along the chromosome). Compared to the genomes of Arabidopsis and sorghum, tomato has fewer high-copy, full length long terminal repeat (LTR) retrotransposons with older average insertion ages (2.8 versus 0.8 million years ago) and fewer high-frequency k-mers. This is consistent with previous studies (Peterson et al.1998; Zamir and Tanksley, 1988) that revealed that the tomato genome is unusual among angiosperms by being largely comprised of low-copy DNA.

Several re-sequencing projects of tomato genome are today under way, such as 150 tomato genome resequencing project (http://www.tomatogenome.net/). This set of sequenced individuals will provide a first step in tomato research to further develop "genotyping-by-sequencing" approaches, and to study LD, diversity, codon usage bias and recombination events at the sequence level.

Integration of different levels

Integration of -omics data allows scientists to investigate the cell networks as a whole and lead to the direct candidate gene discovery. However, relatively few analyses have combined multilevel approaches. Recent examples for the combined application of metabolic and transcriptional profiling in order to identify candidate genes that modify metabolites content has been carried out by Carrari et al. (2006). Mounet et al. (2009) explored transcriptome and metabolome variation in mesocarp and locular tissue of tomato fruit at cell expansion phase to identify regulatory genes involved in the developmental and metabolic processes that may affect fruit quality. Data was integrated through correlation analysis. This integration identified numerous correlations, common to both tissues, between the two levels and allowed the construction of a large network. Garcia et al. (2009) studied three transgenic tomato lines

for genes related to ascorbic acid pathway and their wild type at phenotypic and transcriptome, proteome and metabolome levels at two stages of fruit development. By using Kohonen's self-organizing maps (SOMs) to cluster the biological data, pair-wise Pearson correlation analyses and simultaneous visualization of transcript/protein and metabolites (MapMan), this approach allowed to uncover major relationships between ascorbic acid and other metabolic pathways. Wang et al. (2009) compared transcriptome and targeted metabolome results and uncovered important features of the molecular events underlying pollination-induced and pollination-independent fruit set. Transcriptomic profiling identified a high number of genes common to both types of fruit set. The combined results allow a far greater comprehension of the regulatory and metabolic events controlling early fruit development both in the presence and absence of pollination/fertilization. Enfissi et al. (2010) characterized the tomato fruit at several omic levels in order to study how modulation of DE-ETIOLATED1 (DET1) gene may be used to improve quality traits. Metabolite profiling revealed quantitative increase in carotenoid, tocopherol, phenylpropanoids, flavonoids, and anthocyanidins. Parallel metabolomic and transcriptomic analyses reveal the widespread effects of DET1 down-regulation on diverse sectors of metabolism and sites of synthesis. Correlation analysis of transcripts and metabolites independently indicated strong co-response within and between related pathways/processes. Steinhauser et al. (2010) analysed correlation between enzyme activities, metabolites and transcripts in *S. lycopersicum* and showed little connectivity between the developmental changes of transcripts and enzymes and even less between enzymes and metabolites. Osorio et al. (2011) investigated systems of nonripening (NOR), ripening inhibitor (RIN) and the ethylene receptor Never-ripe (Nr) tomato mutations at the transcriptomic, proteomic, and metabolomic levels during development and ripening. The results showed that changes of the content of metabolites from primary metabolism lead to decreases in metabolic activity during ripening. Integration of metabolomic, transcriptomic and proteomic data identified several aspects of the regulation of metabolism during ripening. The expression of transcripts were not frequently correlated with the abundance of corresponding proteins during early ripening, indicating that post-transcriptional regulatory mechanisms play an important role in these stages. Strong correlations between ripening-associated transcripts and specific metabolite groups, such as organic acids, sugars, and cell wall-related metabolites, underlined the importance of these metabolic pathways during fruit ripening. In conclusion, the system biology approaches integrating 'omic' resources and technologies are revolutionizing biology research and offer new strategies for a better understanding of the molecular and genetic bases of complex traits.

1.3 Objectives and content of this study

Molecular markers are precious tools to improve crops through marker-assisted selection. They allowed the characterisation of molecular diversity and relatedness in genetic resources as well as mapping of QTL in a wide range of plant species. In tomato, either passionate people or research institutes conserve large collections of accessions around the world. In Avignon (South of France), the INRA Research Institute characterizes and maintains more than 1000 heirloom varieties, which have been partly characterized at the molecular level (Ranc et al. 2008). For other collections, few examples report their characterization at the molecular level (Alvarez et al. 2001; Park et al. 2004; Van Deynze et al. 2007; Manoj and Uday 2010). Molecular characterization studies have shown that the rate of polymorphism in tomato is low in cultivated accessions. Domestication and modern breeding have strongly reduced tomato molecular diversity, limiting the possibility of improvement of some traits such as yield or fruit quality. It is thus necessary to come back to wild species and old germplasm and to use them as sources of new loci or new alleles. If RFLP and SSR allowed polymorphism to be revealed in wild species, the old cultivars or cherry tomato accessions were too close to the cultivated accessions to be easily studied with these markers. For this purpose, new molecular markers such as SNP constitute precious tools to saturate the genetic maps and identify QTL and associations in poorly polymorphic species like tomato.
The objectives of this study were to characterize tomato genetic diversity at the molecular and proteome levels and to try to identify QTLs, proteins responsible for fruit quality traits in tomato. For this purpose, three independent studies were conducted leading to the discovery of new SNP markers, their use for association studies and finally the analysis of proteome diversity in relation to phenotypes.

1. **Exploration and comparison of two Next Generation Sequencing platforms to identify polymorphisms in tomato**

One of the challenges in genetics is to deepen our knowledge in molecular evolution and to improve association genetics to use it as a mapping strategy. By combining genome re-sequencing and NGS, the high volume of polymorphisms identified between natural genetic resources is promising in such aims. In order to go further in the characterization of tomato germplasm, it is necessary to identify SNPs extensively. In order to test whether NGS could be appropriate to generate SNPs in tomato, we used two high throughput sequencing

platforms (Genome Analyser and Roche 454) to sequence 198 targeted 10kb genomic regions obtained by Long-Range PCR in two contrasted cultivated varieties (Heinz1706 and LA2675). More than 3000 SNPs were detected, but many differences were observed among the results of the different platforms and analysis software. To validate some of the SNPs, we developed CAPS markers that have been used to generate a map in a F2 segregating population.

2. Phenotypic diversity and association mapping in cultivated tomato and its wild related species

In a previous pilot study focused on one chromosome and 90 tomato accessions (Ranc et al. 2012), we showed that association mapping was possible in tomato (Annexe I). Technologies evolve fast and in the same time, high-throughput genotyping technologies such as SNPlex™ technology has emerged. We thus attempted to develop a SNPlex™ assay (De La Vega et al. 2005; Tobler et al. 2005) of 192 SNPs selected from re-sequencing experiments or from databases (Van Deynze et al. 2007). A large germplasm collection including cultivated, cherry type and wild accessions were characterized for both genetic diversity using the SNPlex™ assay, 20 SSR markers (Ranc et al. 2008) and ten phenotypic traits related to fruit quality. We first describe the phenotypic diversity of the accessions, then compare the genetic structure of the collection based on SSR and SNP markers and finally present the association mapping results. Associations are compared to previously mapped QTL for similar traits. Our work is the first example of an association study carried out using a broad sample of cultivated, cherry type and wild tomato accessions.

3. Systems biology approach focusing on proteomics and metabolomics for the characterization of fruit quality traits in tomato

DNA sequence variation (SNP or Indel) may not affect the traits directly. There are several intermediate levels between DNA genotypes and the phenotypes. The cascade of effects from DNA variation to phenotype is organized in complicated biological networks (Kliebenstein, 2010; Sulpice et al., 2010). Intermediate molecular phenotypes such as transcript and protein abundance also genetically vary in populations and are themselves quantitative traits (Rockman and Kruglyak, 2006). High-throughput approaches have opened new prospects for analyzing biological systems and their complex functions at different levels including genomic, transcriptomic, proteomic, and metabolomic levels. Systems biology approaches

integrating several 'omic' levels offer new strategies for discovering links between co-regulated genes and pathways and ultimately, for predicting gene function and identifying regulatory genes in plants (Saito et al. 2008). It should enable us to understand the biology inside the black-box of quantitative genetics relating genotype to phenotype in terms of causal networks of interacting genes. System approaches have been applied in yeast (Ideker et al., 2001), in the model plant Arabidopsis (Hirai et al., 2007) and in tomato (Mounet et al., 2009), at several levels.

We applied a systems biology approach focused on proteome, metabolome and phenotypic analysis to characterize eight contrasted lines as well as four of their hybrids at two stages, cell expansion and orange red stage in order to identify proteins responsible for fruit quality traits. This study is the starting point of a broader experiment including the development of a MAGIC population derived from the eight lines. The eight lines and hybrids were characterized at several levels from phenotype to metabolites (Xu et al. 2012, Acta Horticulturae; Annexe II). Due to time consideration, I will detail mostly the proteome and metabolome analysis.

The second chapter will briefly present the materials and methods used in each part. Then results will be presented in the three next chapters presented as articles, manuscripts and additional data. Then prospects are presented before a general conclusion.

CHAPTER II: MATERIALS AND METHODS

In this chapter, we summarize the materials and methods used in this work. Detailed procedures are presented in each chapter

CHAPTER II: MATERIALS AND METHODS

2.1 Plant materials

Three sets of plant materials were used in this thesis

Exploration and comparison of two Next Generation Sequencing platforms to identify polymorphisms in tomato

Two contrasted accessions Heinz1706 and LA2675 were selected for re-sequencing. Heinz1706 (*S. lycopersicum* var. *esculentum*) is a processing tomato accession with large and enlongated fruit (**Figure 2-1**). It has been used for the international tomato sequencing project (Sato et al. 2012). LA2675 (*S. lycopersicum* var. *cerasiforme*) is a cherry tomato accession producing small fruits. The F2 segregating population which consisted of 96 individuals was derived from the cross between the two accessions and used for mapping all the developed markers. A set of 23 accessions including these two accessions were chosen from genetic resources maintained at INRA Avignon (France) for marker validation and polymorphism identification.

Phenotypic diversity and association mapping in tomato and related wild species

A total of 188 tomato accessions were selected from a germplasm collection maintained and characterized at INRA Avignon (France) to perform association mapping. The sample consisted of 127 cherry type tomato accessions (*Solanum lycopersicum* var. *cerasiforme,* hereafter named *S. l. cerasiforme*), 44 large fruited accessions (*S. lycopersicum* var. *esculentum*, hereafter named *S. lycopersicum*) and 17 *S. pimpinellifolium* accessions (**Figure 2-2**). Accessions were obtained from various sources, among which the Tomato Genetics Resource Center (Davis, USA), the Centre for Genetic Resources (Wageningen, Netherlands), the North Central Regional Plant Introduction Station (Ames, Iowa, USA) and the N.I. Vavilov Research Institute of Plant Industry (St Petersburg, Russia).

Systems biology approaches focusing on proteomics and metabolomics for the characterization of fruit quality traits in tomato

In the third experiment, eight contrasted lines and four of their corresponding hybrids were used. The eight genotypes represent the largest possible range of tomato genetic diversity in a

Figure 2-1 Fruits from the two contrasted accessions Heinz1706 and LA2675 used in this experiment.

Figure 2-2 Fruits collected from the 188 accessions showing their phenotypic diversity

collection of 360 accessions (Ranc et al. 2008). They include four *Solanum lycopersicum* lines (Levovil, Stupicke Polni Rane, LA0147 and Ferum) and four *S. l. cerasiforme* lines (Cervil, Criollo, Plovdiv24A, and LA1420). Cervil produces very small fruits (less than 10 g). Levovil, Ferum and LA0147 genotypes have large fruits. The others are intermediate tomato (**Figure 2-3**).

2.2 Methods

2.2.1 Methods of the polymorphism identification and validation

The strategy for polymorphism identification and validation is presented in **Figure 2-4**. It is detailed in the following paragraphs.

gDNA isolation

Genomic DNA of the two contrasted accessions Heinz1706 and LA2675 was extracted from leaf tissue of 3-week old plants according to the protocol of Fulton et al. (1995).

For DNA quantification, Long Range PCR amplification, PCR products pooling, sequencing methods of GAII and Roche 454, sequence analysis, validation of SNPs, CAPS marker development and genetic map construction, see materials and methods in Chapter III.

2.2.2 Methods of the association analysis

Association mapping was performed on the whole collection of 188 accessions and on the set of 127 *S.l. cerasiforme* accessions, respectively. **Figure 2-5** shows the strategy used in this study. Details are described in the following sections.

Phenotyping

All 188 tomato accessions (4 plants per accession) were grown in a plastic greenhouse in Avignon (south of France) during summers 2007 and 2008. Three harvests of ten ripe fruits per accession were used as repeats in the phenotypic analysis. Ten fruits were first evaluated for fruit weight (FW), firmness (FIR), locule number (LCN), color components: lightness (L), color on red to green (a*) and on yellow to blue (b*). For FIR, it was recorded at two points on opposite equatorial for each fruit with a durometer Durofel (http://www.setop.fr/). This portable device measures the pressure exerted by a piston on a surface with an arbitrary scale (0 indicates that the piston is fully extended and 100 indicates maximum strength).

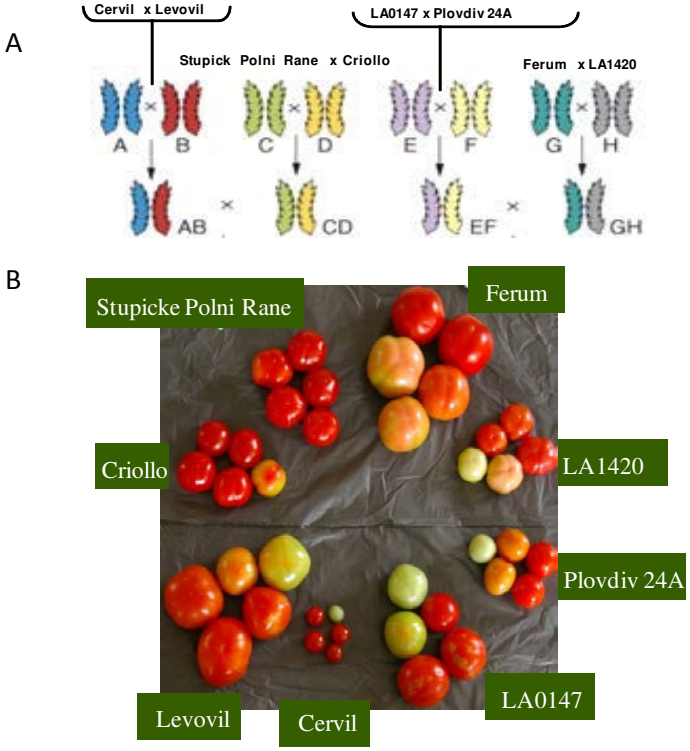


Figure 2-3 Eight divergent lines and four F_1 used in this study.

A: Four F_1 derived from the crosses of the eight lines; B: fruits from the eight lines.

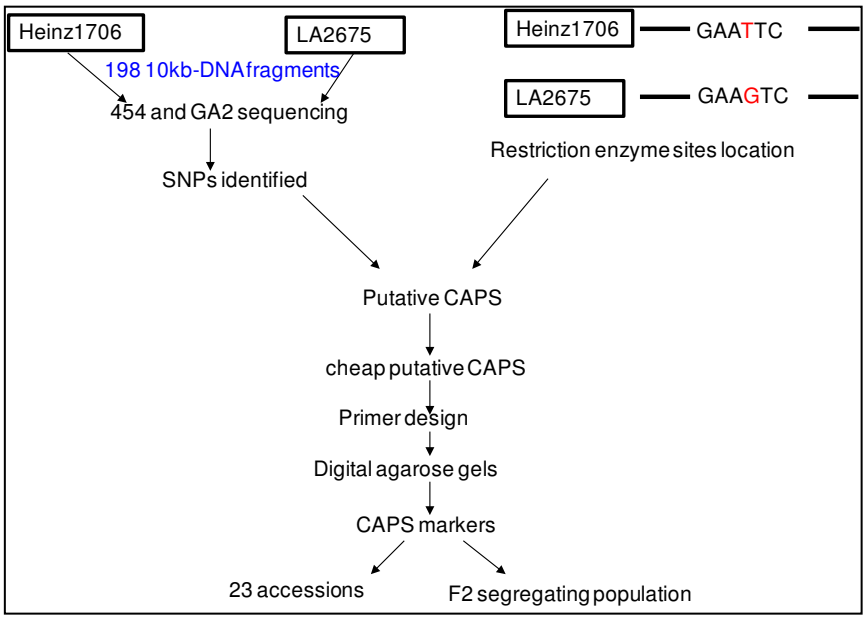


Figure 2-4 Strategy for polymorphism identification and validation. SNPs were detected using both 454 and GA2 technologies. CAPS markers were developed to validate identified SNP and used for diversity analysis of a collection of 23 accessions (Ranc et al. 2008) and for mapping the F2 population derived from the cross of the two contrasted accessions Heinz1706 and LA2675.

Color components were measured by taken at two different points on each fruit with a Konica Minolta CR300 colorimeter. The fruits were then cut transversely in order to count the locule number and were then grounded to obtain a powder. Powders were stored at -20 ° C and were then used to measure soluble solid content (SSC), sugar content (SUG), pH, titratable acidity (TA). SSC was measured using a digital refractometer (Palette PR 101). TA was measured by the amount of NaOH (in mmol / L) necessary to bring the pH of a solution to 8.1 (turning area of the phenolphthalein). The measurements were made using a titrator equipped with an autosampler (Crinson compact titrator). A mass m_p pulp is weighed, diluted in 50 mL of water and NaOH is added to pH threshold. The titratable acidity is then calculated as follows: $C = \left[\left(C_{NaOH} \times V_{NaOH}\right) \times 100\right] / m_p$. The sugar content was quantified by enzymatic assay in 96-well microplates, as detailed in Gomez et al. (2007).

The 274 additional accessions (4 plants per accession) were grown in the same greenhouse conditions during 2009. Ten fruits were evaluated for fruit weight (FW), locule number (LCN), and Soluble Solid Content (SSC).

Statistical analysis

The heritabilities were estimated on the collection of 188 homozygous lines. Heritabilities (h^2) were calculated as $h^2=\sigma^2_g/ (\sigma^2_g+ \sigma^2_{gy}+ \sigma^2_e)$ with σ^2_g, σ^2_{gy} and σ^2_e the genetic, genetic by environment interaction and residual variance, respectively. σ^2_g, σ^2_{gy} and σ^2_e were estimated by (MSc-MScy)/ry, (MScy-MSe)/r and MSe, respectively. MSc represents the accessions mean square, MScy represents the mean square of genotype by year interaction, MSe the residual mean square. r and y represents the number of replicates and the number of years. Associations were tested using the adjusted means of accessions calculated by general linear model. The Pearson correlation coefficients were calculated for all pairs of variables. Analyses were carried out with the R program (R Development Core Team 2005).

SNPlex^{TM} assay design

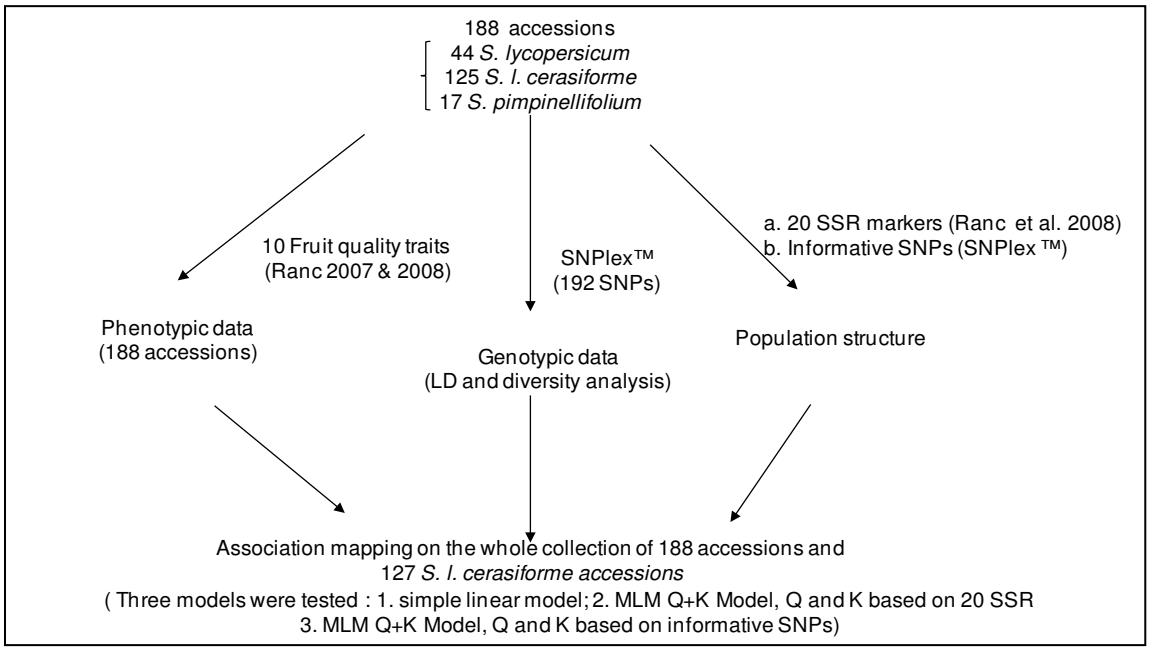


Figure 2-5 Strategy for association mapping studies

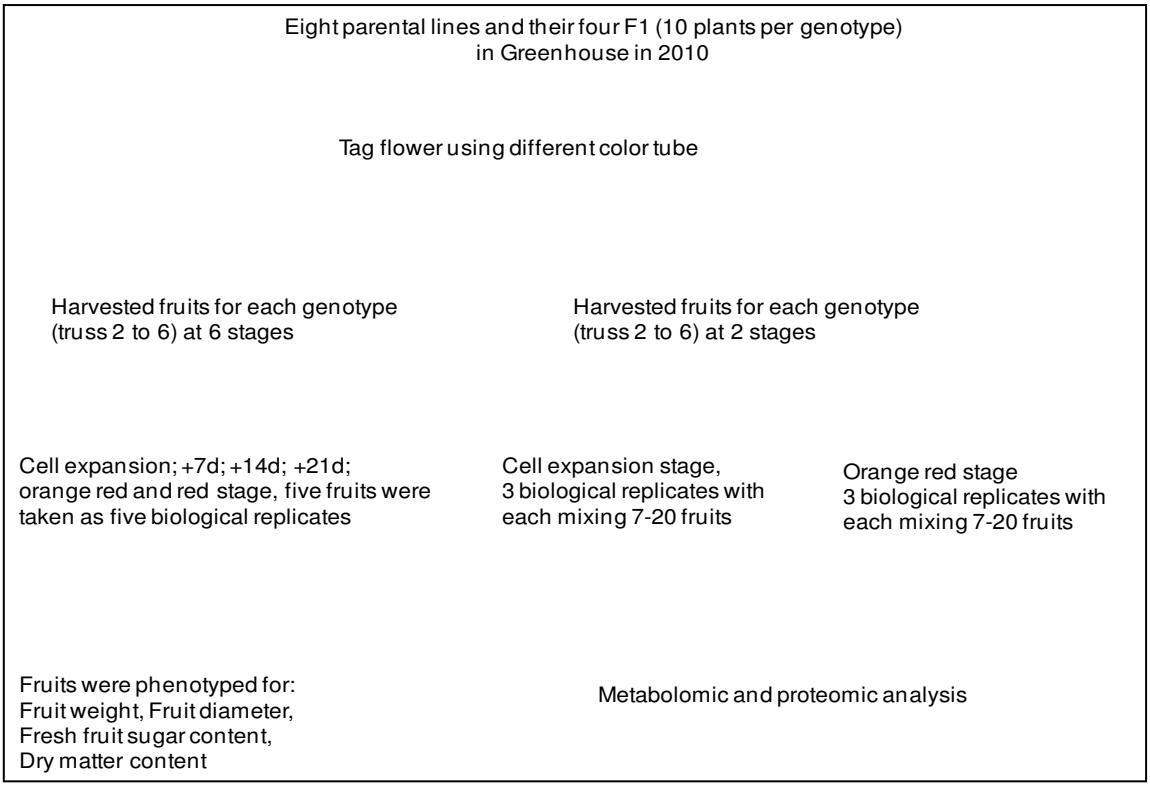


Figure 2-6 Sampling strategy for phenotypic and omic analysis

Allele-specific probes and optimized multiplexed assays using SNPs of interest were designed by an automated multi-step pipeline [Applied Biosystems]. These steps include: (1) entering the sequence containing target SNPs; (2) checking for formatting errors such as non-target polymorphisms near the target SNP or sequence motifs incompatible with the assay; (3) submitting the SNPs that passed the format check for the assay design. The ABI probe design prevents self-complementarity and dimerization, and annealing efficiencies are optimized for ligation. Furthermore, the optimal combination of SNPs to produce the highest yield per multiplex reaction is determined (De La Vega et al. 2005; Tobler et al. 2005).

Association Mapping

Association mapping was performed using the Mixed linear model (Yu et al. 2006) (based on 20 SSR and 121 informative SNPs) and simple linear model on the two sets of samples (188 accessions and 127 *S.l. cerasiforme* accessions). Polymorphisms of the whole collection and the three species were analyzed using 121 informative SNPs. LD was then deduced on the whole genome level and on chromosome 2 for the two samples. Structure of the 188 accessions was compared using the two types of markers. Methods are detailed in Chapter IV.

2.2.3 Methods of the integration of system biology approaches

Sampling strategy for phenotypic and omic analysis is illustrated in **Figure 2-6.** It is detailed as follows.

Phenotypic traits analysis

For phenotypic trait analysis, five fruits were harvested from ten plants of each genotype at the following six stages: cell expansion stage (25, 20 and 14 days after anthesis for large, medium and small fruited tomato, respectively), +7d, +14d, +21d, orange red, then until red ripen stage. They were evaluated for fruit fresh weight, fruit diameter, dry matter content in fresh weight, sugar content. Fruits were first weighted. Fruit diameter was then measured using a caliper. Dry matter content (expressed in g / 100 g FW) was assessed after 5 d in a ventilated oven at 80 ˚C. Sugars were extracted from the pericarp according to the method described in Gomez et al. (2002). The main sugar contents (glucose, fructose and sucrose) were quantified by enzymatic assay in 96-well micro-plates, as detailed in Gomez et al. (2007).

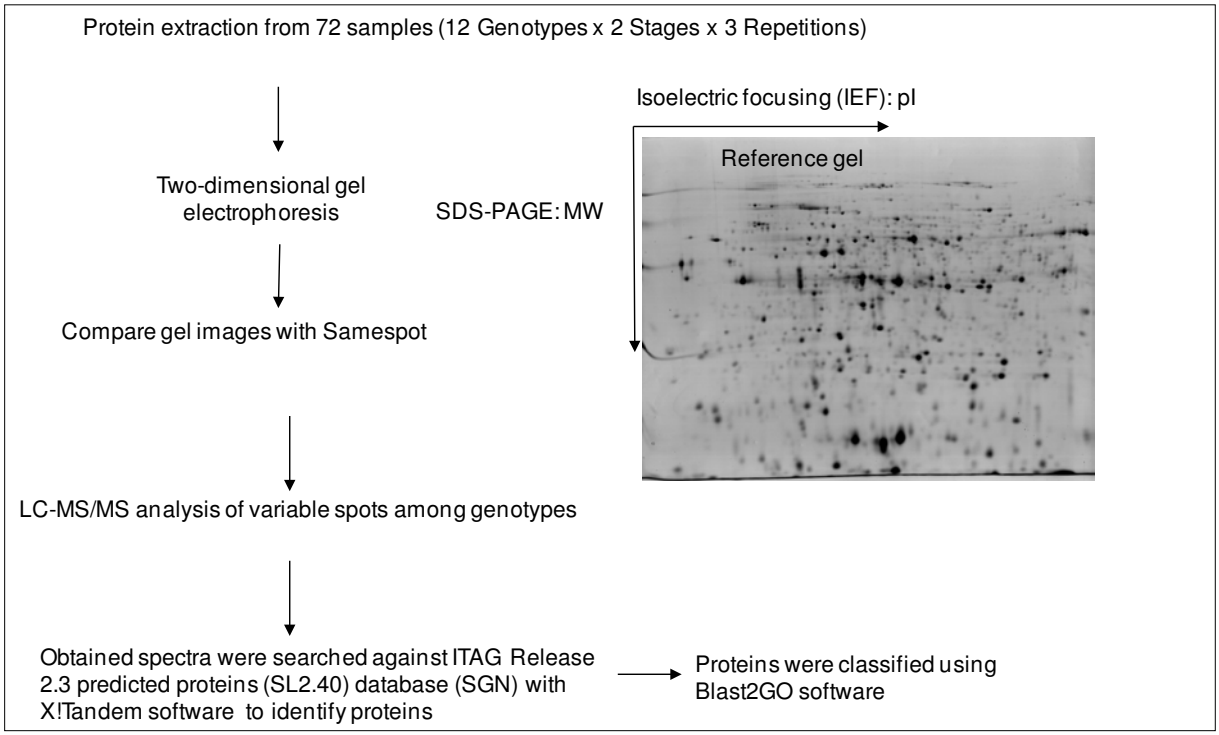


Figure 2-7 General schemes for proteomic analysis

Sugar contents were expressed relative to the pericarp fresh weight (g / 100 g FW) or to the pericarp dry weight (g/ 100 g DW). For proteomic and metabolomic analyses, fruits were collected at two stages of development, cell expansion stage and orange-red stage, according to the fruit colour. For each genotype and stage, three pools of 7 to 20 fruits were made by mixing fruits from truss 2 to 6 of 10 plants, but avoiding the first and last fruit of the truss. Pericarps were collected from each pool, immediately frozen, ground in liquid nitrogen and stored at -80 °C before analysis. Fruit weight curves were fitted to three-parameter sigmoid logistic function.

Proteome analysis

General strategy for proteomic analysis as presented in **Figure 2-7**.

Protein extraction

Proteins were extracted using the phenol extraction method developed by Faurobert et al. (2007). Plant powder was suspended in 3 volumes of extraction buffer (containing 700 mM saccharose, 500 mM Tris, pH 8, 100 mM KCl, 2% [v/v] b-mercaptoethanol, 2 mM phenylmethylsulfonyl fluoride, pH 8.5) and incubated for 10 min on ice. Afterward, an equal volume of Tris-saturated phenol was added. Samples were shaken for 10 min at room temperature and then centrifuged (10 min, 5,525g, 4˚C) to separate the phenolic and the

aqueous phase. The phenolic phase was recovered and re-extracted with the same volume of extraction buffer. Subsequently, centrifugation was repeated and 5 volumes of precipitation solution (0.1 M ammonium acetate in methanol) were added to the recovered phenol phase. Proteins were precipitate at -20˚C overnight. After centrifugation for 10 min (5,525g, 4˚C), the protein pellet was washed three times with the precipitation solution and once with acetone. Each washing step was followed by 5 min of centrifugation as described above. After drying under vacuum, the pellet was resuspended in lysis buffer (9 M urea, 4% [w/v] CHAPS, 0.5% [v/v] Triton X-100, 20 mM dithiothreitol [DTT], 1.2% [v/v] pharmalyte, pH 3–10).

Two-dimensional gel electrophoresis

Protein concentration was measured according to a modified Bradford assay (Ramagli and Rodriguez 1985) in order to load 500 µg of proteins on 24 cm long Immobiline dry strips, pH 4 to 7 (Amersham Bioscience, Uppsala, Sweden). Proteins were first separated according to

their charge after passive rehydration of 24-cm-long Immobiline dry strips, pH 4 to 7 (Amersham Biosciences), with 100 mg of resuspended proteins, 9 mL immobilized pH gradient buffer, pH 4 to 7, and rehydration buffer (8 M urea, 2% [w/v] CHAPS, 0.3% [w/v] DTT, 2% [v/v] pharmalyte, pH 3–10), to a final volume of 450 mL.). Isoelectric focusing was performed with the Multiphor II (Amersham Bioscience, Uppsala, Sweden) according to the following program: 2 h at 150 V, 2 h at 400 V, 2 h to increase the voltage from 400 to 3,500 V, 18 h at 3,500 V. After migration, isoelectric focusing strips were stored at -80 ˚C or immediately incubated in equilibration buffer (6 M urea, 50 mM Tris-HCl, pH 8.8, 30% glycerol, 2% [w/v] SDS with addition of 2% [w/v] DTT in the first equilibration step and 2.5% [w/v] iodoacetamide in the second equilibration step, respectively), for 20 min. SDS-PAGE was carried out with 12.5% acrylamide gels in the Bio-Rad Protean Plus Dodeca cell electrophoresis chamber (45 min at 80 V, 15 h at 120 V).

Gels were stained with Coomassie colloidal blue and gel images were analysed using Progenesis SameSpots v3.0 software (Nonlinear Dynamics Ltd). Spot volumes were normalized according to the total spot volumes per gel to avoid experimental variations among 2-D gels. Detected spots were manually corrected after automatic detection by SameSpots v3.0 software. For statistical analysis, Samespots software was used to detect varying spots using one way ANOVA on normalized spot volume from the three biological pools. A two-way ANOVA then was performed to detect genotype, stage, and interaction effects on the deduced set of spots, a P-value less than 0.05 was considered to be statistically significant. Protein spots which showed significant values in ANOVA were selected and sequenced by nano LC-MS/MS.

Protein identification

This step was performed at the proteome platform of Le Moulon (Gif-sur-Yvette). In-gel digestion was performed with the Progest system (Genomic Solution) according to a standard trypsin protocol. Gel pieces were washed twice by successive separate baths of 10% acetic acid, 40% ethanol, and acetonitrile (ACN). They were then washed twice with successive baths of 25 mM NH4CO3 and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH4CO3. The peptides were extracted successively with 2% trifluoroacetic acid (TFA) and 50% ACN

and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 µL of 0.05% TFA, 0.05% HCOOH, and 2% ACN.

HPLC was performed on an NanoLC-Ultra system (Eksigent). A 4 µL sample was loaded at 7.5 µL/min^{-1} on a precolumn cartridge (stationary phase: C18 Biosphere, 5 µm; column: 100 µm i.d., 2 cm; Nanoseparations) and desalted with 0.1% HCOOH. After 3 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 Biosphere, 3 µm; column: 75 µm i.d., 150 mm; Nanoseparations). Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 11 min at 300 nL/min^{-1}. Including the regeneration step at 95% B and the equilibration step at 95% A, one run took 25 min.

Eluted peptides were analysed on-line with a LTQ XL ion trap (Thermo Electron) using a nanoelectrospray interface. Ionization (1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10 µm i.d.; New Objective). Peptide ions were analysed using Xcalibur 2.07 with the following data-dependent acquisition steps: (1) full MS scan (mass-to-charge ratio (m/z) 300 to 1400, centroid mode) and (2) MS/MS (qz = 0.25, activation time = 30 ms, and collision energy = 35%; centroid mode). Steps 2 was repeated for the three major ions detected in step 1. Dynamic exclusion was set to 30 s.

A database search was performed with XTandem (version 2010.12.01.1) (http://www.thegpm.org/TANDEM/). Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Cys carboxyamidomethylation and Met oxidation were set to static and possible modifications, respectively. Precursor mass and fragment mass tolerance were 2.0 and 0.5, respectively. A refinement search was added with similar parameters except that semi-trypsic peptide and possible N-ter proteins acetylation were searched. The International Tomato Annotation Group (ITAG) database (http://solgenomics.net/, version 2.3) and a contaminant database (trypsin, keratins, ...) were used. Only peptides with a E-value lower than 0.1 were reported.

Identified proteins were filtered and grouped using XTandem Pipeline (http://pappso.inra.fr/bioinfo/xtandempipeline/) according to : (1) A minimum of two different peptides was required with a E value smaller than 0.03, (2) a protein E value (calculated as the product of unique peptide E values) smaller than 10-3 . In case of identification with only two

or three MS/MS spectra, similarity between the experimental and the theoretical MS/MS spectra was visually checked. To take redundancy into account, proteins with at least one peptide in common were grouped. This allowed to group proteins of similar function. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as sub-groups. Methods of protein classification were detailed in Chapter V.

Measurement of metabolites and enzyme activities

Metabolome analyses were performed on the metabolome platform of Bordeaux. Primary metabolites were quantified using quantitative 1H-NMR profiling of polar extracts, as described in Deborde (2009) with minor modifications, on an Avance III 500 MHz spectrometer equipped with an ATMA inverse 5 mm probe. Enzymatic activity profiling was performed as described in Steinhauser et al. (2010).

Proteome, metabolome statistical analysis

Coefficient of variation (CV= standard deviation/mean) were calculated for each phenotypic trait, metabolite, enzyme, protein spot per genotype at each stage. Then variation (ANOVA), inheritance, principal component, correlation and network analysis were performed for all the traits. For detail, see materials and methods in Chapter VI.

CHAPTER III: COMBINED LONG RANGE PCR AND NEXT-GENERATION SEQUENCING FOR THE IDENTIFICATION OF POLYMORPHISMS BETWEEN TWO TOMATO LINES

This chapter, in the form of a manuscript to be submitted to journal of BMC plant biology, presents the comparison of two sequencing technologies Illumina GA2X and Roche 454 pyrosequencing towards the identification of SNPs in two contrasted tomato lines Heinz1706 and LA2675. Both 454 and GA2X gave similar results, when compared with similar depth of coverage. As a result, the yield of sequences obtained from GA2X sequencers is now much higher than Roche 454 and a high quality genome sequence is available in tomato, this technology seems more appropriate to SNP calling within tomato genome. More than 3000 SNP were identified between the two lines. A subset 64 of these SNPs was validated by developing CAPS markers

Combined Long Range PCR and Next-Generation Sequencing for the identification of polymorphisms between two tomato lines

Stéphane Muños (1), Jiaxin Xu (1,2), Maria Tchoumakov (3), Caroline Callot (1), Sophie Rolland (1), Stéphane Schlub (3), Marie-Christine Le Paslier (3), Jean-Paul Bouchet (1), Dominique Brunel (3), Mathilde Causse (1)

(1) INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, Avignon, 84143, France

(2) Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, P.R.China

(3) INRA, UR1279, Unité Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, Evry, 91057, France

Corresponding author
Mathilde Causse
Phone: 33 04 32 72 27 01
Fax: 33 04 32 72 27 02
Email: Mathilde.Causse@avignon.inra.fr

Abstract:

Next-Generation Sequencing (NGS) technologies have revolutionized genomics and genetics by increasing the throughput by an order higher than 100× compared to the classical sequencing techniques (i.e. Sanger) and simultaneously reducing the cost. In order to further characterize tomato germplasm and test whether NGS could be appropriate to generate Single Nucleotide Polymorphisms markers in tomato, we have used GA2X Illumina and 454 Roche sequencing platforms to re-sequence targeted sequences covering about 0.2% of the tomato genome from two contrasted accessions, Heinz1706 and LA2675. We used long-range PCR to reduce genome complexity and amplify a total of 188 10kb-targeted regions spread on the whole tomato genome. We compared two software abilities to map the reads onto the reference genome and search for polymorphisms. We obtained similar results using both GA2X and 454. Discrepancies in the percentage of unmapped reads and in the number of polymorphisms detected were observed between both softwares. Moreover, almost 3000 single nucleotide polymorphisms (SNP) were identified between Heinz1706 and LA2675. The polymorphism rate was highly variable among fragments and chromosomes. Finally, more than 300 SNPs were associated to coding regions from which a subset 64 of these SNPs was validated by developing CAPS markers. Markers were then used to genotype a set of 23 highly diverse tomato accessions and mapped on the F2 segregating population derived from the two sequenced lines. All markers were polymorphic among the 23 accessions. A total of 54 markers were successfully mapped onto the F2 populations.

Keywords: Next-Generation Sequencing; Polymorphism; CAPS marker; tomato

Introduction:

The Sanger sequencing method (Sanger et al. 1977) has been extensively used for DNA sequencing. Firstly used with manually loaded vertical acrylamid gels, it has been combined with improved instruments such as capillary sequencers, which increased sequencing throughput. The successive technical improvements and the increase in the volume of data produced have made it possible to sequence complex genomes like human genome (Venter 2001). In plants, Arabidopsis genome sequence was first published in 2000, after several years of labour from an international consortium (The Arabidopsis Genome Initiative 2000) [3]. It was followed by the rice genome (International Rice Genome Sequencing 2005). Sanger sequencing is still considered as a standard because it produces the longest read length and the most accurate results (Bonetta 2006). However, Next-Generation Sequencing (NGS) techniques now produce much more data at a much lower price. Their throughput is continuously increasing since a few years. Combined together, these arguments have rapidly raised an enthusiasm for these methods in the scientific communities (Edwards et al. 2012). The sequencing machines Roche/454 (Margulies et al. 2005) and Illumina (Genome Analyzer, GA (Bentley et al. 2008) are among the most frequently used and are complementary in term of read length and read number per run (Pareek et al. 2011). They are used for genome sequencing (*de novo* or targeted re-sequencing), for qualitative or quantitative transcriptome analysis, the so-called 'RNA-seq' (Martin and Wang 2011) and for ChIP-sequencing (Chromatin Immuno Precipitation), which allows the identification of sequences recognized by DNA binding proteins (Schmidt et al. 2009).

One of the challenges in genetics is to deepen our knowledge in molecular evolution and to improve association genetics to identify the loci controlling the variation of traits of agronomical interest. By combining genome re-sequencing using NGS, the large number of polymorphisms identified between natural genetic resources is promising for such aims. The molecular characterisation of segregating progenies has allowed the identification of several genes or QTL in many species (Flint and Mott 2001; Alonso-Blanco et al. 2009). However, domestication and modern breeding have reduced molecular diversity, particularly in self-pollinated crops, like tomato (*Solanum lycopersicum*). It is thus necessary to use wild species and distant germplasm, through hybridization, as a source of new alleles to improve adaptation traits and yield.

In tomato, either passionate people or research institutes conserve large collections of accessions around the world. In Avignon (South of France), the INRA Research Institute characterizes and maintains more than 2000 accessions, which have been partly characterized at the molecular level (Ranc et al. 2008) using SSR markers. For other collections, few examples report their characterization at the molecular level (Alvarez et al. 2001; Nuez et al. 2004; van Deynze et al. 2007). Molecular characterization studies have shown that the rate of polymorphism in tomato is low, ranging from 1.66 polymorphism per kb in coding regions to 4.27 in non coding regions in cultivated accessions (Ranc et al. 2012). These rates are two to three times higher in the cherry tomato accessions and the *S. pimpinellifolium* wild relative. In order to further characterize tomato germplasm and test whether NGS could be appropriate to generate Single Nucleotide Polymorphisms (SNPs) markers in tomato, we used GA2X and 454 sequencing platforms to sequence targeted genomic regions being amplified by Long-Range PCR (LR-PCR) in two contrasted cultivated varieties (Heinz1706 and LA2675) which have long genetic distance. The obtained data have been used to compare the sequences from the two platforms and two analysis software packages. We further validated a subset of the identified SNPs by developing CAPS markers, which were mapped in a segregating population.

Material and Methods

Plant material

Heinz1706 (*S. lycopersicum* var. *esculentum*) and LA2675 (*S. lycopersicum* var. *cerasiforme*) are two contrasted lines used for re-sequencing. Heinz1706 is a processing tomato inbred which has been used to generate the high quality sequence of the tomato genome (Sato et al. 2012). LA2675 is a tomato accession producing cherry type fruits, prospected in Peru by Charles Rick and colleagues in 1985 (http://tgrc.ucdavis.edu). The F2 mapping populations derived from the two lines composed of 96 individuals were used to map developed markers.

A core collection of 23 accessions including the two lines above consisted of 10 large fruited accessions (*S. lycopersicum* var. *esculentum*), 7 cherry type tomato accessions (*S. lycopersicum* var. *cerasiforme*), 3 *S. pimpinellifolium* accessions, one *S. galapagense*, one *S. habrochaites* and one *S. pennellii* accessions were used for diversity analysis. They were obtained from the Tomato Genetics Resource Center (Davis, USA), the North Central Regional Plant Introduction Station (Ames, Iowa, USA), from the N.I. Vavilov Research Institute of Plant Industry (St Petersburg, Russia) and from Institut National de Recherche Agronomique (INRA, Avignon, France). **Supplemental table S1** presents the list of the 23 accessions. The three sets of plants were grown for 3 weeks before sampling of the young leaves for genomic DNA isolation.

DNA isolation and quantification

Genomic DNA has been isolated from 2 grams of fresh leaves, which have been crushed in a mortar containing a pinch of sand and 5 ml of extraction buffer (0.35M D-Sorbitol, 0.1M Tris, 5mM EDTA, 2M NaCl and 70mM $Na_2S_2O_5$). After transfer to a 50ml tube, 5ml of lysis buffer (200mM Tris, 50mM EDTA, 2M NaCl and 2% CTAB) and 2 ml of a 5% N lauryl sarcosine solution were added and the solution was homogenised. The solutions were incubated 20 min. at 65°C before adding 15 ml of a 24:1 (v:v) choloroform/isoamil alcohol solution. After homogenization, the tubes were centrifuged during 25 min. at 3,600rpm. The upper aqueous phase was transferred in a new tube where 10 ml of 99% ethanol was added. By slowly inverting the tubes, centrifuged, dried and re-suspended, a pellet appeared and was sampled using a hook. The pellet was then washed in a 75% ethanol-0.2M NaAc solution. The pellet was partially dried and dissolved in 500µl of water. gDNA and PCR product concentration

was assessed using Quant-iT Picogreen dsDNA kit (Invitrogen) according to the manufacturer protocol and using a EnVision microplate reader (PerkinElmer).

Long Range PCR amplification

We first tested seven different Taq polymerases - Phusion (Finnzymes), PrimeSTAR and LA Taq (Takakra) Herculase II Fusion and PfuUltra II (Stratagene), Platinum High Fidelity and AccuPrime Pfx (Invitrogen) - for their ability to amplify long fragments. We tested them on genomic DNA from the two tomato lines with four primer pairs.

To define the primers, we selected 872 BAC sequences available in NCBI database, either fully or partially sequenced. Primers were designed, using primer3 software (Rozen and Skaletsky 2000), on all the BAC sequences and a set of 384 primer pairs were finally selected to cover the entire genome.

Long Range PCR were finally performed using 100 ng of gDNA in a 50µl reaction volume composed of 10µl of 5X Buffer, 1.25µl of the forward primer solution at 10µM, 1.25µl of the reverse primer solution at 10µM, 1.25µl of a 10mM dTNPs mix solution, 1µl of the Herculase II Taq Polymeras (Stratagene) and water up to 50µl. Mastercycler thermal cycler (Eppendorf) was used with a thermal program as suggested by Stratagene (2 min. at 92°C; 30 cycles composed of three steps at 92°C for 10 s, 58°C for 20 s and 68°C for 5 min; and followed by a final step of 8 min at 68°C). The PCR products were kept at 4°C for less than 12 hours and then conserved at -20°C until sequencing. PCR products were checked on agarose gel (1.2 %). The specific amplifications that have been successfully obtained for the two lines were quantified and were equimolarly pooled.

Sequencing

The same DNA samples have been sequenced both on the Roche 454 and the Illumina Genome Analyzer GA2x sequencers. For the 454 sequencing run, 2 gaskets per sample were used. The GA2x run were done using a paired-end kit with a sizing of 500 nucleotides. One lane was used per accession for sequencing. The sequencing was performed in Unité Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, Evry (France).

Read mapping and SNP calling

The sequence reads were cleaned and mapped on the reference sequences extracted from the BAC sequences using two different software packages: CLC Genomics Workbench (CLC bio) and Seqman NGen (DNASTAR). We used the parameters recommended by the software

companies. All reference sequences derived from the Heinz1706's genome sequence (Sato et al. 2012) were used to evaluate the assembly and the sequence quality. SNP were then selected with a minimum of 90% of the reads showing a different allele from the reference sequence, and different depth depending on the platform.

As the fragments were not sequenced with the same depth of coverage with the two NGS plaforms, we randomly reduced the GA2X to 1,807,560 reads (10%) for Heinz1706 and to 1,706,340 for LA2675. With the randomly reduced run of the GA2x, the depth could be compared to the depth obtained with the 454 run.

Validation of SNPs and CAPS marker development

SNP located in restriction enzyme recognition sites were identified using Webcutter 2.0 program (http://rna.lundberg.gu.se/cutter2/). The cheapest restriction enzyme was selected if several restriction sites were affected in the same fragment. Primer pairs with an optimum of 22bp were designed by Primer 3.0 to amplify fragments measuring 700-900 bp and a 200bp window centred on the SNP position was defined as the target region. The CAPS markers were then (1) used to genotype a set of 23 highly diverse tomato accessions and (2) mapped on the F2 segregating population derived from the two sequenced lines. Genomic DNA (2.5 ng) from young leaves were mixed individually with 2.5 µl of 10X Buffer, 1.25 µl of 10 µM dTNPs, 0.5 µl of 10 µM forward primer, 0.5 µl of 10 µM reverse primer, and 0.15 µl Taq polymerase, water up to 25 µl. The PCR reactions were programmed as follows: denaturation at 94°C 5 min, followed by 35 cycles at 94°C for 30 s, 52°C for 30 s, 72°C for 1 min, then a final extension at 72°C for 10 min. PCR product (12.5 µl) was mixed with 2.5 µl of 10X buffer, 2 U restriction enzyme, and water up to 25 µl, incubated at 37°C overnight. Digested products were separated on a 1.2 % agarose gel

Genetic linkage map construction

The program CarthaGéne (de Givry et al. 2005) was used to assign CAPS markers to linkage groups with a LOD score threshold of 3.0. Markers belonging to the same group were ordered by the "kh" command of MAPMAKER. The recombination values were converted into map distances (Kosambi cM). Genetic map was then plotted using Mapchart software (Voorrips 2002).

Figure 1: **Location of 188 long Range PCR fragments onto the tomato genome.** All LR sequences have been aligned on the tomato genome (v2.31) using BLAST. The distance are indicated in Mb. The highlighted LR fragments are the ones for which the expected location of the BAC clones was different from the one found by BLAST.The C12HBa0183M6_LR379 didn't locate to any chromosomes.

Results and Discussion

Reduction of the genome complexity by Long Range PCR

Several methods to reduce genome complexity have been proposed and are compared by (Davey et al. 2011). Long Range PCR can be a valuable strategy to focus on target regions, but it may be tricky to systematically amplify long fragments of DNA. Several Taq polymerases are commercialized and described to be efficient to amplify long DNA fragments with a good fidelity. We tested seven different enzymes. Following the procedure described by the manufacturers, amplification was possible with three of them, Platinum High Fidelity (Invitrogen), LA Taq (Takara) and Herculase II Fusion (Stratagene). However, the yield was higher for the Herculase II Fusion (Stratagene). We thus used this enzyme to amplify the fragments to be resequenced.

A set of 384 primer pairs was first designed to cover the entire genome from the 872 BAC sequences available when the project started. These primers were used to amplify the corresponding genomic regions of two tomato lines, Heinz1706 and LA2675. Finally, 188 PCR fragments were obtained in both lines and selected on their specificity and the yield of the PCR amplifications on the two lines. The success rate (48.9% of the 384 fragments leading to a PCR product in the two lines) was intermediate but quite reproducible in the two genotypes.

The 188 sequences were aligned onto the complete tomato genome sequence (v2.31) once it was available in order to precisely locate each Long Range PCR fragment (**Figure 1**). Ten sequences were not located on their expected chromosomes (highlight in Figure 1). All chromosomes were represented but were not covered with the same number of fragments that ranged from 2 fragments on chromosome 10 to 37 for chromosome 2 (**Table 1**).

NGS data analysis

The fragments size obtained after nebulisation in the Roche 454 process was 588 bp in average. With this technique, we obtained 181,987 reads of 358 bases in average for Heinz1706 and 175,990 reads of 347 bases in average for LA2675 (**Table 2**). The amount of sequence data and the read size were consistent with manufacturer's expection. Compared to the 1.88 Mb of the total Long Range PCR fragment size, the expected sequencing depth should be 34.6X for Heinz1706 and 32.5X for LA2675 with the hypothesis of a perfect stoechiometry of the different fragments and a perfect coverage along each sequence. With

Table 1: Description of LR PCR fragments distribution on the tomato genome.
From The 872 BAC sequences, 384 primer pairs have been designed and tested on the two lines LA2675 and Heinz1706. 188 have been selected for sequencing. These 188 BAC sequences have been mapped by BLAST on the tomato genome.

Chromosome	Available BAC sequences	Number of primer pairs tested for LR PCR	Number of LR PCR after mapping on the genome	Total length (bp)
1	13	8	7	68,310
2	168	75	33	337,544
3	16	16	9	89,820
4	145	55	22	218,787
5	45	22	10	96,703
6	110	37	17	172,346
7	97	35	17	169,150
8	135	30	13	134,137
9	57	49	26	261,225
10	6	4	2	20,044
11	23	23	17	165,638
12	57	30	14	136,653
0			1	9,473
Total	872	384	**188**	**1,879,830**

the GA2X sequencer, a much higher depth was reached, as we obtained 15,364,464 36-bases reads and 17,063,588, for Heinz1706 and LA2675, respectively. The expected average depth was thus 272X for Heinz1706 and 302X for LA2675.

Alignment of sequence reads

The reads were mapped on the 188 reference sequences extracted from the BAC clones sequences (**Supplemental File 1**). From the 454 data, we compared the assembly obtained with two software packages, Seqman NGen (DNASTAR) and CLC Genomics Workbench (CLC bio). A number of reads did not map to any reference sequence. The differences in frequency of unmapped sequences were marked between the two software packages, but not between the NGS platforms (**Table 2**). For CLC Genomics, there was an average of 8.6% of reads that did not map to any reference sequence. For Seqman NGen, 14.9% of the reads did not map to any reference sequence. To compare results with the same depth, we randomly selected the GA2X data and obtained similar proportions of unmapped sequences between the GA2X run and the randomly reduced GA2X data set. Previous studies already identified such differences in the performance of assemblers (Feldmeyer et al. 2011). Mapping depth varied between fragments from 3 to 30 (**Supplemental Table S2**) and fragments were not covered at all and could be considered as absent from the DNA pools. Even though an equimolar quantity of DNA had been pooled, these differences could be explained by differences of PCR product quantity. The depth was similar between software and the differences of depth between GA2X and 454 were similar between fragments.

Polymorphism identification

Once reads were mapped, we identified the polymorphic sites by focusing our analysis on SNPs. The number of SNPs depends on the depth at each site and the frequency of varying alleles. Because depth was highly different between GA2X and 454 data, we did not select the SNPs according to the same criteria. For 454, we selected SNPs with a minimum depth of 3 reads and 90% of the reads showing the polymorphism. For GA2X, we selected SNPs with a minimum of 30 reads and 90% of the reads showing the polymorphism. This difference of criteria can induce a difference of the selected polymorphism sites. We thus used *in silico* reduced GA2X runs to obtain a depth similar to 454 data. For the reduced GA2X runs, we selected the SNPs under similar conditions to 454. The results were then compared and SNPs were considered identical if the position on the reference sequence and the polymorphic nucleotide were identical. Surprisingly, SNPs from Heinz1706 tomato accession were

Table 2: Summary of read mapping on the 188 reference sequences

The sequence reads have been mapped on the 188 reference sequences using CLC or Seqman NGen softwares.
Three sets of data have been used for each sequenced line Heinz1706 and LA2675: 454, GA2x and an in silico reduced GA2x software.

	454 Roche				GA2x				GA2x (in silico reduced runs)			
	Heinz1706		LA2675		Heinz1706		LA2675		Heinz1706		LA2675	
Total number of reads	181,977		175,980		15,364,266		17,063,390		1,807,560		1,706,340	
Average lenght of the reads	358		344		36		36		36		36	
Software	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC
Number of assembled reads	152,214	171,649	152,214	162,422	13,339,230	13,905,759	14,209,403	15,061,091	1 ,577,171	1,635,563	1,421,692	1,505,571
Number of unassembled reads	29,763	10,328	23,766	13,558	2 025,036	1 458,507	2,853,987	2,002,299	230,389	171,997	284,648	200,769
% unassembled sequences	16.36	5.68	13.50	7.70	13.18	9.49	16.73	11.73	12.75	9.52	16.68	11.77

identified although the reference sequences used for mapping were obtained from this accession (Sato et al. 2012). According to the combination of software used for analysis and sequencing technology, the number of SNPs detected for Heinz1706 varied from 645 to 1871 SNPs (**Figure 2**). The four sets of SNPs (deduced from two platforms and two software packages) were compared all together and 259 SNPs were common to all of them. The same strategy, used on the reduced GA2X runs, identified 245 SNPs, among which 202 were also found previously with the full GA2X runs.

For LA2675 accession, we identified from 4119 SNPs to 6174 SNPs. By comparing the four sets of SNPs, we identified 2933 SNPs common to the four analyses with the full GA2X runs and 2955 SNPs with the reduced GA2X runs. A total of 2718 SNPs were found in both lists. Among them, 92 SNPs were common with the 202 SNPs detected from the re-sequencing of Heinz1706. Summarizing all these comparisons, we selected a set of 2626 reliable SNPs specific to LA2675 for further analysis.

The analyses were conducted using two different software packages (CLC Genomic Workbench and Seqman NGen). The results showed that combining a sequencing technology to a specific software could lead to identify specific SNPs. The proportion of specific SNPs was higher for Heinz1706 than for LA2675. The number of specific SNPs was lower when similar criteria for the SNPs selection were applied for both sequencing technologies by decreasing the depth of GA2X to the depth of 454.

When using the 454 sequencing data, Seqman NGen identified less specific SNPs (16.8% for Heinz1706 and 7.9% for LA2675) than CLC Genomics (68.5% for Heinz1706 and 35.2% for LA2675), being more efficient for this analysis. In the same manner, when using the complete set of GA2X data, CLC Genomics identified less specific SNPs (27.7% for Heinz1706 and 12.1% for LA2675) than Seqman NGen (62.2% for Heinz1706 and 28.9% for LA2675) and seems more efficient for this type of data (**Figure 2**). Although sequencing depth has been mentioned as a factor of error in SNP calling (Nielsen et al. 2011), it seems to be much less influencing the final dataset than the choice of the software package.

The 202 SNPs identified for Heinz1706 were located on 17 different BAC sequences but 171 of them (84.6%) were located on four different fragments only and could correspond either to sequencing mistakes in the reference sequence or to sequences with a high number of repeated sequences, which could have been more difficult to assemble. As a consequence, we cannot either reject the hypothesis of small differences in the seeds of the two Heinz1706 sequenced due to residual heterozygosity.

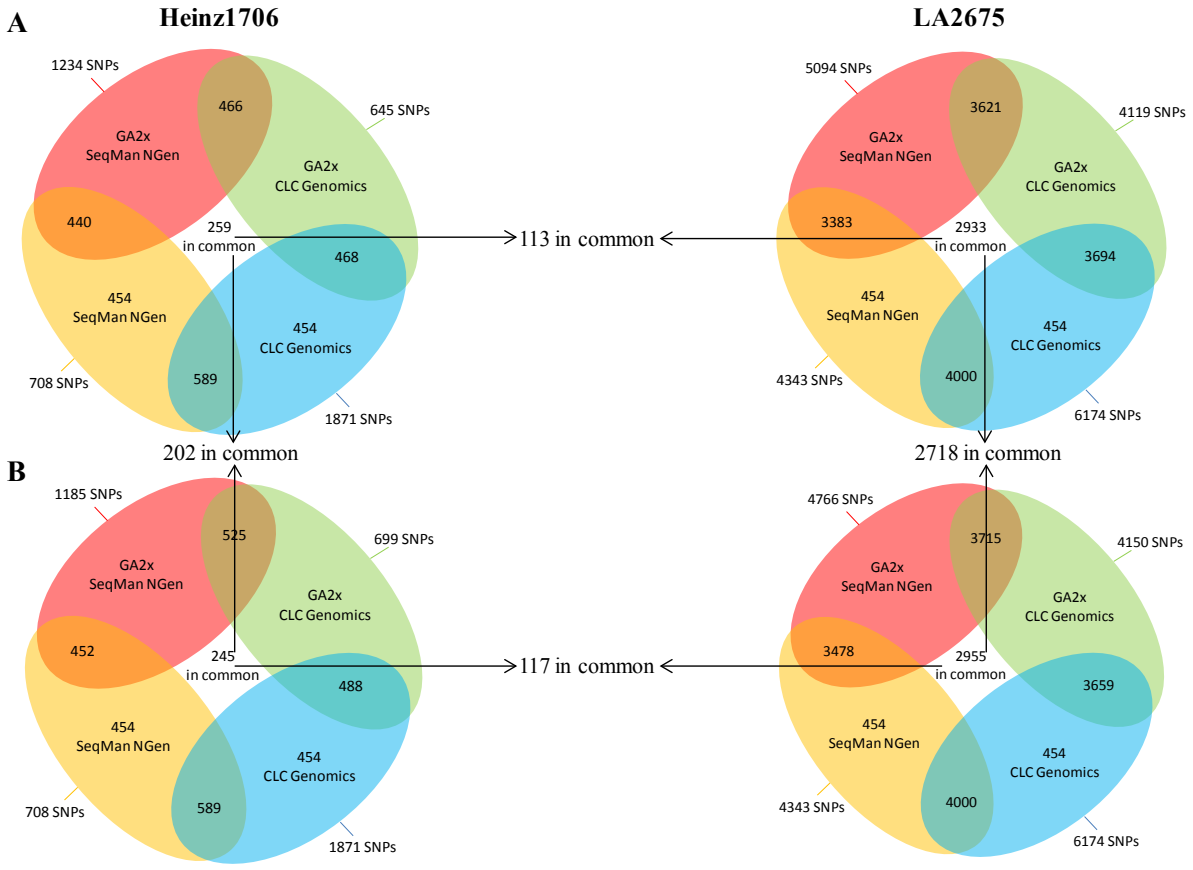


Figure 2: Comparison of SNPs identified by GA2x or 454 and analyzed by SeqMan NGen or CLC

The same 2 gDNA samples have been sequenced by Roche 454 and Illumina GA2x. The sequence data have been analyzed with SeqMan Ngen and CLC softwares with a mapping strategy on reference sequences. A: The complete set of GA2x have been used. SNPs have been selected with a depth of 3 reads for 454 and 30 reads for GA2x and 90% of polymorphic sequences for both. B: GA2x data have been *in silico* reduced to correspond to the 454 depth. For both, SNPs have been selected whith a depth of 3 reads and 90% of polymorphic sequences.

The 2626 SNPs specific to LA2675 were localized on 142 different fragments. A total of 31 fragments were not polymorphic although they were sequenced with a depth higher than 30x using GA2X and 3x using 454. These 31 fragments corresponded to a total sequence length of 312,565 nucleotides. Among the 142 polymorphic fragments, six were not enough sequenced and showed only one SNP for five of them and two SNPs for one of them. We thus calculated the polymorphism rate using the 2619 SNPs located on the 167 fragments with a depth higher than 30 for GA2X and 3 for 454 in all analyses. This rate was highly variable between chromosomes (**Table 3**) while ranging from 0.2 SNPs.kb^{-1} nucleotides for chromosome 1 (1 SNP every 5363 nucleotides) to 2.7 SNPs.kb^{-1} over chromosome 11 (1 SNP every 371 nucleotides). Over the whole fragments, the average polymorphism rate was 1.6 SNPs.kb^{-1} (1 SNP every 1018 nucleotides).

We annotated the genomic regions amplified thanks to the International Tomato Annotation Group effort (ITAG_CDS; http://solgenomics.net/). The 167 fragments sequenced with a sufficient depth were composed of 275 coding sequence (CDS) representing a total size of 299,827 bp of coding sequences. We found 18.26% of coding sequences within these 167 fragments spread over the genome on average but the proportion ranged from 10.07% for chromosome 11 to 31.92% for chromosome 1, which corresponded to the higher and lower SNP rate respectively (**Table 3**). From the 2,619 SNPs, 307 were located in coding sequences corresponding to 107 different CDS (**Supplemental Table S3**). In average, 11.77 % of the SNPs were located in coding sequences, confirming the lower rate of polymorphism in CDS than in intergenic sequence supporting previous observation in tomato by Ranc et al (2012).

SNP validation by molecular marker development

Almost 3000 SNPs were obtained after alignment of the 188 sequenced fragments amplified in the two contrasted accessions. A subgroup of 105 fragments was selected for the validation of polymorphic sites by the development and the genotyping of CAPS markers. We chose to develop CAPS markers because they do not need any specific instruments and can be easily used in any laboratory. We first drew the restriction maps of all reference sequences. We then searched for SNPs that were located in restriction sites and selected the most appropriate enzymes to finally design primers surrounding the SNPs. *In silico* digestion detected 713 putative CAPS according to the SNP sites. Then, only one restriction enzyme was selected and the number of CAPS was restricted to a final set of 215 putative markers. After in silico simulation of digestions, 64 CAPS markers were selected as candidate (**Supplemental table**

Table 2: Summary of read mapping on the 188 reference sequences

The sequence reads have been mapped on the 188 reference sequences using CLC or Seqman NGen softwares.
Three sets of data have been used for each sequenced line Heinz1706 and LA2675: 454, GA2x and an in silico reduced GA2x software.

	454 Roche				GA2x				GA2x (in silico reduced runs)			
	Heinz1706		LA2675		Heinz1706		LA2675		Heinz1706		LA2675	
Total number of reads	181,977		175,980		15,364,266		17,063,390		1,807,560		1,706,340	
Average lenght of the reads	358		344		36		36		36		36	
Software	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC	Seqman	CLC
Number of assembled reads	152,214	171,649	152,214	162,422	13,339,230	13,905,759	14,209,403	15,061,091	1 ,577,171	1,635,563	1,421,692	1,505,571
Number of unassembled reads	29,763	10,328	23,766	13,558	2 025,036	1 458,507	2,853,987	2,002,299	230,389	171,997	284,648	200,769
% unassembled sequences	16.36	5.68	13.50	7.70	13.18	9.49	16.73	11.73	12.75	9.52	16.68	11.77

S4), whereas the other CAPS digested fragments were too small (less than 50bp) to be distinguished by agarose gel.

Candidate CAPS markers were then used to validate these SNPs within a set of 23 accessions containing Heinz1706 and LA2675 and an F2 progenies resulting from the cross between Heinz1706 and LA2675 (**Figure 3**). All markers were found to be polymorphic between the two sequenced lines, validating the SNP previously detected. For 4 markers (LRC005, LRC066, LRC098 and LRC0103) the polymorphism was specific to LA2675 only and for 4 other markers (LRC001, LRC020, LRC055 and LRC063), only LA2675 was homozygous but heterozygous accessions could be identified (**Table 4**). For the remaining 56 markers, the polymorphism was shared with other lines than LA2675 (**Table 4**). Within the set of 23 highly diverse tomato accessions, the allele from Heinz1706 was more frequent in *S. lycopersicum* accessions than in cherry tomato accession. For 13 markers, we observed other alleles for the CAPS markers than the two expected from the sequencing data (**Figure 3**). In the set of the 23 highly diverse accessions, the rate of these extra alleles was particularly high in the two wild species *S. habrochaites* and *S. pennellii* (12.5% of the markers). The rate of missing data was also higher in these two species (21.8% in *S. habrochaites* and 26.5% in *S. pennellii*) due to sequence divergence.

The combination of all the results reported in the present paper (i.e. high-throughput sequencing combined to marker validation) show that the polymorphisms identified between the two lines could be used to design a high-throughput genotyping assay to characterize more accessions that could be interesting for association studies for example. We also developed an F2 segregating population from the cross between Heinz1706 and LA2675. The 96 F2 individuals were genotyped using the 64 CAPS markers and a low density genetic map was obtained (**Figure 4**). For the 54 grouped markers, the genetic map was consistent with the physical map, but the relationships between genetic and physical distances varied greatly according to the intervals (**Figure 4**), as already shown by Sim et al. (2012). Seven markers were not linked to any linkage groups. Three markers contained too much missing data and could not be mapped.

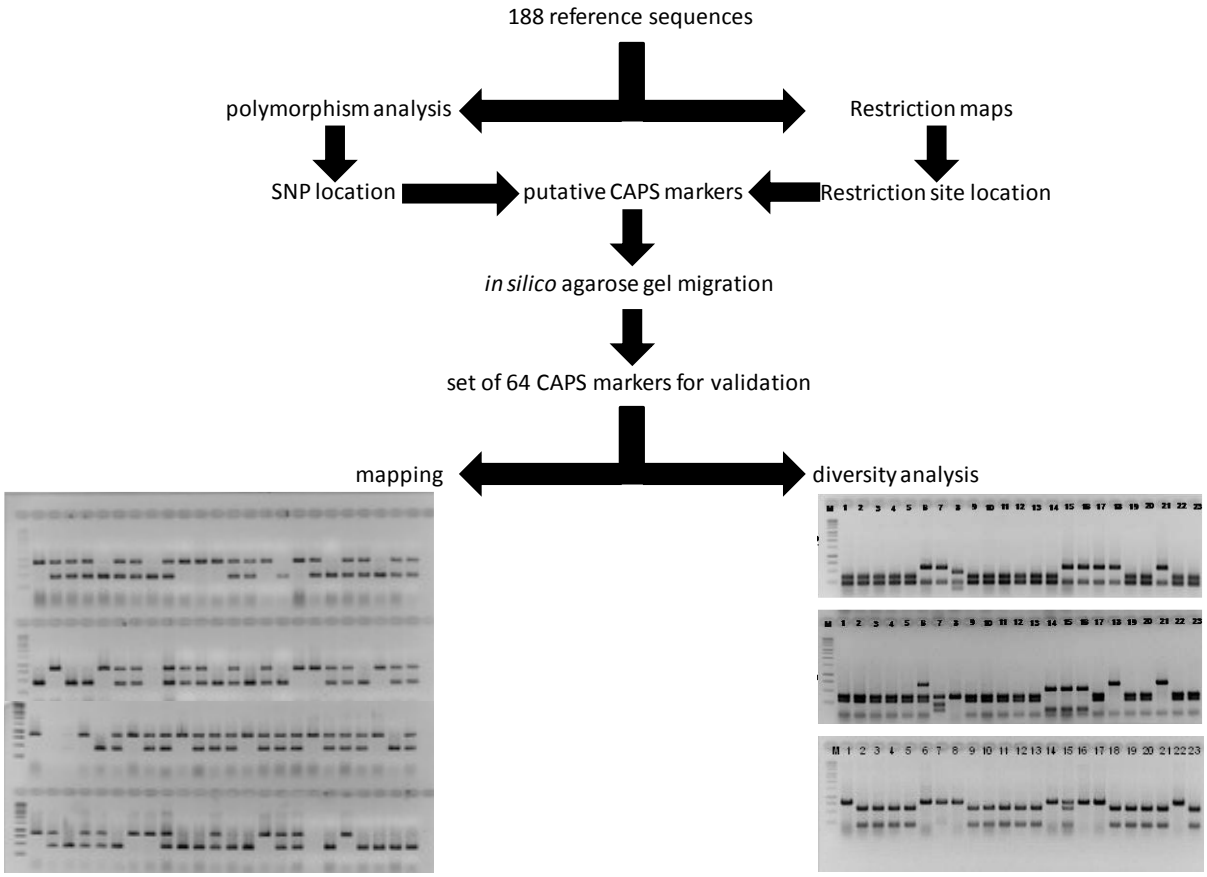


Figure 3: Development of CAPS markers
The strategy for CAPS markers development is described. Examples of genotyping on a core collection for diversity analysis and on a F2 segregating population for mapping are shown.

Table 4: Genotyping of a Core Collection composed of 23 tomato accessions. A set of 64 CAPS markers have been used for genotyping 1 corresponds to homozygous plants with the allele from LA2675, 3 corresponds to homozygous plants with the allele from Heinz1706, 2 to heterozygous plants, m for missing data and o for other alleles.

Accession	species	LRC001	LRC003	LRC005	LRC007	LRC009	LRC010	LRC011	LRC012	LRC014	LRC016	LRC017	LRC019	LRC020	LRC021	LRC023	LRC024	LRC025	LRC026	LRC028	LRC032	LRC035	LRC037	LRC038	LRC040	LRC042	LRC043	LRC044	LRC046	LRC047	LRC050	LRC052	LRC053
LA0147	lyco	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3
LA0409	lyco	3	3	3	3	3	3	3	3	1	3	3	o	3	m	3	3	3	1	3	3	3	3	1	1	1	3	3	3	1	1	1	3
OrangeCocktail	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	m	3	3	3	3	3	3	3	3	3	3	3	3	3
VIR1011	lyco	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Moneymaker	lyco	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Levovil	lyco	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	o	3	3	3	3	3	3	3	m	3	3	3	3	3
Ferum	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	1	1	1	3	3	3	3	3	3	3	3	3	3	o	3	3
M-82	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	o	2	3	3	1	3	3	3	3	3	3	3	3	3
Microtom	lyco	3	1	3	3	3	m	3	3	3	3	3	m	3	3	3	3	3	1	2	1	1	3	1	3	m	1	m	3	3	3	3	2
Heinz1706	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
LA2675	cera	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cervil	cera	3	1	3	1	3	3	1	3	1	3	3	1	3	1	1	1	1	1	o	3	1	3	1	3	3	3	3	1	3	o	3	3
LA1464	cera	3	m	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	1	3	3	3	3	3	3	o	3	3
Tinytim	cera	3	o	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	m	m	1	1	1	3	m	3	3	1	1	3	3	3
Criollo	cera	3	1	3	3	3	3	3	1	3	3	1	m	3	1	1	3	3	1	m	m	1	1	1	3	1	3	3	1	1	1	3	3
Phyra	cera	3	o	3	1	3	1	3	3	1	1	3	1	3	1	3	1	1	1	m	m	1	1	1	3	3	3	3	3	3	3	3	3
LA2137	cera	2	3	3	2	1	3	2	1	2	2	1	2	2	1	3	2	2	1	o	2	2	2	1	1	1	2	3	1	1	1	1	2
LA1589	pimp	2	o	3	2	3	1	3	1	1	1	3	2	2	1	1	3	1	1	o	3	1	1	1	3	3	2	3	1	1	1	3	3
LA1602	pimp	3	o	3	1	3	1	3	1	1	1	3	1	3	1	1	1	1	1	m	3	1	1	1	3	3	3	3	1	3	o	3	3
Wva700	pimp	3	1	3	3	3	3	1	3	3	3	3	m	3	3	3	3	3	1	3	3	3	3	1	3	3	3	3	3	3	3	m	3
LA1401	ches	3	1	3	3	3	3	3	1	3	1	3	m	m	1	3	3	1	1	m	m	1	1	1	3	m	1	3	3	3	o	3	3
PI247087	habr	m	1	m	3	m	3	3	1	1	1	m	m	m	1	m	3	1	o	2	m	1	2	3	3	m	1	1	m	m	o	3	1
LA716	penn	m	3	m	3	1	3	3	o	1	1	m	1	m	1	o	3	1	o	o	m	1	2	3	3	m	1	3	3	m	o	3	1

Accession	species	LRC054	LRC055	LRC057	LRC058	LRC059	LRC060	LRC061	LRC063	LRC064	LRC066	LRC067	LRC068	LRC069	LRC071	LRC074	LRC075	LRC080	LRC081	LRC082	LRC087	LRC088	LRC089	LRC091	LRC095	LRC096	LRC097	LRC098	LRC099	LRC101	LRC102	LRC103	LRC105
LA0147	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3
LA0409	lyco	1	3	1	3	m	1	3	3	3	3	1	3	3	3	1	1	3	3	1	3	1	3	m	3	3	3	3	3	1	m	3	3
OrangeCocktail	lyco	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3
VIR1011	lyco	1	m	3	3	3	3	1	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	1	3	3	3
Moneymaker	lyco	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	m	3	3	3	3	m	3	3	1	3	3	3	3	3	3	3
Levovil	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	m	3	3	3	3	3	3	m	3	1	3	3	3	3	3	3	3
Ferum	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3
M-82	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3
Microtom	lyco	3	3	3	3	3	3	3	3	m	3	3	3	3	3	3	m	2	3	1	3	1	3	3	3	1	1	3	3	3	3	3	3
Heinz1706	lyco	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
LA2675	cera	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cervil	cera	1	3	m	3	3	3	3	3	3	3	3	1	1	3	3	3	1	3	1	3	1	3	1	3	1	3	3	3	1	3	3	3
LA1464	cera	3	3	1	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3
Tinytim	cera	3	3	3	1	3	3	m	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3	1	m	3	3	3	3	3	3
Criollo	cera	1	3	3	1	3	1	3	3	3	3	3	1	3	1	1	1	3	3	1	3	m	3	m	1	1	1	3	3	1	3	3	3
Phyra	cera	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	1	3	3	1	m	3	1	3	1	1	3	3	1	1	3	3
LA2137	cera	1	2	3	3	3	1	3	2	2	3	3	1	1	2	3	1	2	1	3	3	1	1	1	2	m	2	3	2	1	3	3	1
LA1589	pimp	1	3	1	1	1	2	1	3	3	3	m	1	1	3	3	1	1	1	1	1	1	3	1	m	1	1	3	1	3	1	3	m
LA1602	pimp	1	m	1	1	m	3	3	3	3	3	3	1	1	3	1	3	1	3	1	1	1	3	m	3	1	1	3	3	m	1	3	3
Wva700	pimp	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	m	3	3	3	1	3	3	3	3	1	3	3
LA1401	ches	1	3	1	1	1	1	m	3	1	3	1	1	1	3	3	3	1	1	3	1	1	3	1	3	1	m	3	3	m	1	3	3
PI247087	habr	1	3	o	1	1	o	m	3	m	o	1	o	m	3	3	3	1	3	3	1	1	3	3	3	o	1	o	3	3	1	3	1
LA716	penn	m	3	1	1	m	o	m	m	m	o	1	m	m	3	3	m	1	3	3	1	1	3	1	3	o	m	3	3	m	1	3	3

lyco: *S. lycopersicum*; cera: *S. lycopersicum var cerasiforme*; pimp: *S. pimpinellifolium*; ches: *S. chesmaniae* ; habr: *S. habrochaites* ; penn: *S. pennelii*

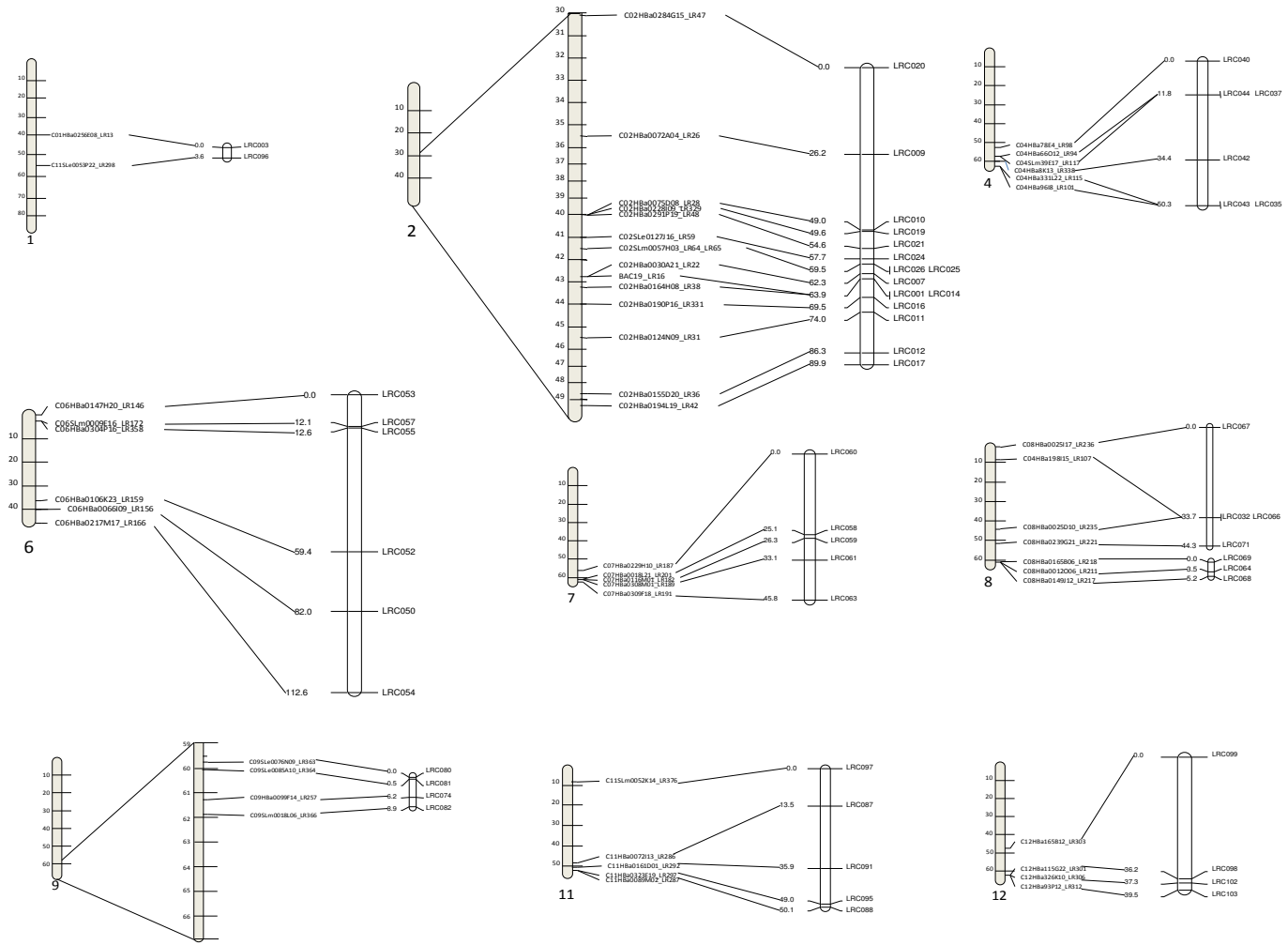


Figure 4: Genetic mapping of SNP derived CAPS markers and relationship with physical distance

A set of 64 CAPS markers have been developped from SNPs. 54 markers have been successfully linked togethers (7 were not linked to any linkage groups and 3 contained too much missing data to be mapped). The genetic mapping has been done with CarthaGene. Genetic map is represented by open bars and the physical map by grey bars.

Conclusion

NGS tools are highly efficient to rapidly detect SNPs to be implemented in various genetic studies. Our results provide complementary SNPs to the recent SNP array developed in tomato from RNA-seq (Sim et al, 2012) which carries more than 7,000 SNP. Both 454 and GA2X gave similar results, when similar depth is compared. As the yield of sequences obtained from Illumina sequencers is now much higher than 454 and a high quality genome sequence is obtained for tomato, this technology seems more appropriate to SNP calling in tomato genome. We showed that the software package might greatly affect the results. Today a large number of methods for NGS assembling and SNP calling are available (Nielsen et al. 2011). Our comparative study led to the identification of more than 2,600 SNPs, which appeared to be highly reliable as part of these were validated through the development of CAPS markers. Nevertheless the resequencing of an higher number of tomato accessions is necessary to obtain a catalogue of tomato SNPs, as already shown in Arabidopsis (Gan et al. 2011) or rice (Subbaiyan et al. 2012). When projected at the genome scale, we can deduce from the SNP number obtained that Heinz1706 and LA2675 may differ by about one and a half million SNPs, but the strong differences across chromosomes justify to resequence several accessions in order to be less specific.

Acknowledgements

This work was funded by INRA AIP BioRessources. Jiaxin Xu was funded by a grant from the China Scholarship Council.

List of supplemental information

Supplemental File 1: Sequences of the 188 fragment sequenced ***(too large to be included in the thesis)***

Supplemental table S1: List of a core collection of 23 accessions

Supplemental table S2: List of SNPs specific of LA2675 detected in common by the two sequencing platforms and the two software packages (too large to be included in the thesis)

Supplemental table S3: List of SNPs specific of LA2675 detected in common by the two sequencing platforms and the two software packages, located in coding regions

Supplemental Table S4: List of CAPS markers

Supplemental table S1: List of a core collection of 23 accessions

Accessions	Genebank Origin	Species
LA0147	TGRC	S. lycopersicum
LA0409	TGRC	S. lycopersicum
Orange Cocktail	INRA	S. lycopersicum
VIR1011	VIR	S. lycopersicum
Moneymaker	INRA	S. lycopersicum
Levovil	INRA	S. lycopersicum
Ferum	INRA	S. lycopersicum
M-82	INRA	S. lycopersicum
Microtom	INRA	S. lycopersicum
Heinz 1706	INRA	S. lycopersicum
LA2675	TGRC	S. lycopersicum var cerasiforme
Cervil	INRA	S. lycopersicum var cerasiforme
LA1464	TGRC	S. lycopersicum var cerasiforme
Tiny tim	INRA	S. lycopersicum var cerasiforme
Criollo	INRA	S. lycopersicum var cerasiforme
Phyra	INRA	S. lycopersicum var cerasiforme
LA2137	TGRC	S. lycopersicum var cerasiforme
LA1589	TGRC	S. pimpinellifolium
LA1602	TGRC	S. pimpinellifolium
Wva700	INRA	S. pimpinellifolium
LA1401	TGRC	S. galapagense
PI247087	USDA	S. habrochaites
LA0716	TGRC	S. pennellii

INRA: Institut National de Recherche Agronomique
VIR: Vavilov Research Institute of Plant Industry
USDA: United State Department of Agriculture (North Central Regional Plant Introduction Station)
TGRC: Tomato Genetics Resource Center at University of California, Davis, USA

Supplemental Table S3: List of the SNPs located in coding sequences

Reference Sequence ID	Chromosome	Position on	Reference base	SNP base	CDS ID	Exon ID
C02HBa0185P07_LR40	2	8795	T	C	Solyc02g070530.1.1	Solyc02g070530.1.1.10
C02HBa0072A04_LR26	2	269	A	C	Solyc02g071300.1.1	Solyc02g071300.1.1.1
C02HBa0072A04_LR26	2	455	T	C	Solyc02g071300.1.1	Solyc02g071300.1.1.1
C02HBa0072A04_LR26	2	851	T	C	Solyc02g071300.1.1	Solyc02g071300.1.1.1
C02HBa0072A04_LR26	2	8915	T	C	Solyc02g071320.1.1	Solyc02g071320.1.1.3
C02HBa0072A04_LR26	2	8918	C	T	Solyc02g071320.1.1	Solyc02g071320.1.1.3
C02HBa0072A04_LR26	2	8950	A	T	Solyc02g071320.1.1	Solyc02g071320.1.1.3
C02HBa0072A04_LR26	2	9124	T	A	Solyc02g071320.1.1	Solyc02g071320.1.1.3
C02HBa0072A04_LR26	2	9306	C	T	Solyc02g071320.1.1	Solyc02g071320.1.1.3
C02HBa0072A04_LR26	2	9566	G	A	Solyc02g071320.1.1	Solyc02g071320.1.1.2
C02HBa0329G05_LR52	2	6972	A	G	Solyc02g079020.1.1	Solyc02g079020.1.1.15
C02HBa0329G05_LR52	2	7020	C	A	Solyc02g079020.1.1	Solyc02g079020.1.1.15
C02HBa0329G05_LR52	2	7061	G	A	Solyc02g079020.1.1	Solyc02g079020.1.1.15
C02HBa0329G05_LR52	2	6544	G	A	Solyc02g079020.1.1	Solyc02g079020.1.1.13
C02HBa0329G05_LR52	2	5478	A	G	Solyc02g079020.1.1	Solyc02g079020.1.1.12
C02HBa0329G05_LR52	2	3491	G	A	Solyc02g079020.1.1	Solyc02g079020.1.1.9
C02HBa0329G05_LR52	2	182	T	A	Solyc02g079020.1.1	Solyc02g079020.1.1.5
C02HBa0329G05_LR52	2	317	A	C	Solyc02g079020.1.1	Solyc02g079020.1.1.5
C02HBa0329G05_LR52	2	402	A	G	Solyc02g079020.1.1	Solyc02g079020.1.1.5
C02HBa0204D01_LR334	2	8073	C	T	Solyc02g080340.1.1	Solyc02g080340.1.1.18
C02HBa0228I09_LR329	2	4006	C	A	Solyc02g081500.1.1	Solyc02g081500.1.1.1
C02HBa0228I09_LR329	2	4483	G	A	Solyc02g081500.1.1	Solyc02g081500.1.1.2
C02HBa0228I09_LR329	2	5598	G	C	Solyc02g081500.1.1	Solyc02g081500.1.1.4
C02HBa0228I09_LR329	2	7926	T	G	Solyc02g081510.1.1	Solyc02g081510.1.1.2
C02HBa0228I09_LR329	2	7975	G	T	Solyc02g081510.1.1	Solyc02g081510.1.1.2
C02HBa0228I09_LR329	2	8371	A	G	Solyc02g081510.1.1	Solyc02g081510.1.1.2
C02HBa0228I09_LR329	2	8545	G	A	Solyc02g081510.1.1	Solyc02g081510.1.1.2
C02HBa0075D08_LR28	2	6401	A	G	Solyc02g081670.1.1	Solyc02g081670.1.1.1
C02HBa0075D08_LR28	2	6435	T	C	Solyc02g081670.1.1	Solyc02g081670.1.1.1
C02HBa0075D08_LR28	2	7089	T	C	Solyc02g081670.1.1	Solyc02g081670.1.1.1
C02HBa0075D08_LR28	2	7346	A	G	Solyc02g081670.1.1	Solyc02g081670.1.1.1
C02HBa0075D08_LR28	2	9417	A	G	Solyc02g081680.1.1	Solyc02g081680.1.1.11
C02HBa0075D08_LR28	2	9429	T	A	Solyc02g081680.1.1	Solyc02g081680.1.1.11
C02HBa0075D08_LR28	2	9528	T	C	Solyc02g081680.1.1	Solyc02g081680.1.1.11
C02HBa0075D08_LR28	2	10071	C	T	Solyc02g081680.1.1	Solyc02g081680.1.1.9
C02HBa0075D08_LR28	2	10369	A	G	Solyc02g081680.1.1	Solyc02g081680.1.1.8
C02HBa0075D08_LR28	2	10522	A	T	Solyc02g081680.1.1	Solyc02g081680.1.1.8
C02HBa0008G02_LR67	2	431	C	T	Solyc02g082340.1.1	Solyc02g082340.1.1.11
C02HBa0008G02_LR67	2	3168	G	C	Solyc02g082340.1.1	Solyc02g082340.1.1.6
C02HBa0008G02_LR67	2	4800	A	G	Solyc02g082340.1.1	Solyc02g082340.1.1.1
C02HBa0291P19_LR48	2	8270	A	G	Solyc02g082540.1.1	Solyc02g082540.1.1.9
C02HBa0291P19_LR48	2	1755	C	A	Solyc02g082540.1.1	Solyc02g082540.1.1.2
C02SLe0127J16_LR59	2	4984	C	A	Solyc02g083340.1.1	Solyc02g083340.1.1.2
C02SLe0127J16_LR59	2	316	G	A	Solyc02g083350.1.1	Solyc02g083350.1.1.11
C02SLm0057H03_LR64_	2	16050	C	G	Solyc02g083930.1.1	Solyc02g083930.1.1.1
C02SLm0057H03_LR64_	2	16482	C	A	Solyc02g083930.1.1	Solyc02g083930.1.1.1
C02SLm0057H03_LR64_	2	12738	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	12978	G	C	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	13005	G	A	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	13486	A	G	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	13557	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	13560	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.1
C02SLm0057H03_LR64_	2	12738	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	12978	G	C	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	13005	G	A	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	13486	A	G	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	13557	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	13560	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.2
C02SLm0057H03_LR64_	2	12103	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.3
C02SLm0057H03_LR64_	2	12113	A	T	Solyc02g083940.1.1	Solyc02g083940.1.1.3
C02SLm0057H03_LR64_	2	12186	T	C	Solyc02g083940.1.1	Solyc02g083940.1.1.3
C02SLm0057H03_LR64_	2	10944	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.4
C02SLm0057H03_LR64_	2	11315	C	A	Solyc02g083940.1.1	Solyc02g083940.1.1.4
C02SLm0057H03_LR64_	2	11446	G	A	Solyc02g083940.1.1	Solyc02g083940.1.1.4
C02SLm0057H03_LR64_	2	10944	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.5
C02SLm0057H03_LR64_	2	11315	C	A	Solyc02g083940.1.1	Solyc02g083940.1.1.5
C02SLm0057H03_LR64_	2	10944	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.6
C02SLm0057H03_LR64_	2	10944	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.7
C02SLm0057H03_LR64_	2	8538	C	T	Solyc02g083940.1.1	Solyc02g083940.1.1.10
C02SLm0057H03_LR64_	2	7226	A	G	Solyc02g083940.1.1	Solyc02g083940.1.1.11
C02SLm0057H03_LR64_	2	7728	G	A	Solyc02g083940.1.1	Solyc02g083940.1.1.11
C02SLm0057H03_LR64_	2	1513	A	T	Solyc02g083950.1.1	Solyc02g083950.1.1.3
C02SLm0057H03_LR64_	2	302	A	G	Solyc02g083950.1.1	Solyc02g083950.1.1.2
C02SLm0057H03_LR64_	2	185	C	T	Solyc02g083950.1.1	Solyc02g083950.1.1.1
C02SLm0057H03_LR64_	2	302	A	G	Solyc02g083950.1.1	Solyc02g083950.1.1.1
C02SLm0057H03_LR336	2	9341	A	G	Solyc02g083980.1.1	Solyc02g083980.1.1.3
C02SLm0057H03_LR336	2	8786	A	G	Solyc02g083980.1.1	Solyc02g083980.1.1.4
C02SLm0057H03_LR336	2	8474	G	A	Solyc02g083980.1.1	Solyc02g083980.1.1.5
C02SLm0057H03_LR336	2	8023	T	C	Solyc02g083980.1.1	Solyc02g083980.1.1.6
C02SLm0057H03_LR336	2	8053	T	C	Solyc02g083980.1.1	Solyc02g083980.1.1.6
C02SLm0057H03_LR336	2	8085	A	G	Solyc02g083980.1.1	Solyc02g083980.1.1.6
C02SLm0057H03_LR336	2	3835	C	A	Solyc02g083990.1.1	Solyc02g083990.1.1.1
C02SLm0057H03_LR336	2	4254	T	C	Solyc02g083990.1.1	Solyc02g083990.1.1.1
C02SLm0057H03_LR336	2	4367	T	C	Solyc02g083990.1.1	Solyc02g083990.1.1.1
C02SLm0057H03_LR336	2	1461	A	G	Solyc02g083990.1.1	Solyc02g083990.1.1.2
C02SLm0057H03_LR336	2	1586	C	T	Solyc02g083990.1.1	Solyc02g083990.1.1.2
C02HBa0030A21_LR22	2	8346	C	G	Solyc02g085150.1.1	Solyc02g085150.1.1.3
C02HBa0030A21_LR22	2	8479	G	A	Solyc02g085150.1.1	Solyc02g085150.1.1.3
C02HBa0030A21_LR22	2	8662	T	A	Solyc02g085150.1.1	Solyc02g085150.1.1.3
C02HBa0030A21_LR22	2	8772	G	A	Solyc02g085150.1.1	Solyc02g085150.1.1.3
C02HBa0030A21_LR22	2	8913	A	G	Solyc02g085150.1.1	Solyc02g085150.1.1.3
C02HBa0030A21_LR22	2	6307	T	G	Solyc02g085160.1.1	Solyc02g085160.1.1.1
BAC19_LR16	2	4212	C	T	Solyc02g085590.1.1	Solyc02g085590.1.1.3
C02HBa0164H08_LR38	2	3368	C	T	Solyc02g085750.1.1	Solyc02g085750.1.1.5
C02HBa0164H08_LR38	2	556	C	T	Solyc02g085760.1.1	Solyc02g085760.1.1.1
C02HBa0164H08_LR38	2	259	A	G	Solyc02g085760.1.1	Solyc02g085760.1.1.2
C02HBa0011O23_LR68	2	2311	C	T	Solyc02g086370.1.1	Solyc02g086370.1.1.4
C02HBa0011O23_LR68	2	590	G	A	Solyc02g086370.1.1	Solyc02g086370.1.1.9
C02HBa0011O23_LR68	2	685	T	C	Solyc02g086370.1.1	Solyc02g086370.1.1.9
C02HBa0190P16_LR331	2	8337	C	T	Solyc02g086820.1.1	Solyc02g086820.1.1.4
C02HBa0016A12_LR19_I	2	5921	C	T	Solyc02g086880.1.1	Solyc02g086880.1.1.1
C02HBa0046M08_LR23	2	8706	A	G	Solyc02g089310.1.1	Solyc02g089310.1.1.1

C02HBa0155D20_LR36	2	8748	T	G	Solyc02g092780.1.1	Solyc02g092780.1.1.5
C02SLe0034H10_LR327	2	7297	G	A	Solyc02g093540.1.1	Solyc02g093540.1.1.2
C02SLe0034H10_LR327	2	7346	A	C	Solyc02g093540.1.1	Solyc02g093540.1.1.2
C02HBa0194L19_LR42	2	3840	G	A	Solyc02g093870.1.1	Solyc02g093870.1.1.3
C03HBa0233O20_LR82	3	6915	C	T	Solyc03g115000.1.1	Solyc03g115000.1.1.3
C03HBa0233O20_LR82	3	8084	A	G	Solyc03g115000.1.1	Solyc03g115000.1.1.3
C03HBa0233O20_LR82	3	6148	T	G	Solyc03g115000.1.1	Solyc03g115000.1.1.4
C03HBa0076J13_LR79	3	2923	T	C	Solyc03g116790.1.1	Solyc03g116790.1.1.9
C03HBa0076J13_LR79	3	4173	A	T	Solyc03g116790.1.1	Solyc03g116790.1.1.10
C03HBa0076J13_LR79	3	4182	T	A	Solyc03g116790.1.1	Solyc03g116790.1.1.10
C03HBa0076J13_LR79	3	4675	G	A	Solyc03g116790.1.1	Solyc03g116790.1.1.11
C03HBa0076J13_LR79	3	5567	A	G	Solyc03g116800.1.1	Solyc03g116800.1.1.2
C03HBa0076J13_LR79	3	5888	C	T	Solyc03g116800.1.1	Solyc03g116800.1.1.1
C03HBa0076J13_LR79	3	5914	C	T	Solyc03g116800.1.1	Solyc03g116800.1.1.1
C03HBa0076J13_LR79	3	9767	T	C	Solyc03g116810.1.1	Solyc03g116810.1.1.6
C03HBa0012D06_LR72	3	8346	T	C	Solyc03g117690.1.1	Solyc03g117690.1.1.2
C03HBa0012D06_LR72	3	6365	G	C	Solyc03g117700.1.1	Solyc03g117700.1.1.6
C03HBa0012D06_LR72	3	6494	G	A	Solyc03g117700.1.1	Solyc03g117700.1.1.6
C03HBa0012D06_LR72	3	6554	A	T	Solyc03g117700.1.1	Solyc03g117700.1.1.6
C03HBa0012D06_LR72	3	6289	T	C	Solyc03g117700.1.1	Solyc03g117700.1.1.5
C03HBa0012D06_LR72	3	5926	G	A	Solyc03g117700.1.1	Solyc03g117700.1.1.4
C03HBa0012D06_LR72	3	5999	A	G	Solyc03g117700.1.1	Solyc03g117700.1.1.4
C03HBa0012D06_LR72	3	5732	G	A	Solyc03g117700.1.1	Solyc03g117700.1.1.3
C03HBa0012D06_LR72	3	3349	G	T	Solyc03g117710.1.1	Solyc03g117710.1.1.1
C04HBa78J4_LR99	4	8687	G	A	Solyc04g008000.1.1	Solyc04g008000.1.1.1
C04HBa78J4_LR99	4	8706	G	A	Solyc04g008000.1.1	Solyc04g008000.1.1.1
C04HBa78J4_LR99	4	8735	A	G	Solyc04g008000.1.1	Solyc04g008000.1.1.1
C04HBa78J4_LR99	4	8746	T	C	Solyc04g008000.1.1	Solyc04g008000.1.1.1
C04HBa78E4_LR98	4	5644	A	G	Solyc04g056560.1.1	Solyc04g056560.1.1.4
C04HBa78E4_LR98	4	5645	T	C	Solyc04g056560.1.1	Solyc04g056560.1.1.4
C04HBa78E4_LR98	4	8623	C	A	Solyc04g056570.1.1	Solyc04g056570.1.1.3
C04HBa78E4_LR98	4	9772	C	A	Solyc04g056570.1.1	Solyc04g056570.1.1.2
C04HBa132O11_LR104_1	4	3655	G	T	Solyc04g064790.1.1	Solyc04g064790.1.1.3
C04HBa66O12_LR94	4	247	T	C	Solyc04g072040.1.1	Solyc04g072040.1.1.1
C04SLm39E17_LR117	4	238	T	C	Solyc04g072260.1.1	Solyc04g072260.1.1.49
C04SLm39E17_LR117	4	1834	T	G	Solyc04g072260.1.1	Solyc04g072260.1.1.46
C04SLm39E17_LR117	4	2085	T	C	Solyc04g072260.1.1	Solyc04g072260.1.1.45
C04HBa331L22_LR115	4	8661	G	A	Solyc04g081440.1.1	Solyc04g081440.1.1.4
C04HBa96I8_LR101	4	8976	G	A	Solyc04g081510.1.1	Solyc04g081510.1.1.1
C04HBa96I8_LR101	4	1310	G	A	Solyc04g081520.1.1	Solyc04g081520.1.1.3
C04HBa96I8_LR101	4	644	G	A	Solyc04g081520.1.1	Solyc04g081520.1.1.5
C04HBa96I8_LR101	4	129	C	T	Solyc04g081520.1.1	Solyc04g081520.1.1.6
C04HBa219H8_LR109	4	4862	T	C	Solyc04g081930.1.1	Solyc04g081930.1.1.1
C04HBa219H8_LR109	4	262	A	G	Solyc04g081950.1.1	Solyc04g081950.1.1.1
C05HBa0145P19_LR136	5	8001	C	T	Solyc05g007610.1.1	Solyc05g007610.1.1.3
C05HBa0145P19_LR136	5	8028	A	T	Solyc05g007610.1.1	Solyc05g007610.1.1.3
C05HBa0145P19_LR136	5	8155	T	C	Solyc05g007610.1.1	Solyc05g007610.1.1.3
C05HBa0145P19_LR136	5	6220	T	G	Solyc05g007610.1.1	Solyc05g007610.1.1.5
C05HBa0145P19_LR136	5	6380	C	T	Solyc05g007610.1.1	Solyc05g007610.1.1.5
C05HBa0145P19_LR136	5	6444	C	T	Solyc05g007610.1.1	Solyc05g007610.1.1.5
C05HBa0145P19_LR136	5	4339	A	T	Solyc05g007620.1.1	Solyc05g007620.1.1.1
C05HBa0145P19_LR136	5	4836	T	G	Solyc05g007620.1.1	Solyc05g007620.1.1.1
C05HBa0057G22_LR130	5	3805	A	G	Solyc05g012700.1.1	Solyc05g012700.1.1.3
C05HBa0057G22_LR130	5	3832	A	G	Solyc05g012700.1.1	Solyc05g012700.1.1.3
C05HBa0057G22_LR130	5	4327	C	G	Solyc05g012700.1.1	Solyc05g012700.1.1.2
C05HBa0057G22_LR130	5	4484	G	A	Solyc05g012700.1.1	Solyc05g012700.1.1.2
C05HBa0057G22_LR130	5	4513	T	C	Solyc05g012700.1.1	Solyc05g012700.1.1.2
C05HBa0057G22_LR130	5	5052	T	A	Solyc05g012700.1.1	Solyc05g012700.1.1.1
C05HBa0135A02_LR134	5	8927	G	A	Solyc05g051680.1.1	Solyc05g051680.1.1.8
C05HBa0138J03_LR135	5	2459	G	A	Solyc05g052190.1.1	Solyc05g052190.1.1.4
C05HBa0138J03_LR135	5	2926	G	A	Solyc05g052190.1.1	Solyc05g052190.1.1.4
C05HBa0138J03_LR135	5	1213	A	G	Solyc05g052190.1.1	Solyc05g052190.1.1.3
C05HBa0131D04_LR133	5	8847	C	T	Solyc05g054620.1.1	Solyc05g054620.1.1.4
C05HBa0131D04_LR133	5	8267	A	G	Solyc05g054620.1.1	Solyc05g054620.1.1.5
C05HBa0131D04_LR133	5	2368	T	C	Solyc05g054630.1.1	Solyc05g054630.1.1.3
C06HBa0147H20_LR146	6	9789	G	C	Solyc06g005660.1.1	Solyc06g005660.1.1.6
C06HBa0147H20_LR146	6	7765	A	T	Solyc06g005660.1.1	Solyc06g005660.1.1.1
C06HBa0304P16_LR358	6	3253	G	C	Solyc06g009530.1.1	Solyc06g009530.1.1.9
C06HBa0106K23_LR159	6	5314	T	C	Solyc06g064550.1.1	Solyc06g064550.1.1.18
C06HBa0106K23_LR159	6	4430	A	G	Solyc06g064560.1.1	Solyc06g064560.1.1.8
C06HBa0106K23_LR159	6	4241	C	A	Solyc06g064560.1.1	Solyc06g064560.1.1.7
C06HBa0106K23_LR159	6	4068	A	G	Solyc06g064560.1.1	Solyc06g064560.1.1.6
C06HBa0106K23_LR159	6	3537	T	C	Solyc06g064560.1.1	Solyc06g064560.1.1.3
C06HBa0106K23_LR159	6	3585	C	T	Solyc06g064560.1.1	Solyc06g064560.1.1.3
C06HBa0106K23_LR159	6	2023	T	C	Solyc06g064560.1.1	Solyc06g064560.1.1.1
C06HBa0106K23_LR159	6	2048	T	C	Solyc06g064560.1.1	Solyc06g064560.1.1.1
C06HBa0066D13_LR353	6	1649	A	G	Solyc06g072050.1.1	Solyc06g072050.1.1.6
C06HBa0066D13_LR353	6	2407	C	T	Solyc06g072050.1.1	Solyc06g072050.1.1.3
C06HBa0066D13_LR353	6	3742	A	G	Solyc06g072050.1.1	Solyc06g072050.1.1.1
C07HBa0229H10_LR187	7	3398	T	C	Solyc07g049500.1.1	Solyc07g049500.1.1.16
C07HBa0229H10_LR187	7	3646	T	C	Solyc07g049500.1.1	Solyc07g049500.1.1.15
C07HBa0229H10_LR187	7	5088	A	T	Solyc07g049500.1.1	Solyc07g049500.1.1.9
C07HBa0229H10_LR187	7	5561	G	A	Solyc07g049500.1.1	Solyc07g049500.1.1.7
C07HBa0018L21_LR201	7	3105	A	G	Solyc07g055970.1.1	Solyc07g055970.1.1.1
C07HBa0116M01_LR182	7	5733	G	A	Solyc07g056230.1.1	Solyc07g056230.1.1.3
C07HBa0116M01_LR182	7	3668	T	C	Solyc07g056240.1.1	Solyc07g056240.1.1.2
C07HBa0116M01_LR182	7	195	T	C	Solyc07g056250.1.1	Solyc07g056250.1.1.2
C07HBa0308M01_LR189	7	8961	T	C	Solyc07g062610.1.1	Solyc07g062610.1.1.3
C07HBa0308M01_LR189	7	7331	A	T	Solyc07g062610.1.1	Solyc07g062610.1.1.5
C07SLm0119A22_LR209	7	7830	G	A	Solyc07g065880.1.1	Solyc07g065880.1.1.1
C07HBa0309F18_LR191	7	2631	T	C	Solyc07g066110.1.1	Solyc07g066110.1.1.1
C07HBa0309F18_LR191	7	3576	G	A	Solyc07g066110.1.1	Solyc07g066110.1.1.1
C11HBa0190J03_LR294	8	10394	C	T	Solyc08g007610.1.1	Solyc08g007610.1.1.1
C11HBa0190J03_LR294	8	10805	G	A	Solyc08g007610.1.1	Solyc08g007610.1.1.1
C11HBa0190J03_LR294	8	6706	T	A	Solyc08g007620.1.1	Solyc08g007620.1.1.1
C11HBa0190J03_LR294	8	7231	A	G	Solyc08g007620.1.1	Solyc08g007620.1.1.1
C11HBa0190J03_LR294	8	4823	A	G	Solyc08g007620.1.1	Solyc08g007620.1.1.3
C11HBa0190J03_LR294	8	4828	C	T	Solyc08g007620.1.1	Solyc08g007620.1.1.3
C08HBa0239G21_LR221	8	1944	C	T	Solyc08g065870.1.1	Solyc08g065870.1.1.2
C08HBa0239G21_LR221	8	4980	G	A	Solyc08g065870.1.1	Solyc08g065870.1.1.5
C08HBa0239G21_LR221	8	8324	A	G	Solyc08g065880.1.1	Solyc08g065880.1.1.2
C08HBa0165B06_LR218	8	9230	G	A	Solyc08g079800.1.1	Solyc08g079800.1.1.2
C08HBa0165B06_LR218	8	9262	T	G	Solyc08g079800.1.1	Solyc08g079800.1.1.2
C08HBa0165B06_LR218	8	8501	C	T	Solyc08g079810.1.1	Solyc08g079810.1.1.1
C08HBa0165B06_LR218	8	4950	C	T	Solyc08g079810.1.1	Solyc08g079810.1.1.3

C08HBa0165B06_LR218	8	4975	A	G	Solyc08g079810.1.1	Solyc08g079810.1.1.3
C08HBa0165B06_LR218	8	6135	T	C	Solyc08g079810.1.1	Solyc08g079810.1.1.3
C08HBa0165B06_LR218	8	4539	A	T	Solyc08g079810.1.1	Solyc08g079810.1.1.4
C08HBa0165B06_LR218	8	1261	G	A	Solyc08g079820.1.1	Solyc08g079820.1.1.1
C08HBa0149J12_LR217	8	10292	G	A	Solyc08g081320.1.1	Solyc08g081320.1.1.14
C08HBa0149J12_LR217	8	7342	G	A	Solyc08g081320.1.1	Solyc08g081320.1.1.6
C08HBa0149J12_LR217	8	4566	T	C	Solyc08g081320.1.1	Solyc08g081320.1.1.2
C08HBa0201M14_LR220	8	6831	C	A	Solyc08g083240.1.1	Solyc08g083240.1.1.5
C09SLe0068C01_LR272	9	5815	C	T	Solyc09g007710.1.1	Solyc09g007710.1.1.1
C09SLe0068C01_LR272	9	5986	T	C	Solyc09g007710.1.1	Solyc09g007710.1.1.1
C09SLe0068C01_LR272	9	4315	A	G	Solyc09g007710.1.1	Solyc09g007710.1.1.2
C09SLe0068C01_LR272	9	3047	T	C	Solyc09g007710.1.1	Solyc09g007710.1.1.3
C09SLe0068C01_LR272	9	3169	C	T	Solyc09g007710.1.1	Solyc09g007710.1.1.3
C09SLe0068C01_LR272	9	3186	T	C	Solyc09g007710.1.1	Solyc09g007710.1.1.3
C09SLe0068C01_LR272	9	1979	T	G	Solyc09g007710.1.1	Solyc09g007710.1.1.4
C09SLe0068C01_LR272	9	1565	T	C	Solyc09g007710.1.1	Solyc09g007710.1.1.5
C09SLe0068C01_LR272	9	1751	G	T	Solyc09g007710.1.1	Solyc09g007710.1.1.5
C09HBa0100J12_LR259_	9	3757	T	A	Solyc09g009120.1.1	Solyc09g009120.1.1.5
C09HBa0100J12_LR259_	9	5639	T	C	Solyc09g009120.1.1	Solyc09g009120.1.1.6
C09HBa0100J12_LR259_	9	6370	G	T	Solyc09g009120.1.1	Solyc09g009120.1.1.6
C09HBa0142I14_LR265	9	4787	G	A	Solyc09g010910.1.1	Solyc09g010910.1.1.2
C09HBa0142I14_LR265	9	3731	C	T	Solyc09g010910.1.1	Solyc09g010910.1.1.3
C09HBa0142I14_LR265	9	3902	C	A	Solyc09g010910.1.1	Solyc09g010910.1.1.3
C09SLe0076N09_LR363_	9	4751	A	G	Solyc09g065950.1.1	Solyc09g065950.1.1.10
C09SLe0076N09_LR363_	9	6618	T	G	Solyc09g065950.1.1	Solyc09g065950.1.1.6
C09HBa0099F14_LR257	9	10168	C	G	Solyc09g074110.1.1	Solyc09g074110.1.1.15
C09SLm0018L06_LR366	9	3379	T	C	Solyc09g074630.1.1	Solyc09g074630.1.1.1
C09SLm0018L06_LR366	9	8005	A	G	Solyc09g074640.1.1	Solyc09g074640.1.1.4
C09HBa0022M02_LR247	9	3597	T	C	Solyc09g091180.1.1	Solyc09g091180.1.1.5
C09SLm0037I08_LR367	9	794	A	C	Solyc09g097910.1.1	Solyc09g097910.1.1.7
C09SLm0037I08_LR367	9	795	A	G	Solyc09g097910.1.1	Solyc09g097910.1.1.7
C09SLm0037I08_LR367	9	3030	C	A	Solyc09g097920.1.1	Solyc09g097920.1.1.1
C09SLm0037I08_LR367	9	3032	C	T	Solyc09g097920.1.1	Solyc09g097920.1.1.1
C09SLm0037I08_LR367	9	3701	A	G	Solyc09g097920.1.1	Solyc09g097920.1.1.1
C09SLm0037I08_LR367	9	4016	T	A	Solyc09g097920.1.1	Solyc09g097920.1.1.1
C09SLm0037I08_LR367	9	4126	G	T	Solyc09g097920.1.1	Solyc09g097920.1.1.1
C09HBa0226D21_LR244_	9	5936	C	T	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	5960	A	T	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	6124	G	C	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	6989	C	T	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	7075	A	G	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	7204	C	A	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C09HBa0226D21_LR244_	9	7430	A	G	Solyc09g098100.1.1	Solyc09g098100.1.1.1
C11HBa0168B23_LR293	11	7581	G	T	Solyc11g005240.1.1	Solyc11g005240.1.1.2
C11HBa0096D22_LR288	11	8945	T	C	Solyc11g008380.1.1	Solyc11g008380.1.1.2
C11HBa0096D22_LR288	11	8948	A	G	Solyc11g008380.1.1	Solyc11g008380.1.1.2
C11HBa0096D22_LR288	11	5177	A	G	Solyc11g008380.1.1	Solyc11g008380.1.1.1
C11HBa0096D22_LR288	11	332	C	A	Solyc11g008390.1.1	Solyc11g008390.1.1.3
C11HBa0096D22_LR288	11	508	T	C	Solyc11g008390.1.1	Solyc11g008390.1.1.3
C11HBa0107K14_LR289	11	1660	A	G	Solyc11g010640.1.1	Solyc11g010640.1.1.2
C11HBa0072I13_LR286	11	8482	A	G	Solyc11g066650.1.1	Solyc11g066650.1.1.4
C11HBa0303G16_LR296	11	4328	A	G	Solyc11g068610.1.1	Solyc11g068610.1.1.4
C11HBa0303G16_LR296	11	4352	G	A	Solyc11g068610.1.1	Solyc11g068610.1.1.4
C11HBa0161D01_LR292	11	367	A	C	Solyc11g069270.1.1	Solyc11g069270.1.1.9
C11HBa0161D01_LR292	11	390	T	C	Solyc11g069270.1.1	Solyc11g069270.1.1.9
C11HBa0161D01_LR292	11	1151	C	T	Solyc11g069270.1.1	Solyc11g069270.1.1.11
C11HBa0161D01_LR292	11	1253	T	A	Solyc11g069270.1.1	Solyc11g069270.1.1.11
C11HBa0161D01_LR292	11	1412	G	A	Solyc11g069270.1.1	Solyc11g069270.1.1.12
C11HBa0161D01_LR292	11	1486	G	A	Solyc11g069270.1.1	Solyc11g069270.1.1.12
C11HBa0161D01_LR292	11	1804	T	C	Solyc11g069270.1.1	Solyc11g069270.1.1.13
C11HBa0161D01_LR292	11	2104	G	C	Solyc11g069270.1.1	Solyc11g069270.1.1.14
C11HBa0161D01_LR292	11	2121	A	G	Solyc11g069270.1.1	Solyc11g069270.1.1.14
C11HBa0161D01_LR292	11	2467	A	T	Solyc11g069270.1.1	Solyc11g069270.1.1.15
C11HBa0161D01_LR292	11	2702	A	G	Solyc11g069270.1.1	Solyc11g069270.1.1.16
C11HBa0161D01_LR292	11	3381	C	G	Solyc11g069270.1.1	Solyc11g069270.1.1.17
C11HBa0161D01_LR292	11	3432	T	C	Solyc11g069270.1.1	Solyc11g069270.1.1.17
C11HBa0161D01_LR292	11	3444	C	T	Solyc11g069270.1.1	Solyc11g069270.1.1.17
C11HBa0161D01_LR292	11	3473	C	G	Solyc11g069270.1.1	Solyc11g069270.1.1.17
C11HBa0161D01_LR292	11	3816	C	T	Solyc11g069270.1.1	Solyc11g069270.1.1.18
C11HBa0161D01_LR292	11	3919	A	T	Solyc11g069270.1.1	Solyc11g069270.1.1.18
C11HBa0161D01_LR292	11	4401	C	T	Solyc11g069270.1.1	Solyc11g069270.1.1.19
C11HBa0161D01_LR292	11	5163	T	A	Solyc11g069280.1.1	Solyc11g069280.1.1.2
C11HBa0161D01_LR292	11	5377	G	A	Solyc11g069280.1.1	Solyc11g069280.1.1.2
C11HBa0161D01_LR292	11	5427	C	A	Solyc11g069280.1.1	Solyc11g069280.1.1.2
C11HBa0161D01_LR292	11	6273	A	G	Solyc11g069280.1.1	Solyc11g069280.1.1.1
C11HBa0089M02_LR287	11	6652	G	A	Solyc11g071820.1.1	Solyc11g071820.1.1.6
C11HBa0089M02_LR287	11	8132	C	T	Solyc11g071830.1.1	Solyc11g071830.1.1.5
C11HBa0089M02_LR287	11	8223	A	G	Solyc11g071830.1.1	Solyc11g071830.1.1.5
C11HBa0089M02_LR287	11	8976	T	G	Solyc11g071830.1.1	Solyc11g071830.1.1.3
C12HBa221M9_LR377	12	6579	A	T	Solyc12g005400.1.1	Solyc12g005400.1.1.8
C12HBa221M9_LR377	12	6618	C	T	Solyc12g005400.1.1	Solyc12g005400.1.1.8
C12HBa221M9_LR377	12	772	C	T	Solyc12g005400.1.1	Solyc12g005400.1.1.1
C12HBa221M9_LR377	12	883	G	T	Solyc12g005400.1.1	Solyc12g005400.1.1.1
C12HBa90D9_LR311	12	9387	T	C	Solyc12g005890.1.1	Solyc12g005890.1.1.2
C12HBa90D9_LR311	12	9443	A	G	Solyc12g005890.1.1	Solyc12g005890.1.1.2
C12HBa90D9_LR311	12	9510	T	G	Solyc12g005890.1.1	Solyc12g005890.1.1.2
C12HBa90D9_LR311	12	8440	C	T	Solyc12g005890.1.1	Solyc12g005890.1.1.1
C12HBa165B12_LR303	12	910	C	T	Solyc12g055890.1.1	Solyc12g055890.1.1.6
C12HBa165B12_LR303	12	1642	A	G	Solyc12g055890.1.1	Solyc12g055890.1.1.3
C12HBa165B12_LR303	12	1746	C	T	Solyc12g055890.1.1	Solyc12g055890.1.1.3
C12HBa165B12_LR303	12	1997	C	T	Solyc12g055890.1.1	Solyc12g055890.1.1.2
C12HBa165B12_LR303	12	2000	A	G	Solyc12g055890.1.1	Solyc12g055890.1.1.2
C12HBa165B12_LR303	12	2557	G	A	Solyc12g055890.1.1	Solyc12g055890.1.1.1
C12HBa165B12_LR303	12	6767	G	A	Solyc12g055900.1.1	Solyc12g055900.1.1.1
C12SLm103K8_LR380	12	1443	G	C	Solyc12g056010.1.1	Solyc12g056010.1.1.2
C12SLm103K8_LR380	12	1610	A	G	Solyc12g056010.1.1	Solyc12g056010.1.1.2
C12HBa115G22_LR301	12	7690	C	A	Solyc12g094630.1.1	Solyc12g094630.1.1.3
C12HBa115G22_LR301	12	2064	A	T	Solyc12g094640.1.1	Solyc12g094640.1.1.8
C12HBa326K10_LR306	12	5180	C	T	Solyc12g095810.1.1	Solyc12g095810.1.1.1
C12HBa93P12_LR312	12	1606	G	T	Solyc12g096610.1.1	Solyc12g096610.1.1.3

Supplemental table S4 : Information of 64 CAPs markers

Marker	Forward primer sequence (5'-3')	Reverse primer sequence (5' -3')	Expected PCR si	SNP position o	Restrction enzyme site
LRC001	CCAATTTATGACCAAAACAGCA	CACAGTAGCTGATGGAACAACC	742	332	1 HindⅢ site at 329
LRC003	CCTGATTGGGCGAAATAAATAG	GTGGACAATATGGAGGGAGAAA	821	535	2 DraI sites at 394 and 534
LRC005	GACGCAAGAAGTTGATGATTTG	TTTCTGTTTGGGAATTTTTACTTTG	750	423	1 AluI site at 426
LRC007	ACTCGCACTTCTCAATTTTGCT	ATACCGCTAGTCATGCCAAAGA	803	415	1 HinfI site at 415
LRC009	GCTACTCCATTCAACCAGCTTC	CATGAGGAACCTAGAACCATCC	765	133	2 EcoRI sites at 135 and 340
LRC010	GACTCACCCCTGCATTCTTATC	GCTCTCTTCTGTCCCTGTTGTT	724	470	5 HinfI sites at 3, 473 ,509, 609 and 663
LRC011	CATGATTCTTGGGCATATTTCA	TGAACAAAGGGAAGTCACAGAG	804	547	2 HaeIII sites at 178 and 550
LRC012	GCAGGTGGAGTACCTTAGCTTG	AAGGGTTGGCGTCTTTAGTTTT	798	458	2 HinfI sites at 231 and 457
LRC014	GTTAAAAGCCTCCACCATTCAC	CAGATTGTTTCCACTCGTACCC	745	555	1 HaeIII site at 557
LRC016	GAAATCAAGATCCCTGGAAACA	CGGGATGACAGAACTACATCAA	860	192	1HinfI site at 190
LRC017	GAAACGGGCACTTAGTCTTCC	CTGTTTCGCCTTTTCTGATTTT	841	265	1 XbaI site at 264
LRC019	AGAGACATGAAGGTGCAAATGA	CCATCTAGGCTCTCCACAAACT	891	369	2 TaqI sites at 186 and 369
LRC020	CAAGTGCCCTATATGATCTCCA	ATGAGTGGGTAGTGGATTGGTT	716	462	1HaeIII site at 462
LRC021	GTTGAGATAGAAGTCGGGTTGG	GGGATGATTGATTGATTGATCTG	796	454	1 BamHI site at 451
LRC023	TTCAATTTTTGCTTTCGATGCT	TCGTTCCATCTTCAACACACTC	843	418	1 SspI site at 416
LRC024	AATCAGCCCATACAACAACTCA	CTTCATGTGGGTCAGCAGTAAA	752	374	2 TaqI sites at 374 and 505
LRC025	CCAATGCCATCAACTACACAAC	AAGGCAAGCTCGTGAAAGATAA	797	477	1 HaeIII site at 477
LRC026	ATTGTCCACCTCCATTCAAAAA	GAGATGGGGCAAACAAATAAAG	760	540	2 DraI sites at 334 and 538
LRC028	GTAGAGCCCCGACAACTTAACC	TTTTATGCACCGTACCAAACTG	701	498	2 NdeI sites at 160 and 499
LRC032	TTCCTCCAGTTTATTTCCAGCA	TGTGTCCTTCCTCGATTTCTTC	755	498	1 NsiI sites at 498
LRC035	TACCAAGTAACCAGCAATGGAC	AGATGGCAAGAGCTAGTTGGAG	712	430	1NsiI site at 434
LRC037	CAGTAGGACTGAATCGGAAACC	ACTGTTCTTTTCTCCACTGTCC	749	514	1 NdeI site at 515
LRC038	TGCTGTGTGTTTTAATGGTTTTG	CCTCCACTATGTCTCCTACTTCC	756	609	1 TaqI site at 610
LRC040	ATTCATGGTAGGGCAATCAACT	ATCAACACTCCACCCCATAAAA	711	340	5 HinfI sites at 104, 163, 188, 341 and 617
LRC042	ATTTTTGGACGGTATGGACACT	AACCAACCACCTCTTGTCAACT	732	378	3 TaqI sites at 71, 378and 488
LRC043	GGTGGGGTCTGAGAAGAGTAGA	GTAAGGTCAAGCTCGGAAAATG	749	359	4 HinfI sites at 360 ,686 ,715 and 724
LRC044	GTAGGCAGACCTTACCCCTACC	ATCCATCTCCCTGAACAGAAAA	739	445	1 MspI site at 443
LRC046	GAGGAATTGAGGGATGAGGTAA	AGAGGAGACACAAGCTGGGTAA	827	569	6 RsaI sites at 112, 195, 302, 312, 485 and 572
LRC047	TCTCCATAAAGCCAGCTACACA	CATTCAAAATGCCAAAGTCAAA	879	603	1 EcoRI at 613
LRC050	AGGGGAAAAAGAGTGGAGAAAG	AAAGTCGTTGATTGATGGAGGT	668	117	2 TaqI sites at 118 and 302
LRC052	GGAGGAGACTTGCAGTAGGTTG	TGAACTTGGTTTTTGAAGCAGA	775	358	1 MspI site at 358
LRC053	TAGCGGAGTCAAGATTCTCACA	GTGAAGAAAGCGTTATCCCAAG	721	362	1 HaeIII site at 361
LRC054	GACATTCCTAGTTCCGCTAAAAA	TCAAACAGACTTCTCCGATGAC	706	436	1 EcoRV site at 438
LRC055	ATTTGGGAATTTTGCTATGAGG	GAGGGAAATTGTGCTGAGAAAG	737	514	1 DraI site at 514
LRC057	GAAACTGAAAGGGTTTCCAAGA	TGTCTTTTCAACAATAAACGGAAA	848	356	3 TaqI sites at 218, 354 and 685
LRC058	TTTTGGATTTGTTGATCCTGTG	ATCTCATTTATTCCGCCATGTT	782	498	1 TaqI site at 499
LRC059	ATCTAAGGTCATGTTCGCCAGT	TTCATTTTCATACCGCTCGATT	897	470	1 ecorI at 466
LRC060	AATTGCATCACTACCGTTGAAG	ATGGTACTCCCTCTGTCCCTAA	729	358	2 TaqI sites st 359 and 658
LRC061	TCGATCAAAATCAGTCAACATTC	GCTTTCACTTATGACTCACTCAACC	772	415	4 HinfI site at 55 ,412 ,531 and 756
LRC063	GTGGGGCTATTTGACTGAAAAC	GAGCAAGAGGGTAAGGTGCTAA	822	536	1 MspI site at 536
LRC064	TTTGCAGTTTTGCTGTCATTTT	GGTCTCTTTTCAATTTTGTATTTCATT	924	363	6 TaqI sites at 48, 365, 534, 583, 632 and 762
LRC066	TCGAACATACTTGCCCTCTCTT	TAGCCGGATCAATATACGAGGT	802	286	3 MspI sites at 286, 375and 796
LRC067	TGCTTTTTCTTTTTAGTCCGTCA	GCTCCATTTTTGGTGGATACTT	702	393	1 MspI site at 391
LRC068	TTGACGGGGAGAATAGAATGAC	TTCCGCAACTAGCTTTTCTAAC	793	679	5 SspI sites at 36, 204, 290, 393 and 677
LRC069	CTTTCCATTCCAACTCTTCCAC	TGGATGGGCTTAATAAAAGCAG	827	556	2 TaqI sites st 44 and 555
LRC071	TGTGAGGGTGTTGATCTTTGTC	GTGATCTCCCTTTCTCCTTCCT	850	642	2 TaqI sites at 96 and 641
LRC074	GAAAGAAGGTTGCTTGAGATGG	AGGAAGATGTACCCGAAAGACA	807	143	1 BamHI site at 142
LRC075	TATCGTAGAACATGCCCTTTTG	TGATGTGGAATTTATCTAAGGTGGT	822	385	3 AseI sites at 383, 455 and 592
LRC080	TTCAGTTCCCTGGATTATTGCT	GCAGGATGGAGATAGATTTTGG	905	577	1 TaqI site at 588
LRC081	GGAGAGTGTGAGGGATAACGAG	TAGGATTTCGAGACTGCCATTT	911	344	3 TaqI sites at 283, 344, 902
LRC082	ATAGTCCTGGATCTCCCAAACA	ATACATCCACTTCCTGCTCCAC	736	469	1NsiI site at 469
LRC087	TGACTTGATTTGATCTGCCAAC	ATAGAGCAACAACCACACCTCA	705	337	1 SspI site at 340
LRC088	TAACCGACCCACTTGTCACTC	TAGCCTCCAACAACAGATTTCA	709	497	2 RsaI sites at 26 and 496
LRC89	AAACCTCATGGAGTAGCATTGG	ATTTCAGGATGCCTTTGAAGAA	817	343	1 PvuII site at 345
LRC091	CCACTTGCTATGTCTTGTGGAA	TCACATTTTGTGTTAGAGTTCGAGT	766	330	1 BamHI site at 326
LRC095	CATTTGTGTTAGCAGGGTCAAA	GACAAAGAGCCCAAAGTTGTAG	853	472	1 EcoRI site at 468
LRC096	CCTCAAGTTTCTTGCTTCCTTC	TTCATGCGCTCTTTATCCTTTT	751	426	1 HindII site at 427
LRC097	TCCATCCACGATTCTAATACACC	GCTTCAAGAAATGAATGGGAAG	814	643	1 EcoRI site at 642
LRC098	TGACCTTACGATATTGGCAACC	GAATGTGAGCATTACCTGCAAA	834	506	4 HinfI sites at 219, 304 ,503 and 695
LRC099	TCTAAGGTTATGGCCTCGGTAA	AAGTAGCCAAGTGAAGCAGGAC	790	409	2 HaeIII sites at 11 and 406
LRC101	AAGGATACAAGCAAAGGGTCAG	TCAACTTCGCCTAAGAATCACA	930	445	2 TaqI sites at 447 and 823
LRC102	AATTTGAAAGGAAGGCGAAAAG	TTCATCAAGTTTGTTCCGTCAA	831	494	2 RsaI sites at 495 and 775
LRC103	CTGCTGACTGTCTAAAGGTGGA	ACTCATTTTGGGGATATGTTGG	722	284	1 AluI site at 285
LRC105	GTTTGAATGCTAGAGGGAGCAA	CGACCTAAAAGCCAAGGTAATG	735	493	1 MspI site at 493

CHAPTER IV: PHENOTYPIC DIVERSITY AND ASSOCIATION MAPPING FOR FRUIT QUALITY TRAITS IN CULTIVATED TOMATO AND RELATED SPECIES

This chapter, in the form of a manuscript accepted for publication in Theoretical and Applied Genetics, presents the phenotypic and molecular characterization of a sample composed of 188 tomato accessions and subsequent association study between phenotypes and genotypes. The collection included a large (127) set of cherry tomato accessions. We genotyped the collection with 192 SNPs unevenly spread on the genome. A larger number of SNPs was designed in chromosome regions where QTL were previously detected. The 121 informative SNPs were used to analyze LD decay and population structure using either SSR markers or SNP. We then compared the associations observed using different structure covariate and samples. Several associations were detected in regions where QTL had previously been mapped, showing the power of the approach, but some other associations were also identified in new regions. This work also provides the community with a new set of SNPs useful for any study. It follows the study presented in Ranc et al. 2012 (ANNEXE 1).

Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species

Jiaxin Xu[(1,2)], Nicolas Ranc[(2)], Stéphane Muños[(2)], Sophie Rolland[(2)], Jean-Paul Bouchet[(2)], Nelly Desplat[(2)], Marie-Christine Le Paslier[(3)], Yan Liang[(1)], Dominique Brunel[(3)], Mathilde Causse[(2)]

(1) Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, P.R.China

(2) INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, Avignon, 84143, France

(3) INRA, UR1279, Unité Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, Evry, 91057, France

Corresponding author

Mathilde Causse

Phone: 33 04 32 72 27 01

Fax: 33 04 32 72 27 02

Email: Mathilde.Causse@avignon.inra.fr

Present address:

NR : Syngenta Seeds SAS, 12, chemin de l'Hobit, B.P. 27, 31790 Saint-Sauveur, France

SM : CNRS-INRA, UMR 2594/441, Laboratoire des Interactions Plantes Micro-organismes (LIPM), 31326 Castanet-Tolosan Cedex, France

SR : INRA-AgroCampus Ouest-Université Rennes1, UMR118, Amélioration des Plantes et Biotechnologies Végétales, BP 35327, 35653 Le Rheu Cedex, France

Abstract

Association mapping has been proposed as an efficient approach to assist in the identification of the molecular basis of agronomical traits in plants. For this purpose, we analyzed the phenotypic and genetic diversity of a large collection of tomato accessions including 44 heirloom and vintage cultivars (*Solanum lycopersicum*), 127 *Solanum lycopersicum* var. *cerasiforme* (cherry tomato) and 17 *S. pimpinellifolium* accessions. The accessions were genotyped using a SNPlex^{TM} assay of 192 SNPs, among which 121 were informative for subsequent analysis. Linkage disequilibrium of pairwise loci and population structure were analyzed, and the association analysis between SNP genotypes and ten fruit quality traits was performed using a mixed linear model. High level of linkage disequilibrium was found in the collection at the whole genome level. It was lower when considering only the 127 *S. lycopersicum* var. *cerasiforme* accessions. Genetic structure analysis showed that the population was structured into two main groups, corresponding to cultivated and wild types and many intermediates. The number of associations detected per trait varied according to the way the structure was taken into account, with zero to 41 associations detected per trait in the whole collection and a maximum of four associations in the *S. lycopersicum* var. *cerasiforme* accessions. A total of 40 associations (30%) were co-localized with previously identified quantitative trait loci. This study thus showed the potential and limits of using association mapping in tomato populations.

Keywords: Association mapping; linkage disequilibrium; mixed linear model; tomato; fruit quality.

Introduction

Genetic dissection of quantitative traits in plants is a major goal for plant breeding. Quantitative trait loci (QTLs) were first mapped in bi-parental populations using linkage mapping approach (Paterson et al. 1991; Saliba-Colombani et al. 2001; Wang et al. 2006; Szalma et al. 2007; Orsini et al. 2012). This approach has several advantages: (i) no population structure in the mapping population; (ii) segregating alleles are at balanced frequency; (iii) it allows the detection of rare alleles and epistasis. However, the linkage mapping approach has several limitations: (i) restricted allelic variation in bi-parental mapping population; (ii) low precision due to limited recombination within the population (Hall et al. 2010).

Nowadays, association mapping or linkage disequilibrium (LD) mapping is proposed as an alternative approach. On one hand, association mapping has several advantages over QTL mapping: (i) it is based on occurring variation in collections of natural genetic resources; (ii) it is more precise because of the recombination events resulting from many lineages; (iii) if LD is sufficiently low, it allows the discovery of the gene controlling the trait of interest. On the other hand, it has several limitations: (i) unbalanced allele frequency in the population; (ii) it is not efficient for the detection of rare alleles; (iii) it requires large population sizes and an efficient control of the population structure. Many plant association studies have been published to date for several traits, such as flowering time and pathogen resistance in Arabidopsis (Aranzana et al. 2005), yield and its components in rice (Agrama et al. 2007), leaf architecture in maize (Tian et al. 2011), iron deficiency chlorosis in soybean (Mamidi et al. 2011). Association mapping approach requires the knowledge of the genetic structure and the extent of LD of the samples studied. Population structure can lead to false positives and must be taken into account (Yu et al. 2006). Several statistical methods have been developed

to deal with structured samples in order to control false associations (Pritchard et al. 2000; Price et al. 2006; Yu et al. 2006). The mixed linear model has been shown to be efficient in maize (Yu et al. 2006) and Arabidopsis (Atwell et al. 2010). LD is the non-random association between alleles at several loci. The extent of LD over the genome will influence association mapping strategy. If LD is high, the resolution will be low but fewer markers will be required and a whole-genome scan approach may be performed. If LD is low, the resolution will be higher as the number of marker required and a candidate gene analysis may be conducted (Rafalski 2002). LD is expected to be higher in average in autogamous species than in allogamous species because the higher homozygozity at a given locus leads to lower rate of efficient recombination than in allogamous species (Flint-Garcia et al. 2003). Nevertheless, a large range of variation in the rate of recombination per Mb exists along the chromosomes (Sim et al, 2012). Thus, the resolution of association mapping in autogamous species is expected to be lower than in allogamous species. Therefore, association mapping was first used in allogamous species or species with wide range of genetic diversity.

Tomato (*Solanum lycopersicum*, formerly *Lycopersicon esculentum*) is a highly autogamous species. It was domesticated from its wild relative *S. pimpinellifolium* with the first domesticated form presumably represented by *S. lycopersicum* var. *cerasiforme* (i.e. the cherry tomato, hereafter *S. l. cerasiforme*). Cultivated tomato shows a low genetic diversity but higher phenotypic diversity compared to *S. pimpinellifolium* (Miller and Tanksley 1990) due to intensive human selection. The higher molecular diversity and the genetic admixture of *S. l. cerasiforme* genome (being a mosaic of cultivated and wild tomato genomes, (Ranc et al. 2008) may be useful to overcome the high LD in this autogamous species. Several association studies have been carried out to dissect morpho-physical and fruit traits in tomato. Mazzucato et al. (2008) studied associations between 29 simple sequence repeat (SSR) markers and 15 morpho – physiological traits in 50 tomato landraces. Nesbitt and Tanksley (2002)

investigated associations between fruit size and genomic sequence of the fw2.2 region which controls fruit weight (Frary et al. 2000) in a collection of 39 cherry tomato accessions. Unfortunately, they failed to find any association, but they demonstrated that the genome of cherry tomato accessions is a mosaic composed of polymorphisms of *S. pimpinellifolium* and *S. lycopersicum*. Munos et al. (2011) used association analysis to identify two single-nucleotide polymorphisms (SNP) located in a small region of chromosome 2 involved in the control of locule number of tomato fruit.

In a previous pilot study focused on one chromosome and 90 tomato accessions (Ranc et al. 2012), we showed that association mapping was possible in tomato. In the present work we developed a SNPlex^{TM} assay (De La Vega et al. 2005; Tobler et al. 2005) of 192 SNPs selected from re-sequencing experiments or from databases (Van Deynze et al. 2007). A large germplasm collection including cultivated, cherry type and wild accessions were characterized for both genetic diversity using the SNPlex^{TM} assay, 20 SSR markers (Ranc et al. 2008) and ten phenotypic traits. We first describe the phenotypic diversity of the accessions, then the genetic structure of the collection based on SSR and SNP markers and finally the association mapping results. Associations are compared to previously mapped QTL. Our work is the first example of an association study carried out using a broad sample of cultivated, cherry type and wild tomato accessions.

Materials and methods

Plant materials

Tomato accessions were selected from a germplasm collection maintained and characterized at INRA Avignon (France) to maximize both genetic and phenotypic diversity. The sample consisted of 127 cherry type tomato accessions, 44 large fruited accessions (*S. lycopersicum* var. *esculentum*, hereafter named *S. lycopersicum*) and 17 *S. pimpinellifolium* accessions (**Supplemental Table S1**). Accessions were obtained from the Tomato Genetics Resource Center (Davis, USA), the Centre for Genetic Resources (Wageningen, Netherlands), the North Central Regional Plant Introduction Station (Ames, Iowa, USA) and from the N.I. Vavilov Research Institute of Plant Industry (St Petersburg, Russia). Genomic DNA of the 188 accessions was isolated from 50 – 100 mg young leaves. After freeze-drying, the leaf material was ground and DNA extraction was performed using the DNeasy 96 plants Mini Kit (Qiagen, Valencia, USA) according to the manufacturer's protocol.

SNPlex^{TM} assay design

Allele-specific probes and optimized multiplexed assays using SNPs of interest were designed by an automated multi-step pipeline [Applied Biosystems, Foster city, USA]. The ABI probe design prevents self-complementarity and dimerization, and annealing efficiencies are optimized for ligation. Furthermore, the optimal combination of SNPs to produce the highest yield per multiplex reaction is determined (De La Vega et al. 2005; Tobler et al. 2005). Four SNPlex each carrying 48 SNPs were developed. Among the 192 SNPs used in the SNPlex^{TM} assay, 131 SNPs were chosen from Sanger resequencing experiments of candidate genes mostly located on chromosome 2 (69 SNPs), 4 (22 SNPs) and 9 (30 SNPs) (Ranc et al. 2012).

The remaining 61 SNPs were chosen from published information for covering the whole genome (Van Deynze et al. 2007). **Supplemental Table S2** presents the characteristics of all the SNPs analyzed.

Genotyping was carried out on fragmented gDNA at a final concentration ranging from 45 to 225 ng and a final volume of 12.5 µl arrayed into 384 – well plates according to the manufacturer's instructions. In each plate, six negative controls (water) and eighteen positive controls (mixed DNA of known genotypes) were included. The allelic discrimination was detected using GeneMapper® Analysis Software v3.7 (Applied Biosystems Foster City, CA, USA) based on the SNPlex_Rules_3730 method. SNP markers with minimum allele frequencies lower than 10% and more than 10% missing data were discarded from statistical analysis, which were thus performed on 121 markers.

Linkage disequilibrium analysis

The LD extent was calculated on two sets of genotypes, the whole collection and the subset of accessions of *S. l. cerasiforme*. GGT 2.0 (van Berloo et al. 2008) software was used to calculate the squared correlation coefficients r^2 (Zhao et al. 2005) between 121 markers throughout the genome. The decay of LD over genetic distance was investigated by plotting pair-wise r^2 values against the distances at the whole genome level and on chromosome 2 (covered by the largest number of SNPs) for all accessions, then for *S. l. cerasiforme* accessions separately.

Inference of population structure

Structure v2.1 program (Pritchard et al. 2000) was used to estimate the number of sub-populations in the complete set of accessions using admixture model for the ancestry of individuals and correlated allele frequencies. Population structure was modeled with a burning of 2.5×10^5 cycles followed by 10^6 Markov Chain Monte Carlo (MCMC) repeats.

Evanno transformation method was then used to infer the most likely number of populations (K) (Evanno et al. 2005). Structure analysis was obtained on two sets of markers: 121 informative SNPs selected from SNPlex^{TM} assay and compared to the structure obtained with 20 SSR markers (Ranc et al. 2008). Distruc1.1 program (Rosenberg 2004) was used to display the graphics of population structure. The kinship matrix was generated by SPAGeDi (Hardy and Vekemans 2002) software based on the two set of markers: 121 informative SNPs and 20 SSR markers. Diagonal of the matrix was set to two and negative values were set to zero, according to Yu et al. (2006).

Phenotyping

All tomato accessions (4 plants per accession) were grown in a plastic greenhouse in Avignon (south of France) during summers 2007 and 2008. Three harvests of ten ripe fruits per accession were used as repeats in the phenotypic analysis. Ten fruits were evaluated for fruit weight (FW), firmness (FIR), soluble solids content (SSC), sugar content (SUG), locule number (LCN), pH, titratable acidity (TA), color components: lightness (L), color on red to green (a*) and on yellow to blue (b*) scales with a Konica Minolta CR-300 chromameter. All measurements were performed as described in Saliba-Colombani et al. (2001).

Statistical analysis

The heritabilities were estimated on the collection of homozygous lines. Heritabilities (h^2) were calculated as $h^2=\sigma^2_g/(\sigma^2_g+\sigma^2_{gy}+\sigma^2_e)$ with σ^2_g, σ^2_{gy} and σ^2_e the genetic, genetic by environment interaction and residual variance, respectively. σ^2_g, σ^2_{gy} and σ^2_e were estimated by (MSc-MScy)/ry, (MScy-MSe)/r and MSe, respectively. MSc represents the accessions mean square, MScy represents the mean square of genotype by year interaction, MSe the residual mean square. r and y represents the number of replicates and the number of years. Associations were tested using the adjusted means of accessions calculated by general linear

model. The Pearson coefficients correlations were calculated for all pairs of variables. Analyses were carried out with the R program (R Development Core Team 2005).

Association Mapping

The association analyses were performed on the whole collection of 188 accessions and on 127 *S. l. cerasiforme* accessions by performing mixed linear model (MLM, K+Q model) as described by Yu et al. (2006) and implemented in Tassel 2.1 software (Bradbury et al. 2007). Two MLM models were used, with structure and kinship based on 20 SSR marker (model A), and on 121 SNP markers (model B). P-values were corrected following the standard Bonferroni procedure. Significant associations were detected with corrected p-value lower than 0.005 (an arbitrary choice of threshold due to the poor correction of the structure) for model A, and 0.05 for model B. After obtaining the significant markers, a general linear model with all fixed-effect terms was used to estimate R^2, the amount of phenotypic variation explained by each marker. The Pearson correlation coefficients were calculated between Q1 value (the probability that an individual belongs to the first subpopulation defined by 20 SSR markers and 121 SNPs) and each phenotypic trait.

Surrounding sequences of the 121 SNPs used in association analysis were blasted against Tomato whole genome sequence Chromosomes (SL2.40) database and ITAG (International Tomato Annotation Group) release 2.3 database (www.solgenomics.net) in order to map them on the physical map and get the annotation of surrounding genes. Physical map locations of the 121 SNPs were converted to genetic map on the tomato EXPEN 2000 map (www.solgenomics.net) based on the closest mapped markers and on the assumption of a local linear relationship between physical and genetic distance. Markers which have been previously mapped in QTL studies (Goldman et al. 1995; Grandillo and Tanksley 1996; Saliba-Colombani et al. 2001) were also aligned using BLAST on the genome sequence or

directly mapped on the EXPEN 2000 map to get their corresponding genetic position. SNPs significantly associated with a trait were considered as co-localizing with previous QTLs if they were located in a window of 20 cM surrounding the QTLs. Genetic map with associations and previously identified QTLs were plotted using MapChart v2.1 software (Voorrips 2002).

Table 1 Phenotypic variation of fruit quality traits in the whole collection (all), *S. lycopersicum* (lyco), *S. l. cerasiforme* (cera) and *S. pimpinellifolium* (pimp) group, respectively. Traits are described by mean, standard deviation (SD), significant level in group and heritability.

Trait	Phenotypic variation				Heritability			
	all (N [a] = 188) Mean ± SD	lyco (N [a] = 44) Mean ± SD	*cera* (N [a] = 127) Mean ± SD	*pimp* (N [a] =17) Mean ± SD	all (N [a] =188)	lyco (N [a] = 44)	cera (N [a] = 127)	pimp (N [a] =17)
a*	15.47 ± 4.98 ***	18.51 ± 4.64 ***	14.32 ± 4.90 ***	16.16 ± 2.63 ***	0.81	0.77	0.82	0.54
b*	13.72 ± 5.47 ***	14.81 ± 4.47 ***	13.75 ± 5.88 ***	10.69 ± 3.33 ***	0.76	0.69	0.79	0.66
L	42.67 ± 3.4 ***	43.35 ± 2.28 ***	42.94 ± 3.58 ***	38.94 ± 2.06 ***	0.61	0.47	0.61	0.36
FIR	53.22 ± 7.43 ***	52.90 ± 8.58 ***	52.12 ± 6.30 ***	62.21 ± 6.10 ***	0.72	0.73	0.66	0.75
FW	33.03 ± 46.56 ***	96.29 ± 61.80 ***	15.19 ± 8.40 ***	2.65 ± 1.03 ***	0.83	0.75	0.86	0.87
LCN	3.22 ± 2.24 ***	5.19 ± 3.87 ***	2.69 ± 0.71 ***	2.09 ± 0.11 ***	0.85	0.81	0.78	0.34
pH	4.08 ± 0.13 ***	4.11 ± 0.12 ***	4.08 ± 0.12 ***	4.06 ± 0.15 ***	0.57	0.58	0.68	0.26
SSC	7.38 ± 1.34 ***	6.37 ± 0.84 ***	7.42 ± 1.09 ***	9.69 ± 1.18 ***	0.73	0.62	0.62	0.58
SUG	4.02 ± 1.54 ***	2.99 ± 0.92 **	4.04 ± 1.22 ***	6.49 ± 2.12 *	0.63	0.55	0.55	0.56
TA	11.10 ± 2.50 ***	9.44 ± 2.05 ***	11.30 ± 2.23 ***	13.84 ± 2.59 ***	0.75	0.73	0.70	0.73

[a] Number of accessions.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

a*, b*, L= color, FIR= firmness, FW= fruit weight, LCN =locule number, pH= pH , SSC = soluble solids content, SUG = sugar content, TA = titratable acidity

Results

Phenotypic variation and correlations of fruit quality traits

All the traits showed a large range of phenotypic variation in the whole collection, and within the three groups composed of 44 *S. lycopersicum*, 127 *S. l. cerasiforme*, and 17 *S. pimpinellifolium* (**Table 1**). Heritabilities were high (ranging from 0.57 to 0.85) in the whole collection with lower values for L, a*, LCN and pH in the *S. pimpinellifolium* group which exhibited a lower genetic variability. Phenotypic correlations among the fruit quality traits in the whole collection and within groups are detailed in **Supplemental Table S3**. Color components L, a* and b* were highly correlated with each other, and moderately correlated with the other traits in the whole collection, and in the *S. l. cerasiforme* and *S. pimpinellifolium* groups. However, a* was not significantly correlated with b* in the *S. lycopersicum* group. Few significant correlations were observed between FIR and the nine other traits in the whole collection and in the *S. l. cerasiforme* group. Correlations were higher between FIR and the nine other traits in *S. lycopersicum* and *S. pimpinellifolium* groups than in the *S. l. cerasiforme* group. FW was strongly positively correlated with LCN, and negatively correlated with SSC and TA in the whole collection and in each group. It was also negatively correlated with SUG in the whole collection and in the *S. l. cerasiforme* and *S. lycopersicum* groups. SSC and SUG were highly significantly correlated with each other in all the groups. pH was negatively correlated with TA in the whole collection and in the three groups. It was significantly correlated with SUG in *S. l. cerasiforme* and *S. pimpinellifolium* group.

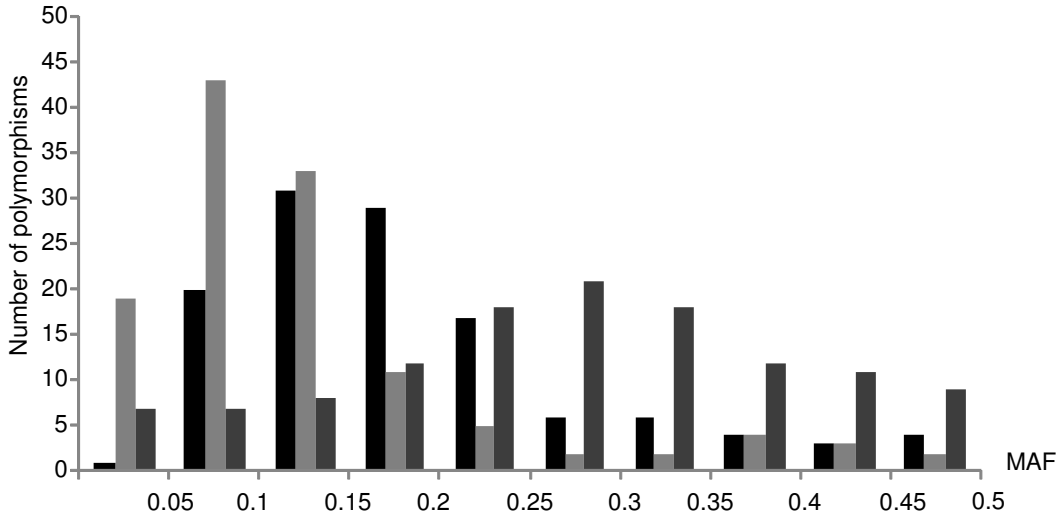


Fig. 1 Distribution of minimum allele frequencies (MAF) of the 121 SNPs in the three tomato groups (*S. l. cerasiforme* (N=127) represented in black, *S. lycopersicum* (N=44) in light gray and *S. pimpinellifolium* (N=17) in dark gray). Polymorphisms with MAF lower than 0.10 in the whole collection were previously discarded

Molecular polymorphism

The 188 accessions were genotyped with 192 SNPs, combined in four 48-plex panels, among which 139 SNPs (73%) were successfully scored. Three SNPs with more than 10% missing data and 15 SNPs with minimum allele frequencies (MAF) lower than 10% were removed from further analysis. Finally, 121 informative SNPs were used for polymorphism and association analysis. The results were very similar among the 4 SNPlex^{TM} panels (**Supplemental Table S4**). The distributions of the MAF were different among the three groups of accessions (**Fig. 1**). The average MAF was 0.26, 0.18 and 0.12 for *S. pimpinellifolium*, *S. l. cerasiforme* and *S. lycopersicum,* respectively.

LD decay was analyzed separately for all markers and for the 50 markers on chromosome 2 (carrying the largest number of markers) for the 188 accessions and for 127 *S. l. cerasiforme* accessions. Pairwise r^2 were plotted according to genetic distance between loci and non-linear regression fitted the decay of LD over genetic distance. LD on the whole genome for all accessions extended on average over 18 cM for $r^2 = 0.3$ (**Fig. 2a**), with the same pattern if only 50 markers of chromosome 2 were analyzed (**Fig. 2b**). The *S. l. cerasiforme* accessions had lower LD (reaching $r^2 = 0.3$ for 10 cM) as illustrated for chromosome 2 (**Fig. 2c**). The same pattern of LD was observed when all the SNPs were taken into account, with a few loci in strong LD with several markers responsible for high LD values.

Population structure

The structure of a collection of 360 accessions, including the 188 accessions studied here, was assessed with 20 SSR markers spread over the genome by Ranc et al. (2008). Four groups were detected, but the subset of 188 accessions studied here revealed two main groups, a cultivated and a wild group and many intermediate types (**Fig. 3a**). The genetic structure of the 188 accessions assessed with the 121 SNP markers revealed the same pattern with two

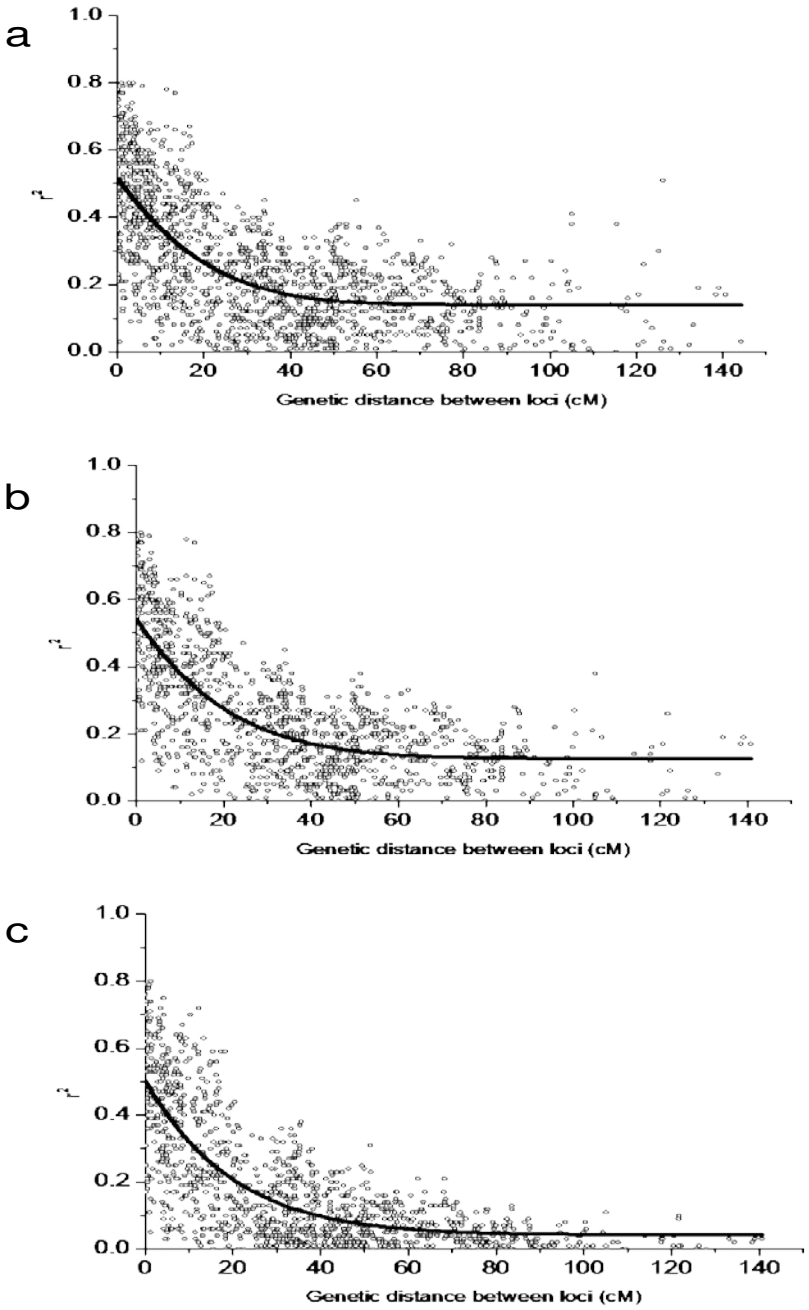


Fig. 2 Decay of linkage disequilibrium (r^2) over genetic distance: (a) on all chromosomes for all accessions, (b) on chromosome 2 for all accessions, (c) on chromosome 2 for S. l. cerasiforme accessions. Each plot of r^2 over genetic distance is fitted by non linear regression (black curve). Genetic distance corresponding to r^2=0.3 is indicated.

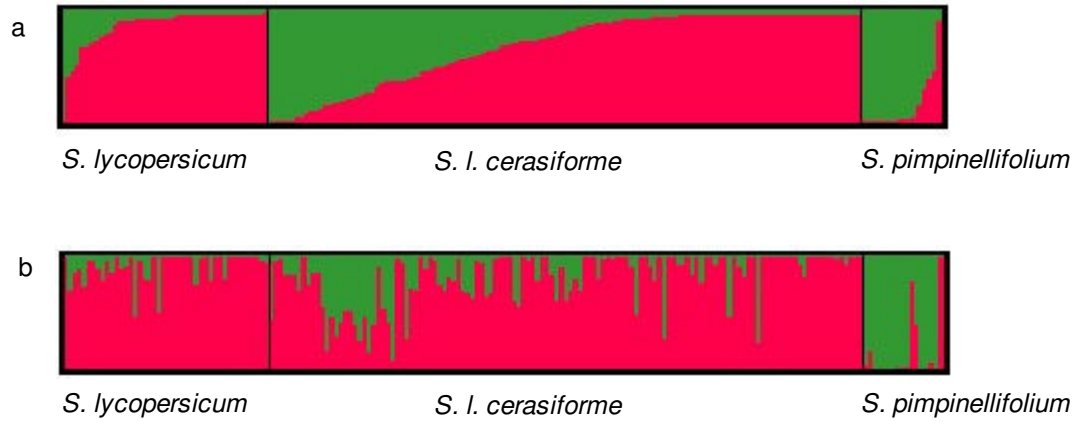


Fig. 3 Comparison of population structure generated from the genotypes with two different types of markers: (a) 20 SSR markers, (b) 121 SNP markers. Accessions are in the same order for the two types of markers.

Table 2 Significant associations for color (a), color (L), firmness (FIR), fruit weight (FW), locule number (LCN), soluble solid content (SSC), sugar content (SUG) and titratable acidity (TA) estimated with K + Q models on 188 accessions. Model A: MLM model, with structure and kinship based on 20 SSR (*p*-values lower than 0.005 are shown with indication on allele effect); model B: MLM model with structure and kinship based on 121 SNP (*p*-value lower than 0.05 are shown). Only the most significant association from each group is shown. See Supplemental Table S5 for detail

Trait (Correlation with Q1 [a])	SNP group [b]	Chromo-some	Locus	Location [c]	Model A		MAF [f]	Model B
					Corrected *p* –value [d]	R^{2} [e]		Corrected *p* –value [d]
a*	2.5	2	TD083-685	133.1	0.042 ns	0.09	0.47	0.011
(0.02 - 0.11)								
L	2.2	2	TD049-528	72.4	0.003	0.09	0.36	ns
(0.29 - 0.43)	2.4	2	Z2117-98	120.5	0.002	0.09	0.22	ns
	4.1	4	TD200-317	6.6	0.004	0.09	0.19	ns
	4.2	4	TD160-458	27.7	4.53×10^{-04}	0.11	0.31	ns
	9.2	9	Z1475-87	51.0	6.47×10^{-04}	0.10	0.17	ns
	9.4	9	TD168-241	99.5	7.78×10^{-04}	0.09	0.13	ns
FIR	1.1	1	CON203-643	44.7	1.86×10^{-04}	0.11	0.22	ns
(0.3 – 0.49)	2.1	2	TD091-657	49.5	0.002	0.07	0.22	ns
	2.4	2	TD113-132	121.6	9.74×10^{-06}	0.14	0.14	ns
	3.2	3	CON174-206	102.3	8.59×10^{-06}	0.13	0.21	ns
	4.3	4	Z1703-106	65.5	1.51×10^{-07}	0.17	0.12	0.194 ns
	4.4	4	TD212-247	82.2	3.27×10^{-11}	0.24	0.17	0.011
	4.5	4	CON219-313	122.0	2.20×10^{-06}	0.15	0.22	ns
	4.5	4	CON105-290	131.7	0.001	0.10	0.12	ns
	5.1	5	CON173-501	68.3	1.16×10^{-07}	0.16	0.23	ns
	5.2	5	CON222-388	75.8	2.72×10^{-04}	0.11	0.19	ns
	6.1	6	TD025-87	20.8	0.002	0.09	0.12	ns
	9.2	9	TD167-449	59.8	6.84×10^{-04}	0.10	0.16	ns
	10.1	10	CON176-455	59.0	1.01×10^{-07}	0.17	0.12	.0.117 ns
	11.1	11	TD247-57	6.4	3.10×10^{-05}	0.13	0.15	ns
	11.2	11	TD251-230	35.4	6.66×10^{-05}	0.12	0.12	ns
	11.3	11	CON141-576	54.8	7.38×10^{-07}	0.16	0.18	ns
	11.4	11	CON50-294	62.3	9.55×10^{-07}	0.16	0.13	ns
	12.2	12	TD156-314	93.7	0.004	0.08	0.12	ns
Log (FW)	1.2	1	TD011-260	97.3	2.68×10^{-04}	0.08	0.13	ns
(0.47 – 0.53)	2.3	2	TD133-395	83.8	7.30×10^{-06}	0.12	0.34	ns
	2.4	2	TD116-707	122.2	7.38×10^{-08}	0.16	0.34	0.007
	2.5	2	TD083-685	133.1	2.42×10^{-04}	0.12	0.47	0.157 ns
	3.1	3	CON57-121	63.5	0.003	0.07	0.13	ns
	3.3	3	TD152-159	167.1	0.002	0.09	0.17	ns
	4.1	4	TD200-317	6.6	0.001	0.08	0.19	ns
	4.3	4	CON300-472	68.6	1.000 ns	0.01	0.47	0.045
	4.4	4	CON130-112	91.8	9.21×10^{-04}	0.10	0.23	0.077 ns
	6.1	6	TD025-87	20.8	0.001	0.08	0.12	ns
	9.1	9	Z1707-100	38.0	0.003	0.08	0.17	ns
	9.2	9	Z1475-87	51.0	1.08×10^{-04}	0.11	0.17	ns
	9.3	9	Z2305-99	69.4	1.55×10^{-05}	0.11	0.22	ns
	9.4	9	TD243-38	114.4	2.32×10^{-06}	0.13	0.31	0.206 ns
	10.1	10	CON369-191	52.6	9.34×10^{-04}	0.09	0.34	ns
	12.3	12	TD008-95	112.5	5.77×10^{-05}	0.09	0.19	ns
Log (LCN)	2.2	2	TD049-528	72.4	1.54×10^{-04}	0.09	0.36	ns
(0.24 – 0.30)	2.3	2	TD133-395	83.8	9.27×10^{-05}	0.14	0.34	0.045
SSC	1.2	1	TD011-260	97.3	0.003	0.06	0.13	ns
(0.41 – 0.47)	1.3	1	Z2300-99	140.6	9.56×10^{-04}	0.10	0.12	ns
	2.3	2	TD280-108	88.6	8.94×10^{-05}	0.13	0.43	0.023
	2.4	2	TD116-707	122.2	0.001	0.09	0.34	ns

Table 2 Continued -1

Trait (Correlation with Q1 [a])	SNP group [b]	Chromo-some	Locus	Location [c]	Model A		MAF [f]	Model B
					Corrected *p* –value [d]	R^2 [e]		Corrected *p* –value [d]
SSC	2.5	2	TD178-104	138.4	0.020 ns	0.12	0.11	0.037
(0.41 – 0.47)	3.1	3	CON57-121	63.5	0.001	0.08	0.13	ns
	4.1	4	TD200-317	6.6	1.80×10^{-05}	0.12	0.19	ns
	6.1	6	TD025-87	20.8	1.71×10^{-04}	0.11	0.12	ns
	9.1	9	Z1707-100	38.0	3.54×10^{-07}	0.16	0.17	ns
	9.2	9	Z1475-87	51.0	0.003	0.07	0.17	ns
	9.3	9	TD237-253	70.0	0.003	0.07	0.14	ns
	9.4	9	TD168-241	99.5	1.02×10^{-04}	0.12	0.13	ns
	10.1	10	CON369-191	52.6	0.003	0.08	0.34	ns
	11.3	11	TD255-218	56.9	2.09×10^{-06}	0.05	0.13	ns
	11.4	11	CON50-294	62.3	0.001	0.07	0.13	ns
	12.1	12	Z2302-103	21.0	0.004	0.08	0.15	ns
SUG	1.3	1	Z2300-99	140.6	1.06×10^{-05}	0.14	0.12	ns
(0.39 – 0.52)	2.1	2	TD091-657	49.5	0.004	0.07	0.22	ns
	2.2	2	TD139-547	72.2	1.37×10^{-04}	0.08	0.20	ns
	2.3	2	TD133-395	83.8	7.66×10^{-07}	0.15	0.34	0.194 ns
	2.4	2	TD114-259	121.7	6.72×10^{-04}	0.09	0.14	ns
	2.5	2	TD178-104	138.4	1.34×10^{-06}	0.14	0.11	0.206 ns
	3.1	3	CON57-121	63.5	0.001	0.08	0.13	ns
	4.1	4	TD200-317	6.6	2.15×10^{-06}	0.14	0.19	ns
	4.5	4	CON105-290	131.7	5.55×10^{-04}	0.08	0.12	ns
	6.1	6	TD025-87	20.8	1.94×10^{-04}	0.11	0.12	ns
	9.1	9	Z1707-100	38.0	3.68×10^{-05}	0.12	0.17	ns
	9.2	9	Z1475-87	51.0	5.37×10^{-04}	0.09	0.17	ns
	9.4	9	TD168-241	99.5	3.57×10^{-04}	0.11	0.13	ns
	10.1	10	CON369-191	52.6	6.59×10^{-04}	0.10	0.34	0.011
	11.3	11	TD255-218	56.9	4.42×10^{-04}	0.07	0.13	ns
	11.4	11	CON50-294	62.3	1.23×10^{-04}	0.09	0.13	ns
	12.1	12	Z2302-103	21.0	3.81×10^{-04}	0.11	0.15	ns
TA	2.3	2	TD275-101	88.5	0.110 ns	0.04	0.13	0.048
(0.26 – 0.22)	4.3	4	Z1370-98	66.8	ns	0.08	0.14	0.003

[a] Q1 is the probability that an individual belongs to the "cultivated" subpopulation generated from STRUCTURE2.1 software (Pritchard, Stephens et al. 2000). Correlations with the Q value defined by 20 SSR markers and 121 SNPs.

[b] Associated SNPs in less than 10 cM on each chromosome were grouped together. SNP which is 10 cM apart from the other SNPs was assigned as an independent group. Groups are detailed in Supplemental Table S5.

[c] Genetic distance of the marker on EXPEN2000 reference map (http://sgn.cornell.edu/).

[d] *p*-values were corrected following the standard Bonferroni procedure. ns: non significant

[e] R^2 were calculated using a Q model.

[f] Minimum allele frequencies (MAF).

groups (**Fig. 3b**). A similar genetic structure was obtained when combining SNPs and SSR markers (data not shown). The Q1 values of each accession (corresponding to the probability to belong to the "cultivated" group) estimated by the two types of markers were correlated (r^2=0.63), but several accessions were clustered in different groups. We thus compared both structure patterns in the mixed linear models to detect associations. L, FIR, FW, SSC and SUG strongly participated to the structure as shown by the highly significant correlations between these traits and the probability to belong to the cultivated group, with r values ranging from 0.30 to 0.59 (**Table 2**).

Association mapping

Associations between polymorphisms and fruit quality traits were determined by taking into account structure and kinship in a K+Q mixed linear model (MLM) model first on the whole collection (**Table 2 and Supplemental Table S5, Fig. 4 and Supplemental Fig. S1)**. The comparison of the probabilities obtained with either a simple linear model or K+Q model with Q based on SSR markers (Model A) or SNP markers (Model B) showed that Model A was intermediate between the simple linear model and Model B for a*, b*, FIR and TA, and was similar to the simple linear model for L, FW, LCN, and SSC. The three models provided very close results for pH **(Fig. 5 and Supplemental Fig. S2)**. Model B thus corrected better for the structure than the other models and revealed much less significant associations. To reduce the false positive associations, we thus described the associations with Model A using a threshold of the corrected p-value of 0.005 while we used a threshold of 0.05 for Model B. For Model A, 132 significant associations were detected for six traits (L, FIR, FW, LCN, SSC and SUG), with a maximum of 41 (for SUG) and a minimum of 7 (for L) associations per trait. No association was detected for a*, b*, pH and TA. A total of 74 SNPs spread on almost all chromosomes except chromosome 7 and 8 were involved in at least one association. The

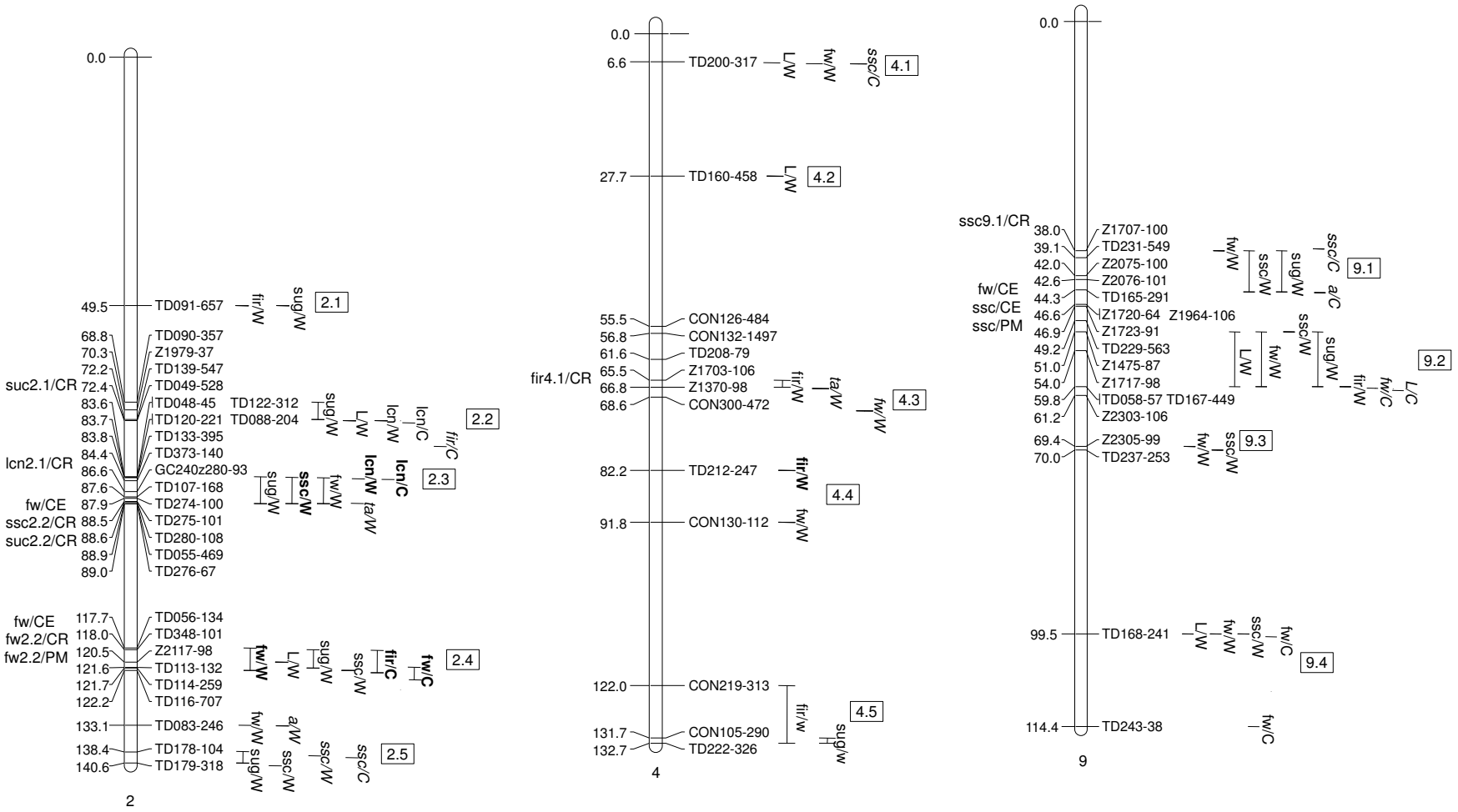


Fig. 4 Comparison of associations and QTLs identified by linkage mapping on chromosomes 2, 4 and 9. SNPs were mapped on tomato EXPEN 2000 reference map (http://sgn.cornell.edu/). Associations detected in the 188 accessions (W) and in 127 *S. l. cerasiforme* accessions (C) are indicated to the right of the chromosomes. Associations were estimated with K + Q model, model A: with structure and kinship based on 20 SSR marker, model B: with structure and kinship based on 121 SNP (common font: associations detected with model A; in italic: associations detected with model B; in bold: associations detected with both models). Horizontal line "-" correspond to the genetic location of associated marker, associations are linked together by a vertical line when linked markers in less than 10 cM are associated to the same trait. Associated SNPs in less than 10 cM on each chromosome were grouped together. SNP which is 10 cM apart from the others were assigned as independent groups. Groups are named as consecutive number according to their genetic location on each chromosome. Traits are: FIR= firmness, FW= fruit weight, SSC= soluble solids content, SUG = to tal sugar content, LCN = locule number, a and L=color, TA = titratableacidity. Only SNPs significantly associated with one trait are represented on chromosome 2 where markers are too dense. QTLs identified by linkage mapping in the populations from crosses of *S. Lycopersicum* × *S. l. cerasiforme* (Saliba-Colombani et al. 2001), *S. Lycopersicum* × *S. pimpinellifolium* (Grandillo et al. 1996) and *S. Lycopersicum* × *S.l. cheesmanii* (Goldman et al. 1995) are shown to the left of the chromosomes (CR= QTL from *S. l. cerasiforme,* CE= QTL from *S.l.cheesmanii,* PM= QTL from *S. pimpinellifolium*). Only QTL co-localizing with an association are show

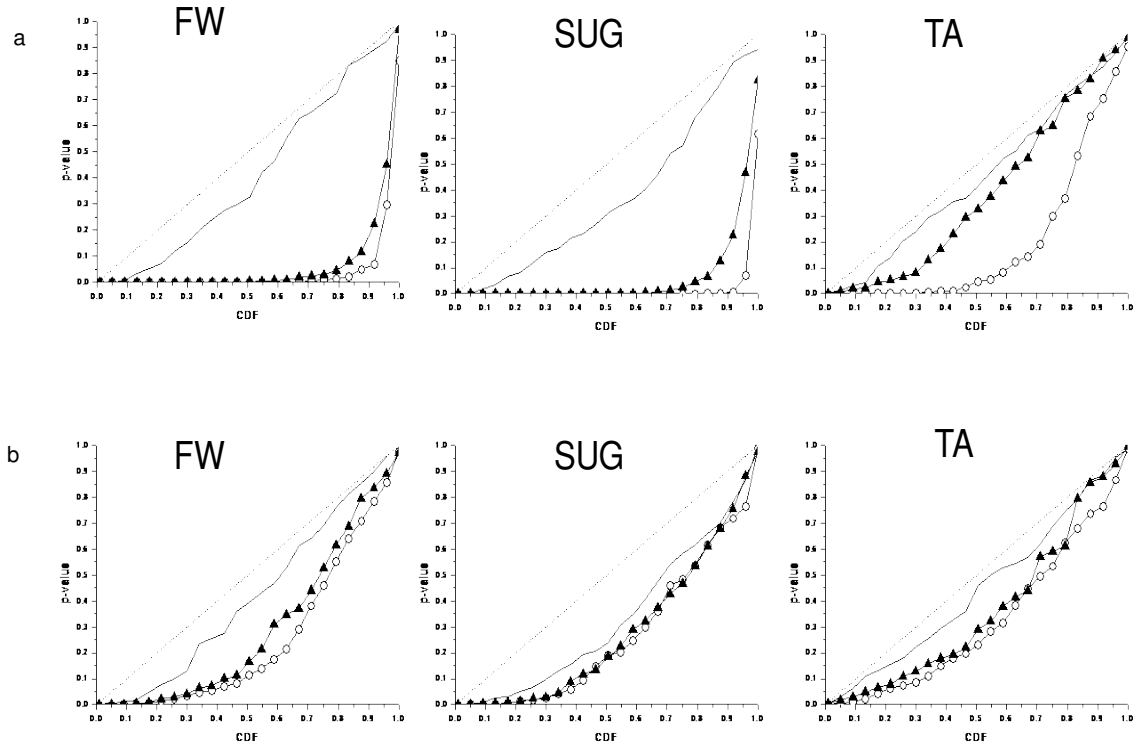


Fig. 5 Cumulative density functions (CDF) using three alternative models of association for fruit weight (FW), sugar content (SUG) and Titratable acidity (TA). Associations are tested for 121 polymorphic sites on 188 accessions (a) and 127 *S. l. cerasiforme* (b). Simple linear model (empty circle) and K+Q models, with structure and kinship based on SSR markers (black triangle), and on 121 SNP markers (black line) were tested. The diagonal indicates uniform distribution of *p*-values under the expectation that random SNPs are unlinked to the polymorphisms controlling these traits (H0: no SNP effect).

percentage of phenotypic variation explained by the SNPs ranged from 4% to 24%. For Model B, 10 associations were significant for 7 traits. Six associations were significant with both models. They concerned FIR on chromosome 4, FW, LCN and SSC on chromosome 2 and SUG on chromosome 10. In order to compare the associations with QTL previously mapped, markers showing significant association and linked in less than 10 cM were grouped together. Thus, 31 groups of association (with one to twenty-seven associations) were defined (**Table 2 and Supplemental Table S5, Fig. 4 and Supplemental Fig. S1).**

For L, six groups of association (involving seven SNPs) were identified with Model A on chromosomes 2 (two SNPs), 4 (two SNPs) and 9 (three SNPs). The two most significant associated markers TD160-458 in group 4.2 on chromosome 4 and Z1475-87 in group 9.2 on chromosome 9 explained 11% and 10% of the phenotypic variation, respectively. No significant association was detected using Model B. For a*, a single association was detected with Model B on chromosome 2 with TD083-685.

For FIR, 16 groups of association (involving 30 SNPs) were detected using Model A on chromosomes 1, 2 (four SNPs), 3 (three SNPs), 4 (six SNPs), 5 (six SNPs), 6, 9, 10, 11 (six SNPs) and 12. The two most significant associations involved markers TD212-247 in group 4.4 on chromosome 4 and CON176-455 in group 10.1, responsible for 24% and 17% of the phenotypic variation, respectively. The association with TD212-247 was also significant with Model B.

For FW, 15 groups of association (involving 23 SNPs) were detected with Model A on chromosomes 1, 2 (eight SNPs), 3 (two SNPs), 4 (two SNPs), 9 (six SNPs), 10 (two SNPs) and 12. Markers TD116-707 in group 2.4 and marker TD243-38 in group 9.4 showed the most significant associations and explained 16% and 13% of the FW variation. Using Model B, TD116-707 and CON300-472 showed significant associations, on chromosomes 2 and 4, respectively.

Table 3 Significant associations for color (a), color (L), firmness (FIR), fruit weight (FW), locule number (LCN), sugar content (SUG), soluble solid content (SSC) estimated with K + Q models on 127 *S. l. cerasiforme* accessions. Model A: MLM model, with structure and kinship based on 20 SSR (only *p*-values lower than 0.005 are shown with indication on allele effect); model B: MLM model with structure and kinship based on 121 SNP (*p*-value lower than 0.05 are shown)

Trait (Correlation with Q1 [a])	Chromo-some	Locus	Location [b]	Model A			Model B
				p –value [c]	R^{2} [d]	MAF [e]	*p* –value [c]
a*	5	CON310-990	71.9	ns	0.01	0.21	0.002
(0.12 – 0.20)	9	Z1723-91	46.9	ns	0.03	0.18	0.008
L	9	TD167-449	59.8	ns	0.03	0.11	0.011
(0.17 – 0.25)							
FIR	2	TD018-103	75.3	ns	0.01	0.23	0.026
(0.25 – 0.35)	2	TG454z273-252	75.5	ns	0.05	0.13	0.040
	2	TD348-101	118.0	0.001	0.12	0.16	ns
	2	TD113-132	121.6	8.13×10^{-4}	0.16	0.10	0.001
	2	TD114-259	121.7	0.002	0.15	0.15	0.018
Log (FW)	2	Z2117-98	120.5	0.001	0.18	0.19	ns
(0.11 – 0.20)	2	TD116-707	122.2	6.75×10^{-7}	0.26	0.36	0.002
	9	TD058-57	59.8	ns	0.05	0.15	3.88×10^{-4}
	9	TD167-449	59.8	ns	0.04	0.11	0.025
	9	TD168-241	99.5	0.004	0.13	0.08	ns
	9	TD243-38	114.4	8.22×10^{-4}	0.15	0.30	ns
Log (LCN)	2	TD049-528	72.4	4.00×10^{-4}	0.16	0.36	ns
(0.13 – 0.18	2	TD120-221	83.7	0.009 ns	0.12	0.19	0.007
	2	TD133-395	83.8	6.67×10^{-5}	0.20	0.33	0.009
SSC	2	TD178-104	138.4	0.028 ns	0.10	0.08	0.045
(0.02 - 0.03)	4	TD200-317	6.6	ns	0.08	0.15	0.015
	5	TD032-112	72.3	0.002	0.17	0.19	0.009
	9	Z1707-100	38.0	0.051 ns	0.10	0.12	0.016
SUG	5	TD032-112	72.3	ns	0.10	0.19	0.033
(0.02 – 0.14)	11	TD247-57	6.4	ns	0.08	0.12	0.015

[a] Q1 is the probability that an individual belongs to the "cultivated" subpopulation generated from STRUCTURE2.1 software (Pritchard, Stephens et al. 2000). Correlations with the Q value defined by 20 SSR markers and 121 SNPs.

[b] Genetic distance of the marker on EXPEN2000 reference map (http://sgn.cornell.edu/)

[c] *p*-values were corrected following the standard Bonferroni procedure. ns: non significant

[d] R^{2} were calculated using a Q model.

[e] Minimum allele frequencies (MAF).

For LCN, two groups of association (involving two SNPs) were identified on chromosome 2. The two association involved marker TD133-395 in group 2.3 (also significant with Model B) and TD049-528 in group 2.2, explained 14% and 9% of the phenotypic variation.

For SSC, 16 groups of association (involving 28 SNPs) were detected with Model A on chromosomes 1 (two SNPs), 2 (ten SNPs), 3, 4, 9 (eight SNPs), 10 (two SNPs), 11 (two SNPs) and 12. The most significant associations involved markers Z1707-100 in group 9.1 and TD255-218 in group 11.3, explained 16% and 5% of the soluble solid variation. Two associations on chromosome 2 were significant with Model B.

For SUG, 17 groups of association were identified on chromosomes 1, 2 (22 SNPs), 3, 4 (three SNPs), 6, 9 (eight SNPs), 10 (two SNPs), 11 (two SNPs) and 12 with Model A. The strongest association involved marker TD133-395 in group 2.3 and TD178-104 in group 2.5 on chromosome 2, which explained 15% and 14% of the sugar content variation. The only significant association with Model B involved CON369-191 on chromosome 10.

We then performed association analysis using Model A and B on the subset of 127 *S. l. cerasiforme* accessions (**Table 3, Fig. 4 and Supplemental Fig. S1**). Population structure accounted for much less variation of all the traits, with the highest correlation between Q-values and FIR (**Table 3**). The comparison of the probabilities associated to the tests using simple linear model, Model A and Model B showed that Model A was still intermediate between simple linear model and Model B for L, FW, FIR and was similar to the simple linear model for LCN. The three models were very close for a*, b*, SUG, SSC, pH and TA (**Fig. 5 and Supplemental Fig. S3**). For Model A, ten significant associations were found for FIR, FW, LCN and SSC. Population structure accounted for 25%, 11%, 13% and 2% of the phenotypic variation for these traits, respectively. Eight significant associations were common with associations found with 188 accessions. Two new associations were observed between marker Z1117-98 and FW, TD032-112 and SSC, responsible for 18% and 17% variation,

respectively. No significant associations were found for a*, b* and L, pH and TA. With Model B, 18 significant associations were detected for seven traits. Three associations were common with associations found in 188 accessions. Five associations were detected with both models, for FIR and for LCN on chromosome 2, for FW on chromosome 2 and 9 and for SSC on chromosome 5. Only two associations were common to the two Models and the two samples, one for FW with TD116-707 and one for LCN with TD133-395.

Discussion

We herein present the phenotypic and genetic diversity of a large collection of tomato accessions representing wild relatives, intermediate and cultivated types characterized using a SNPlex™ genotyping assay. The percentage of SNPs successfully scored (73%) is consistent with the success rate reported by Pindo et al. (2008) and Berard et al. (2009). The results suggest that this assay is reliable, flexible and cost-effective for medium-throughput SNP detection. This pioneering technology opened the way to new technologies more flexible or with higher throughput like the one proposed by Fluidigm (Moonsamy et al. 2011) or Illumina GoldenGate™. Although SNPlex™ assay is no more used, the SNPs used in this assay may be adapted to other genotyping platforms to be used by tomato breeders.

A source of phenotypic variability

The sample consisted of 127 cherry type tomato accessions *S. l. cerasiforme*, 44 *S. lycopersicum* large fruited accessions and 17 *S. pimpinellifolium* accessions. The genome structure of *S. l. cerasiforme* accessions was previously described as a mosaic of *S. lycopersicum* and *S. pimpinellifolium* genomes (Ranc et al. 2008). Compared to *S. lycopersicum*, *S. l. cerasiforme* and *S. pimpinellifolium* fruits are much smaller (less than 20 g for cherry types, less than 5 g for wild types) with only 2 or 3 locules and higher sugar content, soluble solid content and titratable acidity. The same trend of variation was already observed in smaller samples (Davies and Hobson 1981; Causse et al. 2003). Correlations among traits were quite homogenous in the whole collection and among the three groups. *S. pimpinellifolium* and cherry type accessions may be useful sources of alleles for tomato fruit quality improvement, particularly to improve firmness, or the content in sugars and acids, but the strong negative correlation between fruit size and soluble solids or sugar content may

hamper the simultaneous improvement of both traits. A better knowledge of the loci controlling these traits may thus help breeders to use these resources.

LD decay and population structure

Association mapping requires a thorough understanding of LD and population structure in the collection. In tomato, LD remains high over genetic distance. Robbins et al. (2011) observed that LD decayed at 6-8 cM in a collection of 102 tomato varieties, 6-14 cM within 39 processing varieties, and 3-16 cM within 24 fresh market varieties. In our study, slightly higher level of LD was observed in the whole collection, although it was lower in *S. l. cerasiforme.* This result is consistent with van Berloo et al. (2008) who found LD extent to 15 to 20 cM using AFLP markers in a sample of 18 cherry tomato accessions. Such extent of LD will allow the identification of regions carrying QTL rather than direct associations with candidate genes. Nevertheless, at the physical scale, some SNPs may appear in complete equilibrium with their neighbors (Munos et al, 2011). At the physical scale, recombination hotspots may be detected with very low LD in short distances (Ranc et al, 2012). When considering LD among chromosomes, only a few pairs of markers (less than 10 loci) exhibited high LD (data not shown). SSR and SNP markers revealed similar structure patterns with two main groups and many intermediates. This result is consistent with Hamblin et al. (2007) who compared the structure based on 89 SSR to the structure based on 847 SNPs in a set of 259 maize lines. The SSRs performed better to cluster the germplasm into populations but the population structure assessed by both marker types was similar. Laval et al. (2002) stated that *k-1* times more bi-allelic markers are needed to obtain the same genetic distance accuracy as a set of microsatellites with *k* alleles. In the present study, the average number of alleles per SSR locus was about 7, thus 20 SSR markers should correspond to 120 biallelic SNP markers and should thus provide the same accuracy. Nevertheless several individuals were not classified in the same groups and both structures did not correct for structure the

same way (fig. 5). The SSR were less efficient than SNP markers. This result may be due to the fact that the SNPs revealed more loci than SSR and were for a large extent chosen to be located in regions where QTL were previously detected in crosses between one wild or cherry accession and a cultivated one, and thus may be linked to the polymorphisms responsible of the structure.

Associations confirmed previously identified QTLs and detected new candidate polymorphisms

Compared with previous association analysis (Nesbitt and Tanksley 2002; Mazzucato et al. 2008; Munos et al. 2011), it is the first time that associations are analysed between more than hundred SNP markers and ten tomato fruit quality traits in a large collection. Ranc et al (2012), in a pilot study on chromosome 2 and 90 accessions, showed that association mapping permitted to map QTLs that were already cloned. They showed that to get just a location of major QTLs, a few thousands SNPs will be sufficient, while their precise characterisation and the identification of mutations which have evolved under balancing selection and introgressed into many accessions (like *Lcn2.1)* may require a much larger number of SNPs (more than 50,000). The way we take into account the population structure influences the results, as population structure may cause false association results (Mezmouk et al. (2011). Statistical methods have been developed to deal with the effect of population structure (Pritchard et al. 2000; Price et al. 2006; Yu et al. 2006). The MLM model has been shown to efficiently correct for the effects of population structure by including the structure and a matrix of genetic similarity among the accessions (Yu et al. 2006; Atwell et al. 2010). Population structure accounted for a large part of the phenotypic variation for several traits in the 188 tomato accessions and accounted for much less phenotypic variation in the 127 *S. l. cerasiforme* accessions. Although a structure with two subgroups was detected with both SSR and SNP markers, the classification of some individuals changed. It thus appeared that model

A did not correct well for the structure, leading to a large number of associations, particularly for FW, which is strongly correlated with structure. To reduce the false positive associations, a more stringent p-value threshold was used in Model A. In the collection of 127 *S. l. cerasiforme* accessions the structure is less significantly correlated with the trait values and the number of associations detected with both models were less different. Associations were found for most of the traits. With Model A, 132 and 10 SNPs were associated with the traits, for all accessions and *S. l. cerasiforme* accessions, respectively, while with Model B, these numbers were 11 and 18. Only eight and three associations were detected in both sets of accessions when using Model A and B, respectively. Associations between markers and fruit quality traits were mostly localized on chromosome 2, 4 and 9, but this is partly due to the higher number of markers representing these chromosomes (50, 12 and 18 markers, respectively) than the other chromosomes. Almost all marker groups, except group 4.2, were associated with two or more traits. For example, group 2.3 was associated with sugar, soluble solid content, fruit weight, locule number and titratable acidity. Such co-localization of associations for several traits was found in several studies (Zhao et al. 2011; Bergelson and Roux 2010). Co-localized associations for soluble solid content and sugar content, fruit weight and locule number were also frequent. Such co-localization might be related to the pleiotropic effects of the same genes or due to genetic linkage, as already shown for QTL (Lecomte et al, 2004).

In tomato, QTLs for fruit size, shape and quality traits have been mapped in several bi-parental populations involving one wild species (Paterson et al. 1991; Goldman et al. 1995; Eshed and Zamir 1995; Grandillo and Tanksley 1996; Frary et al. 2000; van der Knaap and Tanksley 2001; Causse et al. 2002; van der Knaap and Tanksley 2003; Barrero and Tanksley 2004; Causse et al. 2004; Lecomte et al. 2004). A few genes controlling fruit trait QTL have been cloned, like *FW2.2* which controls fruit weight (Frary et al. 2000), *Lin5* which is

responsible for fruit sugar content (Fridman et al. 2000) or Lcn2.1 which controls locule number (Munos et al. 2011).

The localization of associations for eight quality traits (a*, L, FIR, FW, LCN, SUG, SSC and TA) were compared with those of QTLs previously detected from populations derived from crosses of *S. lycopersicum* × *S. l. cerasiforme* (hereafter named EC × CR) (Saliba-Colombani et al. 2001), *S. lycopersicum* × *S. pimpinellifolium* (hereafter named EC × PM) (Grandillo and Tanksley 1996) and *S. lycopersicum* × *S. l. cheesmanii* (hereafter named EC × CE), another species closely related to the cultivated tomato (Goldman et al. 1995). An association was considered to be in the same region as a QTL when it mapped within a 20 cM region of the tomato EXPEN 2000 map (www.solgenomics.net) around the QTL. On average, 30% of the associations were localised in a region where a QTL for the same trait has been mapped. With model A, 40 associations (two for FIR, ten for FW, two for LCN, 12 for SUG and 14 for SSC) were co-localized with previously identified QTL or known genes, and many other associations for these traits and associations for a*, L and TA were detected in regions where no known QTL have been located to date (**Fig. 4 and Supplemental Fig. S1**). More Associations were found to be co-localized with previously identified QTL with model A than with model B because on one hand model B revealed less associations and on another hand, the structure being better taken into account, the QTL responsible of this structure may be more difficult to detect.

For FW, the markers of group 2.4 associated with FW shared the same position as the major QTL, fw2.2. Actually, TD056-134 corresponds to a polymorphism in fw2.2 promoter. Nesbitt and Tanksley (2002) failed to detect any association in the region around fw2.2 in a small set of *S. l. cerasiforme* accessions, but using a larger sample, Ranc et al (2012) could detect one and identify the accessions that carry the large fruit allele in this region. In the present study, the association was detected with both Models with TD116, a marker less than 2 cM far from fw2.2. Association of group 2.3, group 3.1 and group 9.1 were also co-localized with QTL for

FW detected in EC × CE population (Goldman et al. 1995). For LCN, association of group 2.3 co-localized with lcn2.1, a QTL controlling LCN identified in the EC × CR population (Saliba-Colombani et al. 2001) and recently cloned by Munos et al. (2011). Sequencing 1800 bp around the lcn2.1 locus in 90 accessions allowed the identification of two SNPs strongly associated with the variation of locule number of tomato fruit. These SNPs were also associated in the 188 accessions. TD133-395 was in less than two cM from these two SNP. These results show that our resolution do not allow the precise localisation of the responsible genes, but we may detect regions carrying relevant QTLs. The combination of QTL fine mapping and association study is much more efficient for this purpose.

For SUG, associations of group 2.2 and group 2.3 co-localized with two QTLs detected in the EC × CR population (Saliba-Colombani et al. 2001). Marker TD274 (group 2.3) was also strongly associated with FW and SSC. It was defined in the 5' region of the gene *Solyc02g085170.2* coding for a Glucose transporter protein. Marker TD055-469 (group 2.3) was also found to be associated with FW and SSC. It was designed in the 5' region of a gene *Solyc02g085500.2* coding for the *ovate* protein. The Ovate locus is responsible for pear fruit shape and no effect of this locus on FW, SSC or SUG has been reported before. This polymorphism could thus only be linked to a causative polymorphism. Association of group 10.1 with sugar content was co-localized with a QTL detected for sugar content in EC × CR population (Saliba-Colombani et al. 2001). For SSC, association of group 2.3 was located in the same region as QTL for SSC detected in the EC × CR population (Saliba-Colombani et al. 2001). Association of group 3.1 was found to be co-localized with a QTL for SSC in the EC × PM population (Grandillo and Tanksley 1996). Association of group 9.1 was co-localized with a QTL detected in the three studies. Marker Z1707-100 in this group was closely linked to the previously cloned QTL *lin5* (Fridman et al. 2000) and found in association with soluble

solid content. For FIR, the main association of group 4.3 was located in the same region as a QTL detected in the EC × CR population (Saliba-Colombani et al. 2001).

In conclusion, we identified several associations between SNP markers and fruit traits in a large sample of tomato accessions. The large LD and frequently low MAF in the cultivated group may hamper association discovery in this group. The *S. l. cerasiforme* accessions represent intermediate type between cultivated and wild species with various degrees of introgression as shown by the admixture structure of these accessions. This group exhibited higher MAF on average than cultivated group, lower LD and a less structured pattern. Association mapping should thus be easier with this group. About half of the associations detected with Model B were also detected with Model A in both sets of accessions. Around 30% of the associations detected with Model A were localized in regions where QTLs were previously mapped. We thus presented these results although we are aware that several of these associations may be false positives. This approach will thus have to be combined with QTL fine mapping to identify the relevant polymorphisms as suggested by Nemri et al (2010). Model B allowed the detection of 25 associations. Nevertheless, the density of SNP is too low to identify SNPs in candidate genes. The availability of large panels of SNPs (Sim et al, 2012) will soon allow whole genome scan for association.

Acknowledgements

The research was financed by INRA AIP Bioresources project. Jiaxin Xu was supported by China Scholarship Council. We thank Hélène Burck for characterizing and maintaining the INRA tomato Genetic Resources collection. We are grateful to Yolande Carretero, Esther Pelpoir and Laure David for their help with growing and phenotyping cherry tomato accessions.

Supplemental Data

Supplemental Table S1 List of accession origins, genotypic and phenotypic data.***(too large to be included in the thesis)***

Supplemental Table S2 Characteristics of all the SNPs analyzed ***(too large to be included in the thesis)***

Supplemental Table S3 Phenotypic correlations among quality traits in the whole collection and three groups of accessions

Supplemental Table S4 Summary of SNPs analyzed on the 4 SNPlex panels

Supplemental Table S5 Significant associations for eight fruit quality traits estimated with K + Q models on 188 accessions

Supplemental Fig. S1 Comparison of QTLs identified by linkage mapping and association studies on chromosomes 1, 3, 5, 6, 10, 11 and 12

Supplemental Fig. S2 Cumulative density functions (CDF) using three alternative models of association in the 188 collection

Supplemental Fig. S3 Cumulative density functions (CDF) using three alternative models of association in the 127 *S. l. cerasiforme* accessions

Supplemental Table S3 Phenotypic correlations among quality traits in the whole collection (188 accessions), *S. lycopersicum* (44 accessions), *S. l. cerasiforme* (127 accessions), and *S. pimpinellifolium* (17 accessions) groups, respectively.

Trait	a		b		L		FIR		FW [a]		LCN [a]		pH		SSC		SUG		TA	
a*	All	Lyco																		
	Cera	Pimp																		
b*	-0.50**	ns	All	Lyco																
	-0.72**	-0.71**	Cera	Pimp																
L	-0.55**	-0.17*	0.86**	0.76**	All	Lyco														
	-0.73**	-0.46**	0.89**	0.70**	Cera	Pimp														
FIR	0.22**	0.37**	ns	0.60**	ns	0.53**	All	Lyco												
	0.20*	-0.46**	ns	ns	ns	-0.20**	Cera	Pimp												
FW [a]	0.25**	0.20*	0.23**	0.18*	0.33**	ns	-0.21*	ns	All	Lyco										
	ns	0.34**	0.17*	-0.29*	0.29**	ns	ns	-0.29**	Cera	Pimp										
LCN [a]	0.17*	ns	ns	-0.19*	ns	ns	-0.22**	-0.39**	0.66**	0.61**	All	Lyco								
	ns	0.39**	ns	-0.56**	0.19*	ns	ns	-0.32**	0.51**	0.58**	Cera	Pimp								
pH	ns	0.36**	ns	ns	ns	ns	ns	ns	ns	0.29**	ns	ns	All	Lyco						
	ns	ns	ns	ns	ns	-0.20*	ns	0.35**	ns	ns	-0.26**	ns	Cera	Pimp						
SSC	-0.17*	ns	-0.23**	ns	-0.30**	ns	ns	ns	-0.67**	-0.37**	-0.38**	ns	ns	ns	All	Lyco				
	-0.19*	0.34*	ns	-0.36**	-0.18*	-0.39**	-0.20*	0.19*	-0.49**	-0.31**	-0.29**	ns	0.24*	0.41**	Cera	Pimp				
SUG	-0.20*	ns	ns	ns	-0.18*	ns	0.20*	0.18*	-0.58**	-0.47**	-0.36**	-0.23**	ns	Ns	0.86**	0.85**	All	Lyco		
	-0.28**	0.20*	ns	-0.34**	ns	-0.33**	ns	0.20*	-0.32**	ns	-0.23*	ns	0.29**	0.54**	0.80**	0.88**	Cera	Pimp		
TA	ns	-0.27**	-0.24**	ns	-0.24**	ns	ns	ns	-0.52**	-0.35**	ns	0.24**	-0.68**	-0.73**	0.53**	0.57**	0.36**	0.39**	All	Lyco
	ns	0.26*	-0.23**	ns	-0.19*	0.32**	ns	ns	-0.24**	-0.24**	ns	0.18*	-0.71**	-0.56**	0.29**	ns	ns	ns	Cera	Pimp

[a] $\log_{10}$ transformed.

* $P < 0.05$, ** $P < 0.01$.

ns: non-significant.

All= the whole collection, Lyco= S. lycopersicum, Cera= S. l. cerasiforme, Pimp= S. pimpinellifolium. a*, b*, L= color, FIR= firmness, FW= fruit weight, LCN =locule number, pH= pH , SSC = soluble solids content, SUG = sugar content, TA = titratable acidity. Only significant correlations are shown

Supplemental Table S4 Summary of SNPs analyzed on the 4 SNPlex panels

SNPlex panel	SNP number	SNPs that passed quality control (No.)	SNPs with missing data > 10% (No.)	SNPs with MAF < 10% (No.)	Informative SNPs (No.)
Panel 1	48	37 (77%)	1	3	33 (89%)
Panel 2	48	32(67%)	0	2	30 (94%)
Panel 3	48	34 (71%)	0	5	29 (85%)
Panel 4	48	36(75%)	2	5	30 (81%)
Total	192	139(73%)	3	15	121 (87%)

Supplemental Table S5 Significant associations for eight fruit quality traits estimated with K + Q models on 188 accessions. Color (a*), color (L), firmness (FIR), fruit weight (FW), locule number (LCN) and soluble solid content (SSC), sugar content (SUG) and titratable acidity (TA). Model A: MLM model, with structure and kinship based on 20 SSR (only *p*-values lower than 0.005 are shown with indication on allele effect); Model B: MLM model with structure and kinship based on 121 SNP (*p*-values lower than 0.05 are shown)

Trait (Correlation with Q1 [a])	SNP group [b]	Chromo-some	Locus	Location [c]	Model A		MAF [f]	Model B
					Corrected *p* –value [d]	R^2 [e]		Corrected *p* –value [d]
a* (0.02 - 0.11)	2.5	2	TD083-685	133.1	0.042 ns	0.09	0.47	0.011
L (0.29 - 0.43)	2.2	2	TD049-528	72.4	0.003	0.09	0.36	ns
	2.4	2	Z2117-98	120.5	0.002	0.09	0.22	ns
	4.1	4	TD200-317	6.6	0.004	0.09	0.19	ns
	4.2	4	TD160-458	27.7	4.53×10^{-04}	0.11	0.31	ns
	9.2	9	Z1475-87	51.0	6.47×10^{-04}	0.10	0.17	ns
	9.2	9	TD167-449	59.8	0.004	0.09	0.16	ns
	9.4	9	TD168-241	99.5	7.78×10^{-04}	0.09	0.13	ns
FIR (0.3 – 0.49)	1.1	1	CON203-643	44.7	1.86×10^{-04}	0.11	0.22	ns
	2.1	2	TD091-657	49.5	0.002	0.07	0.22	ns
	2.4	2	TD348-101	118.0	3.94×10^{-04}	0.10	0.18	ns
	2.4	2	TD113-132	121.6	9.74×10^{-06}	0.14	0.14	ns
	2.4	2	TD114-259	121.7	1.18×10^{-04}	0.12	0.14	ns
	3.2	3	CON29-566	99.7	2.35×10^{-04}	0.10	0.25	ns
	3.2	3	CON174-206	102.3	8.59×10^{-06}	0.13	0.21	ns
	3.2	3	CON171-373	103.8	0.002	0.09	0.11	ns
	4.3	4	Z1703-106	65.5	1.51×10^{-07}	0.17	0.12	0.194 ns
	4.3	4	Z1370-98	66.8	1.41×10^{-04}	0.10	0.14	0.629 ns
	4.4	4	TD212-247	82.2	3.27×10^{-11}	0.24	0.17	0.011
	4.5	4	CON219-313	122.0	2.20×10^{-06}	0.15	0.22	ns
	4.5	4	CON105-290	131.7	0.001	0.10	0.12	ns
	4.5	4	TD222-326	132.7	0.001	0.09	0.17	ns
	5.1	5	CON187-220	64.6	0.001	0.10	0.26	ns
	5.1	5	CON173-501	68.3	1.16×10^{-07}	0.16	0.23	0.448 ns
	5.1	5	CON110-323	69.2	1.79×10^{-07}	0.14	0.23	0.811 ns
	5.1	5	CON310-990	71.9	5.30×10^{-05}	0.13	0.23	ns
	5.1	5	CON134-526	72.0	1.46×10^{-06}	0.16	0.23	ns
	5.2	5	CON222-388	75.8	2.72×10^{-04}	0.11	0.19	ns
	6.1	6	TD025-87	20.8	0.002	0.09	0.12	ns
	9.2	9	TD167-449	59.8	6.84×10^{-04}	0.10	0.16	ns
	10.1	10	CON176-455	59.0	1.01×10^{-07}	0.17	0.12	0.117 ns
	11.1	11	TD247-57	6.4	3.10×10^{-05}	0.13	0.15	ns
	11.2	11	TD251-230	35.4	6.66×10^{-05}	0.12	0.12	ns
	11.3	11	CON336-388	49.3	5.71×10^{-06}	0.14	0.21	ns
	11.3	11	CON141-576	54.8	7.38×10^{-07}	0.16	0.18	0.908 ns
	11.3	11	TD255-218	56.9	5.51×10^{-05}	0.12	0.13	ns
	11.4	11	CON50-294	62.3	9.55×10^{-07}	0.16	0.13	0.545 ns
	12.2	12	TD156-314	93.7	0.004	0.08	0.12	ns
Log(FW) (0.47 – 0.53)	1.2	1	TD011-260	97.3	2.68×10^{-04}	0.08	0.13	ns
	2.3	2	TD133-395	83.8	7.30×10^{-06}	0.12	0.34	0.774 ns
	2.3	2	TD274-100	87.9	0.004	0.06	0.20	ns
	2.3	2	TD280-108	88.6	1.12×10^{-04}	0.12	0.43	0.145 ns
	2.3	2	TD055-469	88.9	4.37×10^{-04}	0.09	0.18	ns
	2.3	2	TD276-67	89.0	7.44×10^{-04}	0.09	0.19	ns
	2.4	2	TD056-134	117.7	6.86×10^{-04}	0.04	0.27	0.859 ns
	2.4	2	TD116-707	122.2	7.38×10^{-08}	0.16	0.34	0.007
	2.5	2	TD083-685	133.1	2.42×10^{-04}	0.12	0.47	0.157 ns
	3.1	3	CON57-121	63.5	0.003	0.07	0.13	ns
	3.3	3	TD152-159	167.1	0.002	0.09	0.17	ns
	4.1	4	TD200-317	6.6	0.001	0.08	0.19	ns

Supplemental Table S5 Continued -1

Trait (Correlation with Q1 [a])	SNP group [b]	Chromo-some	Locus	Location [c]	Model A		MAF [f]	Model B
					Corrected *p* –value [d]	R^{2} [e]		Corrected *p* –value [d]
Log(FW)	4.3	4	CON300-472	68.6	ns	0.01	0.47	0.045
(0.47 – 0.53)	4.4	4	CON130-112	91.8	9.21×10^{-04}	0.10	0.23	0.077 ns
	6.1	6	TD025-87	20.8	0.001	0.08	0.12	ns
	9.1	9	Z1707-100	38.0	0.003	0.08	0.17	ns
	9.2	9	Z1475-87	51.0	1.08×10^{-04}	0.11	0.17	ns
	9.2	9	TD167-449	59.8	0.002	0.09	0.16	ns
	9.3	9	Z2305-99	69.4	1.55×10^{-05}	0.11	0.22	ns
	9.4	9	TD168-241	99.5	0.001	0.05	0.13	ns
	9.4	9	TD243-38	114.4	2.32×10^{-06}	0.13	0.31	0.206 ns
	10.1	10	CON369-191	52.6	9.34×10^{-04}	0.09	0.34	0.545 ns
	10.1	10	TD003-417	52.6	0.001	0.08	0.33	0.629 ns
	12.3	12	TD008-95	112.5	5.77×10^{-05}	0.09	0.19	ns
	9.4	9	TD243-38	114.4	2.32×10^{-06}	0.13	0.31	0.206 ns
Log(LCN)	2.2	2	TD049-528	72.4	1.54×10^{-04}	0.09	0.36	0.750 ns
(0.24 – 0.30)	2.3	2	TD133-395	83.8	9.27×10^{-05}	0.14	0.34	0.045
SSC	1.2	1	TD011-260	97.3	0.003	0.06	0.13	ns
(0.41 – 0.47)	1.3	1	Z2300-99	140.6	9.56×10^{-04}	0.10	0.12	ns
	2.3	2	TD120-221	83.7	0.004	0.07	0.23	ns
	2.3	2	TD133-395	83.8	1.07×10^{-05}	0.13	0.34	0.883 ns
	2.3	2	GC240z280-93	86.6	4.09×10^{-04}	0.09	0.21	ns
	2.3	2	TD107-168	87.6	1.95×10^{-04}	0.09	0.17	ns
	2.3	2	TD274-100	87.9	3.41×10^{-04}	0.10	0.20	ns
	2.3	2	TD280-108	88.6	8.94×10^{-05}	0.13	0.43	0.023
	2.3	2	TD055-469	88.9	3.35×10^{-05}	0.12	0.18	ns
	2.3	2	TD276-67	89.0	2.52×10^{-04}	0.11	0.19	ns
	2.4	2	TD116-707	122.2	0.001	0.09	0.34	ns
	2.5	2	TD178-104	138.4	0.020 ns	0.12	0.11	0.037
	2.5	2	TD179-318	140.6	0.002	0.09	0.11	0.581 ns
	3.1	3	CON57-121	63.5	0.001	0.08	0.13	0.399 ns
	4.1	4	TD200-317	6.6	1.80×10^{-05}	0.12	0.19	ns
	6.1	6	TD025-87	20.8	1.71×10^{-04}	0.11	0.12	ns
	9.1	9	Z1707-100	38.0	3.54×10^{-07}	0.16	0.17	0.787 ns
	9.1	9	TD231-549	39.1	1.34×10^{-04}	0.11	0.19	ns
	9.1	9	TD165-291	44.3	5.47×10^{-04}	0.09	0.14	ns
	9.1	9	Z1964-106	46.6	0.002	0.07	0.13	ns
	9.2	9	Z1475-87	51.0	0.003	0.07	0.17	ns
	9.3	9	TD237-253	70.0	0.003	0.07	0.14	ns
	9.4	9	TD168-241	99.5	1.02×10^{-04}	0.12	0.13	ns
	9.4	9	TD243-38	114.4	0.002	0.08	0.31	ns
	10.1	10	CON369-191	52.6	0.003	0.08	0.34	0.194 ns
	10.1	10	TD003-417	52.6	0.004	0.08	0.33	0.230 ns
	11.3	11	TD255-218	56.9	2.09×10^{-06}	0.05	0.13	ns
	11.4	11	CON50-294	62.3	0.001	0.07	0.13	ns
	12.1	12	Z2302-103	21.0	0.004	0.08	0.15	ns
SUG	1.3	1	Z2300-99	140.6	1.06×10^{-05}	0.14	0.12	ns
(0.39 – 0.52)	2.1	2	TD091-657	49.5	0.004	0.07	0.22	ns
	2.2	2	TD090-357	68.8	8.42×10^{-04}	0.08	0.21	ns
	2.2	2	Z1979-37	70.3	1.40×10^{-04}	0.07	0.21	ns
	2.2	2	TD139-547	72.2	1.37×10^{-04}	0.08	0.20	ns
	2.3	2	TD048-45	83.6	0.002	0.06	0.21	ns
	2.3	2	TD122-312	83.6	8.69×10^{-04}	0.07	0.25	ns
	2.3	2	TD120-221	83.7	3.30×10^{-05}	0.11	0.23	ns
	2.3	2	TD088-204	83.7	7.71×10^{-04}	0.10	0.22	ns
	2.3	2	TD133-395	83.8	7.66×10^{-07}	0.15	0.34	0.194 ns
	2.3	2	TD373-140	84.4	1.46×10^{-04}	0.10	0.23	ns
	2.3	2	GC240z280-93	86.6	1.19×10^{-05}	0.12	0.21	ns
	2.3	2	TD107-168	87.6	7.04×10^{-06}	0.11	0.17	ns
	2.3	2	TD274-100	87.9	5.60×10^{-05}	0.11	0.20	ns

Supplemental Table S5 Continued- 2

Trait (Correlation with Q1 [a])	SNP group [b]	Chromo-some	Locus	Location [c]	Model A		MAF [f]	Model B
					Corrected *p* –value [d]	R^2 [e]		Corrected *p* –value [d]
SUG	2.3	2	TD275-101	88.5	4.51×10^{-04}	0.09	0.13	ns
(0.39 – 0.52)	2.3	2	TD280-108	88.6	0.001	0.10	0.43	0.303 ns
	2.3	2	TD055-469	88.9	9.40×10^{-06}	0.13	0.18	ns
	2.3	2	TD276-67	89.0	1.45×10^{-04}	0.11	0.19	ns
	2.4	2	TD348-101	118.0	9.22×10^{-04}	0.09	0.18	ns
	2.4	2	TD113-132	121.6	7.13×10^{-04}	0.09	0.14	ns
	2.4	2	TD114-259	121.7	6.72×10^{-04}	0.09	0.14	ns
	2.5	2	TD178-104	138.4	1.34×10^{-06}	0.14	0.11	0.206 ns
	2.5	2	TD179-318	140.6	8.11×10^{-05}	0.11	0.11	0.932 ns
	3.1	3	CON57-121	63.5	0.001	0.08	0.13	ns
	4.1	4	TD200-317	6.6	2.15×10^{-06}	0.14	0.19	ns
	4.5	4	CON105-290	131.7	5.55×10^{-04}	0.08	0.12	ns
	4.5	4	TD222-326	132.7	7.10×10^{-04}	0.09	0.17	ns
	6.1	6	TD025-87	20.8	1.94×10^{-04}	0.11	0.12	ns
	9.1	9	Z1707-100	38.0	3.68×10^{-05}	0.12	0.17	ns
	9.1	9	TD231-549	39.1	2.36×10^{-04}	0.10	0.19	ns
	9.1	9	TD165-291	44.3	2.15×10^{-04}	0.10	0.14	ns
	9.1	9	Z1964-106	46.6	5.97×10^{-05}	0.10	0.13	ns
	9.2	9	Z1475-87	51.0	5.37×10^{-04}	0.09	0.17	ns
	9.2	9	Z1717-98	54.0	0.002	0.08	0.16	ns
	9.2	9	TD167-449	59.8	0.003	0.08	0.16	ns
	9.4	9	TD168-241	99.5	3.57×10^{-04}	0.11	0.13	ns
	10.1	10	CON369-191	52.6	6.59×10^{-04}	0.10	0.34	0.011
	10.1	10	TD003-417	52.6	6.09×10^{-04}	0.10	0.33	0.010
	11.3	11	TD255-218	56.9	4.42×10^{-04}	0.07	0.13	ns
	11.4	11	CON50-294	62.3	1.23×10^{-04}	0.09	0.13	ns
	12.1	12	Z2302-103	21.0	3.81×10^{-04}	0.11	0.15	ns
TA	2.3	2	TD275-101	88.5	0.110 ns	0.04	0.13	0.048
(0.26 – 0.22)	4.3	4	Z1370-98	66.8	ns	0.08	0.14	0.003

[a] Q1 is the probability that an individual belongs to a subpopulation which generated from STRUCTURE2.1 software (Pritchard, Stephens et al. 2000). Correlations with the Q value defined by 20 SSR markers and 121 SNPs.

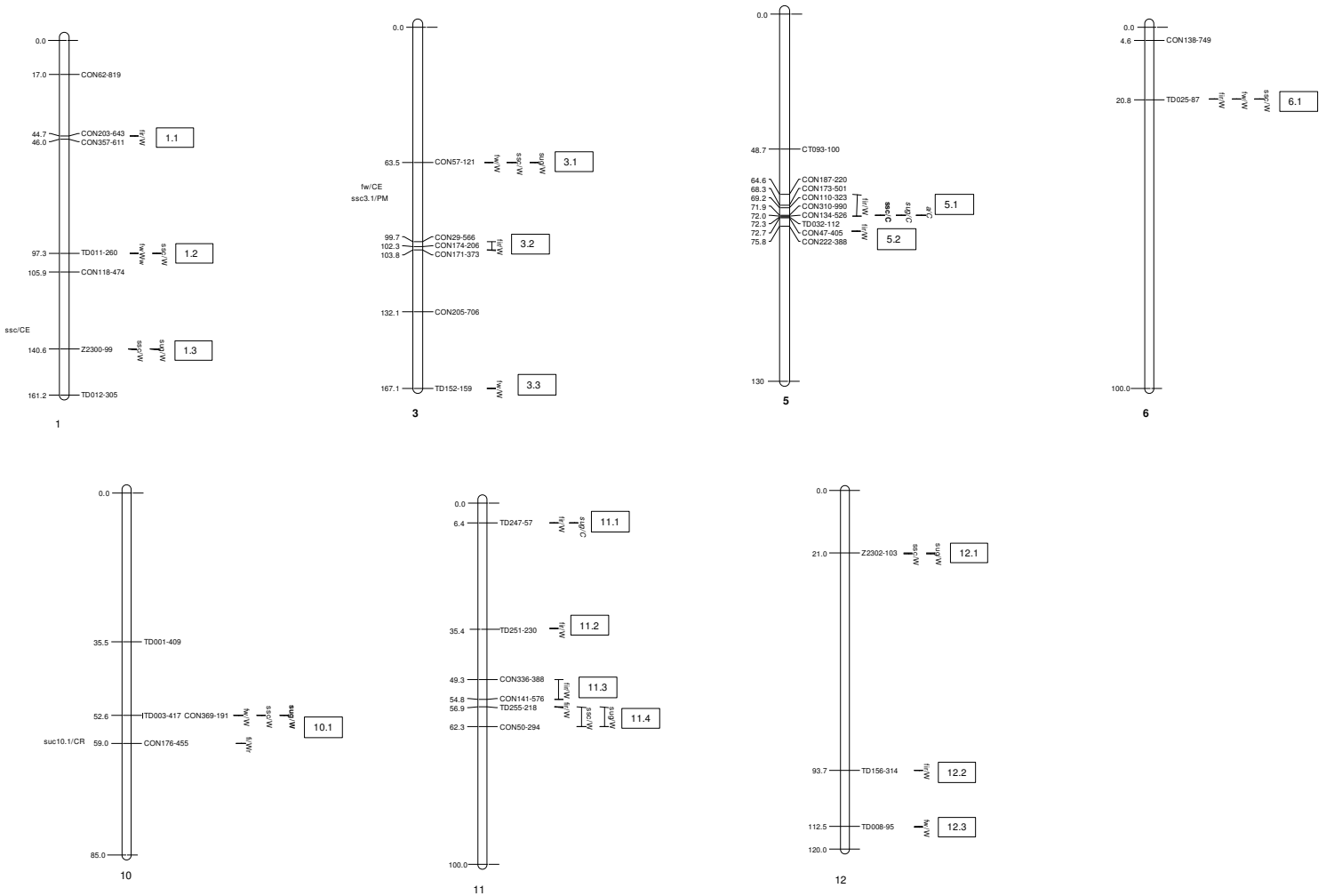
[b] Associated SNPs in less than 10 cM on each chromosome were grouped together. SNP which is 10 cM apart from the other SNPs was assigned as an independent group.

[c] Genetic distance of the marker on EXPEN2000 reference map (http://sgn.cornell.edu/).

[d] *p*-values were corrected following the standard Bonferroni procedure. ns: non significant

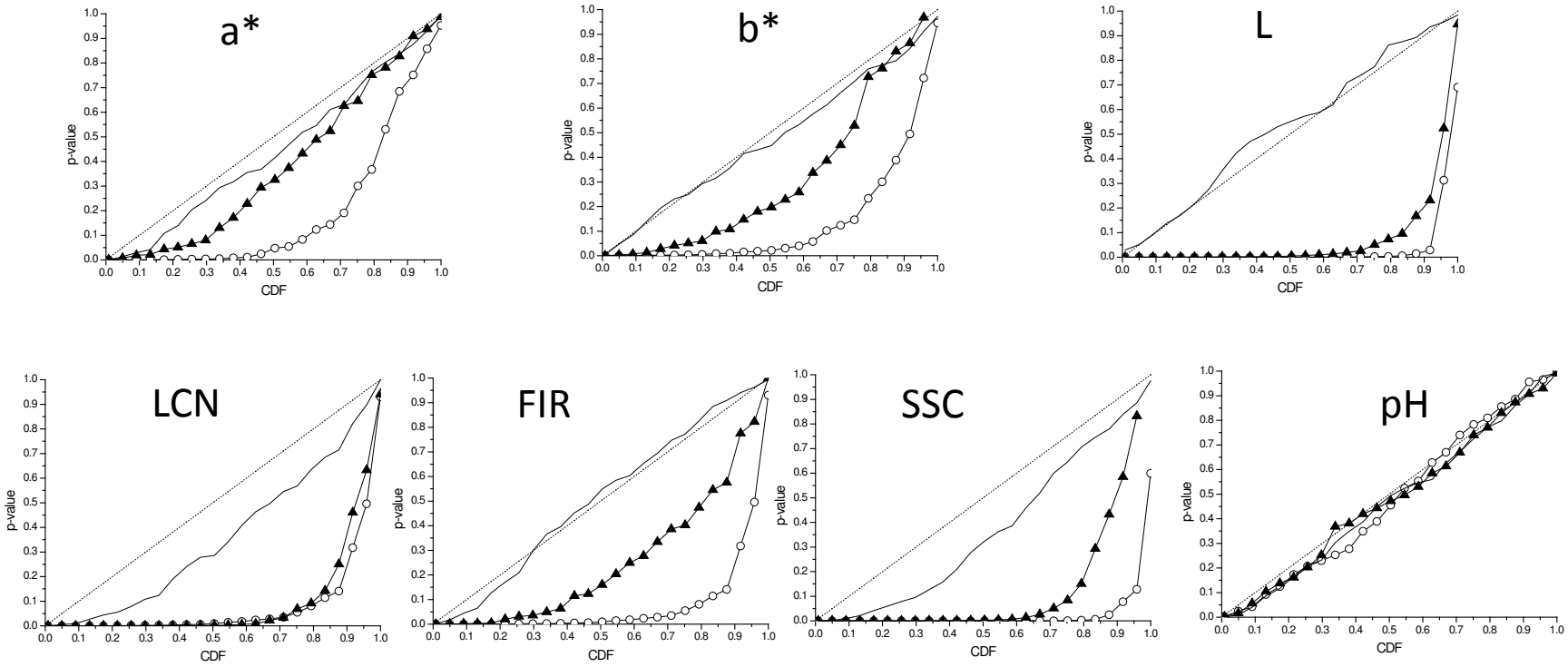
[e] R^2 were calculated using a Q model.

[f] Minimum allele frequencies (MAF).

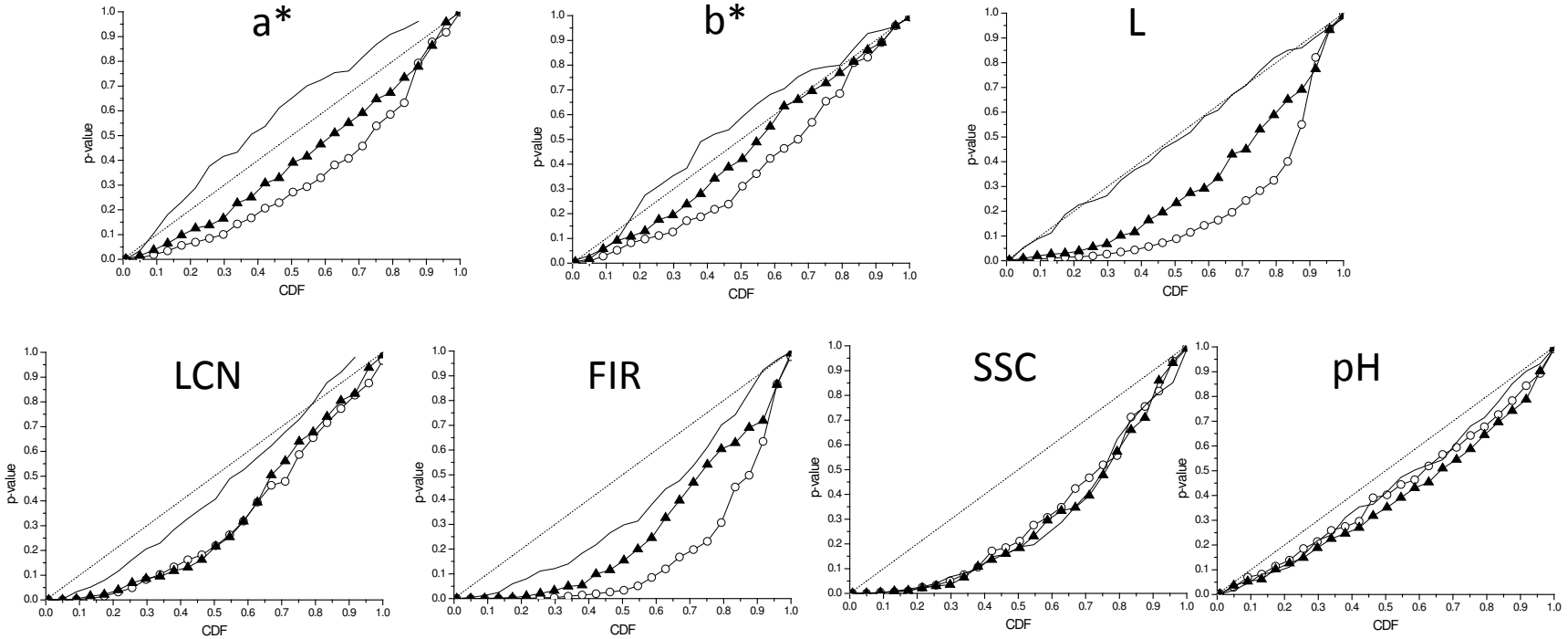


Supplemental Fig. S1 Comparison of QTLs identified by linkage mapping and association studies on chromosomes 1, 3, 5, 6, 10, 11 and 12. SNPs were mapped on tomato EXPEN 2000 reference map (http://sgn.cornell.edu/). Associations detected in the 188 accessions (W) and in 127 *S. l. cerasiforme* accessions (C) are indicated to the right of the chromosomes. Associations were estimated with K + Q model, model A: MLM model, with structure based on 20 SSR marker, model B: MLM model with structure based on 121 SNP (common font: associations detected in model A; in italic: associations detected in model B; in bold, associations detected in both models). Horizontal lines "-" correspond to the genetic location of associated marker, associations are linked together by a vertical line when linked markers in less than 10 cM are associated to the same trait. Associated SNPs in less than 10 cM on each chromosome were grouped together. SNP which is 10 cM apart from the others were assigned as independent groups. Groups are named as consecutive number according to their genetic location on each chromosome. Traits are : FIR= firmness, FW= fruit weight, SSC= soluble solids content, SUG = total sugar content, LCN= locule number, a and L=color, TA = titratable acidity.

QTLs identified by linkage mapping in the populations from crosses of *S. Lycopersicum* × *S. l. cerasiforme* (Saliba-Colombani et al. 2001), *S. Lycopersicum* × *S. pimpinellifolium* (Grandillo et al. 1996) and *S. Lycopersicum* × *S. l. cheesmanii* (Goldman et al. 1995) are shown to the left of the chromosomes (CR= QTL from *S. l. cerasiforme,* CE= QTL from *S. l. cheesmanii* , PM= QTL from *S. pimpinellifolium*). Only QTL co-localizing with an association are shown



Supplemental Fig. S2 Cumulative density functions (CDF) using several alternative models of association for color (a*, b*and L), locule number (LCN), firmnes (FIR), pH, soluble solids content (SSC). Associations are tested for 121 polymorphic sites on 188 accessions. Simple linear model (empty circle) and K+Q models, with structure and kinship based on SSR markers (black triangle)), and on 121SNP markers (black line) were tested. The diagonal indicates uniform distribution of *p*-values under the expectation that random SNPs are unlinked to the polymorphisms controlling these traits (H0: no SNP effect).



Supplemental Fig. S3 Cumulative density functions (CDF) using several alternative models of association for color (a*, b*and L), locule number (LCN), firmnes (FIR), pH, soluble solids content (SSC). Associations are tested for 121 polymorphic sites on 127 *S. l. cerasiforme* accessions. Simple linear model (empty circle) and K+Q models, with structure and kinship based on SSR markers (black triangle), and on 121SNP markers (black line) were tested. The diagonal indicates uniform distribution of *p*-values under the expectation that random SNPs are unlinked to the polymorphisms controlling these traits (H0: no SNP effect)

CHAPTER V: AN EXTENSIVE PROTEOME MAP OF THE TOMATO (*SOLANUM LYCOPERSICUM*) FRUIT PERICARP

This chapter is a manuscript to be submitted to the section Dataset Brief of the Journal of Proteomics. As a dataset brief, the text is less than 3000 words. It presents the first comprehensive proteome reference map of the tomato fruit pericarp at two developmental stages from 12 genotypes representing a large phenotypic and genotypic diversity. We identified 506 spots representing 333 proteins expressed in fruit. We described the physiological function of identified proteins. These data provide experimental evidence for tomato fruit proteins that had only been predicted by genome annotation and are valuable tools for comparative studies of protein expression.

An extensive proteome map of tomato (*Solanum lycopersicum*) fruit pericarp

Jiaxin Xu[1,2], Laura Pascual-Banuls[1], Rémy Aurand[1,3], Jean-Paul Bouchet[1], Yan Liang[2], Benoît Valot[4], Michel Zivy[4], Mathilde Causse[1] and Mireille Faurobert[1]

(1) INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, CS 60094 - 84143 MONTFAVET CEDEX, France

(2) Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, P.R. China

(3) INRA, UR1115 Plantes et Systèmes de Culture Horticoles. F-84914 Avignon, France

(4) INRA/Université Paris-Sud/CNRS, Plateforme d'Analyse Protéomique de Paris Sud-Ouest, UMR 0320/UMR 8120 de Génétique Végétale, Gif sur Yvette, France

Corresponding author :

Mathilde Causse

Phone: 33 04 32 72 27 01

Fax: 33 04 32 72 27 02

Email: Mathilde.Causse@avignon.inra.fr

Abstract

Tomato (*Solanum lycopersicum*) is the model species for studying fleshy fruit development. In this manuscript, an extensive proteome map of the fruit pericarp is described in light of the newly available genome sequence. The proteomes of fruit pericarp from 12 tomato genotypes at two developmental stages (cell expansion and orange-red) were analyzed. The two-dimensional gel electrophoresis reference map included 506 spots identified by nano-LC/MS and ITAG Database searching. A total of 425 spots corresponded to a unique protein. Thirty four spots resulted from the transcription of genes belonging to multi-gene families involving two to six genes. A total of 47 spots corresponded to a mixture of different proteins. The whole protein set was classified according to Gene Ontology annotation. The quantitative protein variation was analyzed in relation to genotype and developmental stage. This tomato fruit proteome dataset is currently the largest available and constitutes a valuable tool for comparative genetic studies of tomato genome expression at the protein level.

Keywords: Tomato / Fruit pericarp proteome / genome sequence / Two-dimensional gel electrophoresis / Liquid chromatography–mass spectrometry

Introduction

Proteomics is a valuable approach for studying the biology of living organisms and their interaction with the environment in the post genomic era (Maghuly et al. 2011). Protein expression integrates post-transcriptional and post-translational modifications that modulate the quantity, the localization and the efficiency of the final product within the cell. The plasticity of a phenotype is driven by these altered levels of proteins and metabolites (Weckwerth 2008). Recently protein metabolism and especially protein stability were suggested to play a major role in plant growth, yield and heterosis (Goff 2011). Knowledge about the fruit proteome is a challenging area of research, as reviewed by Palma et al (2011).

Apart from being one of the most important vegetables consumed worldwide, tomato (*Solanum lycopersicum*) is also the model species for studying fleshy fruit development (Giovannoni 2004). This self-pollinated species exhibits a large range of phenotypes but a limited molecular diversity. Studies on the tomato proteome have been recently reviewed by Faurobert et al. (Faurobert et al. 2012). Most of these studies were carried out on one or two genotypes and identified a relatively small number of proteins, failing to provide an insight into the extent of natural genetic variation (Iwahashi and Hosoda 2000; Page et al. 2010; Manaa et al. 2011; Marjanovic et al. 2012; Sheoran et al. 2005; Rocco et al. 2006; Faurobert et al. 2007a). Improvement of protein identification technologies and the availability of the complete high quality tomato genome sequence (Sato et al. 2012) has increased the efficiency of protein identification. This study constitutes an extensive characterization of the tomato fruit proteome and completes the study of Faurobert et al. (2007a) in light of the genome sequence. The proteome map presented here is based on two experiments. It consists of 506 protein spots identified from 11,692 peptide sequences. Such resources may help to improve genome sequence and annotation.

Materials and methods

In the first experiment, we compared eight genotypes, which represent the largest range of tomato genetic and phenotypic diversity identified in a collection of 360 accessions (Ranc et al. 2008). It describes the proteomes of four *S. lycopersicum* lines (Levovil, Stupicke Polni Rane, LA0147, and Ferum) and four *S. lycopersicum* var *cerasiforme* lines (Cervil, Criollo, Plovdiv 24A, and LA1420). Cervil produces small fruits (less than 10 g). Levovil, Ferum and LA0147 genotypes have large fruits. Stupicke Polni Rane, Criollo, Plovdiv 24A, and LA1420 have intermediate fruit size. In the second experiment, three parental lines (Cervil, Levovil and VilB, a large fruited tomato) and three lines with intermediate fruit size carrying introgressed chromosome fragments from Cervil in Levovil (L4 and L9) and VilB backgrounds (B9) (Chaib et al. 2007) were compared. Plants were grown in 2010 under greenhouse conditions (16/20°C) in Avignon for the first experiment and in Bellegarde for the second one (South of France). Given the large effect of developmental stage on fruit proteome which has been previously demonstrated (Faurobert et al. 2007a), fruits were collected at two stages of development in both experiments, cell expansion (25, 20 and 14 days after anthesis for large, intermediate and small fruited lines, respectively) and orange-red stage, according to the fruit color. Three to four biological pools of 5 to 20 fruits per genotype and per developmental stage were harvested. Fruit pericarp was isolated and immediately frozen, then ground in liquid nitrogen and stored at -80 ˚C. Proteins were extracted using a phenol extraction method developed by Faurobert et al. (2007b) and separated by two-dimensional electrophoresis (2-DE) according to Page et al. (2010). After Coomassie colloidal staining, image analysis was performed with Samespot software and the normalized spot volumes were assessed and compared through two-way ANOVA (testing the differences between genotypes, stages and their interaction) using "stats" package. A principal component analysis (PCA) was carried out on spot abundance for the first experiment using “pcaMethods” package. Statistical analyses were performed using the R program (R Development Core Team 2005).

Within each experiment the varying spots ($P < 0.05$) were picked for identification by nano-LC-MS/MS method as detailed in **Supplemental text S1**. FASTA sequences of the identified proteins were used to re-annotate the proteins using the Blast2GO package (Conesa et al. 2005). Sequences were compared against the NCBI-NR database of non-redundant protein sequences using BLASTX with the default setting.

Table 1 Summary of the protein spots studied and identified

	Number
Total number of protein spots detected	2307
Number of spots identified by MS	833
Non redundantly identified spots across experiments	506
Number of identified spectra for the 506 spots	16,598
Number of unique peptides for the 506 spots	11,692
Number of spots resulting from protein mix	47
Number of spots encoded by a unique gene locus	333
Number of spots corresponding to multi-gene family	34
Number of loci contributing to one spot	1 to 6
Number of protein functions displaying multiple spots	63

Results and discussion

A total of 1230 spots were detected in the first experiment comparing eight genotypes at cell expansion and orange-red stages, while 1077 spots were detected in the second experiment comparing six genotypes at the two stages (**Table 1**). After ANOVA analysis, a total of 424 and 419 spots showing volume variation were detected for the first and the second experiment, respectively. In the first experiment, 333 spots showed significant changes in protein abundance according to the developmental stage, 321 according to the genotype and 215 according to a developmental stage and genotype interaction. In the second experiment, these numbers were 351, 251 and 176, respectively **(see Supplemental table S1).**

The large variation in the tomato pericarp proteome during fruit development agrees with the results of Faurobert et al. (Faurobert et al. 2007a) and other studies on fruits reviewed by Palma et al. (Palma et al. 2011). In both experiments, developmental stage was the major discriminating factor, but the effect of the genotype was also very high at both stages, as illustrated in **Figure 1** for the first experiment. The first PCA plane accounted for 30% of the variation at the cell expansion stage and 27% at the orange-red stage **(Figure 1).** The three biological replicates of each genotype appeared well clustered. The genotypes showed the same pattern of clustering for the two developmental stages. Cervil was distinct from all other genotypes. The *S. lycopersicum* genotypes were more tightly clustered than the *S. lycopersicum* var *cerasiforme* genotypes. The possibility of discriminating genotypes according to their proteome profile has already been demonstrated in roots under salt stress condition by Manaa et al. (2011). They showed that the effect of genotype on the root proteome variations was much higher than the effect of salt.

Finally, the 424 and 419 variable spots from the two experiments were submitted to mass spectrometry. For both experiments, a total of 16,598 spectra were identified by nano LC-MS/MS (corresponding to 11,692 unique peptides) leading to the identification of 506 spots. 337 spots were common to both experiments and corresponded to the same identification of proteins **(Table 1).** Their position on 2D gels is illustrated in **Supplemental Figure S1**. Ten spots could not be identified. This very low level of failure was due to the high sensitivity of mass spectrometry and the high quality of the genome assembly. On average, the detected peptides covered 54% of protein sequences. **Supplemental table S1** presents the list of the spots and their main features.

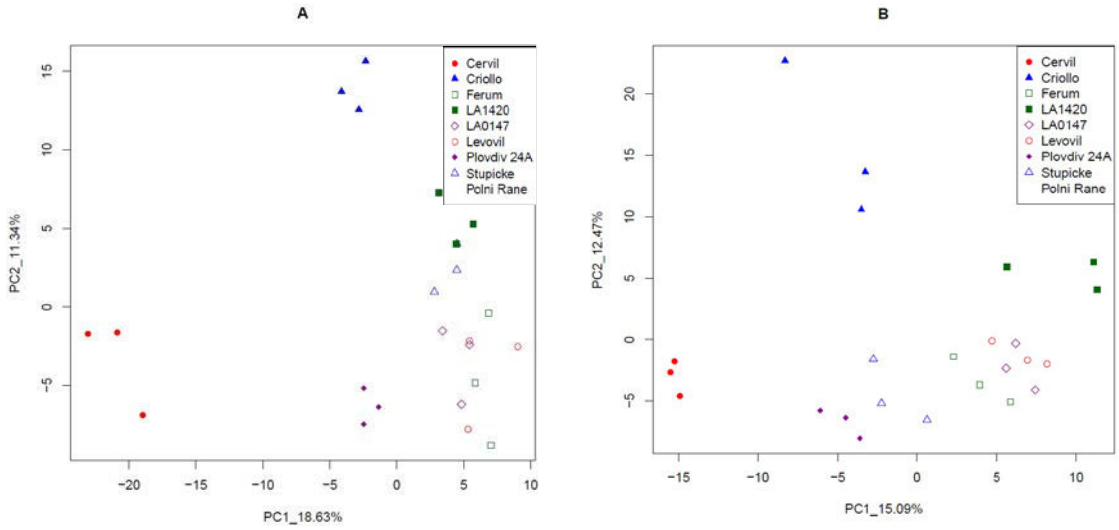


Figure 1. Principal Component analysis based on the volume of 424 variable spots detected at cell expansion (A) and orange-red (B) fruit stages, in the first experiment. The 8 genotypes included 4 large fruited lines (LA0147, Levovil, Ferum and Stupicke Polni Rane) and 4 cherry tomato lines (Cervil, Criollo, Plovdiv 24A and LA1420). Each dot corresponds to a biological replicate.). The percentages of total variation accounted for each component are indicated along the axes.

Among the 506 spots, 425 spots corresponded to a unique protein. A total of 47 spots were identified as containing a mix of proteins with diverse functions (**Supplemental table S1**). The remaining 34 spots corresponded to the translation of several members of a multi-gene family (**Supplemental Table S2**). The number of loci that contributed to one protein spot varied from two, in 20 cases, to a maximum of six loci in the case of spot 74. This particular spot was identified as a heat shock protein of high molecular weight (around 70 kDa). The multigenic nature of heat shock proteins is well known. In eight cases, two or more identified genes were located on the same chromosome in a very narrow region (available on the Solanaceae Genomic Network site) and corresponded to duplicated genes in tandem. For example, spot 36 corresponded to 1-aminocyclopropane-1-carboxylate-oxidase resulting from the transcription of two neighbor loci, *Solyc07g049530.2.1* and *Solyc07g049550.2.1*. Alternatively, these two loci could correspond to an incorrect annotation as, based on a relative abundance index, it was possible to show that the first locus contributed more to the final protein abundance. On the other hand, for four spots, several loci contributed equally to protein abundance (*e.g.* spot 30 corresponds to an Actin protein encoded by two different genes, one located on chromosome 3, the other on chromosome 11). Besides the spots resulting from the transcription of several genes, 63 proteins were present in several spots (from two to five). In most cases (56 spots), they were close together on the 2-D gel. For instance, the metacaspase 7 protein (*Solyc09g098150.2.1*) was represented by two closely located spots (spot 125 and spot 142) (Supplemental Fig. 1). In seven cases, a protein was represented by both closely and distantly located spots. For example, the acid beta-fructofuranosidase (*Solyc03g083910.2.1*) was represented by the closely located spots 51 and 87 and by spots located further away: 237, 239 and 284. The presence of multiple spots may be due to post-translational modifications, splice variants, protein degradation or allelic variation and has already been reported for acid beta-fructofuranosidase by Faurobert et al. (2007a). For instance, an enolase protein (*Solyc09g009020.2.1*) was represented by five closely located spots (115, 205, 214, 256 and 323). Each spot was picked from one specific genotype and they showed significant quantitative differences according to the genotypes.

Many proteins have pleiotropic functions within the cell and it is therefore complicated to draw major conclusions from basic protein functional classification. However, we propose here a functional classification of the proteins according to the GO terms description obtained after re-annotating gene sequences with Blast2Go based on biological process term level 3. Proteins were assigned to 16 categories **(Figure 2).** Primary metabolic process represented the

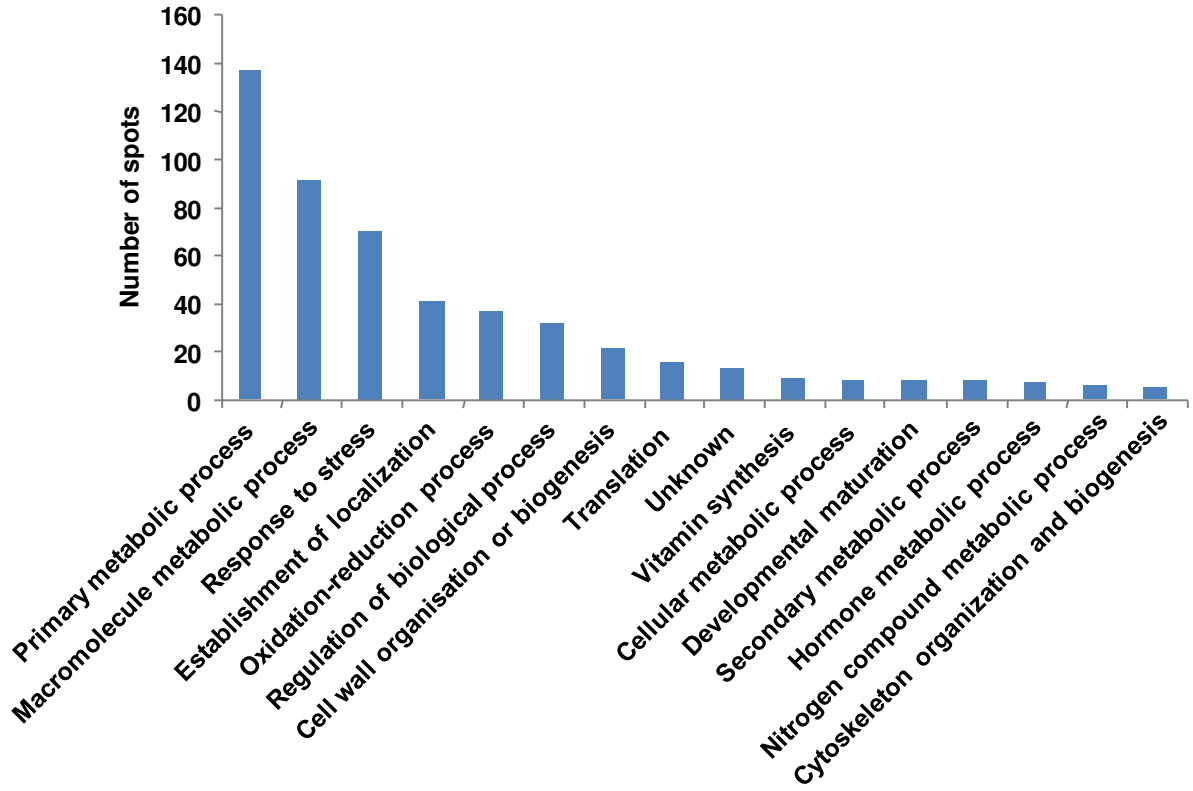


Figure 2. Distribution of the 506 tomato protein spots identified by MS into biological classes according to gene ontology classification.

largest class (27%). Several enzymes were involved in sugar synthesis, such as fructokinase (spots 15, 218 and 363), glucose-6-phosphate 1-dehydrogenase (spot 329), fructose-bisphosphate aldolase (spot 241), UTP-glucose-1-phosphate uridylyltransferase (spots 344 and 403), or in cysteine synthesis, such as cysteine synthase (spots 26, 179 and 402). Malate dehydrogenase (spots 4 and 177) is involved in the tricarboxylic acid cycle which provides the necessary energy for the ripening process and was also identified in the primary metabolism subgroup. Proteins involved in macromolecular metabolic process accounted for 18% of all identified proteins. Chaperonin (spots 152, 181, 201, 249, 330, 382, 390 and 414) and proteasome subunits (spots 114, 182, 336, 375, 387, RA311 and RA 394) were identified in this subgroup. Proteins related to stress responses represented the third largest subgroup (14%). This group was clearly dominated by heat shock proteins (23 spots), but also included some enzymes such as glutathione S-transferase (spots 2, 80, 112, 204, 293 and 374) which has a role in detoxification processes. Heat shock proteins are known to be modulated by a wide range of environmental stresses but also during fruit development (Wang et al. 2004; Neta-Sharir et al. 2005; Faurobert et al. 2007a). Other smaller functional groups which represented a relatively low number of proteins were also identified. Finally a total of 17 spots were classified into unknown category.

Conclusion

In conclusion, we provide the first comprehensive proteome reference map of the tomato fruit pericarp at two developmental stages from 12 genotypes representing a large phenotypic and genotypic diversity. We identified 506 spots representing 333 proteins expressed in fruit. We have described the physiological function of the identified proteins. These data provide experimental evidence for tomato fruit proteins that had only been predicted by genome annotation and are valuable tools for comparative studies of protein expression. Furthermore, they will be available for further investigation of genetic variation. We are now characterizing the fruit proteome of hybrids from the genotypes used in this study to assess the inheritance pattern of all identified proteins.

The spectrometry dataset is available at PRIDE (http://ebi.ac.uk/pride/), with the accession number XXX.

This work was supported by ANR Genomic project MAGICTomSNP project.

We thank Caroline Callot and Karine Leyre for technical help. We also thank Esther Pelpoir for her help in fruit sampling and Yolande Carretero for taking care of the plants in the greenhouse. Many thanks to Rebecca Stevens for English revising. Finally, we thank CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes) from Balandran for plant cultivation.

The authors have declared no conflict of interest.

List of supplemental data

Supplemental text S1: Method for protein identification by nano-LC-MS/MS

Supplemental Figure S1: Proteome map of with 506 tomato proteins at two developmental stages from 12 genotypes

Supplemental Table S1: List of identified spots.

Supplemental Table S2: List of spots resulting from multi-gene families

Supplemental text S1: Method for protein identification by nano-LC-MS/MS

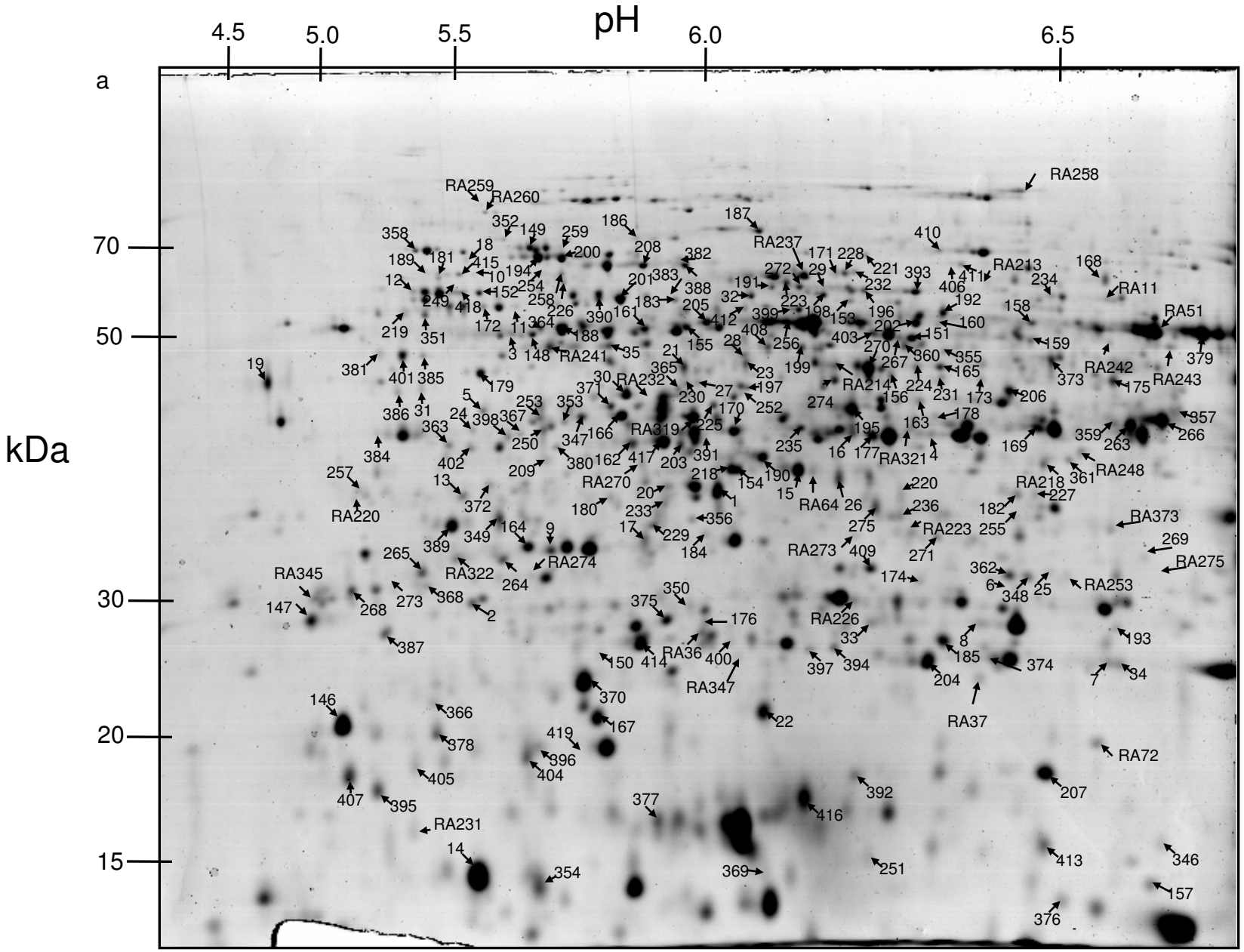
In-gel digestion was performed with the Progest system (Genomic Solution) according to a standard trypsin protocol. Gel pieces were washed twice by successive separate baths of 10% acetic acid, 40% ethanol, and acetonitrile (ACN). They were then washed twice with successive baths of 25 mM NH4CO3 and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modi?ed trypsin (Promega) dissolved in 20% methanol and 20 mM NH4CO3. The peptides were extracted successively with 2% tri?uoroacetic acid (TFA) and 50% ACN and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 µL of 0.05% TFA, 0.05% HCOOH, and 2% ACN.

HPLC was performed on an NanoLC-Ultra system (Eksigent). A 4 µL sample was loaded at 7.5 µL/min-1 on a precolumn cartridge (stationary phase: C18 Biosphere, 5 µm; column: 100 µm i.d., 2 cm; Nanoseparations) and desalted with 0.1% HCOOH. After 3 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 Biosphere, 3 µm; column: 75 µm i.d., 150 mm; Nanoseparations). Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 11 min at 300 nL/min-1. Including the regeneration step at 95% B and the equilibration step at 95% A, one run took 25 min.

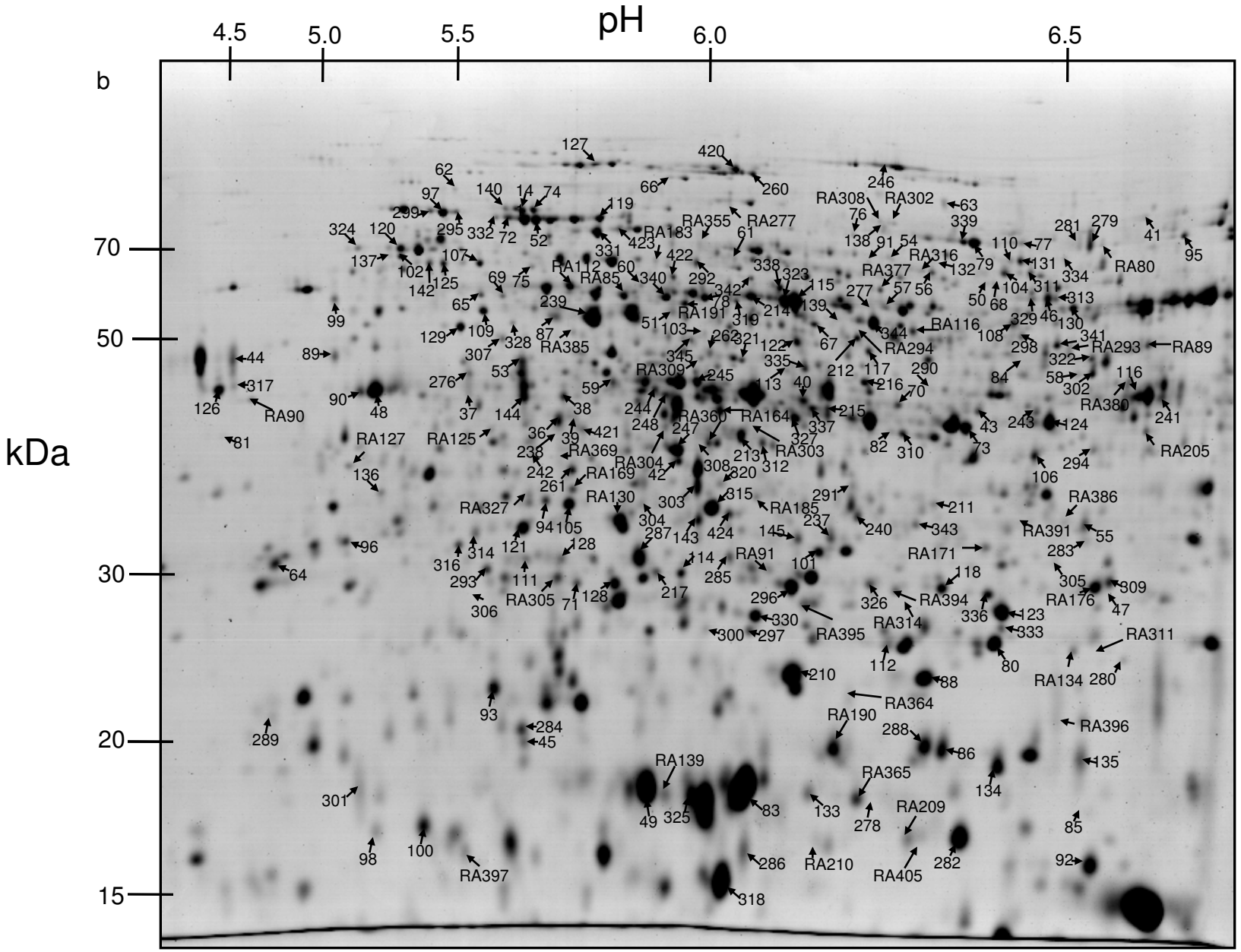
Eluted peptides were analysed on-line with a LTQ XL ion trap (Thermo Electron) using a nanoelectrospray interface. Ionization (1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10 µm i.d.; New Objective). Peptide ions were analysed using Xcalibur 2.07 with the following data-dependent acquisition steps: (1) full MS scan (mass-to-charge ratio (m/z) 300 to 1400, centroid mode) and (2) MS/MS (qz = 0.25, activation time = 30 ms, and collision energy = 35%; centroid mode). Steps 2 was repeated for the three major ions detected in step 1. Dynamic exclusion was set to 30 s.

A database search was performed with XTandem (version 2010.12.01.1) (http://www.thegpm.org/TANDEM/). Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Cys carboxyamidomethylation and Met oxidation were set to static and possible modifications, respectively. Precursor mass and fragment mass tolerance were 2.0 and 0.5, respectively. A refinement search was added with similar parameters except that semi-trypsic peptide and possible N-ter proteins acetylation were searched. The International Tomato Annotation Group (ITAG) Release 2.3 predicted proteins (SL2.40) database (http://solgenomics.net/) and a contaminant database (trypsin, keratins, ...) were used. Only peptides with a E value smaller than 0.1 were reported.

Identified proteins were filtered and grouped using XTandem Pipeline (http://pappso.inra.fr/bioinfo/xtandempipeline/) according to : (1) A minimum of two different peptides was required with a E value smaller than 0.03, (2) a protein E value (calculated as the product of unique peptide E values) smaller than 10-3 . In case of identification with only two or three MS/MS spectra, similarity between the experimental and the theoretical MS/MS spectra was visually checked. To take redundancy into account, proteins with at least one peptide in common were grouped. This allowed to group proteins of similar function. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as sub-groups.



Supplemental Figure S1 Representative two-dimensional electrophoresis gel of tomato pericap proteins of Criollo at cell expansion stage (a) and orange red stage (b). The positions and numbers of 424 identified protein spots are indicated by arrows.



Supplemental Figure S1 Representative two-dimensional electrophoresis gel of tomato pericap proteins of Criollo at cell expansion stage (a) and orange red stage (b). The positions and numbers of 424 identified protein spots are indicated by arrows.

Supplemental table S1 List of identified protein spots

Note	Spot ID	ITAG accession number	Position on genome (bp)	Protein function description	Protein classification
	1	Solyc02g080630.2.1	39392539-39397791	Lactoylglutathione lyase	primary metabolic process
	2	Solyc09g007150.2.1	775043-778675	Glutathione S-transferase	response to stress
	3	Solyc12g010020.1.1	3152019-3154797	Leucyl aminopeptidase	response to stress
	4	Solyc07g062650.2.1	62561877-62565142	Malate dehydrogenase	primary metabolic process
	5	Solyc01g057830.2.1	56934925-56939761	30S ribosomal protein S1	translation
	6	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis
	7	Solyc05g054760.2.1	63753656-63758079	Dehydroascorbate reductase	vitamin synthesis
	8	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process
	9	Solyc02g065400.2.1	31151784-31154764	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process
	10	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization
	11	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process
	12	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process
	13	Solyc08g065900.2.1	51450546-51456933	Charged multivesicular body protein 4b	establishment of localization
#	14	Solyc07g064880.2.1	64080741-64083209	Small ubiquitin-related modifier	macromolecule metabolic process
	15	Solyc06g073190.2.1	41483219-41485991	Fructokinase-like	primary metabolic process
	16	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress
	17	Solyc07g064160.2.1	63644226-63646402	Thiazole biosynthetic enzyme	vitamin synthesis
	18	Solyc05g018700.2.1	22827490-22832816	Protein disulfide isomerase	macromolecule metabolic process
	19	Solyc02g079500.2.1	38618009-38620468	Peroxidase	cell wall organisation or biogenesis
	20	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization
	21	Solyc12g099000.1.1	64658262-64659443	S-adenosylmethionine synthase	hormone metabolic process
	22	Solyc01g111300.2.1	89348859-89349828	Cold shock protein-1	regulation of biological process
	23	Solyc01g101060.2.1	82678819-82681422	S-adenosylmethionine synthase	hormone metabolic process
	24	Solyc09g007940.2.1	1440058-1444221	Adenosine kinase	primary metabolic process
	25	Solyc08g067160.2.1	53325912-53329825	Acyl-protein thioesterase 2	macromolecule metabolic process
	26	Solyc09g082060.2.1	63298831-63303187	Cysteine synthase	primary metabolic process
	27	Solyc08g076990.2.1	58097110-58102792	Acetylornithine deacetylase	primary metabolic process
	28	Solyc02g080420.2.1	39222034-39228232	RNA Binding Protein 45	regulation of biological process
*	29	Solyc08g075160.2.1	56470989-56477178	Bifunctional purine biosynthesis protein purH	primary metabolic process
#	30	Solyc03g078400.2.1	44381441-44383329	Actin	cytoskeleton organization and biogenesis
	31	Solyc08g076970.2.1	58078222-58080829	Acetylornithine deacetylase or succinyl-diaminopimelate desuccinylase	macromolecule metabolic process
	32	Solyc01g094200.2.1	77521221-77531982	NAD-dependent malic enzyme 2	primary metabolic process
	33	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process
	34	Solyc05g054760.2.1	63753656-63758079	Dehydroascorbate reductase	vitamin synthesis
	35	Solyc02g078540.2.1	37780626-37784558	Unknown Protein	Unknown
#	36	Solyc07g049530.2.1	57154220-57156391	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation
*	37	Solyc07g053310.2.1	59102795-59108445	Adaptin ear-binding coat-associated protein 1	establishment of localization
	38	Solyc02g036350.2.1	21279767-21281954	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation
	39	Solyc07g049530.2.1	57154220-57156391	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation
	40	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation
	41	Solyc09g013080.2.1	5484876-5491347	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	primary metabolic process
	42	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization
	43	Solyc11g013110.1.1	5961440-5965680	Anthocyanidin synthase	secondary metabolic process
	44	Solyc02g079500.2.1	38618009-38620468	Peroxidase	cell wall organisation or biogenesis
*	45	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	46	Solyc10g080210.1.1	60883700-60890335	Polygalacturonase A	macromolecule metabolic process
	47	Solyc03g097270.2.1	53034151-53040369	Cysteine proteinase inhibitor	regulation of biological process
	48	Solyc04g082200.2.1	63550865-63552237	Dehydrin	response to stress
	49	Solyc06g076570.1.1	43954001-43954465	class I heat shock protein	response to stress
	50	Solyc02g081160.2.1	39802521-39809198	Diphosphate-fructose-6-phosphate 1-phosphotransferase	primary metabolic process
	51	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	52	Solyc04g011440.2.1	3894918-3898067	heat shock protein	response to stress
*	53	Solyc01g090700.2.1	76099361-76107876	Enoyl-CoA-hydratase	primary metabolic process
*	54	Solyc05g056310.2.1	64792047-64799309	T-complex protein 1 subunit gamma	macromolecule metabolic process
	55	Solyc02g079930.2.1	38895029-38898734	Phosphosulfolactate synthase	response to stress
*	56	Solyc01g099190.2.1	81251679-81256014	Lipoxygenase	primary metabolic process
*	57	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process
	58	Solyc01g005560.2.1	394402-399248	Isocitrate dehydrogenase	primary metabolic process
	59	Solyc06g083790.2.1	45398741-45407021	Succinyl-CoA ligase	primary metabolic process
*	60	Solyc02g078360.2.1	37652619-37657447	Thioredoxin family protein	oxidation-reduction process
	61	Solyc05g013990.2.1	7492470-7499923	T-complex protein 1 subunit epsilon	macromolecule metabolic process
	62	Solyc10g005650.2.1	519710-532795	Peroxisomal targeting signal 1 receptor	establishment of localization
	63	Solyc02g091840.2.1	47648877-47652967	Receptor like kinase. RLK	regulation of biological process

	64	Solyc03g025950.2.1	7764256-7767163	Membrane-associated progesterone receptor component 1	regulation of biological process
	65	Solyc01g111760.2.1	89701090-89706839	V-type ATP synthase beta chain	establishment of localization
§	66	Solyc12g042060.1.1	42533149-42537576	ATP-dependent clp protease ATP-binding subunit	establishment of localization
	67	Solyc07g049450.2.1	57040844-57048137	Thioredoxin/protein disulfide isomerase	macromolecule metabolic process
	68	Solyc02g081160.2.1	39802521-39809198	Diphosphate-fructose-6-phosphate 1-phosphotransferase	primary metabolic process
	69	Solyc01g111760.2.1	89701090-89706839	V-type ATP synthase beta chain	establishment of localization
	70	Solyc02g083590.2.1	41515715-41520837	Dehydroquinate synthase	primary metabolic process
§	71	Solyc05g008450.2.1	2800495-2805716	Oxidoreductase FAD/NAD	oxidation-reduction process
#	72	Solyc11g066100.1.1	48856641-48858939	heat shock protein	response to stress
	73	Solyc09g090140.2.1	65031288-65034684	Malate dehydrogenase	primary metabolic process
#	74	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress
	75	Solyc04g076820.1.1	59289748-59291172	Octicosapeptide/Phox/Bem1p domain-containing protein	Unknown
§	76	Solyc12g044600.2.1	45079934-45087825	NADP-dependent malic enzyme. chloroplastic	primary metabolic process
	77	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress
	78	Solyc12g008630.1.1	2007849-2015328	Mitochondrial processing peptidase alpha subunit	macromolecule metabolic process
	79	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress
	80	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress
	81	Solyc10g081240.1.1	61667783-61677144	Protein grpE	primary metabolic process
	82	Solyc09g082720.2.1	63816287-63819690	Aldo/keto reductase family protein	oxidation-reduction process
#	83	Solyc06g076560.1.1	43949614-43950078	class I heat shock protein	response to stress
	84	Solyc10g078930.1.1	59896781-59900058	Activator of heat shock protein ATPase homolog 1	response to stress
	85	Solyc09g015000.2.1	7427223-7428264	class I heat shock protein	response to stress
	86	Solyc03g113930.1.1	58031590-58032156	class IV heat shock protein	response to stress
	87	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	88	Solyc03g111720.2.1	56431489-56432545	Peptide methionine sulfoxide reductase msrA	oxidation-reduction process
	89	Solyc01g111040.2.1	89209383-89212670	EF-Hand containing protein-like	Unknown
	90	Solyc04g082200.2.1	63550865-63552237	Dehydrin	response to stress
	91	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process
	92	Solyc08g062340.2.1	48122173-48122945	Class II small heat shock protein Le-HSP17.6	response to stress
	93	Solyc08g078700.2.1	59635844-59637072	Heat shock protein 22	response to stress
	94	Solyc02g090030.2.1	46276521-46278207	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process
	95	Solyc12g088720.1.1	62418023-62421876	Polyadenylate-binding protein 2	macromolecule metabolic process
	96	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	regulation of biological process
	97	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress
	98	Solyc10g083650.1.1	62763114-62763800	Peroxiredoxin ahpC/TSA family	oxidation-reduction process
	99	Solyc02g077710.1.1	37172109-37173137	E6-2 protein kinase	regulation of biological process
	100	Solyc10g083650.1.1	62763114-62763800	Peroxiredoxin ahpC/TSA family	oxidation-reduction process
	101	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis
	102	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process
*	103	Solyc08g075210.1.1	56497821-56499095	Acyltransferase-like protein	regulation of biological process
	104	Solyc01g104950.2.1	85031292-85035396	Alpha-L-arabinofuranosidase/beta-D-xylosidase	cell wall organisation or biogenesis
	105	Solyc02g065400.2.1	31151784-31154764	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process
	106	Solyc01g110450.2.1	88910585-88913966	NADP dependent sorbitol 6-phosphate dehydrogenase	primary metabolic process
	107	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process
#	108	Solyc05g055160.2.1	64069121-64073511	DNAJ chaperone	regulation of biological process
	109	Solyc09g011030.2.1	4368272-4373182	Hsp70 nucleotide exchange factor fes1	regulation of biological process
	110	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process
	111	Solyc12g055830.1.1	47184732-47187107	Inorganic pyrophosphatase	primary metabolic process
*	112	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress
*	113	Solyc07g066600.2.1	65205704-65208684	Phosphoglycerate kinase	primary metabolic process
	114	Solyc02g081700.1.1	40134968-40135714	Proteasome subunit alpha type	macromolecule metabolic process
	115	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process
	116	Solyc02g086880.2.1	44063088-44067604	Formate dehydrogenase	oxidation-reduction process
	117	Solyc07g064800.2.1	64009800-64015982	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	macromolecule metabolic process
	118	Solyc06g005150.2.1	170218-173338	Ascorbate peroxidase	oxidation-reduction process
	119	Solyc06g083190.2.1	45004847-45008910	Peptidyl-prolyl cis-trans isomerase	macromolecule metabolic process
	120	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process
	121	Solyc10g005100.2.1	92148-94188	Salt stress root protein RS1	response to stress
	122	Solyc12g005080.1.1	38442-41171	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	macromolecule metabolic process
	123	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process
	124	Solyc05g014470.2.1	8322381-8324881	Glyceraldehyde 3-phosphate dehydrogenase	primary metabolic process
	125	Solyc09g098150.2.1	67309562-67312625	Metacaspase 7	macromolecule metabolic process
	126	Solyc01g104170.2.1	84393496-84398307	Ankyrin repeat domain-containing protein 2	regulation of biological process
#	127	Solyc10g084050.1.1	63048965-63053841	26S protease regulatory subunit 6B homolog	macromolecule metabolic process
	128	Solyc01g111120.2.1	89253979-89260194	Triosephosphate isomerase	primary metabolic process

	129	Solyc01g099760.2.1	81668247-81673265	26S protease regulatory subunit 6A homolog	macromolecule metabolic process
	130	Solyc05g005700.2.1	514498-518886	Aldehyde dehydrogenase 1	primary metabolic process
§	131	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process
	132	Solyc02g091100.2.1	47093169-47096847	Oxalyl-CoA decarboxylase	primary metabolic process
	133	Solyc03g115650.2.1	59340664-59343595	Eukaryotic translation initiation factor 5A	translation
	134	Solyc01g102960.2.1	83371681-83372560	class IV heat shock protein	response to stress
	135	Solyc01g057000.2.1	50798585-50800986	Universal stress protein family protein	response to stress
	136	Solyc02g081170.2.1	39810752-39812645	Plastid-lipid-associated protein. chloroplastic	response to stress
	137	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process
*	138	Solyc02g092670.1.1	48256556-48258820	Subtilisin-like protease	macromolecule metabolic process
*	139	Solyc05g056230.2.1	64732387-64738654	Calreticulin 2 calcium-binding protein	macromolecule metabolic process
#	140	Solyc03g082920.2.1	46341372-46345339	Heat shock protein	response to stress
#	141	Solyc03g082920.2.1	46341372-46345339	Heat shock protein	response to stress
	142	Solyc09g098150.2.1	67309562-67312625	Metacaspase 7	macromolecule metabolic process
	143	Solyc04g040180.2.1	31093421-31095617	S-adenosylmethionine-dependent methyltransferase	hormone metabolic process
*	144	Solyc01g108540.2.1	87591739-87592913	Acetyl esterase	primary metabolic process
	145	Solyc02g031950.2.1	17912940-17915049	Pathogenesis-related protein-like protein	response to stress
	146	no identification			
	147	Solyc10g081030.1.1	61534917-61538284	Nascent polypeptide-associated complex alpha subunit-like protein	establishment of localization
	148	Solyc00g187050.2.1	18478269-18481034	Leucyl aminopeptidase	macromolecule metabolic process
#	149	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress
	150	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process
	151	Solyc10g085550.1.1	63996015-64000623	Enolase	primary metabolic process
	152	Solyc06g075010.2.1	42927697-42932331	chaperonin	macromolecule metabolic process
	153	Solyc03g121640.2.1	63788759-63796170	chaperonin	macromolecule metabolic process
	154	Solyc03g115990.1.1	59594180-59595418	Malate dehydrogenase	primary metabolic process
	155	Solyc03g114500.2.1	58538109-58542525	Enolase	primary metabolic process
	156	Solyc01g097340.2.1	79999429-80002633	NAD-dependent epimerase/dehydratase family protein-like protein	cell wall organisation or biogenesis
	157	Solyc08g082430.2.1	62404261-62407246	Nucleoside diphosphate kinase	primary metabolic process
	158	Solyc12g098420.1.1	64267844-64273580	Ubiquitin carboxyl-terminal hydrolase	macromolecule metabolic process
	159	Solyc12g096190.1.1	63553620-63556459	Tryptophan synthase beta chain	primary metabolic process
	160	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process
	161	Solyc03g114500.2.1	58538109-58542525	Enolase	primary metabolic process
	162	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	163	Solyc04g082630.2.1	63844019-63847151	Glyceraldehyde-3-phosphate dehydrogenase B	primary metabolic process
	164	Solyc01g106430.2.1	86088310-86092110	Inorganic pyrophosphatase family protein	primary metabolic process
	165	Solyc10g078930.1.1	59896781-59900058	Activator of heat shock protein ATPase homolog 1	response to stress
	166	Solyc10g086580.1.1	64688483-64690961	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process
	167	Solyc11g072450.1.1	52705683-52710449	Mitochondrial F0 ATP synthase D chain	establishment of localization
	168	Solyc01g104950.2.1	85031292-85035396	Alpha-L-arabinofuranosidase/beta-D-xylosidase	cell wall organisation or biogenesis
	169	Solyc05g014470.2.1	8322381-8324881	Glyceraldehyde 3-phosphate dehydrogenase	primary metabolic process
#	170	Solyc04g005340.2.1	248903-251735	Alpha-1 4-glucan protein synthase	cell wall organisation or biogenesis
	171	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process
*	172	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process
	173	Solyc09g082990.2.1	64083974-64088590	GDP-D-mannose-3,5-epimerase	vitamin synthesis
	174	Solyc12g006870.1.1	1315304-1322819	Acyl-protein thioesterase 2	macromolecule metabolic process
	175	Solyc09g064370.2.1	57150099-57154745	Alcohol dehydrogenase	cellular metabolic process
	176	Solyc07g041490.1.1	50632738-50633541	Stress responsive alpha-beta barrel domain protein	response to stress
	177	Solyc09g090140.2.1	65031288-65034684	Malate dehydrogenase	primary metabolic process
	178	Solyc12g009400.1.1	2682210-2685916	Pyruvate dehydrogenase E1 component alpha subunit	primary metabolic process
	179	Solyc08g014340.2.1	4063851-4071314	Cysteine synthase	primary metabolic process
	180	Solyc03g120090.1.1	62641034-62641951	Pyridoxal biosynthesis lyase pdxS	vitamin synthesis
	181	Solyc11g069790.1.1	51502451-51507359	chaperonin	macromolecule metabolic process
	182	Solyc04g079200.2.1	61339502-61344701	26S proteasome regulatory subunit	macromolecule metabolic process
	183	Solyc03g113800.2.1	57894787-57902443	Betaine aldehyde dehydrogenase	primary metabolic process
	184	Solyc09g005740.1.1	508767-509831	Chloroplast lumen common family protein	Unknown
	185	Solyc07g061790.2.1	61964267-61965352	Heme-binding protein 2	Unknown
	186	Solyc02g078120.1.1	37474003-37475736	Eukaryotic translation initiation factor 3 subunit 7	translation
	187	Solyc11g011470.1.1	4520459-4526103	NADH-ubiquinone oxidoreductase subunit	oxidation-reduction process
	188	Solyc05g008460.2.1	2805741-2809779	ATP synthase subunit beta	regulation of biological process
*	189	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization
	190	Solyc04g055170.2.1	52933311-52937146	Annexin 2	establishment of localization
	191	Solyc12g043020.1.1	44067590-44080483	Dihydroxy-acid dehydratase	primary metabolic process
	192	Solyc11g039980.1.1	28095562-28096080	ATP synthase subunit alpha	establishment of localization
	193	Solyc01g066480.2.1	66874829-66881864	Fumarylacetoacetate hydrolase domain-containing protein 1	primary metabolic process
	194	Solyc11g066060.1.1	48824058-48826931	heat shock protein	response to stress

	195	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress
	196	Solyc11g069000.1.1	50652125-50657452	T-complex protein 1 subunit beta	macromolecule metabolic process
#	197	Solyc06g071790.2.1	40601620-40603444	Elongation factor Tu	translation
	198	Solyc05g012480.2.1	5715224-5720976	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process
	199	Solyc03g031720.2.1	8453484-8459499	RNA Binding Protein 45	regulation of biological process
§	200	Solyc09g010630.2.1	3965253-3968837	heat shock protein	response to stress
	201	Solyc01g028810.2.1	33327242-33332361	chaperonin	macromolecule metabolic process
	202	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process
	203	Solyc03g083390.2.1	46783589-46788273	Nuclear movement protein nudc	regulation of biological process
	204	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress
	205	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process
	206	Solyc05g017760.2.1	18312063-18317108	Acetyl-CoA C-acetyltransferase	primary metabolic process
	207	Solyc07g008720.2.1	3694316-3696017	Nascent polypeptide-associated complex subunit beta	establishment of localization
§	208	Solyc01g106210.2.1	85918285-85922197	Chaperone DnaK	response to stress
	209	Solyc02g081140.2.1	39782470-39786000	UBX domain-containing protein	regulation of biological process
	210	Solyc03g082420.2.1	45898742-45899828	Heat shock protein	response to stress
*	211	Solyc09g072560.2.1	60593208-60595291	Legumin 11S-globulin	macromolecule metabolic process
	212	Solyc09g082860.2.1	63941988-63946072	Sulfate adenylyltransferase	primary metabolic process
	213	Solyc02g062460.2.1	28653823-28656374	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	214	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process
	215	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress
	216	Solyc06g073280.2.1	41539950-41545003	LL-diaminopimelate aminotransferase	primary metabolic process
	217	Solyc02g082920.2.1	41128832-41129990	Endochitinase	macromolecule metabolic process
	218	Solyc06g073190.2.1	41483219-41485991	Fructokinase-like	primary metabolic process
	219	Solyc02g082800.2.1	41021911-41031456	Ubiquilin-1	macromolecule metabolic process
	220	Solyc03g121720.2.1	63848472-63854092	2-hydroxy-3-oxopropionate reductase	primary metabolic process
	221	Solyc06g050550.2.1	29742430-29749862	Sorting nexin 1	establishment of localization
	222	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	223	Solyc01g106320.2.1	86005942-86007808	Octicosapeptide/Phox/Bem1p domain-containing protein	Unknown
	224	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process
	225	Solyc09g011080.2.1	4413327-4416467	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process
	226	Solyc08g068310.2.1	54615972-54621953	RNA-binding La domain protein	regulation of biological process
	227	Solyc01g110450.2.1	88910585-88913966	NADP dependent sorbitol 6-phosphate dehydrogenase	primary metabolic process
	228	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process
	229	Solyc06g073090.2.1	41417554-41421363	Ribosomal subunit interface protein	macromolecule metabolic process
	230	Solyc08g080370.2.1	60840952-60844640	Acetylornithine aminotransferase	primary metabolic process
*	231	Solyc08g074410.2.1	55693755-55701940	Tryptophanyl-tRNA synthetase	macromolecule metabolic process
	232	Solyc05g056310.2.1	64792047-64799309	T-complex protein 1 subunit gamma	macromolecule metabolic process
	233	Solyc01g006980.2.1	1552093-1555934	Malonyl CoA-acyl carrier protein transacylase containing protein expressed	primary metabolic process
	234	Solyc05g053300.2.1	62572547-62577624	Dihydrolipoyl dehydrogenase	primary metabolic process
	235	Solyc08g022210.2.1	14105526-14113435	Methylthioribose-1-phosphate isomerase	primary metabolic process
	236	Solyc06g081980.1.1	44239265-44240194	Pyridoxal biosynthesis lyase pdxS	vitamin synthesis
	237	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	238	Solyc02g062500.2.1	28750335-28754482	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	239	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	240	Solyc07g062970.2.1	62819985-62823101	Serine/threonine phosphatase family protein	macromolecule metabolic process
	241	Solyc09g009260.2.1	2643600-2645801	Fructose-bisphosphate aldolase	primary metabolic process
	242	Solyc09g064940.2.1	58086206-58088531	Phenazine biosynthesis protein PhzF family	secondary metabolic process
	243	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis
	244	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation
	245	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation
	246	Solyc12g005860.1.1	490745-498964	3-isopropylmalate dehydratase large subunit	primary metabolic process
	247	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization
*	248	Solyc04g005340.2.1	248903-251735	Alpha-1 4-glucan protein synthase	cell wall organisation or biogenesis
	249	Solyc11g069790.1.1	51502451-51507359	chaperonin	macromolecule metabolic process
	250	Solyc05g056490.2.1	64907109-64913076	3+apos	primary metabolic process
	251	Solyc04g076200.2.1	58724666-58728561	Universal stress protein family protein	response to stress
#	252	Solyc06g071790.2.1	40601620-40603444	Elongation factor Tu	translation
	253	Solyc10g005890.2.1	683370-690823	DNA-damage inducible protein DDI1-like	macromolecule metabolic process
*	254	Solyc01g079680.2.1	71306007-71308978	Ran GTPase activating protein	establishment of localization
	255	Solyc01g099240.2.1	81297091-81298341	3-hydroxyisobutyrate dehydrogenase	primary metabolic process
	256	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process
	257	Solyc07g063680.2.1	63306131-63309625	CHP-rich zinc finger protein-like	response to stress
§	258	Solyc07g055320.2.1	60717008-60721284	ATP-dependent Zn protease cell division protein FtsH homolog	macromolecule metabolic process
#	259	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress

	260	Solyc02g088610.2.1	45222192-45229259	ATP-dependent chaperone ClpB	macromolecule metabolic process
	261	Solyc04g016470.2.1	7303341-7305043	Beta-1 3-glucanase	cell wall organisation or biogenesis
*	262	Solyc06g068860.2.1	39073381-39083550	Alpha-mannosidase	primary metabolic process
	263	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis
	264	Solyc10g047950.1.1	38620947-38627409	Inorganic pyrophosphatase family protein	primary metabolic process
	265	Solyc06g082120.2.1	44333171-44336113	Ran GTPase binding protein	establishment of localization
	266	Solyc02g086880.2.1	44063088-44067604	Formate dehydrogenase	oxidation-reduction process
*	267	Solyc10g085550.1.1	63996015-64000623	Enolase	primary metabolic process
	268	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process
	269	Solyc05g050800.2.1	60131934-60136497	Phosphoglycerate mutase family protein	primary metabolic process
	270	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process
	271	Solyc08g080140.2.1	60644754-60646888	dTDP-4-dehydrorhamnose reductase	primary metabolic process
	272	Solyc05g012480.2.1	5715224-5720976	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process
	273	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	translation
	274	Solyc04g009200.2.1	2695320-2699163	Glutamate-1-semialdehyde-2 1-aminomutase	nitrogen compound metabolic process
	275	Solyc04g016360.2.1	7158486-7165657	S-formylglutathione hydrolase	primary metabolic process
	276	Solyc03g095900.2.1	51028612-51030467	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation
#	277	Solyc10g080940.1.1	61453818-61456401	Tubulin beta chain	cytoskeleton organization and biogenesis
	278	Solyc09g015020.1.1	7440133-7440597	class I heat shock protein 3	response to stress
	279	Solyc08g079870.1.1	60466441-60468678	Subtilisin-like protease	macromolecule metabolic process
#	280	Solyc01g059980.2.1	62201815-62226187	Beta-glucanase	cell wall organisation or biogenesis
	281	Solyc08g079870.1.1	60466441-60468678	Subtilisin-like protease	macromolecule metabolic process
	282	Solyc08g062450.1.1	48318166-48318642	class II heat shock protein	response to stress
	283	Solyc10g055810.1.1	52891942-52893092	Endochitinase	macromolecule metabolic process
	284	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	285	Solyc05g005490.2.1	355897-361496	Carbonic anhydrase	nitrogen compound metabolic process
	286	Solyc04g054980.2.1	52664826-52665968	Lipoxygenase homology domain-containing protein 1	primary metabolic process
	287	Solyc07g041900.2.1	51828739-51831068	Cathepsin L-like cysteine proteinase	macromolecule metabolic process
	288	Solyc12g010320.1.1	3377663-3380422	Outer membrane lipoprotein blc	response to stress
	289	Solyc01g079220.2.1	70815571-70820929	NifU like protein	cellular metabolic process
	290	Solyc06g073280.2.1	41539950-41545003	LL-diaminopimelate aminotransferase	primary metabolic process
	291	Solyc02g080630.2.1	39392539-39397791	Lactoylglutathione lyase	primary metabolic process
	292	Solyc08g014130.2.1	3734998-3744536	2-isopropylmalate synthase 1	primary metabolic process
	293	Solyc10g084400.1.1	63281886-63284225	Glutathione S-transferase	response to stress
	294	Solyc03g116110.2.1	59675160-59679638	Alpha/beta hydrolase fold protein	primary metabolic process
	295	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress
	296	Solyc09g007270.2.1	865197-869322	Ascorbate peroxidase	oxidation-reduction process
	297	Solyc10g006650.2.1	1157432-1161954	Flavoprotein wrbA	oxidation-reduction process
	298	Solyc12g096190.1.1	63553620-63556459	Tryptophan synthase beta chain	primary metabolic process
	299	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress
*	300	Solyc01g100520.2.1	82287668-82292399	ATP-dependent Clp protease proteolytic subunit	establishment of localization
*	301	Solyc07g051850.2.1	57713505-57718748	Aspartic proteinase	macromolecule metabolic process
	302	Solyc01g005560.2.1	394402-399248	Isocitrate dehydrogenase	primary metabolic process
	303	Solyc07g051850.2.1	57713505-57718748	Aspartic proteinase	macromolecule metabolic process
	304	Solyc03g025850.2.1	7627656-7630095	Remorin 1	Unknown
§	305	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process
	306	Solyc07g064590.2.1	63891191-63895627	Ubiquitin thioesterase OTU1	primary metabolic process
	307	Solyc04g072400.2.1	57015439-57020648	Endoribonuclease E-like protein	macromolecule metabolic process
	308	Solyc02g062460.2.1	28653823-28656374	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	309	Solyc03g097270.2.1	53034151-53040369	Cysteine proteinase inhibitor	regulation of biological process
	310	Solyc09g082720.2.1	63816287-63819690	Aldo/keto reductase family protein	oxidation-reduction process
	311	Solyc12g044740.1.1	45436676-45448004	Ubiquitin carboxyl-terminal hydrolase	primary metabolic process
	312	Solyc01g107700.2.1	86915524-86922302	Kynurenine formamidase	secondary metabolic process
	313	Solyc06g071000.2.1	40003245-40009711	N-succinylglutamate 5-semialdehyde dehydrogenase	primary metabolic process
	314	Solyc09g090330.2.1	65188896-65193572	Harpin binding protein 1	response to stress
	315	Solyc03g025850.2.1	7627656-7630095	Remorin 1	Unknown
	316	Solyc01g097460.2.1	80067637-80074843	Ribose-5-phosphate isomerase	primary metabolic process
*	317	Solyc01g104170.2.1	84393496-84398307	Ankyrin repeat domain-containing protein 2	regulation of biological process
	318	Solyc01g067740.2.1	69074003-69078931	Superoxide dismutase	oxidation-reduction process
	319	Solyc06g063090.2.1	36234795-36239218	Alanine aminotransferase	primary metabolic process
	320	Solyc12g009060.1.1	2350521-2354721	Charged multivesicular body protein 2a	establishment of localization
#	321	Solyc03g112150.1.1	56698575-56700008	Elongation factor Tu	translation
	322	Solyc01g112280.2.1	90105382-90109100	Succinyl-diaminopimelate desuccinylase	primary metabolic process
	323	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process
	324	Solyc07g008170.2.1	2882629-2893614	Methyl binding domain protein	regulation of biological process
#	325	Solyc01g011000.2.1	6848050-6851207	Eukaryotic translation initiation factor 5A	translation
#	326	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process
	327	Solyc05g005480.2.1	352211-355615	Oxidoreductase zinc-binding dehydrogenase	oxidation-reduction process

	328	Solyc03g007670.2.1	2200919-2205536	SGT1	response to stress
*	329	Solyc02g093830.2.1	49113784-49119691	Glucose-6-phosphate 1-dehydrogenase	primary metabolic process
	330	Solyc12g009250.1.1	2529842-2532163	chaperonin	macromolecule metabolic process
	331	Solyc12g055800.1.1	47132031-47139706	V-type ATP synthase alpha chain	establishment of localization
#	332	Solyc06g076020.2.1	43582389-43585486	heat shock protein	response to stress
	333	Solyc03g115820.2.1	59463075-59469502	Ribulose-5-phosphate-3-epimerase	primary metabolic process
§	334	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process
	335	Solyc09g009390.2.1	2835367-2840425	Monodehydroascorbate reductase	vitamin synthesis
	336	Solyc12g009140.1.1	2453127-2457231	Proteasome subunit alpha type	macromolecule metabolic process
#	337	Solyc05g012070.2.1	5295089-5298245	Alpha-1 4-glucan-protein synthase	cell wall organisation or biogenesis
	338	Solyc02g088700.2.1	45268178-45273518	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process
	339	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress
	340	Solyc01g007320.2.1	1849252-1851757	ATP synthase subunit beta chloroplastic	establishment of localization
	341	Solyc06g082090.2.1	44317047-44320702	Methionine aminopeptidase	macromolecule metabolic process
	342	Solyc07g064810.2.1	64017771-64025505	Imidazole glycerol phosphate synthase subunit hisF	primary metabolic process
	343	Solyc02g087300.1.1	44325520-44326428	Protein transport SEC13-like protein	establishment of localization
	344	Solyc11g011960.1.1	4912805-4919067	UTP-glucose 1 phosphate uridylyltransferase	primary metabolic process
*	345	Solyc02g081400.2.1	39962544-39967049	Intracellular protease PfpI family protein expressed	macromolecule metabolic process
	346	Solyc09g011670.2.1	4913443-4917118	Universal stress protein family protein	response to stress
§	347	Solyc10g086580.1.1	64688483-64690961	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process
*	348	Solyc09g011140.2.1	4499977-4502723	Tropinone reductase I	secondary metabolic process
	349	Solyc07g064160.2.1	63644226-63646402	Thiazole biosynthetic enzyme	vitamin synthesis
	350	Solyc03g098700.1.1	54427774-54428433	Cysteine protease inhibitor 8	regulation of biological process
	351	Solyc02g082800.2.1	41021911-41031456	Ubiquilin-1	primary metabolic process
	352	Solyc03g120230.2.1	62741263-62746412	MAR-binding filament-like protein 1	cytoskeleton organization and biogenesis
	353	Solyc10g008740.2.1	2822974-2826287	Magnesium chelatase ATPase subunit I	primary metabolic process
*	354	Solyc01g109660.2.1	88321513-88322795	Glycine-rich RNA-binding protein	response to stress
	355	Solyc06g007200.2.1	1270929-1276535	Methionine aminopeptidase	macromolecule metabolic process
	356	Solyc01g005520.2.1	349916-352687	Tetratricopeptide TPR_2 repeat protein	response to stress
*	357	Solyc02g082830.1.1	41056853-41058151	Phosphoserine aminotransferase	primary metabolic process
	358	Solyc01g103450.2.1	83821528-83826037	Chaperone DnaK	response to stress
	359	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis
*	360	Solyc05g056230.2.1	64732387-64738654	Calreticulin 2 calcium-binding protein	macromolecule metabolic process
	361	Solyc09g011240.2.1	4573233-4578442	Reductase 2	oxidation-reduction process
	362	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis
	363	Solyc02g091490.2.1	47344370-47349179	Fructokinase 3	primary metabolic process
#	364	Solyc05g008460.2.1	2805741-2809779	ATP synthase subunit beta	establishment of localization
	365	Solyc01g009420.2.1	3621816-3626739	Bifunctional polymyxin resistance arnA protein	macromolecule metabolic process
	366	Solyc12g056830.1.1	48264644-48265396	ATP synthase delta subunit	establishment of localization
*	367	Solyc08g076220.2.1	57405726-57410850	Phosphoribulokinase/uridine kinase	primary metabolic process
	368	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	translation
	369	Solyc12g042650.1.1	43448722-43449792	40S ribosomal protein S12	translation
	370	Solyc07g044860.2.1	55244985-55247025	Oxygen-evolving enhancer protein 2. chloroplastic	regulation of biological process
	371	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process
	372	Solyc09g005700.2.1	488084-495891	Diaminopimelate epimerase family protein	primary metabolic process
	373	Solyc10g083970.1.1	62989745-62990917	S-adenosylmethionine synthase	hormone metabolic process
	374	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress
§	375	Solyc10g008010.2.1	2177130-2182889	Proteasome subunit alpha type	macromolecule metabolic process
#	376	Solyc07g062570.2.1	62472541-62478080	Ubiquitin-conjugating enzyme E2 N	primary metabolic process
	377	Solyc09g090980.2.1	65698002-65700023	Major allergen Mal d 1	response to stress
	378	Solyc05g052150.2.1	61597939-61598511	ATP synthase subunit delta+apos;. mitochondrial	establishment of localization
	379	Solyc05g053810.2.1	62993932-62998448	Serine hydroxymethyltransferase	primary metabolic process
	380	Solyc02g014150.2.1	6519093-6531932	Photosystem II stability/assembly factor Ycf48-like protein	regulation of biological process
#	381	Solyc01g099760.2.1	81668247-81673265	26S protease regulatory subunit 6A homolog	macromolecule metabolic process
*	382	Solyc06g062950.1.1	36119748-36122078	Subtilisin-like protease	macromolecule metabolic process
	383	Solyc01g028810.2.1	33327242-33332361	chaperonin	macromolecule metabolic process
	384	Solyc02g086910.2.1	44078232-44082017	Peptidyl-prolyl cis-trans isomerase cyclophilin-type	macromolecule metabolic process
	385	Solyc04g007120.2.1	832713-843272	UV excision repair protein RAD23	macromolecule metabolic process
	386	Solyc11g020300.1.1	10728385-10741180	Translocon Tic40	establishment of localization
#	387	Solyc02g070510.2.1	34815271-34822336	Proteasome subunit alpha type	macromolecule metabolic process
*	388	Solyc01g106260.2.1	85976768-85981140	Chaperone DnaK	response to stress
	389	Solyc08g062660.2.1	48903019-48905978	Ran GTPase binding protein	establishment of localization
	390	Solyc05g053470.2.1	62706122-62712174	chaperonin	macromolecule metabolic process
	391	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
*	392	Solyc01g105060.2.1	85094784-85097712		macromolecule metabolic process
	393	Solyc08g074620.1.1	55911248-55913011	Polyphenol oxidase	secondary metabolic process
	394	Solyc06g065270.2.1	37095259-37099538	Adenylate kinase	primary metabolic process

#	395	Solyc02g086730.1.1	43937531-43938103	50S ribosomal protein L12-C	translation
	396	Solyc01g100030.2.1	81869800-81871185	Deoxyuridine 5+apos-triphosphate nucleotidohydrolase	macromolecule metabolic process
	397	Solyc08g079020.2.1	59830669-59836240	Adenine phosphoribosyltransferase-like protein	primary metabolic process
	398	no identification			
	399	Solyc02g088700.2.1	45268178-45273518	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process
*	400	Solyc03g007520.2.1	2089415-2095197	Proline-rich cell wall protein-like	cell wall organisation or biogenesis
	401	Solyc02g063130.2.1	29750711-29757113	UV excision repair protein RAD23	macromolecule metabolic process
	402	Solyc08g014340.2.1	4063851-4071314	Cysteine synthase	primary metabolic process
	403	Solyc11g011960.1.1	4912805-4919067	UTP-glucose 1 phosphate uridylyltransferase	primary metabolic process
	404	Solyc12g056230.1.1	47552627-47555380	Glutathione peroxidase	response to stress
	405	Solyc01g097450.2.1	80067623-80073617	Thioredoxin-like protein 1	oxidation-reduction process
§	406	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process
	407	Solyc01g099770.2.1	81675491-81677536	Translationally-controlled tumor protein homolog	regulation of biological process
#	408	Solyc06g082630.2.1	44672378-44677754	26S protease regulatory subunit 6B	macromolecule metabolic process
	409	Solyc06g060260.2.1	34615474-34625952	Stromal ascorbate peroxidase 7	oxidation-reduction process
	410	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress
	411	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process
*	412	Solyc05g051850.2.1	61409236-61413326	Inositol-3-phosphate synthase	primary metabolic process
#	413	Solyc01g007860.2.1	2018307-2022494	Ubiquitin-conjugating enzyme family protein-like	primary metabolic process
	414	Solyc07g042250.2.1	52684617-52687267	chaperonin	macromolecule metabolic process
	415	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization
§	416	Solyc12g010060.1.1	3203778-3206878	Eukaryotic translation initiation factor 5A	translation
	417	Solyc02g086830.2.1	44017964-44022130	Protease Do-like	macromolecule metabolic process
	418	Solyc04g071620.2.1	56178656-56180338	ASR4	response to stress
*	419	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process
	420	Solyc01g080460.2.1	72206073-72214485	Pyruvate phosphate dikinase	primary metabolic process
#	421	Solyc02g062500.2.1	28750335-28754482	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process
	422	Solyc09g083410.2.1	64466144-64472295	Amidase hydantoinase/carbamoylase family protein expressed	primary metabolic process
	423	Solyc12g055800.1.1	47132031-47139706	V-type ATP synthase alpha chain	establishment of localization
	424	Solyc02g031950.2.1	17912940-17915049	Pathogenesis-related protein-like protein	response to stress
	RA011	Solyc08g077180.2.1	58254085-58261125	Pyruvate kinase	primary metabolic process
	RA036	Solyc12g009180.1.1	2477766-2480583	Cupin superfamily	Unknown
	RA037	Solyc06g083810.2.1	45406955-45411773	Glycolipid transfer protein domain-containing protein 1	establishment of localization
	RA051	Solyc01g007330.2.1	1851938-1853783	Ribulose bisphosphate carboxylase large chain	cellular metabolic process
	RA064	Solyc04g073970.2.1	57593055-57595973	cDNA clone J033025P19 full insert sequence	Unknown
	RA072	Solyc07g053690.2.1	59452115-59454460	OB-fold nucleic acid binding domain containing protein	Unknown
§	RA080	Solyc04g058070.2.1	54328964-54342003	UDP-N-acetylglucosamine-pyrophosphorylase	primary metabolic process
*	RA085	Solyc07g006650.2.1	1517595-1522933	Xylose isomerase	cell wall organisation or biogenesis
	RA089	Solyc09g075450.2.1	62645088-62653706	Fumarate hydratase class II	primary metabolic process
	RA090	Solyc08g076480.2.1	57653458-57656810	Plastid lipid-associated protein 3, chloroplastic	regulation of biological process
	RA091	Solyc02g065240.2.1	30994199-30996053	Hydrolase alpha/beta fold family protein	response to stress
#	RA112	Solyc01g060470.2.1	63867613-63872933	Importin alpha-1b subunit	establishment of localization
	RA116	Solyc06g063140.2.1	36272039-36278473	26S protease regulatory subunit 7	macromolecule metabolic process
	RA125	Solyc05g026050.2.1	39648364-39651972	Charged multivesicular body protein 4b	establishment of localization
	RA127	Solyc06g036110.1.1	22246752-22247498	C2 domain-containing protein	regulation of biological process
#	RA130	Solyc03g082580.2.1	46048220-46052563	Glucosamine-6-phosphate deaminase	primary metabolic process
	RA134	Solyc02g079750.2.1	38774685-38777292	Flavoprotein wrbA	oxidation-reduction process
*	RA139	Solyc04g008280.2.1	1954623-1957620	Regulator of ribonuclease activity A	regulation of biological process
	RA164	Solyc09g082730.2.1	63820711-63824566	Aldo/keto reductase family protein	oxidation-reduction process
	RA169	Solyc06g053200.2.1	32435446-32437176	6-phosphogluconolactonase	primary metabolic process
*	RA171	Solyc07g021540.2.1	19310213-19316385	FIP1	macromolecule metabolic process
	RA176	Solyc00g009020.2.1	8742084-8746574	Mitochondrial ATP synthase	establishment of localization
	RA183	Solyc08g074390.2.1	55669339-55671622	Glucan endo-1 3-beta-glucosidase 6	cell wall organisation or biogenesis
*	RA185	Solyc12g005200.1.1	131226-135857	Electron transfer flavoprotein alpha subunit	oxidation-reduction process
	RA190	Solyc11g020330.1.1	10856316-10856888	class IV heat shock protein	response to stress
	RA191	Solyc03g117430.2.1	60652978-60660774	Cobalamin synthesis protein P	vitamin synthesis
	RA205	Solyc11g067160.1.1	49978029-49981767	Aldo/keto reductase family protein	oxidation-reduction process
	RA209	Solyc06g005360.2.1	372206-374775	Actin depolymerizing factor 3	cytoskeleton organization and biogenesis
	RA210	Solyc09g091000.2.1	65708207-65709306	Major allergen Mal d 1	response to stress
	RA213	Solyc10g049890.1.1	42590722-42594292	Phosphoglycerate dehydrogenase	primary metabolic process
	RA214	Solyc03g005260.2.1	148411-153101	Sulfate adenylyltransferase	primary metabolic process
	RA218	Solyc07g017900.2.1	8283617-8294010	Aldose 1-epimerase family protein	primary metabolic process
	RA220	Solyc02g088790.2.1	45342858-45345771	RNA-binding protein	macromolecule metabolic process
#	RA223	Solyc02g083810.2.1	41646719-41650004	Ferredoxin--NADP reductase	oxidation-reduction process
	RA226	Solyc09g010380.2.1	3768116-3770648	Inorganic pyrophosphatase-like protein	primary metabolic process
	RA231	Solyc12g005630.1.1	334079-336443	Cytochrome b6-f complex iron-sulfur subunit	cellular metabolic process

	RA232	Solyc03g025340.1.1	7153156-7154205	C2 domain-containing protein	regulation of biological process
	RA237	Solyc07g044840.2.1	55217173-55222925	2 3-bisphosphoglycerate-independent phosphoglycerate mutase	primary metabolic process
	RA241	Solyc04g077020.2.1	59583881-59586708	Tubulin alpha-3 chain	cytoskeleton organization and biogenesis
	RA242	Solyc11g067240.1.1	50060665-50065903	Vacuolar sorting protein 4b	establishment of localization
	RA243	Solyc05g013380.2.1	6451782-6457884	Alanine aminotransferase 2	nitrogen compound metabolic process
	RA248	Solyc02g080540.1.1	39325443-39326576	ATP synthase gamma chain	establishment of localization
	RA253	Solyc12g056110.1.1	47434718-47440701	V-type proton ATPase subunit E	establishment of localization
	RA258	Solyc08g065220.2.1	50293331-50300381	Glycine dehydrogenase P protein	primary metabolic process
	RA259	Solyc02g082840.2.1	41059268-41070574	Protein GRIP	establishment of localization
	RA260	Solyc07g047790.2.1	56327803-56334029	Chaperone protein htpG	macromolecule metabolic process
	RA270	Solyc02g062340.2.1	28530062-28532470	Fructose-bisphosphate aldolase	primary metabolic process
	RA273	Solyc04g081970.2.1	63392195-63394296	Thioredoxin	oxidation-reduction process
	RA274	Solyc07g066270.2.1	64973786-64976784	Glucosamine-6-phosphate deaminase	primary metabolic process
	RA275	Solyc09g018790.2.1	17001628-17004264	Gamma hydroxybutyrate dehydrogenase-like protein	secondary metabolic process
	RA277	Solyc01g081610.2.1	73313897-73316705	Beta-hexosaminidase 1	establishment of localization
*	RA293	Solyc07g043310.2.1	54313638-54320300	Aminotransferase	nitrogen compound metabolic process
	RA294	Solyc06g073060.2.1	41401609-41405851	Iaa-amino acid hydrolase 6	macromolecule metabolic process
*	RA302	Solyc09g065540.2.1	59258116-59272301	Methylcrotonoyl-CoA carboxylase alpha chain	primary metabolic process
	RA303	Solyc11g018550.2.1	8668547-8674999	Thylakoid-bound ascorbate peroxidase 6	oxidation-reduction process
	RA304	Solyc10g005110.2.1	96242-103626	Coproporphyrinogen III oxidase aerobic	secondary metabolic process
	RA305	Solyc11g011250.1.1	4291341-4296684	Chloride intracellular channel 6	establishment of localization
	RA308	Solyc01g104920.2.1	85004629-85013296	26S protease regulatory subunit 8 homolog	macromolecule metabolic process
	RA309	Solyc01g108880.2.1	87786634-87788179	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation
	RA311	Solyc09g082320.2.1	63473246-63478374	Proteasome subunit beta type	macromolecule metabolic process
	RA314	Solyc02g067750.2.1	32439990-32442359	Carbonic anhydrase	nitrogen compound metabolic process
#	RA316	Solyc07g061940.2.1	62037698-62039615	Acetolactate synthase	primary metabolic process
	RA319	Solyc11g011380.1.1	4456053-4460206	Glutamine synthetase	nitrogen compound metabolic process
	RA321	Solyc02g071700.2.1	35644262-35647481	GDSL esterase/lipase At1g29670	primary metabolic process
*	RA322	Solyc02g079060.2.1	38263644-38270218	Eukaryotic translation initiation factor 3 subunit J	translation
	RA327	Solyc07g064170.2.1	63653756-63656317	Pectinesterase	cell wall organisation or biogenesis
	RA345	Solyc02g066930.2.1	31762776-31764561	RNA-binding protein	macromolecule metabolic process
	RA347	Solyc12g042830.1.1	43709595-43709834	class I heat shock protein	response to stress
*	RA355	Solyc03g007950.2.1	2439038-2441907	Glycoside hydrolase family 28 protein/polygalacturonase family protein	cell wall organisation or biogenesis
	RA360	Solyc10g078740.1.1	59792276-59795246	Enoyl reductase	regulation of biological process
	RA364	Solyc07g005210.2.1	202269-203546	Outer membrane lipoprotein blc	response to stress
	RA365	Solyc07g005560.2.1	446870-450798	Eukaryotic translation initiation factor 5A	translation
*	RA369	Solyc01g098380.2.1	80665607-80671287	Dihydrodipicolinate reductase family protein	primary metabolic process
	RA373	Solyc02g077990.2.1	37374128-37377628	30S ribosomal protein S5	translation
	RA377	Solyc12g099100.1.1	64711180-64716181	Dihydrolipoyl dehydrogenase	primary metabolic process
	RA380	Solyc05g056540.2.1	64952197-64957183	Alcohol dehydrogenase-like protein	cellular metabolic process
	RA385	Solyc01g109300.2.1	88049244-88054027	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	secondary metabolic process
	RA386	Solyc08g078510.2.1	59454948-59456978	GRAM-containing/ABA-responsive protein	response to stress
	RA391	Solyc02g082250.2.1	40521256-40525895	Thioredoxin reductase	oxidation-reduction process
	RA394	Solyc10g077030.1.1	59270503-59275704	Proteasome subunit alpha type	macromolecule metabolic process
	RA395	Solyc10g085040.1.1	63684986-63685585	Soul heme-binding family protein	Unknown
	RA396	Solyc05g014280.2.1	8089580-8092146	Heat shock protein	response to stress
	RA397	Solyc01g096270.2.1	79110769-79111451	Unknown Protein	Unknown
	RA405	Solyc01g100370.2.1	82186051-82188666	Universal stress protein	response to stress

"#" Spot cumulating the expression of several loci

"*" Spot corresponding to mix of protein of diverse functions. The most abundant function is given here.

"§" Not variable spots

[a] Genotype effect (two way ANOVA)

[b] Stage effect (two way-Anova)

[c] Genotype and stage interaction effect of two way ANOVA

[d] Genotype effect at green stage (one way ANOVA)

[e] Genotype effect at Orange/red stage (one way ANOVA)

Unique spectra : number of spectra corresponding to peptides in the sequence

Specific spectra: in case of multiple loci, corresponds to the number of peptides that are specific for this gene product only

Supplemental table S2 List of spots resulting from multi-gene families

ID	Number of translated loci	Chromosome	Genome region in bp	Description	log.E.value.	Coverage	MW	Total number of spectra	Specific peptides for the locus	Number of different peptides	PAI
JX014	3	7	64080741-6408	Solyc07g064880.2.1_	-22,92	68	11,10	6	3	5	1,20
		12	619162-621481	Solyc12g006010.1.1_	-20,43	68	11,00	4	1	4	0,80
		9	52556197-5255	Solyc09g059970.2.1_	-5,21	16	11,80	2	1	2	0,40
JX030	2	3	44381441-4438	Solyc03g078400.2.1_	-223,96	70	41,60	71	4	40	4,93
		11	263971-265417	Solyc11g005330.1.1_	-223,96	70	41,70	71	4	40	4,93
JX036	2	7	57154220-5715	Solyc07g049530.2.1_	-185,85	65	35,70	54	15	31	3,80
		7	57170936-5717	Solyc07g049550.2.1_	-127,11	49	35,90	40	1	22	3,42
JX072	4	11	48856641-4885	Solyc11g066100.1.1_	-174,35	52	71,30	47	18	35	2,00
		4	3894918-38980	Solyc04g011440.2.1_	-149,46	43	71,20	34	6	27	1,59
		11	48824058-4882	Solyc11g066060.1.1_	-154,09	41	77,00	33	3	28	1,43
		6	43582389-4358	Solyc06g076020.2.1_	-127,43	32	70,90	29	2	24	1,32
JX074	6	8	62655311-6265	Solyc08g082820.2.1_	-179,28	53	73,10	47	10	33	2,04
		3	46341372-4634	Solyc03g082920.2.1_	-144,01	41	73,30	35	5	26	1,58
		6	32202991-3220	Solyc06g052050.2.1_	-100,34	27	67,40	25	1	16	1,08
		4	3894918-38980	Solyc04g011440.2.1_	-110,56	40	71,20	22	6	19	1,00
		11	48824058-4882	Solyc11g066060.1.1_	-70,75	25	77,00	16	2	14	0,70
		11	48856641-4885	Solyc11g066100.1.1_	-61,06	21	71,30	14	1	11	0,58
JX083	2	6	43949614-4395	Solyc06g076560.1.1_	-99,44	70	17,60	43	10	21	7,67
		6	43936690-4393	Solyc06g076520.1.1_	-88,06	65	17,60	40	7	19	7,17
JX108	3	5	64069121-6407	Solyc05g055160.2.1_	-76,26	43	46,60	16	4	16	1,07
		4	3081668-30862	Solyc04g009770.2.1_	-65,04	36	46,50	14	8	14	1,27
		11	1149792-11528	Solyc11g006460.1.1_	-52,88	26	46,50	11	1	11	0,73
JX127	4	10	63048965-6305	Solyc10g084050.1.1_	-198,96	42	89,60	38	11	34	1,18
		6	42868689-4287	Solyc06g074980.2.1_	-169,18	40	89,30	31	3	30	0,86
		11	51412028-5141	Solyc11g069720.1.1_	-149,20	37	89,40	29	2	28	0,81
		3	57029718-5704	Solyc03g112590.2.1_	-80,65	23	92,70	16	5	15	0,41
JX140	4	3	46341372-4634	Solyc03g082920.2.1_	-153,41	44	73,30	28	6	26	1,25
		8	62655311-6265	Solyc08g082820.2.1_	-131,94	39	73,10	25	1	23	1,04
		6	32202991-3220	Solyc06g052050.2.1_	-112,57	30	67,40	19	1	17	0,79
		11	48856641-4885	Solyc11g066100.1.1_	-18,61	9	71,30	5	3	5	0,21
JX141	3	3	46341372-4634	Solyc03g082920.2.1_	-171,27	45	73,30	38	9	30	1,79
		8	62655311-6265	Solyc08g082820.2.1_	-143,24	38	73,10	33	4	26	1,48
		11	48856641-4885	Solyc11g066100.1.1_	-57,54	24	71,30	15	11	14	0,71
JX149	3	8	62655311-6265	Solyc08g082820.2.1_	-223,71	58	73,10	58	13	37	2,44
		3	46341372-4634	Solyc03g082920.2.1_	-168,07	47	73,30	40	4	27	1,83
		6	32202991-3220	Solyc06g052050.2.1_	-140,03	33	67,40	33	1	19	1,46
JX170	3	4	248903-251735	Solyc04g005340.2.1_	-160,42	72	40,50	46	12	32	2,18
		5	5295089-52982	Solyc05g012070.2.1_	-92,14	48	41,10	30	1	21	1,65
		3	58802749-5880	Solyc03g114860.2.1_	-91,63	45	40,00	30	1	19	1,74
JX197	2	6	40601620-4060	Solyc06g071790.2.1_	-128,17	48	48,70	34	12	25	1,52
		3	56698575-5670	Solyc03g112150.1.1_	-132,84	49	51,70	30	8	25	1,30
JX252	2	6	40601620-4060	Solyc06g071790.2.1_	-104,90	51	48,70	29	15	22	1,26
		3	56698575-5670	Solyc03g112150.1.1_	-97,04	44	51,70	19	5	17	0,83
JX259	2	8	62655311-6265	Solyc08g082820.2.1_	-223,03	57	73,10	53	15	38	2,20
		3	46341372-4634	Solyc03g082920.2.1_	-174,23	40	73,30	40	2	29	1,75
JX277	5	10	61453818-6145	Solyc10g080940.1.1_	-142,00	62	50,80	33	6	24	1,65
		6	22016405-2201	Solyc06g035970.2.1_	-132,89	61	50,10	33	4	23	1,65
		4	63035466-6303	Solyc04g081490.2.1_	-127,62	63	50,50	31	7	23	1,55
		10	64808095-6480	Solyc10g086760.1.1_	-122,54	60	50,30	31	1	22	1,55
		10	63679058-6368	Solyc10g085020.1.1_	-122,77	59	51,00	31	1	22	1,55
JX280	2	1	62201815-6222	Solyc01g059980.2.1_	-77,69	39	39,60	26	0	15	2,08
		1	62224181-6222	Solyc01g060020.2.1_	-77,69	39	39,60	26	0	15	2,08
JX321	2	3	56698575-5670	Solyc03g112150.1.1_	-164,78	54	51,70	40	13	30	1,83
		6	40601620-4060	Solyc06g071790.2.1_	-114,25	40	48,70	29	2	22	1,30
JX325	3	1	6848050-68512	Solyc01g011000.2.1_	-56,98	69	17,30	25	7	13	3,71
		4	342519-345154	Solyc04g005510.2.1_	-56,98	69	17,30	25	7	13	3,71

		5	1496349-14968	Solyc05g006890.1.1_	-55,10	58	17,40	19	1	12	2,86
JX326	2	6	182627-185280	Solyc06g005160.2.1_	-91,95	60	27,30	21	9	16	2,27
		6	170218-173338	Solyc06g005150.2.1_	-90,78	64	27,20	20	8	17	2,18
JX332	4	6	43582389-4358	Solyc06g076020.2.1_	-160,99	48	70,90	37	16	30	1,86
		11	48856641-4885	Solyc11g066100.1.1_	-139,21	47	71,30	27	7	24	1,17
		11	48824058-4882	Solyc11g066060.1.1_	-128,20	37	77,00	27	4	22	1,26
		4	3894918-38980	Solyc04g011440.2.1_	-110,68	35	71,20	25	5	20	1,18
JX337	3	5	5295089-52982	Solyc05g012070.2.1_	-123,12	67	41,10	37	8	28	1,85
		4	248903-251735	Solyc04g005340.2.1_	-81,42	43	40,50	25	1	18	1,14
		2	45629291-4563	Solyc02g089170.2.1_	-68,94	39	40,20	21	1	16	1,05
JX364	3	5	2805741-28097	Solyc05g008460.2.1_	-184,46	57	59,50	47	15	29	1,60
		2	47122405-4712	Solyc02g091130.2.1_	-128,47	37	59,30	33	1	21	1,18
		1	1849252-18517	Solyc01g007320.2.1_	-33,00	10	65,40	7	3	7	0,25
JX375	2	10	2177130-21828	Solyc10g008010.2.1_	-190,89	80	25,60	58	23	32	4,29
		7	60553200-6055	Solyc07g055080.2.1_	-154,04	80	25,60	48	13	27	3,71
JX376	2	7	62472541-6247	Solyc07g062570.2.1_	-73,24	65	17,10	23	7	16	3,43
		10	1673007-16782	Solyc10g007260.2.1_	-71,06	65	17,10	19	3	14	2,86
JX381	2	1	81668247-8167	Solyc01g099760.2.1_	-103,65	58	47,40	25	14	18	1,19
		8	10475967-1047	Solyc08g021990.1.1_	-32,74	37	21,50	10	1	6	0,91
JX387	2	2	34815271-3482	Solyc02g070510.2.1_	-148,82	73	26,00	40	8	26	3,00
		8	7616130-76222	Solyc08g016510.2.1_	-137,10	73	26,00	36	4	24	2,71
JX395	2	2	43937531-4393	Solyc02g086730.1.1_	-107,79	40	19,80	20	12	13	2,22
		2	43941640-4394	Solyc02g086740.1.1_	-65,19	36	19,40	13	5	9	1,86
JX408	2	6	44672378-4467	Solyc06g082630.2.1_	-140,97	62	46,50	32	0	25	1,80
		6	44701223-4470	Solyc06g082660.2.1_	-140,97	62	46,50	32	0	25	1,80
JX413	2	1	2018307-20224	Solyc01g007860.2.1_	-66,87	77	16,50	26	3	12	2,89
		10	62310434-6231	Solyc10g083120.1.1_	-66,37	77	16,50	26	3	13	2,89
JX421	2	2	28750335-2875	Solyc02g062500.2.1_	-49,01	36	36,40	12	5	11	1,00
		2	28653823-2865	Solyc02g062460.2.1_	-34,73	26	36,40	9	2	7	0,59
RA130	2	7	64973786-6497	Solyc07g066270.2.1_	-46,74	27	28,00	7	2	6	0,70
		3	46048220-4605	Solyc03g082580.2.1_	-91,96	36	34,80	15	10	15	1,25
RA223	2	2	41646719-4165	Solyc02g083810.2.1_	-77,63	30	41,50	13	9	11	0,81
		2	28283170-2828	Solyc02g062130.2.1_	-33,63	16	40,50	6	2	4	0,40
RA316	2	7	62037698-6203	Solyc07g061940.2.1_	-41,30	23	56,30	9	2	9	0,45
		3	12823169-1282	Solyc03g044330.1.1_	-35,34	14	71,80	8	1	8	0,32

CHAPTER VI: GENETIC DIVERSITY AND INHERITANCE OF PROTEOMIC, ENZYMATIC AND METABOLOMIC PROFILES IN TOMATO FRUIT PERICARP

This chapter, in the form of a manuscript to be submitted to Plant Physiology, presents the use of System biology approaches for the dissection of the genetic variation of phenotypic traits. We conducted parallel characterisation of the proteome, the metabolome, enzymatic profiles and phenotypes of eight contrasted accessions and their four corresponding F1s, at two fruit developmental stages. Genetic variability was analysed for all the traits and inheritance modes were assessed for all variable traits in each cross. Correlations were then studied within and between levels of expression.

Genetic diversity and inheritance of proteomic, enzymatic and metabolomic profiles in tomato fruit pericarp

Jiaxin Xu (1,2), Laura Pascual (1), Benoit Biais (3), M. Maucourt (4), P. Ballias (3), C. Deborde (3), Aurore Desgroux (1), Mireille Faurobert (1), Jean-Paul Bouchet (1), Yves Gibon (3), Annick Moing (3), Mathilde Causse (1)

(1) INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, F-84143 Avignon, France

(2) Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, P.R. China

(3) INRA-UMR 619 Biologie du Fruit, Centre de Bordeaux, F–33140 Villenave d'Ornon, France

(4) Université de Bordeaux, UMR1332 Biologie du Fruit et Pathologie, BP 81, F-33140 Villenave d'Ornon, France

Corresponding author

Mathilde Causse

Phone: 33 04 32 72 27 01

Fax: 33 04 32 72 27 02

Email: Mathilde.Causse@avignon.inra.fr

Abstract:

Systems biology proposes new approaches of the genetic variation of phenotypic traits. In an effort to characterize the genomic and genetic variation and the physiological bases of quantitative traits in tomato fruit, we conducted parallel characterisation of the proteome, the metabolome, enzymatic profiles and phenotypes of eight contrasted accessions and their four corresponding F1s, at two fruit developmental stages. Genetic variability was analysed for all the traits and inheritance modes were assessed for all variable traits in each cross. Correlations were then studied within and between levels of expression. A large range of variability was observed for every trait. The number of variable traits varied among crosses in relation to the genetic distance between the parental lines. In average, more than 60% of the traits showed an additive mode of inheritance, but differences were shown among crosses. A significant number of traits exhibited an over-dominant or over-recessive mode of inheritance, without any specific trend towards an excess of dominance or recessivity. Pair-wise Pearson correlations were calculated within each group of traits at each stage. Then networks were constructed based on sparse partial least squares regressions between each pair of groups of traits. Within each group, the number of significant correlations was always much higher than expected by chance and a large tendency to an excess of positive correlations is shown, notably for metabolites and enzyme activities. Correlations between metabolite contents and enzyme activities were higher and more frequent at orange-red stage than at cell expansion stage with almost no negative correlations. A large number of both positive and negative correlations were observed between metabolite contents and protein amounts, with the same ratios in each class. On the contrary, between enzyme activities and protein amounts, there were not much significant correlations compared to random at cell expansion stage, on the contrary to the orange-red stage, where the number of correlations was higher, with a clear tendency to an excess of positive correlations with the proteins corresponding to primary metabolism and stress response. The central role of a few proteins was identified. This systems biology approach provides better understanding of networks of elements (proteins, enzymes, metabolites and phenotypic traits) in tomato fruits.

Introduction

Genetic and molecular dissection of quantitative trait variation is a major objective in biology. Attempts to identify genetic variants underlying quantitative traits have been achieved by traditional linkage mapping and genome wide association studies using molecular markers. However, QTLs only contribute to a little proportion of the variability of a quantitative trait. Much of the variation and heritability remains unexplained for most quantitative traits due to their polygenetic nature (King et al. 2012). Furthermore, DNA sequence variation (SNP or Indel) may not affect the traits directly. There are several intermediate levels between the genotypes and the phenotypes. Intermediate molecular phenotypes such as protein abundance and metabolite concentration also genetically vary in populations and are themselves quantitatively inherited (Rockman and Kruglyak 2006). Therefore, integration of the genome expression products at different levels is needed in order to understand the genetic variation of a given quantitative trait. Rapid technological advances in high-throughput experiment such as next-generation sequencing (NGS), RNA expression analysis through microarray or RNAseq, mass spectrometry (MS) coupled to gas chromatography (GCMS) or to liquid chromatography (LCMS) and nuclear magnetic resource (NMR) enable scientists to obtain large exhaustive datasets and analyze biological systems as a whole. Systems biology proposes novel approaches studying the behaviour of all the elements in a biological system (Gutierez et al. 2008; Saito and Matsuda 2010). These approaches integrate 'omic' resources (genomic, transcriptomic, proteomic, and metabolomic) and large physiological datasets, together with statistical network analysis. They allow the identification of candidate genes underlying phenotypes and of complex networks of regulation (Kliebenstein 2010). Systems biology was first applied to yeast by combining DNA microarrays and quantitative proteomics to describe the galactose utilization pathway (Ideker et al. 2001). It was then applied to gene expression analysis in *E.coli* (Rosenfeld et al. 2002). Hirai et al. (2005) elucidated gene to gene and metabolite to gene networks in Arabidopsis by integrating metabolomic and transcriptomic data. Recently, system biology approaches have been used to study the natural genetic variation at different levels, such as metabolomics (Keurentjes 2009; Kliebenstein 2009a), proteomics (Stylianou et al. 2008) and transcriptomics (Keurentjes et al. 2008a; Kliebenstein 2009b) in Arabidopsis.

Tomato (*Solanum lycopersicum*) is the model species for genetic and genomic studies of Solanaceous plants (Matsukura et al. 2008). It is a self-pollinating species and derived from its closest wild ancestor *S. pimpinellifolium* (Nesbitt and Tanksley 2002). Cherry tomato accessions (*S. lycopersicum* var *cerasiforme*) have an intermediate position between these two species, as their genome is a mosaic of those from *S. lycopersicum* and *S. pimpinellifolium* (Ranc et al. 2008). During tomato domestication, fruit size has considerably increased. The diversification of fruit aspect (shape, color) as well as the adaptation to a wide range of environmental conditions was simultaneous to a strong reduction of molecular diversity as revealed by different molecular markers (Miller and Tanksley 1990; Williams et al. 1993; Saliba-Colombani et al. 2000). Nevertheless many QTL studies were performed in tomato (Alpert et al. 1995; Grandillo and Tanksley 1996; Bernacchi et al. 1998; Saliba-Colombani et al. 2001; Causse et al. 2001; Causse et al. 2002; Causse et al. 2004; Lecomte et al. 2004). Lack of genetic variation in cultivated species leaded to study trait variation mostly in crosses involving one wild species and thus limited the exploitation of intra-specific genetic variation. Today the availability of the tomato genome sequence (Sato et al. 2012) and of a large number of SNP (Sim et al. 2012) allow a re-examination of the variation and inheritance of agronomical and fruit traits at the intra-specific level. Furthermore, systems biology provides a new context to relate the genetic variation analysed at different levels of expression from phenotype to metabolome and proteome levels.

Systems biology approaches have been applied to tomato at several levels. Carrari et al. (2006) and Mounet et al. (2009) analysed transcriptome and metabolome variation along fruit development. Garcia et al. (2009) combined phenotype, metabolome, transcriptome and proteome profiles to study genes related to ascorbic acid pathway in three transgenic lines. Wang et al. (2009) compared transcriptome and metabolome to uncover the molecular events underlying fruit set. Enfissi et al. (2010) characterized metabolites and transcripts modifications due to the modulation of one gene (DE-ETIOLATED1), while Osorio et al. (2011) compared enzyme activity, metabolite and transcript profiles to analyse the connectivity between these groups of traits in fruit ripening mutants. These studies were focused on the study of a few mutants or on the effect of introgression in *S. lycopersicum* of wild species alleles. Little is known on the variation in metabolic, enzymatic, proteomic profiles contributing to phenotypic trait variation in cherry type (*Solanum lycopersicum* var.

cerasiforme) accessions. Faurobert et al. (Faurobert et al. 2007) provided the first characterization of the cherry tomato proteome along fruit development by combining 2D-gel electrophoresis and mass spectrometry. This study was carried out on one genotype and identified a relatively small number of proteins, failing to provide an insight of the extent of natural genetic variation. Recently we extended this analysis to 12 tomato lines and identified 506 protein spots (Xu et al., submitted). On another hand, there has been an increasing interest in the mode of inheritance of omic traits. Metabolite variability and inheritance patterns have been investigated in corn (Lisec et al. 2011). The mode of inheritance of metabolite QTL and enzyme activity QTL has been studied in tomato (Schauer et al. 2008; Steinhauser et al. 2011). However, no study of inheritance at the proteome level has yet been carried out in tomato.

In the present study, we aimed at deciphering the complex relationships between multiple levels of omics data to characterize the genetic variation and physiological bases of quantitative traits in tomato fruit. For this purpose, we first compared the genetic variation of 12 genotypes, including four *S. lycopersicum* accessions and four *S. l.* var *cerasiforme* accessions representing a large range of phenotypic and genotypic diversity and four of their corresponding hybrids at phenotype, metabolome, enzyme activity and proteome levels. Physiological modifications during fruit development have a major impact on the overall quality of fruit. Therefore we analyzed genetic variability at two stages of fruit development. We then assessed the inheritance patterns of traits which were significantly different among genotypes. Correlations were then studied within and between levels of expression.

Materials and methods

Plant materials

Eight tomato lines and four of their corresponding F1 hybrids were used in this study. They included four *Solanum lycopersicum* accessions (Levovil, Stupicke Polni Rane, LA0147, and Ferum) and four *S. l. cerasiforme* accessions (Cervil, Criollo, Plovdiv24A, and LA1420). Cervil produces very small fruits (less than 10 g). Levovil, Ferum and LA0147 accessions produce large fruits. The others have intermediate fruit size. Fruit size diversity of the eight parental lines and the position of F1 is illustrated in **Supplemental figure S1**.

Growth condition and sampling

Ten plants of each accession and F1s were grown in 2010 under greenhouse conditions (16/20°C) in Avignon (south of France). For plant trait measurement, height of the plant was recorded twice a week in order to follow the dynamic model of the plant growth. Stem diameter was measured with a caliper square every week, at a level corresponding to a one-week growth (indicated each week on the string). For fruit phenotypic trait measurements, five fruits were harvested from ten plants of each genotype at the following six stages: cell expansion stage (25, 20 and 14 days after anthesis for large, medium and small fruited accessions, respectively), +7d, +14d, +21d, orange-red, then until red ripen stage. Fruits were evaluated for fresh weight, fruit diameter (measured using a caliper) and dry matter content. Dry matter content (expressed in g / 100 g FW) was assessed after 5 d in a ventilated oven at 80 ˚C. For proteome, metabolome and enzymatic measurement, fruits were collected at two stages of development, cell expansion stage (25, 20 and 14 days after anthesis for large, intermediate, small fruited accession) and orange-red stage, according to the fruit color. Three biological pools of 7 to 20 fruits per accession were obtained at each stage. Pericarps were collected from each pool, immediately frozen, ground in liquid nitrogen and stored at -80 °C before analysis.

Proteome analysis

Methods for protein extraction, two-dimensional gel electrophoresis (2-DE), protein identification and classification were as detailed in Xu et al. (2012, submitted). Briefly, proteins were extracted using the phenol extraction method developed by Faurobert et al. (2007). Later, proteins were separated by 2-DE. After coomassie colloidal staining, image analysis was performed with Samespot software and the normalized spot volumes were obtained. Protein identification of the variable spots was performed at the proteome platform of Le Moulon (Gif-sur-Yvette) using nano-LC-MS/MS method following the procedure that described in Xu et al. (2012, submitted). The database search was run against the International Tomato Annotation Group (ITAG) Release 2.3 of predicted proteins (SL2.40) database (http://solgenomics.net/) with X!Tandem software. Fasta sequence of the identified proteins was employed to re-annotate the proteins using the Blast2GO package (Conesa et al. 2005). Sequences were compared against the NCBI-NR database of non-redundant protein sequence using BLASTX with the default setting.

Metabolome analysis

Metabolome analyses were performed at the metabolome platform of Bordeaux. Primary metabolites were quantified using quantitative 1H-NMR profiling of polar extracts, as described in Deborde et al. (2009) with minor modifications. Briefly, polar metabolites were extracted on lyophilized powder (50 mg DW per biological replicate) with an ethanol–water series at 80°C. The lyophilized extracts were titrated to pH 6 and lyophilized again. Each dried titrated extract was solubilized in 0.5 mL D2O with (trimethylsilyl)propionic-2,2,3,3-d4 acid (TSP) sodium salt (0.01% final concentration) for chemical shift calibration and ethylene diamine tetraacetic acid (EDTA) disodium salt (5 mM final concentration for cell expansion stage and 2 mM for orange-red stage). 1H-NMR spectra were recorded at 500.162 MHz on a Bruker Avance III spectrometer (Bruker, Karlsruhe, Germany) using a ATMA inverse 5 mm probe and an electronic reference for quantification. Sixty-four scans of 32 K data points each were acquired with a 90° pulse angle, a 6000 Hz spectral width, a 2.73 s acquisition time and a 25 s recycle delay. Two technological replicates were used per biological replicate. Preliminary data processing was conducted with TOPSPIN 3.0 software (Bruker Biospin, Wissembourg, France). The assignments of metabolites in the NMR spectra were made by

comparing the proton chemical shifts with values of the MeRy-B metabolomic database (Ferry-Dumazet et al. 2011), by comparison with spectra of authentic compounds recorded under the same solvent conditions and/or by spiking the samples. The metabolite concentrations were calculated using AMIX (version 3.9.7, Bruker) software. Ascorbic acid was measured using a spectrofluorometric method and values expressed as total ascorbate (ascorbic acid + dehydroascorbate) in mg/100 g fresh weight as previously described by Stevens et al. (2007). Starch was extracted using a methanol-choroform mix (Gomez et al. 2002) and quantified by enzymatic assay 96-well microplates (Gomez et al. 2007).

Secondary metabolites were quantified by LC-QTOF-MS profiling of semi polar extracts. LC-QTOF-MS profiling of aqueous-methanol-0.1% formic acid extracts was performed from lyophilized powder (20 mg in 1 ml). For each biological replicate, two extractions were performed and two injections per extract were used. An Ultimate 3000 HPLC (Dionex, Sunnyvale, CA, USA) was used to separate metabolites on a reversed phase C18 column (150 x 2.0 mm, 3 µm; Phenomenex, Torrance, CA, USA) using a 30 min linear gradient from 3 to 95% acetonitrile in water acidified with 0.1% formic acid. Metabolites were detected using a quadrupole time-of-flight (QTOF) mass spectrometer (Bruker, Bremen, Germany). Electrospray ionization in positive mode was used to ionize the compounds. Scan rate for ions at m/z range 100-1500 was fixed at 2 spectra per second. Methyl vanillate was spiked in the extraction solvent and used as an internal standard. One sample was used as a QC sample and injected each ten injections. Raw data were processed in a targeted manner using QuantAnalysis 2.0 software (Bruker, Bremen, Germany). This resulted in a total of 11 compounds putatively identified based on accurate mass measurement and comparison with data from Gomez-Romero et al. (2010), and two unknown compounds. The maximum activity (Vmax) of 26 enzymes of the primary metabolism was assessed as described in Steinhauser et al. (2010). **Supplemental Table 1 and Table 2** present the list of primary, secondary metabolites and enzyme activities analyzed.

Statistical analysis and inheritance analysis

All the metabolites and enzyme activities were expressed on a fresh weight basis to be comparable. Means and coefficients of variation (CV= standard deviation/mean) were then calculated for each trait (phenotype, metabolite amount, enzyme activity, protein spot

amount) in each genotype and stage. Metabolite concentrations, enzymatic activities and protein volumes were submitted to a two way ANOVA with genotype, stage and interaction effect and one way ANOVA with genotype effect for each cross and stage. Significantly different ($P< 0.05$) traits in at least one cross were selected at each stage to estimate additive (A) and dominance (D) components of genetic variation. A is equivalent to half of the parental line difference (the line with large fruits was systematically the first parent in a cross) and D is equivalent to the difference between hybrid value and parental mean. Inheritance pattern of all the traits was then assessed by dominance/additivity (D/A) ratio and classified as over-recessive (OR; $D/A <-1.2$), recessive (R; $-1.2 <D/A < -0.8$), additive (A; $-0.8 < D/A < 0.8$), dominant (D; $0.8 \leq D/A \leq 1.2$), over-dominant (OD; $1.2 <D/A$).

Mean of metabolite concentrations, enzyme activities and protein spot volumes were centered and scaled to variance unit and used for subsequent principal component analysis (PCA), correlation and network analysis. Pairwise metabolites and enzyme activities correlations and p-values were calculated by Pearson's correlation coefficient and plotted using "corplot" package. PCA were performed for protein spot volumes with the "pcaMethods" package. Correlation networks were reconstructed using sparse partial least squares regression (sPLS) analysis with the "MixOmics" package. All the analyses were performed using R program (R Development Core Team 2005).

Results

Molecular and phenotypic diversity of the eight tomato accessions

The eight tomato accessions were selected to maximise the genetic diversity among 360 accessions including large fruited accessions and *S. l. cerasiforme* accessions, characterised with 20 SSR data (Ranc et al. 2008). These lines and Heinz1706, which has been sequenced by the international tomato genome sequencing consortium (Sato et al. 2012), were further genotyped using a set of 139 single nucleotide polymorphism (SNP) markers (Xu et al. 2012). Polymorphism analysis indicated that the 8 parental lines captured 96% of the molecular diversity (133 polymorphic SNPs/139 SNPs). Polymorphism rate was high between parental lines and Heinz1706, ranging from 27% to 82% (**Supplemental table S3**). An unrooted neighbour joining tree was developed using MEGA5 program based on 139 SNP markers to assess the genetic relatedness between the nine accessions. The nine accessions were grouped into two clusters, one large cluster with eight accessions and one small cluster with only Cervil accession. The four crosses studied result from the cross between one cherry and one large fruited accession (Cervil × Levovil), two medium fruited accessions (Stupicke Polni Rane × Criollo), one large fruited and one medium fruited accession (LA0147 × Plovdiv24A and Ferum × LA1420). The four F1 hybrids thus corresponded to different distances among parents (**Supplemental figure S2**), with Cervil and Levovil as the two most distant accessions (82% SNP polymorphic), followed by the cross LA0147 x Plovdiv (40%), Stupicke Polni Rane x Criollo (34%) and Ferum x LA1420 (27%).

The final fruit weight of the eight parental lines and the four hybrids was quite diverse. Fruit weights from the four hybrids were intermediate between the values of their parental lines along fruit development (**Supplemental figure figure S3**). Based on these curves, we harvested fruits at two developmental stages for proteomic, enzymatic and metabolomic analysis: cell expansion stage (25, 20 and 14 days after anthesis for large, intermediate and small fruited accessions, respectively) and orange-red stage, according to the fruit color. Fruit diameter were very close to fruit weight as most of the accessions were round. Plant height at the sixth truss varied from 98 cm to 171 cm. Plant stem diameter was followed as a marker of plant vigor. It was quite stable along plant growth.

Table 1 Analysis of variation for 34 metabolite content and 5 phenotypic traits

	CE				OR				Global analysis			CE/OR	
Trait	F^G	min	max	max/min	F^G	min	max	max/min	F^G	F^S	F^{GxS}	min	max
Gluc	***	6968.6	13168.2	1.9	***	9164.9	15772.1	1.7	***	***	***	0.45	1.27
Suc	***	469.9	961.6	2.0	***	371.5	2636.7	7.1	***	***	***	0.33	1.53
Fru	***	6651.2	13041.8	2.0	***	9755.4	16420.9	1.7	***	***	***	0.41	1.21
Inos	***	233.3	516.4	2.2	***	110.6	308.1	2.8	***	***	***	1.61	2.98
Ala	***	21.1	137.0	6.5	***	10.5	24.1	2.3	***	***	***	1.12	5.92
Asn	***	55.2	248.7	4.5	***	73.5	288.0	3.9	***	***	***	0.40	1.11
Asp	***	38.2	129.1	3.4	***	81.0	316.9	3.9	***	***	***	0.24	0.54
Gaba	***	317.6	714.5	2.2	***	116.8	374.0	3.2	***	***	***	1.05	2.92
Gln	***	438.1	2652.7	6.1	***	404.3	1774.5	4.4	***	***	***	0.56	1.69
Ileu	***	12.7	87.4	6.9	***	13.8	46.7	3.4	***	***	***	0.67	3.54
Leu	***	23.0	93.7	4.1	***	24.6	69.2	2.8	***	***	***	0.47	1.35
Phe	***	90.7	438.8	4.8	***	105.4	504.5	4.8	***	***	***	0.41	1.13
Tyr	***	11.5	76.8	6.7	***	9.6	44.6	4.6	***	***	***	0.59	2.36
Val	***	9.2	67.5	7.4	***	11.8	46.7	3.9	***	*	***	0.44	2.28
Trigo	***	18.2	149.9	8.2	***	22.1	89.2	4.0	***	***	***	0.77	1.98
Cit	***	8.4	22.6	2.7	***	8.2	24.8	3.0	***	***	***	0.68	1.46
Mal	***	1622.4	4840.3	3.0	***	2421.8	9101.5	3.8	***	***	***	0.43	0.75
Fum	***	965.7	1988.5	2.1	***	201.7	1932.9	9.6	***	***	***	0.78	5.35
Stch	***	0.4	1.4	3.5	***	0.0	0.8	0	***	***	***	1.14	NA
Trigo	***	1.2	7.6	6.2	***	0.0	1.1	362.9	***	***	***	6.17	397.56
Ade	***	16.5	67.6	4.1	***	13.0	61.6	4.7	***	***	***	0.98	1.75
Chol	***	6.2	10.3	1.7	***	3.9	8.2	2.1	***	***	***	0.82	2.21
sChlorAc	***	41.1	85.8	2.1	***	34.9	76.0	2.2	***	***	***	0.94	1.44
sTCQ	***	4036.5	11257.3	2.8	***	2721.1	6449.1	2.4	***	***	***	1.17	2.39
sAto	ns	50.0	127.7	2.6	***	708.2	5144.8	7.3	***	***	***	0.01	0.11
sCry	***	60664.9	460156.9	7.6	***	880.2	5885.2	6.7	***	***	***	27.96	218.19
sDHT	*	484.2	1569.5	3.2	ns	316.8	900.0	2.8	ns	***	*	0.76	4.95
sRutp	***	12129.5	267130.1	22.0	***	417.7	7599.0	18.2	***	***	***	11.55	148.02
sPpa	***	1459.0	5146.7	3.5	***	1234.7	7445.1	6.0	***	ns	***	0.52	1.41
sCou	***	434.9	803.9	1.8	***	73.6	607.5	8.3	***	***	***	1.02	9.15
sNna	***	746.9	3088.2	4.1	***	683.2	3223.3	4.7	***	ns	***	0.42	2.80
sOHl	*	35.5	189.0	5.3	***	144.4	64355.1	445.6	***	***	***	0.00	0.43
sRut	***	434.8	708.7	1.6	***	298.7	2794.9	9.4	***	***	***	0.25	1.46

Table 1 Continued

	CE				OR				Global analysis			CE/OR	
Trait	F^G	min	max	max/min	F^G	min	max	max/min	F^G	F^S	F^{GxS}	min	max
xfw	***	1.2	29.5	29.93	***	5.4	124.9	23.0	***	***	***	0.10	0.37
xfd	***	13.8	41.8	3.02	***	23.0	71.1	3.1	***	***	***	0.46	0.70
xdmc	***	6.2	10.5	1.7	***	4.9	8.6	1.8	***	***	***	1.15	1.42
pheight	***	63.0	112.0	1.8	***	98.5	173.5	1.8	***	*	***	0.60	0.70
pstem	**	11.8	16.8	1.4	***	8.6	15.6	1.8	***	***	***	0.96	1.48

F^G: significance level of the ANOVA with genotype factor
F^S: significance level of the ANOVA with the stage factor
F^{GxS}: significance level of the ANOVA with the interaction between genotype and stage
Min: minimum values for each variable among the 12 genotypes and three repetitions
Max: maximum values for each variable among the 12 genotypes and three repetitions
CE/ OR ratio value to cell expansion stages and orange-red (min and max) for each genotype
*: 0.01 <P <0.05. **: 0.001 <P <0.01. ***: P<0.001. ns: P> 0.0

Table 2 Analysis of variation for activity of 26 enzymes

	CE				OR				Global analysis			CE/OR	
Trait	F^G	min	max	max/min	F^G	min	max	max/min	F^G	F^S	F^{GxS}	min	max
eAlAT	ns	843.9	2040.8	2.4	ns	239.2	3218.7	13.5	ns	ns	ns	0.43	3.66
eAsAT	ns	2317.7	6548.2	2.8	ns	1490.2	3970.1	2.7	ns	***	ns	0.90	3.56
eSKDH	**	121.6	262.9	2.2	***	37.6	208.0	5.5	***	***	ns	1.26	3.63
ePGK	ns	3737.8	7930.8	2.1	ns	679.9	4436.0	6.5	ns	***	ns	0.84	9.56
eTPI	ns	41801.3	96038.3	2.3	ns	10801.2	42117.5	3.9	ns	***	ns	2.14	8.07
eEno	**	254.8	494.3	1.9	***	47.3	202.7	4.3	***	***	**	2.33	9.13
eG6PDH	ns	142.8	222.6	1.6	***	49.9	140.5	2.8	ns	***	ns	1.52	4.24
eGAPDHd	ns	1617.4	3100.6	1.9	ns	406.6	1799.9	4.4	ns	***	ns	1.55	6.45
eGAPDHdp	ns	328.0	1126.6	3.4	ns	65.6	1293.8	19.7	ns	***	ns	0.52	9.48
eIDH	ns	97.5	232.0	2.4	***	92.3	378.8	4.1	***	ns	***	0.42	1.98
ePepC	ns	387.0	789.4	2.0	*	94.6	274.2	2.9	ns	***	*	1.75	6.60
ePFKa	ns	84.0	123.0	1.5	***	42.3	121.8	2.9	ns	***	**	0.76	2.91
ePFKp	ns	455.4	985.0	2.2	***	113.1	599.0	5.3	***	***	ns	1.53	7.86
ePyrK	ns	314.6	548.0	1.7	***	109.2	454.8	4.2	**	***	ns	1.05	4.01
eFRK	**	73.1	373.5	5.1	ns	16.4	89.8	5.5	**	***	***	1.52	12.73
eGlK	*	63.5	144.2	2.3	**	9.2	38.6	4.2	***	***	ns	3.00	9.99
eInvA	ns	143.5	519.3	3.6	***	183.5	1351.7	7.4	**	***	ns	0.15	1.46
eInvN	ns	71.5	380.3	5.3	***	93.2	783.2	8.4	***	***	**	0.15	2.26
eSus	ns	187.6	951.0	5.1	ns	87.7	431.4	4.9	ns	***	ns	0.78	6.17
eFbpA	*	1440.3	3596.9	2.5	***	286.0	2257.3	7.9	***	***	ns	1.40	8.97
ePgm	ns	1283.0	2335.3	1.8	***	415.5	1322.6	3.2	***	***	ns	1.68	4.71
eAco	ns	68.8	285.6	4.1	***	7.9	122.0	15.4	*	***	ns	1.00	15.45
eFumase	ns	180.6	522.6	2.9	ns	18.1	398.5	22.0	ns	**	ns	0.77	13.91
eMDH	ns	15130.0	29662.6	2.0	***	4436.9	17568.2	4.0	***	***	ns	1.69	6.07
eMed	ns	349.6	853.2	2.4	ns	197.2	462.8	2.3	ns	***	ns	1.15	3.59
eMedp	***	116.3	476.0	4.1	***	47.0	272.0	5.8	***	***	*	0.83	5.04

F^G: significance level of the ANOVA with genotype factor
F^S: significance level of the ANOVA with the stage factor
F^{GxS}: significance level of the ANOVA with the interaction between genotype and stage
Min: minimum values for each variable among the 12 genotypes and three repetitions
Max: maximum values for each variable among the 12 genotypes and three repetitions
CE/ OR ratio value to cell expansion stages and orange-red (min and max) for each genotype
*: 0.01 <P <0.05. **: 0.001 <P <0.01. ***: P<0.001. ns: P> 0.05

Genetic variability and evolution during fruit development of metabolic and protein traits

^{1}H-NMR profiles allowed the quantification of 24 major metabolites from the central carbon metabolism (**Figure 1**), LC-QTOF-MS profiles identified 11 secondary metabolites (phenolic compounds, glycoalkaloids and penthotenic acid) as listed in **Supplemental Table 1**. The maximum activities of 26 enzymes were obtained by enzyme assays (**Supplemental Table 2**). 2-DE revealed 424 variable protein spots which were identified by nano LC-MS/MS (Xu et al. submitted). Mean values of each trait from each group are presented in **Supplemental table S4**. A large range of variability was observed among the 12 genotypes at each stage and between the two stages for the 34 primary and secondary metabolites and 5 phenotypic traits (**Table 1**). The content of sugars (glucose and fructose), citrate, malate, aspartate and phenylalanine increased from cell expansion to orange-red stage while the starch content decreased. The ratios of the means at two stages reflect that average genetic variability tended to decrease for most of the compounds, except for acids (citrate and malate) (**Table 1**). Almost all metabolites and phenotypic traits were significantly different between genotypes at the two stages except one secondary metabolite, dehydrotomatine (sDHT), which was non-significant. The interactions between stage and genotype were generally significant for the metabolites. Therefore, they were subsequently analysed stage by stage. At each stage, all the metabolites were significantly different among genotypes except one secondary metabolite (a-tomatine, sAto), which was not significant at cell expansion stage and another one (sDHT), which was not significant at orange-red stage (**Table 1**).

The activity of 26 enzymes from the central carbon metabolism also exhibited a large range of variation (**Table 2**). The activity of sucrose breakdown related enzymes, phosphofructokinase and sucrose synthase, decreased during the transition from cell expansion stage to orange-red stage while the activity of invertases (both acid and neutral) increased. The activity of aconitase decreased from cell expansion to orange-red stage, whereas citrate synthase activity significantly decreased. The activity of NAD-dependent malic enzyme (eMed) and NADP-dependent malic enzyme (eMedp), which are related to the metabolism of malate in the Krebs cycle significantly decreased along the two stages. The aspartate aminotransferase (eAsAT) is involved in the metabolism of the amino acid aspartate. Its activity decreased at orange-red stage (**Table 2**). Fourteen enzyme activities were significantly different among the 12

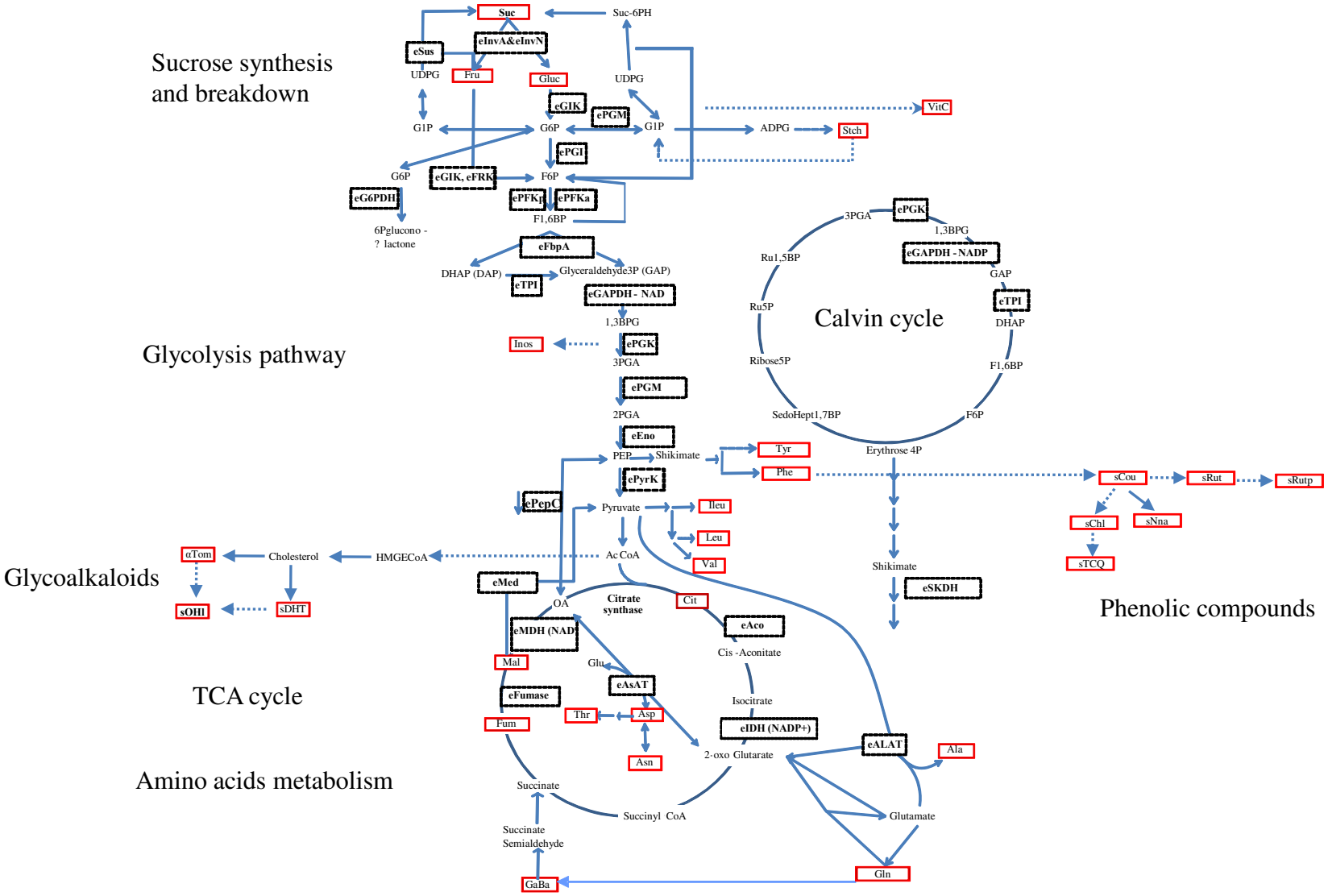


Figure 1. Assignment of metabolites and enzymes to Pathways. A total of 28 metabolites are indicated in red squares, 26 enzymes are highlighted in dot squares. Metabolites Trigo, Ade, Chol, sCry and sPpa are not indicated on the pathway.

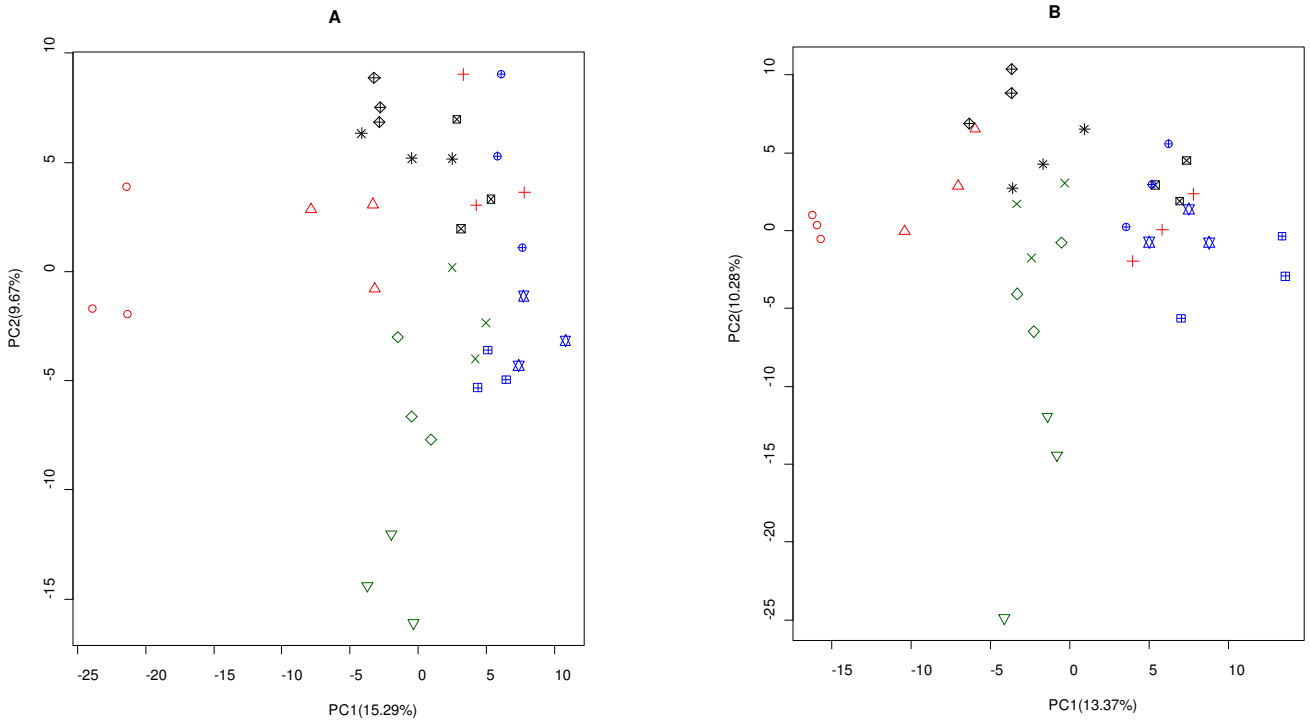


Figure 2. Principal Component analysis based on 424 spots at green (A) and Orange/red (B) fruit stage.
Values along axes indicate the percentage of total variation accounted for each component. Genotypes were indicated as different symbols : Cervil, red circle; CervilxLevovil, red triangle point up; Levovil, red plus; Stupicke Polni Rane, dark green cross; Stupicke Polni Rane x Criollo, dark green diamond; Criollo, dark green triangle point down; LAO147, black square cross; LAO147x Plovdiv 24A, black star; Plovdiv 24A, black diamond plus; Ferum, blue circle plus ; Ferum x LA1420, blue triangles up and down ; LA1420, blue square plus.

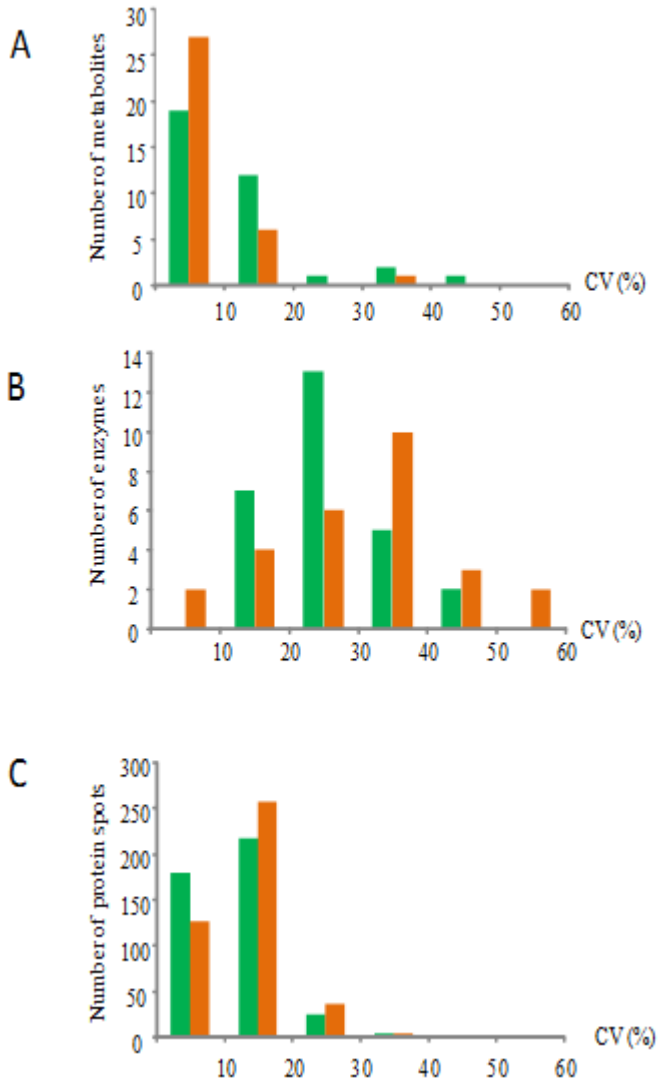


Figure 3. Distribution of the average coefficient of variation per genotype of 34 metabolites (A), 26 enzyme activities (B) and 424 protein spot amounts at cell expansion stage (green) and Orange red stage (orange).

genotypes at the two stages. Six enzyme activities showed significant difference among genotypes at cell expansion stage and 16 at orange-red stage.

A total of 1230 protein spots were detected on the 2DE-gels using Samespots software, among which the abundance of 566 spots was significantly different between genotypes or stages. 424 spots were sequenced by LC-MS/MS and 422 proteins were identified. The amount of the 424 protein spots was compared in the 12 genotypes. A large range of variability was observed among genotypes and between stage for all the protein spot amounts (**Supplemental table S4**). The main differences were observed between stages but the genotype also revealed a large range of variation as illustrated in **Figure 2**. The first plan of the PCA accounted for 25% of the variation at cell expansion stage and 24% at orange-red stage (**Figure 2**). The three biological replicates of the same genotype were well clustered. The groups of genotypes showed the same pattern for the two stages. Cervil was separated from all other genotypes. The hybrids were located in between their parental genotypes for the four crosses. A total of 341 spots showed significant changes in protein abundance according to the genotype, 357 according to the stage and 242 according to the stage by genotype interaction. One way ANOVA showed that 256 spot amounts were significantly different among genotypes at cell expansion stage and 188 at the orange-red stage (117 were common to both stages). These spot amounts were used for inheritance and correlation analysis. We have identified the function of most of these spot (Xu et al, in preparation; **Supplemental table S5**). For 74 genes (corresponding to 187 spots), two to seven different spots were detected. For example, the acid invertase (*Solyc03g083910.2.1*) corresponded to seven spots, and phosphoglucomutase (*Solyc04g045340.2.1*) and enolase (*Solyc09g009020.2.1*) corresponded to five spots. These spots may correspond to post-transcriptional and post-traductional modifications or to allelic variations (Xu et al, submitted). A total of 71 and 60 significant spots classified to primary metabolic process (65 and 58 spots) and vitamin synthesis (6 and 2 spots) classes, at cell expansion and orange-red stage (**Supplemental table S5**), were selected to illustrate the correlation networks between protein spots and metabolites and enzyme activities. Among the 424 protein spots, 190 were in average in lower amount at the orange-red stage and 234 were in higher amount.

Variation of metabolite contents, enzyme activities and protein spot amounts within and among genotypes

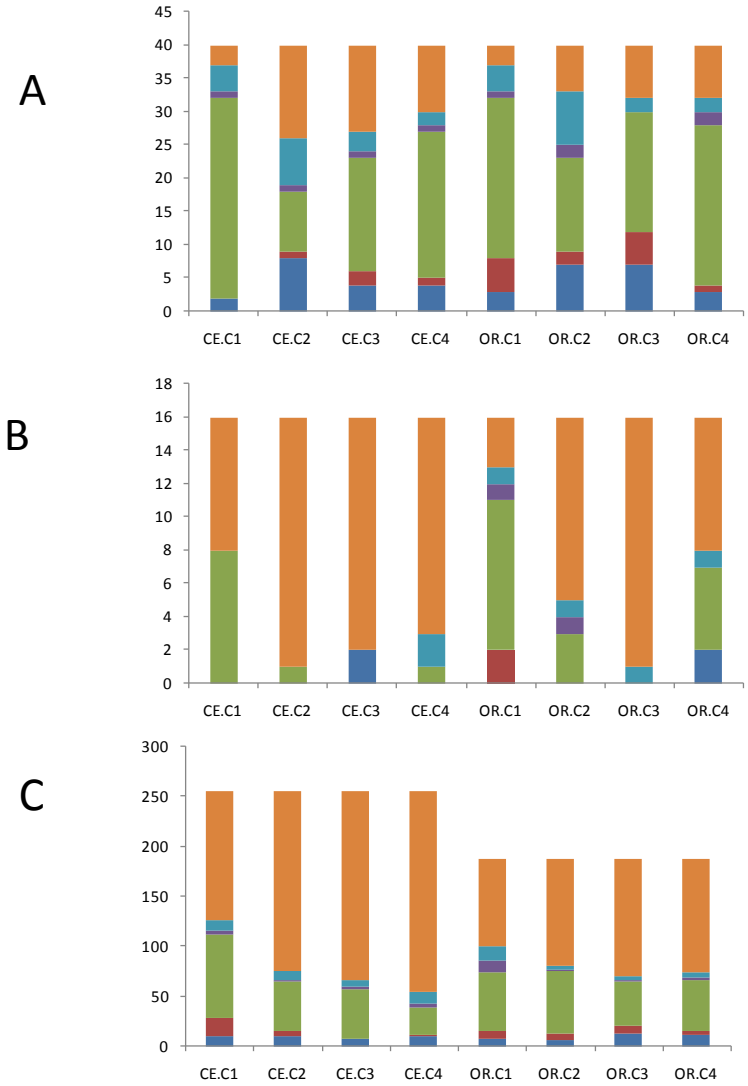


Figure 4. Classification of the mode of inheritance for 34 metabolites content and 5 phenotypic traits (A), 16 enzyme activities (B) and 256 protein spots for cell expansion stage and 188 protein spots for orange red stage (C) in the four crosses at the two stages. C1: Cervil × Levovil; C2: Stupicke Polni Rane × Criollo, C3: LA0147 × Plovdiv 24A, C4: Ferum × LA1420. CE: cell expansion stage, OR: orange-red stage. Colors of the histogram are showing different inheritances: over-dominant, dark blue; dominant, red; additive, light green; recessive, purple; over-recessive, light blue; non-significant, orange.

Coefficients of variation (CV) were calculated for each genotype per stage, and then an average CV was obtained for each trait at each stage (**Figure 3 and Supplemental table S6**). CV was low (less than 0.2) for most of the metabolites (88%) at cell expansion stage, except for one secondary metabolite (coumaric acid, sCou) with an intermediate CV (0.2-0.3) and three phenolic compounds with high CV (0.36, 0.31, 0.42 for sTCQ, sCry and sNna). Similar results were obtained for metabolites at orange-red stage, most of them (94%) showing low CV values. However, Crypto-chlorogenic acid showed high CV (**Figure 3A and Supplemental table S6**). In contrast to metabolites, at cell expansion stage, seven enzymes (26%) showed low CV value (eSKDH , eEno , ePFKa , ePgm , eFbpA , ePepC and eGlK). A total of 13 enzymes (48%) showed intermediate CV and seven enzymes (26%) showed high CV. At orange-red stage, six enzymes (22%) showed low CV, six enzymes (22%) showed intermediate CV and 15 enzymes showed high CV (**Figure 3B and Supplemental table S6**). Most of the protein spot (93%) volumes showed low CV, and only 4 spots showed a high CV. Similar results were obtained for all the protein spot volumes at orange-red stage (**Figure 3C and Supplemental table S6**). The CV were low for the 5 phenotypic traits at both stages (**Supplemental table S6**).

Comparison of CV for phenotypic traits, metabolites, enzyme activities and protein spots in 4 *S. lycopersicum* accessions, 4 *S. l. cerasiforme* and 4 hybrids at cell expansion and orange-red stage was illustrated in **Supplementl figure S4** and **Supplemental table S6**. Compared to low CV values (less than 0.2) for phenotypic traits, metabolites and protein spots, CV were in average higher for the enzyme activities in the three groups (0.21, 0.23 and 0.35 for 4 *S. lycopersicum*, 4 *S. l. cerasiforme* and 4 hybrids, respectively) at cell expansion stage. It was even higher at orange-red stage (0.33, 0.31 and 0.27) for the three groups. There were not many differences among the groups of genotypes.

Mode of inheritance of phenotypes, metabolites, enzyme activities and proteins

The mode of inheritance was assessed for all traits in each cross (**Figure 4 A**). The number of significantly variable traits varied among crosses in relation to the genetic distance between the parental lines. Plant height and stem diameter were systematically inherited in a dominant or overdominant mode. On the contrary, fruit weight and diameter were additive or partially recessive. At cell expansion stage, 28% of the metabolites were not significantly different in the four crosses. In average over the four crosses, the mode of inheritance of the metabolites

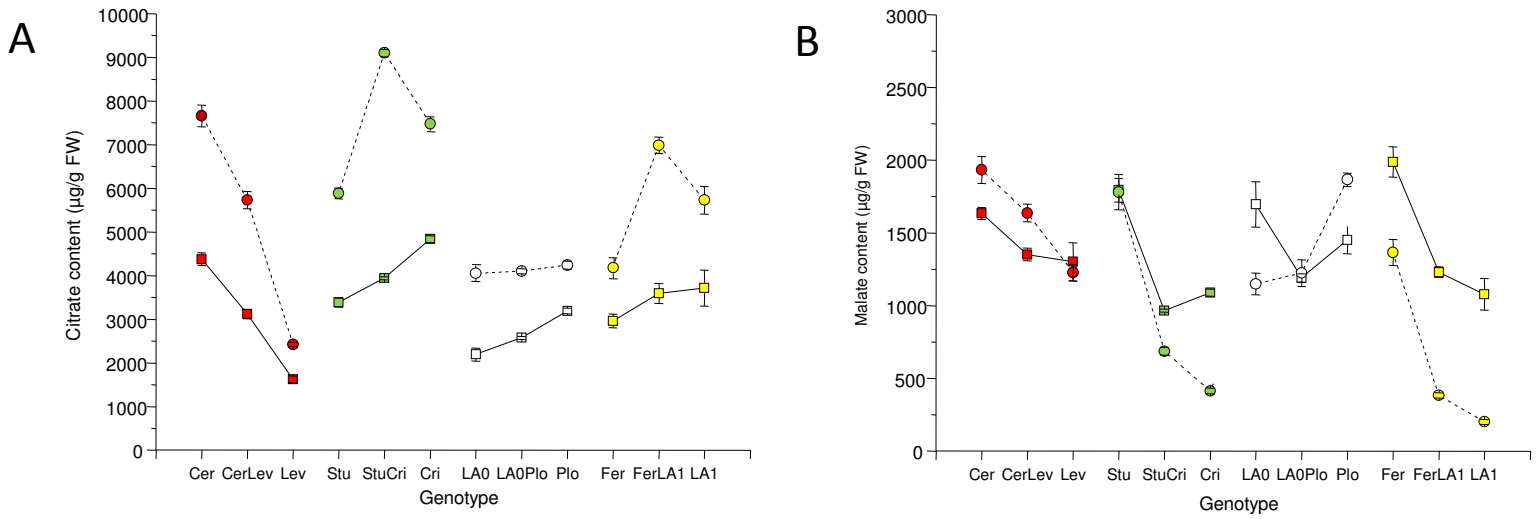


Figure 5. Inheritance pattern of citrate (A) and malate (B) content at cell expansion stage (dash line, solid circle) and orange red stage (black line, empty square). Cer: Cervil, CerLev: Cervil × Levovil, Lev: Levovil, Stu: Stupicke Polni Rane, StuCri: Stupicke Polni Rane × Criollo, Criollo: Cri, LA0: LA0147, LA0 ×Plo: LA0147 × Plovdiv24A , Plo: Plovdiv24A, Fer: Ferum, FerLA1: Ferum × LA1420 and LA1: LA1420. Genotypes from the same cross are shown in the same color.

is not fairly distributed between over-dominance (11%), dominant (4%), additive (45%), recessive (3%) and over-recessive (10%). The same pattern was observed for the orange-red stage. A large number of additive traits (29/40 and 21/40) for cell expansion stage and orange-red stage, respectively) was found in cross between the most distant lines (Cervil x Levovil). However, most of the traits exhibited different inheritance mode at the two stages in the same cross or in different crosses. For example, citrate content showed an additive inheritance at cell expansion stage in Stupicke Polni Rane × Criollo and Ferum ×LA1420 cross, but a recessive inheritance at orange-red stage (**Figure 5A**). Malate content showed over-recessive inheritance at cell expansion stage in Stupicke Polni Rane × Criollo and LA0147 × Plovdiv24A cross but additive inheritance at orange-red stage (**Figure 5B**).

Inheritance mode of enzyme activities was assessed on 16 enzyme activities which were significantly variable at both stages (**Figure 4B**). At cell expansion stage, in average, a small proportion of enzymes (3%) were over-dominantly and over-recessively inherited, 16% of traits were additively inherited; no traits were dominantly and recessively inherited. At orange-red stage, the proportions were 3%, 3%, 27%, 3%, 6% for over-dominant, dominant, additive, recessive and over-recessive traits, respectively (**Figure 4B**).

Distribution of inheritance of protein amounts was also different depending on the stage and the cross (**Figure 4C**). At cell expansion stage, 4 % of protein spots were over-dominant, 2% were dominant, 21% were additive, 1% were recessive, 4% were over-recessive (with 68% non-significantly variable spots). At orange stage, the proportions were 5%, 3%, 28%, 3%, 4% and 57% for the five inheritance modes and for non-significant spots within cross. The largest number of variable protein spots were observed in the first cross (Cervil × Levovil) at cell expansion stage (126/256 spots) and orange-red stage (101/188 spots). Like for the metabolites, a protein may show different inheritance patterns at two stages for the same cross or in different crosses.

Correlations among traits at each level of expression

To analyse the relationships among the 34 metabolites, 5 phenotypic traits, 26 enzyme activities and a subset of significant protein spots (256 for cell expansion and 188 for orange-red stage), pairwise Pearson correlations were first calculated within each group of traits at each stage. Then networks were constructed by sPLS regressions between each pair of groups

Table 3. Overview of the number of significant correlations (|r|>0.71; p<0.01) among phenotypic traits, metabolite contents, enzyme activities and amounts of protein spot (classified into three functional classes) and between each two groups.

	Nb. of traits	Phenotypes			Metabolites			Enzymes			Proteins		
		r>0.71	r<-0.71	Expected by chance	r>0.71	r<-0.71	Expected by chance	r>0.71	r<-0.71	Expected by chance	r>0.71	r<-0.71	Expected by chance
Cell expansion stage													
Phenotypes	5	1	0	0.1	-	-	-	-	-	-	-	-	-
Metabolites	34	16	8	2	117	22	6	-	-	-	-	-	-
Enzymes	26	0	1	1.3	22	0	9	50	0	3	-	-	-
Proteins													
Misc. Pr.	146	26	22	7	232	166	50	39	22	38	-	-	-
Metab. Pr.	71	10	5	4	87	77	24	24	11	19	-	-	-
Stress Pr.	39	7	11	2	93	15	13	3	2	10	-	-	-
Total	256	43	38	13	412	258	87	66	35	67	916	753	326
Orange red stage													
Phenotypes	5	1	1	0.1	-	-	-	-	-	-	-	-	-
Metabolites	34	18	11	2	73	1	6	-	-	-	-	-	-
Enzymes	26	7	7	1.3	99	1	9	90	0	3	-	-	-
Proteins													
Misc. Pr.	98	14	14	5	133	58	33	61	54	25	-	-	-
Metab. Pr.	60	19	8	3	89	39	20	70	4	16	-	-	-
Stress Pr.	30	5	7	2	59	6	10	55	1	8	-	-	-
Total	188	38	29	10	281	103	63	186	59	49	544	361	176

Protein functional Misc.: miscellaneous; Metab.: primary metabolic process and vitamin synthesis; Stress: stress response

of traits. The numbers of correlations observed and expected by chance in each group of analysis are presented in **Table 3**. Within each class, the number of significant correlations is always much higher than expected by chance and a large tendency to an excess of positive correlations is shown for all the class of traits, particularly for metabolites and enzymes. Between metabolite contents and enzyme activities, a larger number of correlations were observed at orange-red stage than at cell expansion stage with almost no negative correlations. Between metabolite and protein amounts a large number of both positive and negative correlations were observed, with the same ratios in each class. On the contrary, between enzyme activities and protein amounts, there was not much significant correlations compared to random at cell expansion stage, on the contrary to the orange-red stage, where the numbers of correlations were higher, with a clear tendency to an excess of positive correlations with the proteins corresponding to primary metabolism and stress response.

Correlations among metabolites and phenotypic traits

For the 34 metabolites and 5 phenotypic traits, at cell expansion stage, among the 741 possible pairwise correlations, 164 and 105 were significant ($p<0.01$) at cell expansion stage (**Figure 6 A**) and orange-red stage (**Figure 6B**). At cell expansion stage, the content of fructose in the 12 genotypes was strongly correlated with the content of glucose, but not with that of sucrose. The content of sucrose was positively correlated with sugar alcohol (Inos). Sugar (glucose, fructose) content was negatively correlated with the contents of amino acids, Trigo and glycoalkaloids (sAto, sDHT). Fructose content also negatively correlated with vitamin C, organic acids (Cit) and one phenolic compound (sCHI) but positively correlated with fruit diameter. Amino acid contents were positively correlated with each other, with starch, with vitamin C, with glycoalkaloids (sAto and SDHT), with phenolic compounds (sChl and sRut) and fruit dry matter content. Organic acid content (Cit) was negatively correlated with fruit weight and fruit diameter. The other one (Fum) was negatively correlated with plant height. Secondary metabolites were positively correlated with each other but negatively correlated with fruit weight and fruit diameter. Fruit weight was positively correlated with fruit diameter. No significant correlation was detected between plant traits (stem diameter) and all the metabolites and the other phenotypic traits (fruit weight, fruit diameter and plant height) (**Figure 6A**).

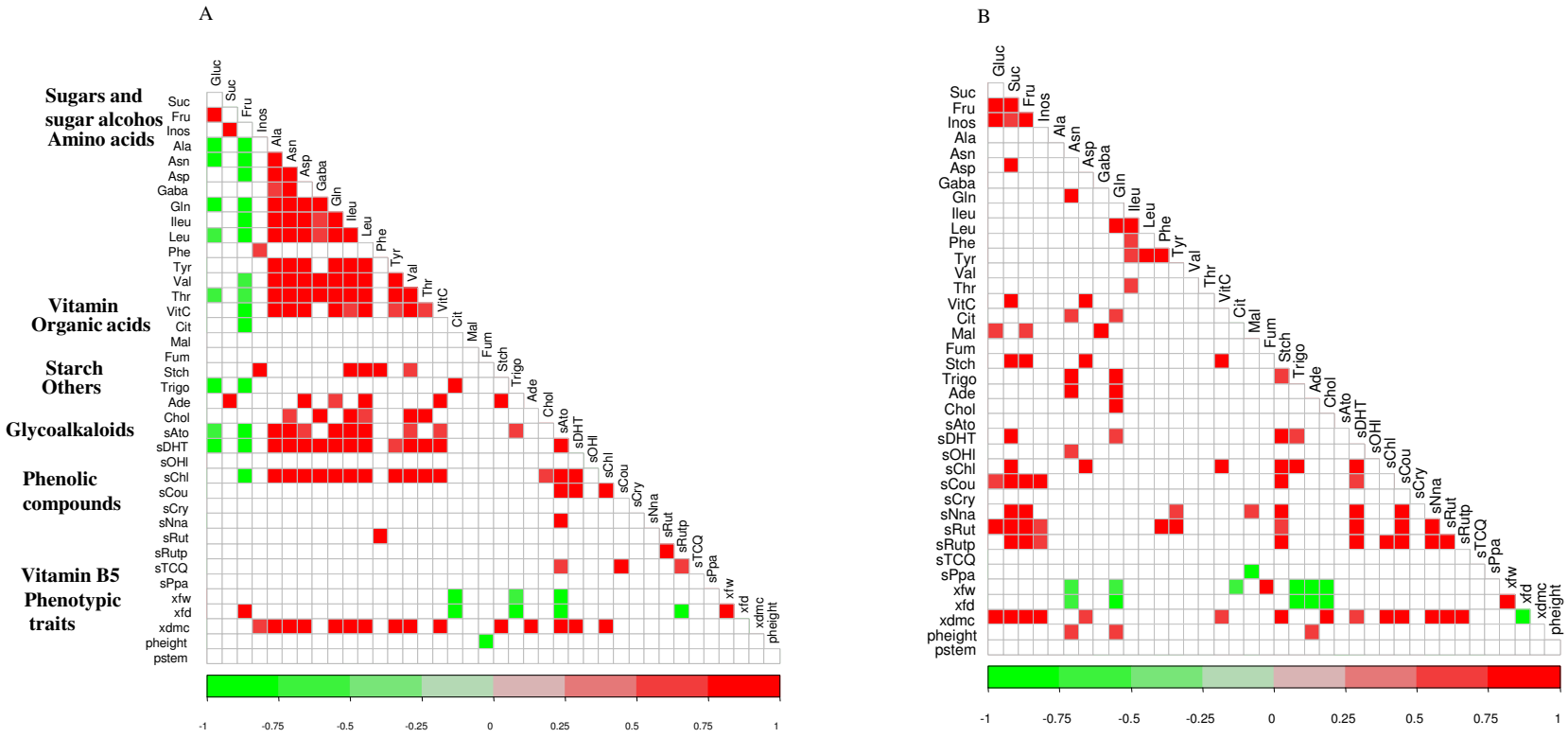


Figure 6. Correlations among 34 metabolites content and 5 phenotypic traits at cell expansion stage (A) and orange red stage (B). Only significant correlations (P<0.01) are shown. Positive and negative correlations are presented in red and green, respectively.

Compared to cell expansion stage, relatively less significant correlations were detected between metabolites and phenotypic traits at orange-red stage (**Figure 6B**). Sugars and sugar alcohol were highly and positively correlated with each other, with two amino acids (Asp and Gln), vitamin C, glycoalkaloids (sDHT), phenolic compounds (sNna, sRut and sRutp) and dry matter content. Amino acids were less significantly correlated together. They were positively correlated with vitamin C and organic acids (Cit and Mal) but negatively correlated with fruit weight and fruit diameter. Glycoalkaloids (sDHT) was positively correlated with phenolic compounds. Phenolic compounds were positively correlated with each other and with dry matter content. Fruit weight and fruit diameter were still positively correlated and they were negatively correlated with dry matter content, Trigo, Ade and Chol.

Correlations between enzyme activities

Among 325 possible pairwise correlations, 50 and 90 significant ($p<0.01$) correlations were obtained at cell expansion stage (**Figure 7A**) and orange-red stage (**Figure 7B**), respectively. Only positive correlations were found at the two stages. No significant correlations were detected among enzymes related to amino acid metabolism and Calvin cycle at the two stages. At cell expansion stage, enzymes involved in the glycolysis pathway were positively correlated with each other and with enzymes related to sucrose breakdown and sucrose synthesis that were also positively correlated together (**Figure 7A**). Few correlations were found among enzymes related to TCA cycle. More correlations were found at orange-red stage between enzymes involved in glycolysis pathway (highly correlated with each other) and enzymes related to sucrose breakdown synthesis and TCA cycle (**Figure 7B**).

Correlations between proteins

A total of 256 and 188 spots significantly varying among genotypes ($p<0.01$) at cell expansion stage and orange-red were used for correlation analysis (Supplemental table S5). Among the 32,640 and 17,578 possible pairwise correlations 1,699 (5 %) and 905 (5 %) significant correlations ($p<0.01$) were detected at cell expansion and orange-red stage, respectively. Correlations were detected within and between functional classes with similar frequences (**Figure 8**, **Supplemental figure S5**). A few correlations were stable at the two developmental stages. However, many changes were observed from cell expansion stage to orange-red stage.

Correlations between enzyme activities and their corresponding protein spot amounts

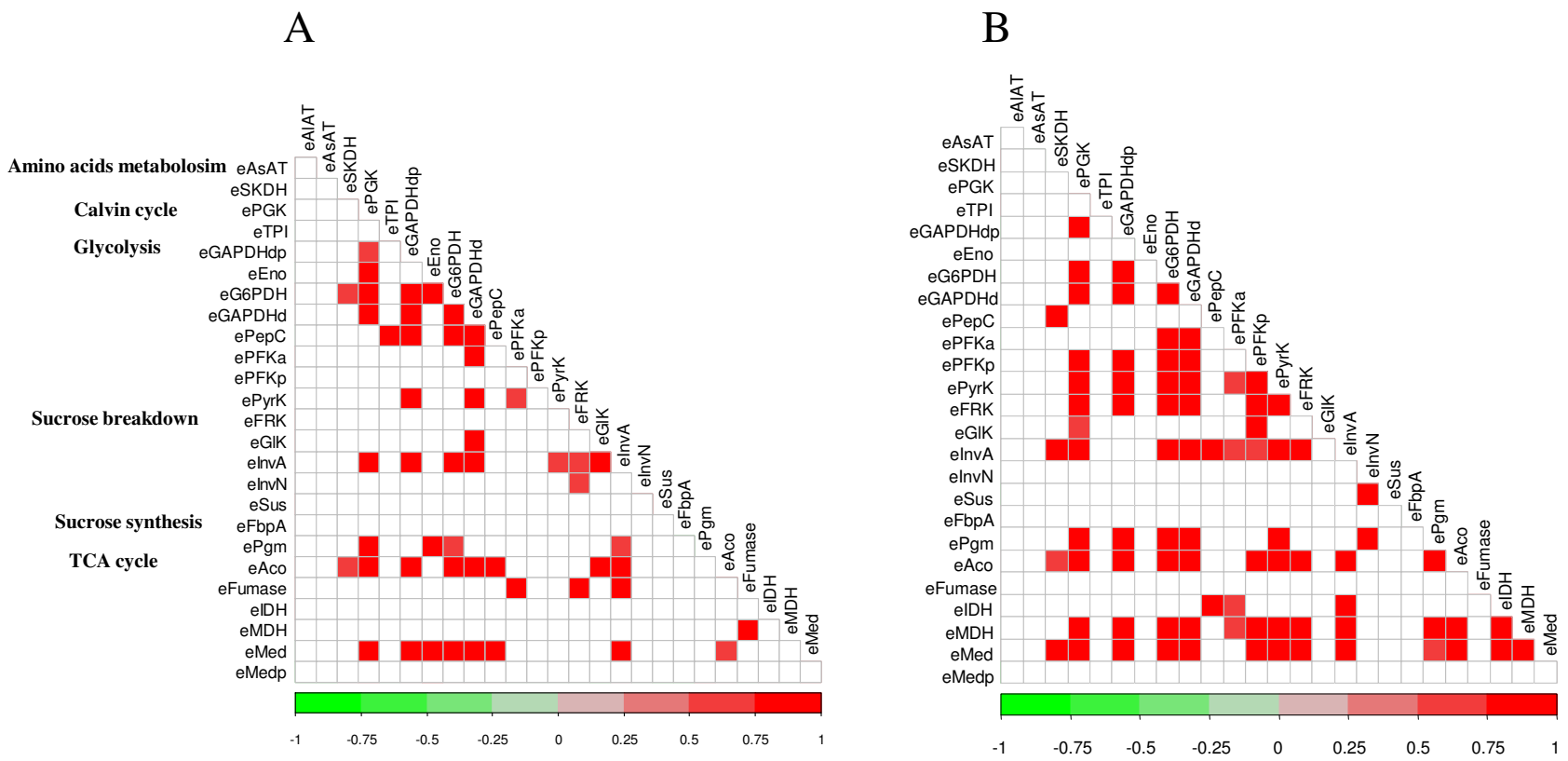


Figure 7. Correlations among 26 enzyme activities at cell expansion stage (A) and orange red stage (B). Only significant correlations are shown ($p<0.01$). Positive and negative correlations are presented in red and green, respectively.

At cell expansion, 19 spots corresponding to 8 enzymes whose activity was quantified were detected as variable. Significant correlations were detected between the amount of JX087 (acid invertases), JX015 (fructokinase), JX154 (malate dehydrogenase) and the activity of the corresponding enzymes (r=0.56, 0.78, 0.76, respectively). The last two spots were also correlated with several other enzymes (GAPDH, GlK, PGM, FBPA, TPI, Eno, MDH, SKDH for the fructokinase and with SKDH, GAPDH and FBPA for MDH) A significant negative correlation was observed between JX228 and eMed (r=-0.67).

At the orange red stages, 20 spots corresponding to 8 enzymes whose activity was quantified were detected as variable. Significant correlations were detected between the amounts of JX419 (acid invertase), four spots corresponding to enolase (JX161, JX214, JX167 and JX323), JX218 (fructokinase), two spots corresponding to isocitrate dehydrogenase (JX058 and JX302), two spots corresponding to MDH (JX004 and JX177) and JX091 (malic enzyme) and the activity of the corresponding enzymes (r=-0.63 with invA, 0.56 to 0.66 with Eno, 0.58 with FRK, 0.72 and 0.63 with IDH, 0.74 and 0.67 with MDH, 0.54 with ME). This last spot and JX004 were highly correlated together and with several other enzymes.

Network analysis between levels of expression

One important problem for data integration is that data sets are collected from different levels. Correlation analyses contain noise, while the number of replications is also a limiting factor. Failure to acknowledge any of these factors may cause unwanted systematic effects, which are introduced during data collection, and create false connections between variables, ultimately affecting biological conclusions. Many alternative strategies, including sparse partial least square regression (sPLS) are proposed in the literature for integrating data from parallel sources, as reviewed by Joyce and Palsson (2006). The sPLS is a bidirectional multivariate regression method that allows separate modeling of covariance between two data sets. The main advantage of sparse methods over non-spare methods is that it sets the contribution of noise variables to zero to improve the prediction or classification performance. Filzmoser et al. (2012) reviewed recent sparse methods for regression and classification and underlined the value of sPLS to integrate datasets from different levels.

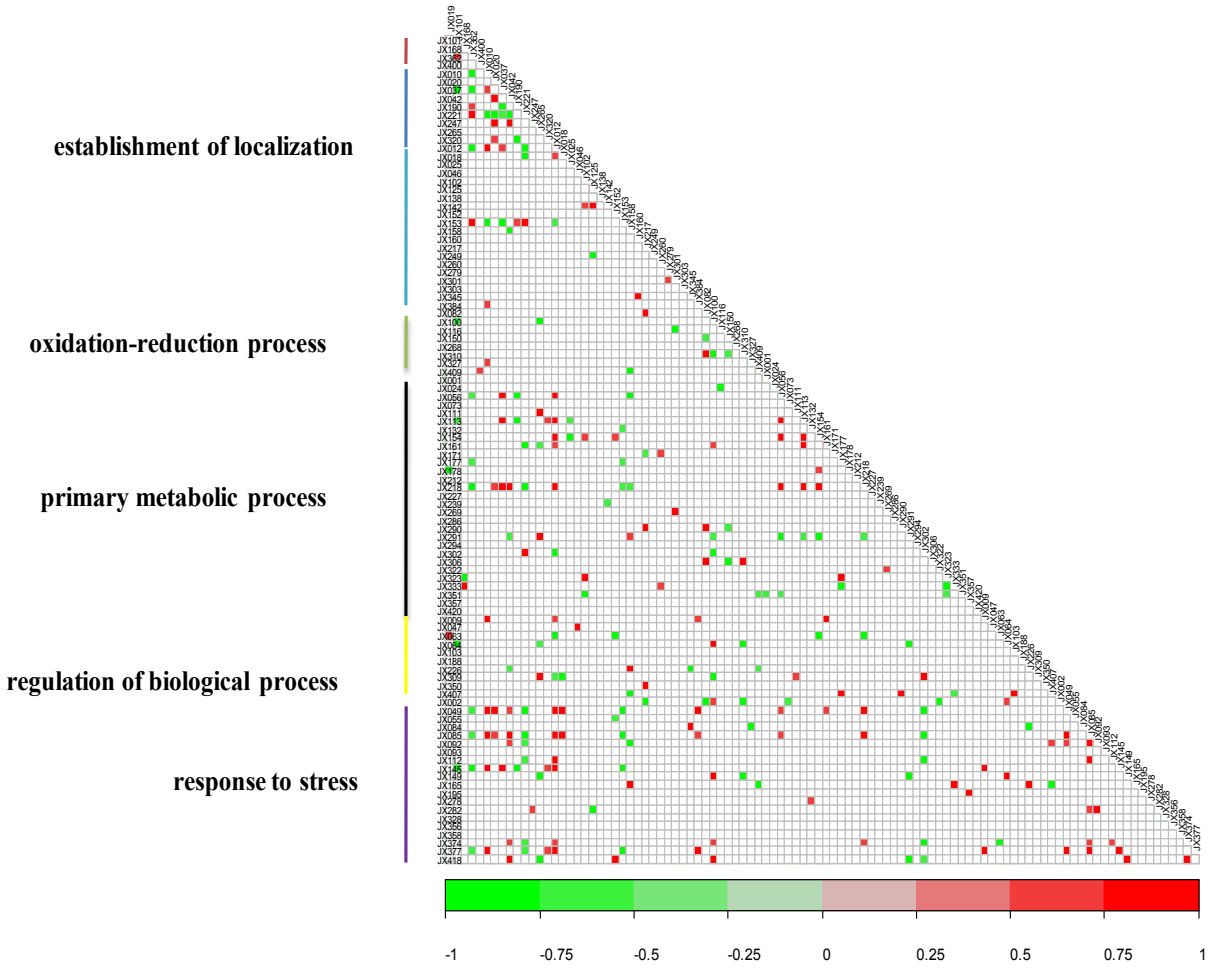


Figure 8. Correlations among the amounts of 101 protein spots at cell expansion stage.
Based on variation analysis, 117 protein spots were significantly different among genotypes at both cell expansion stage and orange red stage. These protein spots were then grouped according to their functional classes. Seven groups (with more than 5 protein spots) with a total of 101 protein spots were kept for representing the correlations among protein spots. Only significant correlations ($P<0.01$) are shown. Positive and negative correlations are presented in red and green, respectively.

A global sPLS network was thus obtained between each two groups of data and each stage. As the final state of metabolites and traits may depend on enzymes and proteins acting earlier, network analyses were also constructed between 71 protein spot amounts at cell expansion stage and 34 metabolites and 5 phenotypic traits measured at orange-red stage, and between 26 enzyme activities at cell expansion stage and 34 metabolites and 5 phenotypic traits at orange-red stage. Only significant ($p<0.01$) correlations were extracted from trait-trait correlations. Metabolites, phenotypic traits, enzymes, protein spots were represented as different nodes and significant correlations as links connecting nodes.

Network analysis of phenotypic, metabolic and enzymatic traits

Between 34 metabolites and 5 phenotypic traits on one hand and 26 enzyme activities on the other hand at cell expansion, only a few connections were detected. One cluster with 7 nodes including 3 enzymes, three secondary metabolites and fruit diameter was detected (**Supplemental figure S6**). eSKDH, an enzyme related to cyclic amino acid metabolism was correlated with rutin content. Another cluster connected eFbpA, an enzyme related to sucrose synthesis, and VitC.

A higher number of connections was detected for the same traits at orange-red stage than at cell expansion stage. One large cluster with a total of 27 nodes (12 metabolites, one phenotypic and 14 enzymes) was detected at this stage (**Figure 9**). In this cluster, many expected correlations were found. Sucrose and starch were connected to several enzymes related to sucrose synthesis, sucrose breakdown and to glycolysis pathway. The dry matter content was connected to 8 enzyme activities. Aspartate was also connected to most of the enzymes. Chlorogenic acid had a central role as it was connected to 13 enzyme activities, together with Dehydrotomatine and Rutin pentoside (**Figure 9**).

Network analysis involving metabolites, phenotypic traits and protein spots

At cell expansion stage, two clusters were detected between 34 metabolites, 5 phenotypic traits on one hand and 71 protein spots related to carbon metabolism on the other hand. One large cluster contained 40 nodes (involving 18 protein spots, 20 metabolites and fruit diameter and weight) and one small cluster with two nodes (one protein spot JX239, an acid invertase and plant height) (**Figure 10**). In the large cluster, both positive and negative connections were observed. Interesting positive correlations were found between a fructokinase spots (JX363 Protein Fructokinase 3) and fructose. Two other fructokinase spots corresponding to

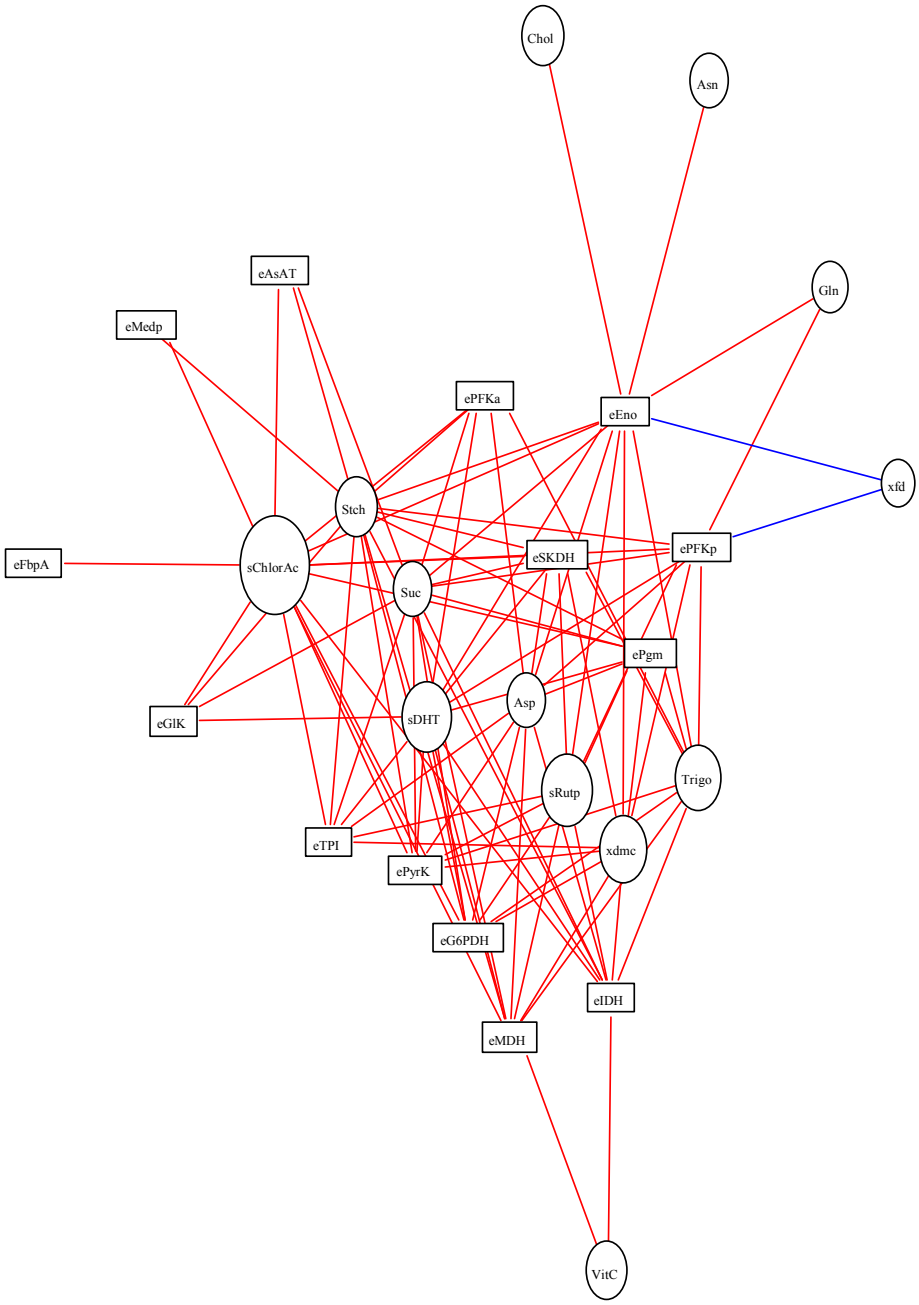


Figure 9. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and 26 enzyme activities at orange red stage. Metabolites, phenotypic traits, and enzymes were represented as different nodes (circle for metabolites and phenotypic traits, square for enzymes). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ((|r|>0.71, p<0.01).

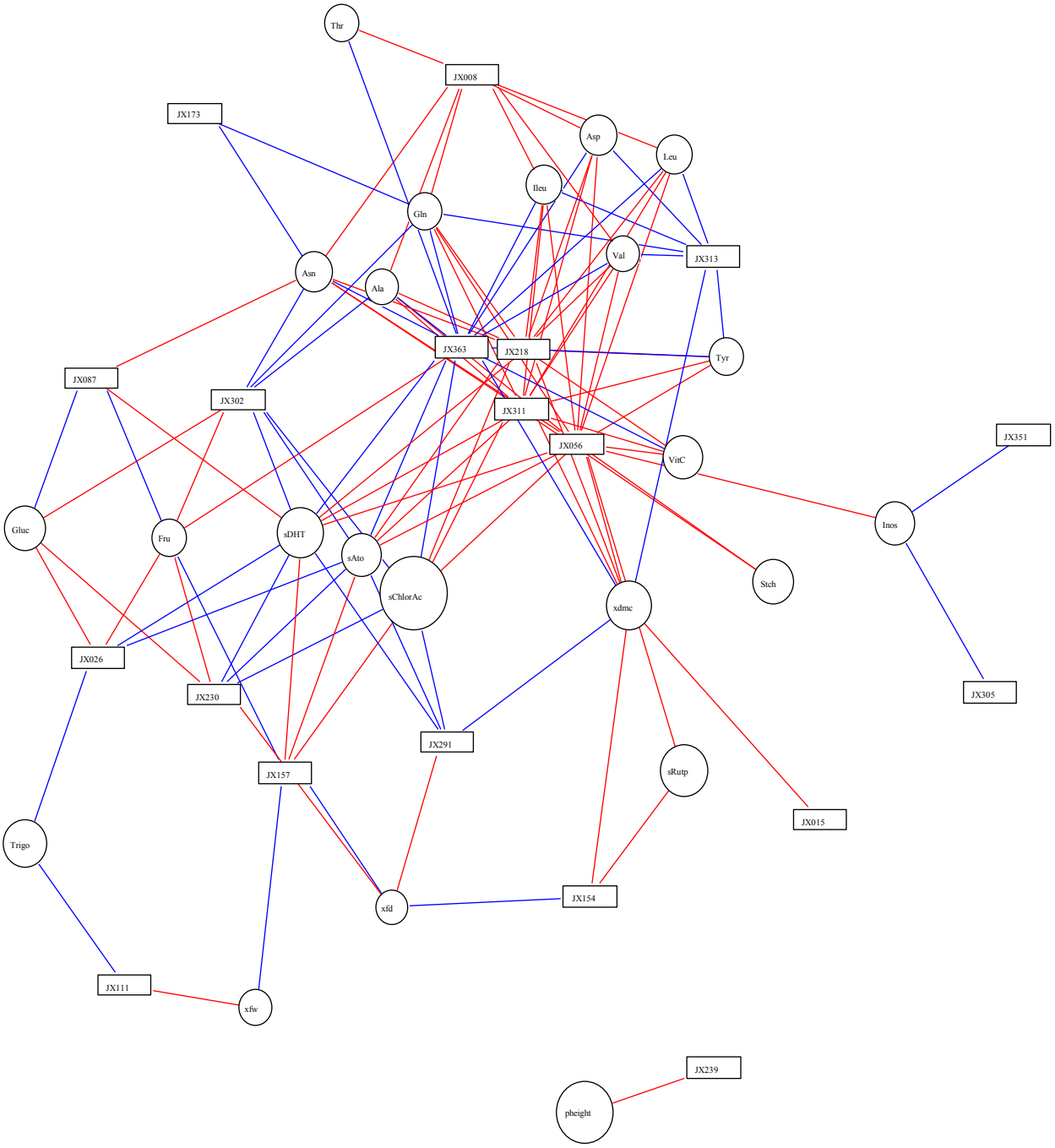


Figure 10. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and the amounts of 71 protein spots at cell expansion stage. The 71 protein spots were significant among genotypes at cell expansion stage and were classified to primary metabolic (65 spots) and vitamin (6 spots) synthesis classes. Metabolites, phenotypic traits, and protein spots were represented as different nodes (circle for metabolites and phenotypic traits, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($|r|>0.71$; $p<0.01$). JX008: Triosephosphate isomerase ; JX015: Fructokinase-like; JX026: Cysteine synthase ; JX056: Lipoxygenase; JX087: Acid beta-fructofuranosidase ; JX111: Inorganic pyrophosphatase ; JX154: Malate dehydrogenase ; JX157: Nucleoside diphosphate kinase ; JX173: GDP-D-mannose-epimerase 2; JX218: Fructokinase-like; JX230: Acetylornithine aminotransferase; JX239: Acid beta-fructofuranosidase ; JX291: Lactoylglutathione lyase; JX302: Isocitrate dehydrogenase; JX305: Triosephosphate isomerase; JX311: Ubiquitin carboxyl-terminal hydrolase ; JX313 N-succinylglutamate 5-semialdehyde dehydrogenase; JX351:Ubiquilin-1; JX363:Fructokinase 3

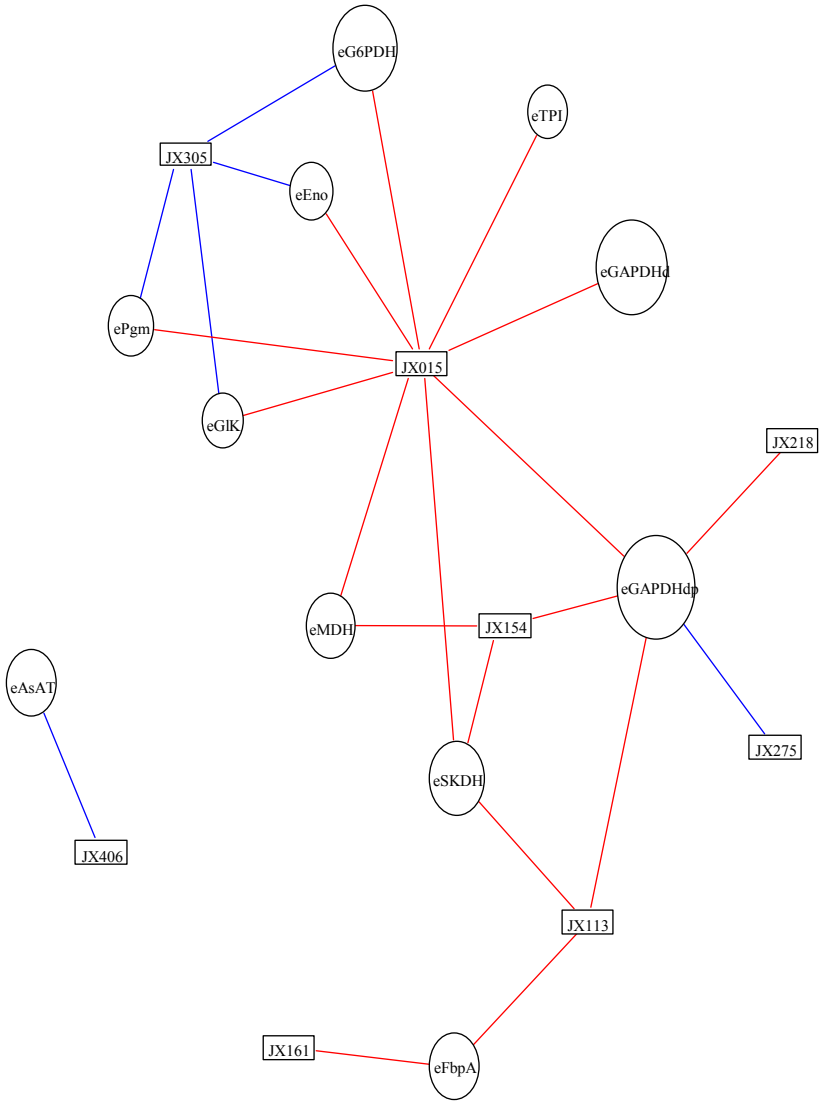


Figure 11. Network analysis based on sPLS analysis of 26 enzyme activities and the amounts of 71 protein spots at cell expansion stage. The 71 protein spots were significant among genotypes at cell expansion stage and were classified to primary metabolic (65 spots) and vitamin (6 spots) synthesis classes. Enzymes and protein spots were represented as different nodes (circle for enzyme, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ((|r|>0.71; $p<0.01$). JX015: Fructokinase-like; JX113: Phosphoglycerate kinase; JX154: Malate dehydrogenase ; JX161:Enolase; JX218: Fructokinase-like; JX275: S-formylglutathione hydrolase ; JX305: Triosephosphate isomerase; JX406: Phosphoglucomutase

different genes (JX218 and JX311) were positively connected with several amino acids and to dry matter content. Spot JX087 (another spot corresponding to the acid invertase gene Solyc03g83910 like JX239) was negatively connected with glucose and fructose, while an isocitrate dehydrogenase spot (JX302) was positively connected to hexose sugars. A cysteine synthase (JX026) was connected to glucose and fructose. An inorganic pyrophosphatase (JX111) was connected with fruit weight. A lactoyl gluthathione lyase (JX291) was positively connected to fruit diameter and negatively to dry matter content, chlorogenic acid and glycoalkaloid content (**Figure 10**).

A less dense cluster with 50 nodes including 24 protein spots, 23 metabolites and three phenotypes was observed at orange-red stage (**Supplemental figure S7**). In this cluster, both positive and negative connections were observed. Two protein spots, JX091 (Malic enzyme) and JX239 (acid invertase) were positively correlated with sugars and sugar alcohol (Gluc, Fru, sucrose, starch, dry matter and Inositol and secondary metabolites (sNna, sRut and sRutp). Another spot corresponding to the same acid invertase gene (JX051) was also connected to many traits among which plant height and fruit weight. Protein spots JX351 (Ubiquilin-1), JX292 (2-isopropylmalate synthase) and JX372 (Diaminopimelate epimerase family protein) were negatively correlated with sugars (Gluc, Fru) and with inositol. Six spots were conneted to fruit weight and diameter among which the lactoyl gluthathione lyase (JX291). Two spots (JX323 and JX256) corresponding to the same gene coding an enolase enzyme (Solyc09g009020) were connected with aspartate; the first one was connected with three other amino acids. A spot corresponding to an adenylate kinase (JX394) was connected to six traits among which the content in adenosine.

Network analysis of enzyme activities and protein spot amounts

Two clusters were constructed by grouping the correlations between 26 enzyme activities and 71 protein spots at cell expansion stage. One large cluster with 17 nodes (10 enzyme activities and 7 protein spots) was obtained (**Figure 11**). In this cluster, protein spot JX015 (Fructokinase-like) was connected with 9 enzyme activities related to several levels of the pathway. The protein spot JX154 (Malate dehydrogenase protein) was correlated with eMDH. Protein spots JX161 (Enolase) and JX113 (Phosphoglycerate kinase) were highly correlated with eFbpA, an enzyme related to sucrose synthesis pathway. A small other cluster involved one protein spot JX406 (Phosphoglucomutase) and one enzyme eAsAT related to amino acids metabolism (**Figure 11**).

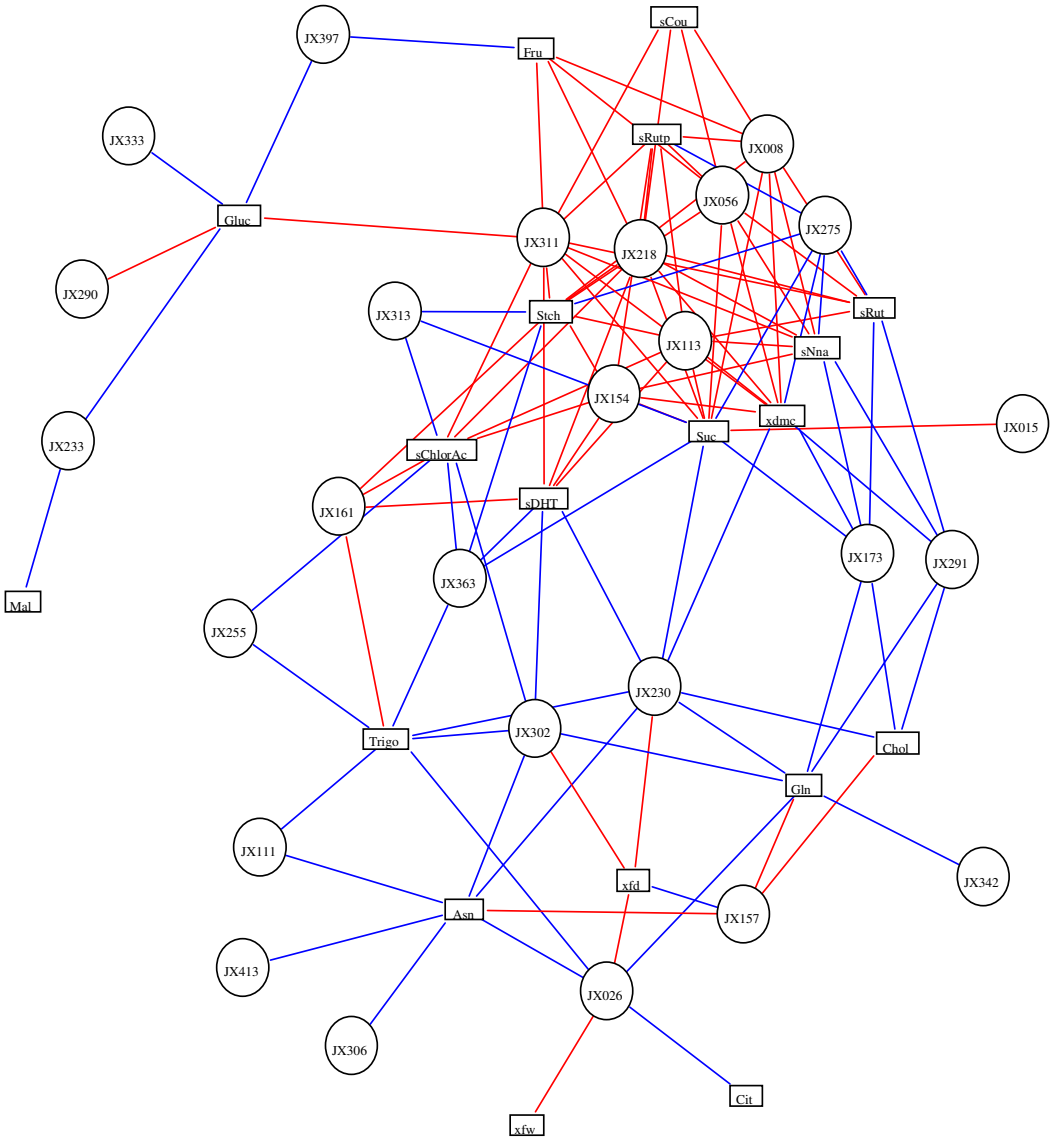


Figure 12. Network analysis based on sPLS analysis of the amounts of 71 protein spots at cell expansion stage and 34 metabolites, and 5 phenotypic traits at orange red stage. The 71 protein spots were significant among genotypes at cell expansion stage and were classified to primary metabolic (65 spots) and vitamin (6 spots) synthesis classes. Metabolites, phenotypic traits and protein spots were represented as different nodes (circle for protein spots, square for metabolites and phenotypic traits). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($|r|>0.71$; $p<0.01$). JX008: Triosephosphate isomerase; JX015: Fructokinase-like; JX026: Cysteine synthase; JX056: Lipoxygenase; JX111: Inorganic pyrophosphatase ; JX113:Phosphoglycerate kinase; JX154: Malate dehydrogenase; JX157: Nucleoside diphosphate kinase; JX161: Enolase; JX173: GDP-D-mannose-epimerase 2; JX218: Fructokinase-like; JX230: Acetylornithine aminotransferase; JX233: Malonyl CoA-acyl carrier protein transacylase containing protein expressed; JX255: 3-hydroxyisobutyrate dehydrogenase; JX275: S-formylglutathione hydrolase; JX290: LL-diaminopimelate aminotransferase; JX291: Lactoylglutathione lyase; JX302: Isocitrate dehydrogenase; JX306: Ubiquitin thioesterase OTU1; JX311: Ubiquitin carboxyl-terminal hydrolase; JX313: N-succinylglutamate 5-semialdehyde dehydrogenase; JX333: Ribulose-5-phosphate-3-epimerase; JX342: Imidazole glycerol phosphate synthase subunit hisF; JX363: Fructokinase 3; JX397: Adenine phosphoribosyltransferase-like protein; JX413: Ubiquitin-conjugating enzyme family protein-like

At orange-red stage, one large cluster with 22 nodes (7 protein spots and 15 enzyme activities) was identified (**Supplemental figure S8**). Only positive connections were observed in this cluster. Protein spot Protein spots JX004 (Malate dehydrogenase) was correlated with eMDH and JX214 corresponding to an enolase was connected with the activity of the corresponding enzyme. Three other protein spots (JX051, acid invertase, JX091 malic enzyme and JX132 an oxalyl CoA decarboxylase were connected with the activities of many different enzymes. In the small cluster, protein spot JX180 (Pyridoxal biosynthesis lyase) was negatively connected to eMed (enzyme related to TCA cycle) and positively correlated with eGAPHdp (enzyme related to Glycolysis pathway). (**Supplementary figure S8**).

Network analysis across stages

We then tested if there were connectivities between protein content or enzyme activities at early stage (cell expansion) and fruit composition at orange-red stage. (**Figure 12 and supplemental figure S9**). In the first network analysis involving 71 protein spots, one large cluster with 45 nodes (26 protein spots, 17 metabolites and two phenotypes) was obtained. Both positive and negative connections were detected. Protein spot JX363 (FRK3) was negatively connected with sucrose and starch, while the protein spot JX218 (Fructokinase-like) protein was positively correlated with fructose, starch, sucrose, dry matter and many other secondary compounds. Protein spot JX173 (GDP-D-mannose-epimerase 2), which is related to vitamin synthesis, was negatively connected to sucrose. Protein spot JX397 (Adenine phosphoribosyltransferase-like protein) was negatively connected to sugars (Fru and Glu). Protein spot JX233 (Malonyl CoA-acyl carrier protein) was negatively connected to sugars (Gluc) and organic acid (Mal). Fruit weight and diameter were again connected with the cysteine synthase (JX026) spot

Three small clusters were detected between enzyme activities at cell expansion stage and metabolites and phenotypic traits at orange-red stage. One 2-node cluster connected Aspartate aminotransferase activity to the content in threonine (which derives from aspartate). One cluster with 6 nodes (2 enzymes and 4 metabolites) was observed. eFbpA was positively connected to Suc, starch and Chloro. In the other small cluster, eSus was connected to sPpa (**Supplemental figure S9**).

Discussion

The variation of tomato fruit composition has been widely studied during the last ten years, due to its role in sensory and nutritional value. For sensory value, the rates of sugars, acids, aromas and in some cases amino acids are the most important traits (Causse et al. 2002; Klee and Giovannoni 2011; Tieman et al. 2012). For the nutritional value, apart from carotenoids, the contents of vitamin C and phenolic compounds have a major role (Frusciante et al. 2007). Glycoalkaloids are also important as they can have a positive action as chemical barrier against a broad range of pathogens and a variety of pharmacological and nutritional properties in animals and humans (Friedman 2002).

The variation of metabolic compounds is usually studied in tomato either along fruit development or according to environmental perturbations in one accession. Results are subsequently supposed to represent the variation of the species. Genetic variation is also frequently studied in isogenic lines resulting from the modification of a gene in a unique genetic background. Genetic variation was mainly studied in a progeny of introgression lines carrying unique genome fragments of a wild species (Schauer et al, 2006; Steinhauser et al, 2011; Do et al, 2010). Knowledge on the fruit proteome is a challenging area of research, as reviewed by Palma et al (2011) and Faurobert et al (2012). Proteome in tomato fruit is poorly documented, partly because of technical issues. While custom-made or commercial arrays are available for transcript profiling and widely used techniques, like NMR, GCMS or LCMS, are available for metabolite profiling, it is still a technical challenge to obtain quantitative information about large numbers of proteins (Rose et al. 2004; Baerenfaller et al. 2008) . Proteomic in tomato was used to study the response of the plant to several environmental stresses (Iwahashi and Hosoda 2000; Page et al. 2010; Manaa et al. 2011; Marjanovic et al. 2012) and proteome variation along seed development and germination (Sheoran et al. 2005) and along fruit development and ripening (Faurobert et al. 2007). Osorio et al. (2011) analysed fruit proteome of ripening mutants and observed 158 differentially expressed proteins. To our knowledge, this is the first comprehensive description of genetic variation of proteome in tomato fruit pericarp. Here we compared 8 accessions selected to represent a large part of the phenotypic and molecular diversity of *S. lycopersicum* (Ranc et al. 2008) and we explored the inheritance of the fruit traits expressed at different levels.

A large range of variation is detected at the intraspecific level for phenotypic, metabolomic, enzymatic and proteomic profiles

A wide range of variation was observed for all the traits at least at one stage, except for nine enzyme activities for which we detected significant variations among stages but genetic variation was not significant. This is rather due to a low repeatability of measurements, as shown by some high coefficients of variation than to a real lack of variation, as the range of variation among the 12 genotypes was quite high. Steinhauser et al. (2010) also identified several enzyme activities with a low heritability. The ratio of maximum to minimum values was in the range of 2 to 3 for many traits, with a larger range of variation at cell expansion than at orange-red stage for the primary metabolite contents and the contrary for the enzyme activities. The secondary metabolites showed extreme range of variation, particularly the glycoalkaloids and flavonoids that may be present in one line and almost absent in another one.

The Cervil accession is distant from all other lines at the molecular level. It revealed a very specific profile for every trait, leading to most of the extreme values (lowest fruit weight, highest dry matter, sugars and acids contents). It is also specific in terms of secondary metabolites, with high content in chlorogenic acid, dehydrotomatin, rutin and coumaric acid, and lowest content in Penthotenic acid (aPpa). The other accessions have less contrasted profiles, each accession showing high and low trait values.

The range of variation is consistent with those observed in *S. pennellii* introgression lines for primary metabolic compounds by Schauer et al, (2006) and for enzyme activities by Steinhauser et al. (2010). For secondary compounds, a number of compounds are identified in tomato, as reviewed by Slimestad and Verheul (2009). A large range of variation has been reported for most of these compounds, which were not subjected to any conscious selection (Martinez-Valverde et al. 2002).

We identified 424 protein spots that were variable among stages or genotypes across more than 1,000 spots (Xu et al, 2012). Faurobert et al (2007) described the proteome of tomato pericarp in Cervil line along fruit development. They identified 148 variable spots among more than 1700 and could identify the function of 90 spots. The analysis of additional accessions thus largely extended the number of variable spots and the tomato genome sequence allowed the identification of the function of almost 100% of the spots. The ratio of the highest to the lowest amount ranged from 1 to 7 at cell expansion stage and from 1 to 17

at orange-red stage, showing a wide range of variation. As 187 spots revealing protein modifications corresponded to 70 genes, the 424 spots corresponded to a total of 307 unique genes. 190 spots were in average in lower amount at the orange-red stage and 234 were in higher amount. Most (68%) of the 117 spots, which were variable at both stages, showed the same tendency (increase or decrease along fruit development) in all the genotypes. Nevertheless 242 spots revealed significant Genotype by Stage interactions, showing that the trends observed in one genotype at the physiological level may change in another genotype.

The same diversity in the modes of inheritance is observed for every trait levels and crosses

Hybrids are widely used in modern agriculture, either for heterosis (the advantage of hybrid compared to both parents) or for the combination of dominant traits. Agronomical traits often show hybrid vigor or heterosis in the F1 when distant cultivated accessions or cultivated and wild species are crossed (Springer and Stupar 2007; Li et al. 2008; Meyer et al. 2010). The molecular origin of heterosis has been studied for years and is usually related to a combination of dominance or over-dominance effects and to epistatic interactions (Stuber 2010). In tomato, on the contrary to the highly heterotic crops like maize, few traits show a systematic heterosis trend. In our study, plant height was the only trait showing systematic heterosis in every cross. Fruit weight was either recessive or additive and dry matter content was additive in the four crosses.

The analysis of QTL inheritance in introgression lines derived from the wild species *S. pennellii* allowed the study of the molecular basis of heterosis in metabolic traits. Lipman and Zamir (2007) showed that traits related to fitness exhibited a higher rate of dominant or overdominant QTL. Many QTL controlling metabolic traits had a dominant (50%) or overdominant (9%) inheritance, and only 19% of them were additive (Schauer et al. 2008). Steinhauser et al (2011) studied the same population for enzyme activities and a similar proportion of QTL (approximately 30%) showed additive, recessive and dominant modes of inheritance, with only 5% showing overdominance. In our case, it is difficult to predict the inheritance of one trait from only one cross. The number of traits significantly variable within each family (one hybrid and his two parents) differed from one cross to the other in relation to the genetic distance between the parents. A maximum of traits were variable in Cervil x Levovil and its two parental lines. In average, more than 60% (64 % at cell expansion and 62% at orange red stage) of the traits showed an additive inheritance. This rate was a little

higher for enzymes at cell expansion (71%) and lower for metabolites at orange red stage (53%). A number of traits exhibited an over-dominant or over-recessive mode of inheritance, but no specific trend towards an excess of dominance or recessive inheritance could be observed. The higher rate of additivity in our study compared to the study of *S. pennellii* introgression lines may result from the lower distance between the parental lines which are all from the same species. Nevertheless the inheritance mode in one cross was not systematically the same in another cross, suggesting complex and diverse inheritance modes of the traits studied. Among the 424 protein spots, 163 were variable at both stages and 70 spots varied at only one stage but in at least 2 crosses. Inheritance mode appeared relatively independent from one stage to the other as 70% of the spots revealed at least two different modes of inheritance in the first group and 55% in the second.

Recent heterosis models focus on regulatory networks derived from systems biology approaches. De Vienne et al. (2001) proposed a metabolic heterosis model, suggesting that heterosis may results from an optimal combination of enzyme quantities leading to maximal metabolic fluxes in hybrids. Goff (2011) underlined the role of protein synthesis and degradation in heterosis. In roots from maize inbred and hybrid lines, the metabolite contents of the hybrids display lower variability than the metabolite contents of inbred lines, suggesting the presence of optimal metabolite levels in the hybrid population (Lisec et al. 2011). In our case, we could not evidence differences in the variability of metabolic traits in hybrids compared to the parents.

Systems approach of the pericarp variability reveals complex connectivity among different levels of analysis

We dissected the variation of pericarp composition at several levels from phenotypes to proteome profiles. As we studied eight unrelated accessions, correlations may reveal the effect of polymorphic gene acting on two related traits, but also fortuitous associations. For instance correlation between the composition in a specific compound and plant height may not be due to a causal relationship but to linkage disequilibrium. Nevertheless the closer the functions the most meaningful may be the relationships observed. Osorio et al, (2011) in a similar approach describes covarying genes or proteins as “guilty by association”.

The enzyme activities we assessed correspond to Vmax, and may thus mainly reflect the corresponding protein amount (Steinhauser et al, 2010). One might hypothesizes that proteins and enzyme activities may have simpler genetic control than metabolites or phenotypic traits.

A polymorphism in a gene coding for an enzyme or in its regulation sequences may directly causes a variation in its amount. The protein amount can also be modulated by trans acting factors acting on its synthesis, its stability or its degradation. Metabolites or phenotypic traits result from many more processes and pathways and their variation could be more complex. Actually, it was shown in tomato (Steinhauser et al. 2011), maize (Huang et al. 2010) and Arabidopsis (Keurentjes et al. 2008b) that the heritability of enzyme activities was often lower than metabolite heritability and even for high heritability traits, most of the QTL were acting in trans, corresponding to regulating factors. Steinhauser et al. (2011) could nevertheless identify a few QTLs which colocalised with their structural gene.

Each group of traits is strongly connected

Cell expansion and orange-red stages represent very distinct physiological processes (Gillaspy et al. 1993; Giovannoni 2004; Faurobert et al. 2007) and this difference is also observed in the very different patterns of protein expression, of metabolite content as well as in the differences in networks of correlations. This suggests that the genes controlling these expressions are at least partly different. Much more significant correlations were detected than expected by chance. Some were expected based on earlier results. For instance, several amino acids or sugars (fructose and glucose) varied in a coordinated manner, particularly at the cell expansion stage, as already shown in several studies (Schauer et al. 2008; Do et al. 2010). Very few correlations were observed between primary and secondary metabolisms at the early stage, while they were more frequent at orange-red stage. On the contrary to metabolism, enzyme activities showed a higher number of correlations at the orange-red stage. The majority of positive correlations among enzyme activities and among metabolites suggest a coordinated regulation.

Correlations between proteins were less skewed in favour of positive correlations. A clear stage effect was shown corresponding to the coordinated fruit development. Faurobert et al. (2007) also observed such a clear evolution in the pattern of proteins differentially along fruit development. Nevertheless, we could not establish clear links between the functional classification of protein spots and the groups of correlated proteins. Many correlations were detected across functional classes.

Relationships between enzyme activities and the corresponding spot amounts

We measured the Vmax for enzyme activities, which are thus expected to be correlated to the enzyme content. We found a few correlations between enzymes and protein spot amount corresponding to the same function at each stage. They involved invertase, enolase, fructokinase, MDH and malic enzyme. The absence of correlation may be related to (1) enzyme activities that may result from a combination of several proteins (subunits) and not all the proteins corresponding to the enzymes were detected; (2) most of the primary metabolism enzymes result from multigene families, which may be highly specific to one cell compartment. Protein spots may also be the product of complex posttranslational modifications (Faurobert et al. 2007). This could explain some negative correlations, when a spot is for instance modified in an inactive form.

Network analyses among different levels of expression reveal complex patterns of connectivity

Networks among groups of traits were constructed using sPLS regression (Le Cao et al. 2008). One problem when dealing with omic data is that the number of traits is much larger than the number of samples. Sparse methods were developed for dealing with high-dimensional data. The main advantage of sparse methods over non-spare methods is that they set the contribution of noise variables to zero and thus improve the prediction or classification performance (Filzmoser et al. 2012). Most of the networks between two levels of expression showed complex patterns of connectivity, relating several nodes together and different pathways or metabolisms. In each network, a few hubs could be identified relating many different compounds or proteins.

The key role of a few proteins could be underlined

The main role of a few key proteins and enzymes could be underlined. The role of invertase in sucrose breakdown has already been documented (Faurobert et al. 2007). We detected seven spots corresponding to this enzyme, most of them corresponding to posttranslational modifications. One spot was strongly related to the content in sugars at orange-red stage, while at cell expansion the amount of another spot was correlated to GABA and Asn content. The role of a polymorphism in the promoter of this gene in fruit sugar content has been recently shown (Moy et al. 2012). Several correlations with enzyme activity and metabolite contents involved one of the protein spots coding for fructokinase, an enzyme participating to the sugars phosphorylation. Fructokinase may play a role in sugar import and in starch

biosynthesis (Dai et al. 2002). Several isoforms were detected, both being correlated with the variation of sugars. A QTL controlling sugar content was also detected in the region of FRK3 in the Cervil x Levovil progeny and could be related to the correlations observed here. Fine mapping experiments are underway to confirm this relationship. Finally another interesting correlation involved a lactoyl gluthatione lyase whose amount was correlated to fruit weight. In an association study, Ranc et al. (2012) detected an association between FW and the gene coding for this protein. This gene also colocalized with a QTL for FW variation in the mapping population derived from Cervil x Levovil cross (Saliba-Colombani et al. 2001). The putative impact of this protein on plant cell proliferation (Paulus et al. 1993), together with these associations, confirms its role as an important candidate.

Protein amounts or enzyme activities and related metabolisms

Some metabolites showed significant correlations with compounds in the same pathway and with other outside of their pathways. This is consistent with the observations of Carrari et al. (2006) and Osorio et al. (2011). We limited the sPLS analysis to the relationships with spots related to primary metabolism and vitamins. This approach should thus be extended to the other spots as a number of proteins related to stress (heat shock proteins), to protein degradation (ubiquitin) or to protein conformation (chaperon) appeared correlated with several metabolic traits.

Conclusion

Finally, the systems biology approach combining proteome, metabolome and phenotypic analysis gave insights into the dioversity and relationships at easch level. Each level revealed the same complexity in terms of variation, of inheritance mode and connectivity. In parallel, we have studied the transcriptome profiles of the 12 genotypes. The eight parental lines were also resequenced and more than 4 million SNPs identified. In the near future, these data will be combined to test the polymorphisms in the enzymes and key proteins detected here. This will allow us to integrate every step from genome to phenotypes.

This study is the starting point of a broader experiment including the development of a MAGIC population derived from the eight parental lines. The MAGIC population will be used to map QTL. The knowledge of the parental lines will thus be useful to relate QTLs to the parental variations.

Acknowledgement:

We thank Caroline Callot and Karine Leyre for technical help. We also thank Esther Pelpoir for her help in fruit sampling and Yolande Carretero for taking care of the plants in the greenhouse. This work was supported by ANR Genomic project MAGICTomSNP project. Jiaxin Xu was supported by China Scholarship Council.

SUPPLEMENTAL DATA:

Supplemental Figure S1. Eight divergent lines and four F_1 used in this study.
A: Four F_1 derived from the crosses of the eight lines; B: fruits from the eight lines.

Supplemental Figure S2. Un-rooted neighbor joining tree of the eight parental lines and Heinz1706 based on 139 SNP markers

Supplemental figure S3 Fruit fresh weight increase of the four crosses during fruit ageing. Curves were fitted to three-parameter Sigmoid logistic function.

Supplemental Figure S4. Distribution of average coefficient of variation of 34 metabolites (A), 26 enzyme activities (B) and 424 protein spot amounts (C) in 4 *S. lycopersicum* (blue), 4 *S. l. cerasiforme* (dark red) and 4 hybrids (green) at cell expansion (CE) and orange red (OR) stage.

Supplemental Figure S5 Correlation among the amounts of 101 protein spots at orange red stage.
Based on variation analysis, a total of 117 protein spots were significantly different among genotypes at both cell expansion and orange-red stage. These protein spots were then grouped according to their functional classes. Seven groups (which have more than 5 protein spots) with a total of 101 protein spots were kept for representing the correlations among protein spots. Only significant correlations ($P<0.01$) are shown. Positive and negative correlations are presented in red and green, respectively.

Supplemental figure S6. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and 26 enzyme activities at cell expansion stage. Metabolites, phenotypic traits, and enzymes were represented as different nodes (circle for metabolites and phenotypic traits, square for enzymes). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown (($|r|>0.71$, $p<0.01$).

Supplemental figure S7. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and the amounts of 60 protein spots at orange red stage. The 60 protein spots were significant among genotypes at cell expansion stage and were classified to primary metabolic (58 spots) and vitamin (2 spots) synthesis classes. Metabolites, phenotypic traits, and protein spots were represented as different nodes (circle for metabolites and phenotypic traits, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($|r|>0.71$; $p<0.01$). JX004: Malate dehydrogenase; JX024:Adenosine kinase; JX051: acid invertase; JX059: Succinyl-CoA ligase; JX091: Malic enzyme; JX110: Phosphoglucomutase; JX111: Inorganic pyrophosphatase; JX154: Malate dehydrogenase; JX161: Enolase; JX164: Inorganic pyrophosphatase family protein : JX178: Pyruvate

dehydrogenase E1 component alpha subunit: JX239: Acid beta-fructofuranosidase ; JX256: Enolase; JX262: Alpha-mannosidase; JX286: Lipoxygenase homology domain-containing protein 1 JX290: LL-diaminopimelate aminotransferase; JX291: Lactoylglutathione lyase; JX292: 2-isopropylmalate synthase 1; JX323: Enolase; JX333: Ribulose-5-phosphate-3-epimerase; JX351: Ubiquilin-1; JX367: Phosphoribulokinase/uridine kinase; JX372: Diaminopimelate epimerase family protein; JX394: Adenylate kinase

Supplemental Figure S8. Network analysis based on sPLS analysis of 26 enzyme activities and the amounts of 60 protein spots at orange red stage. The 60 protein spots were significant among genotypes at orange red stage and were classified to primary metabolic (58 spots) and vitamin (2 spots) synthesis classes. Enzymes and protein spots were represented as different nodes (circle for enzyme, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ((|r|>0.71; p<0.01). JX004: Malate dehydrogenase; JX051: Acid beta-fructofuranosidase; JX091: Malic enzyme ; JX132: Oxalyl-CoA decarboxylase; JX178: Pyruvate dehydrogenase E1 component alpha subunit; JX180: Pyridoxal biosynthesis lyase; JX214: Enolase; JX394: Adenylate kinase

Supplemental Figure S9. Network analysis based on sPLS analysis of 26 enzyme activities at cell expansion stage and 34 metabolites, 5 phenotypic traits and at orange red stage. Metabolites, phenotypic traits, and enzymes were represented as different nodes (circle for enzymes, square for metabolites and phenotypic traits). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ((|r|>0.71, p<0.01).

Supplemental Table S1. List of 24 primary metabolites and 11 secondary metabolites analyzed, with abbreviations and assignment to their metabolic process

Supplemental Table S2. List of the enzymes analysed, with abbreviations and assignment to their respective pathways

Supplemental table S3. Polymorphism rate between eight parental lines and Heinz1706 with 139 SNP. Parental lines from the same cross are indicated in the same color

Supplemental table S4A Mean values of 34 metabolite content, phenotypic traits and enzymatic activities at cell expansion stage and orange red stage

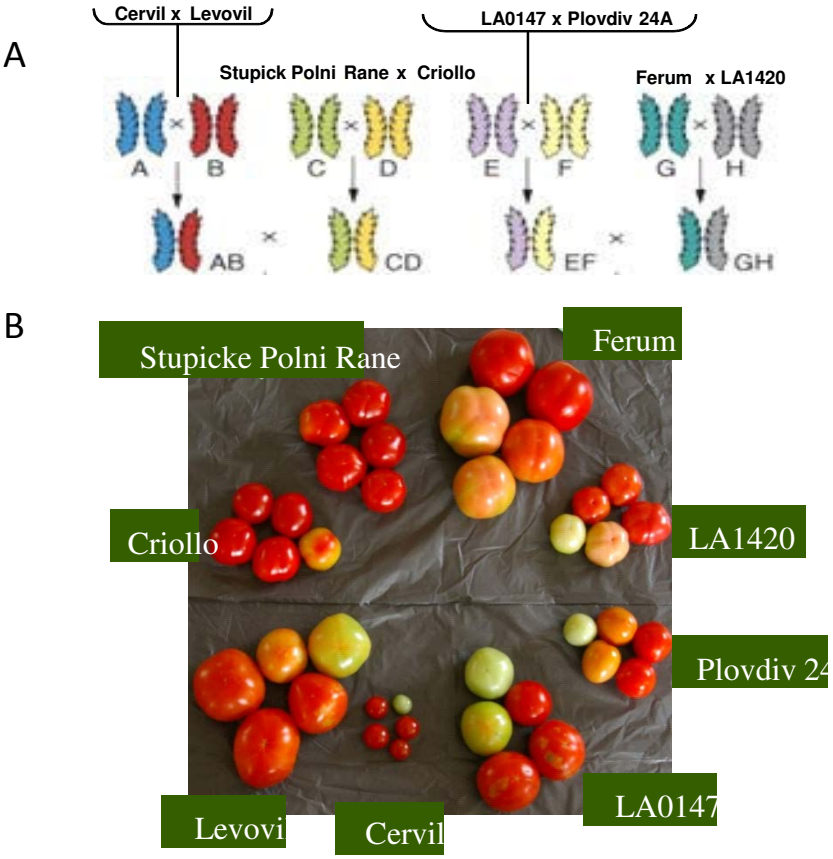
Supplemental table S4B Mean values of the smounts of 256 and 188 significant protein spots among genotypes at cell expansion stage and orange red stage, respectively

Supplemental table S5 Function description and variation analysis of 424 protein spots among 12 genotypes at cell expansion stage (CE) and orange red (OR) stage. Significant spots are highlight in yellow.

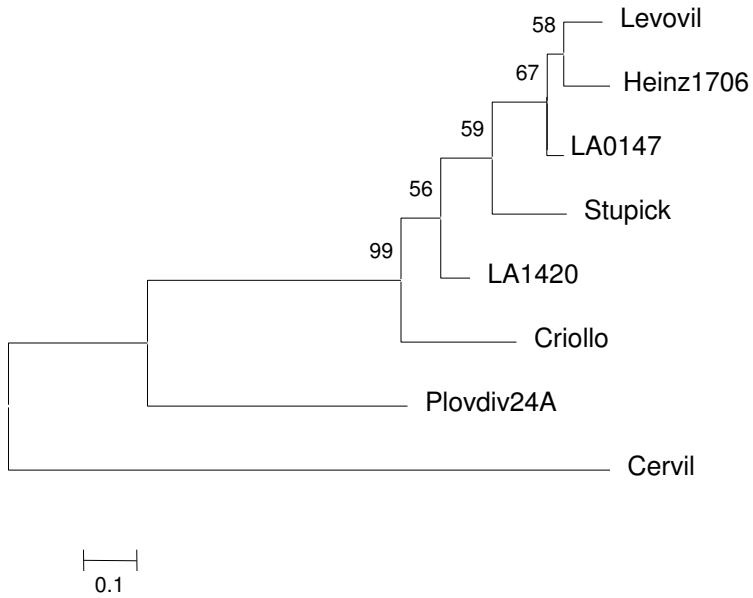
Supplemental table S6 CV within each genotype and in average at cell expansion stage (CE) and orange red stage (OR). CVs in red (0.2<CV<0.3), in green (CV>0.3)

Supplemental table S7A Pairwise pearson coefficient R values and pvalues of 34 metabolites, 5 phenotypic traits, 26 enzyme activities and 256 significant protein spots at cell expansion stage *(too large to be included in the thesis)*

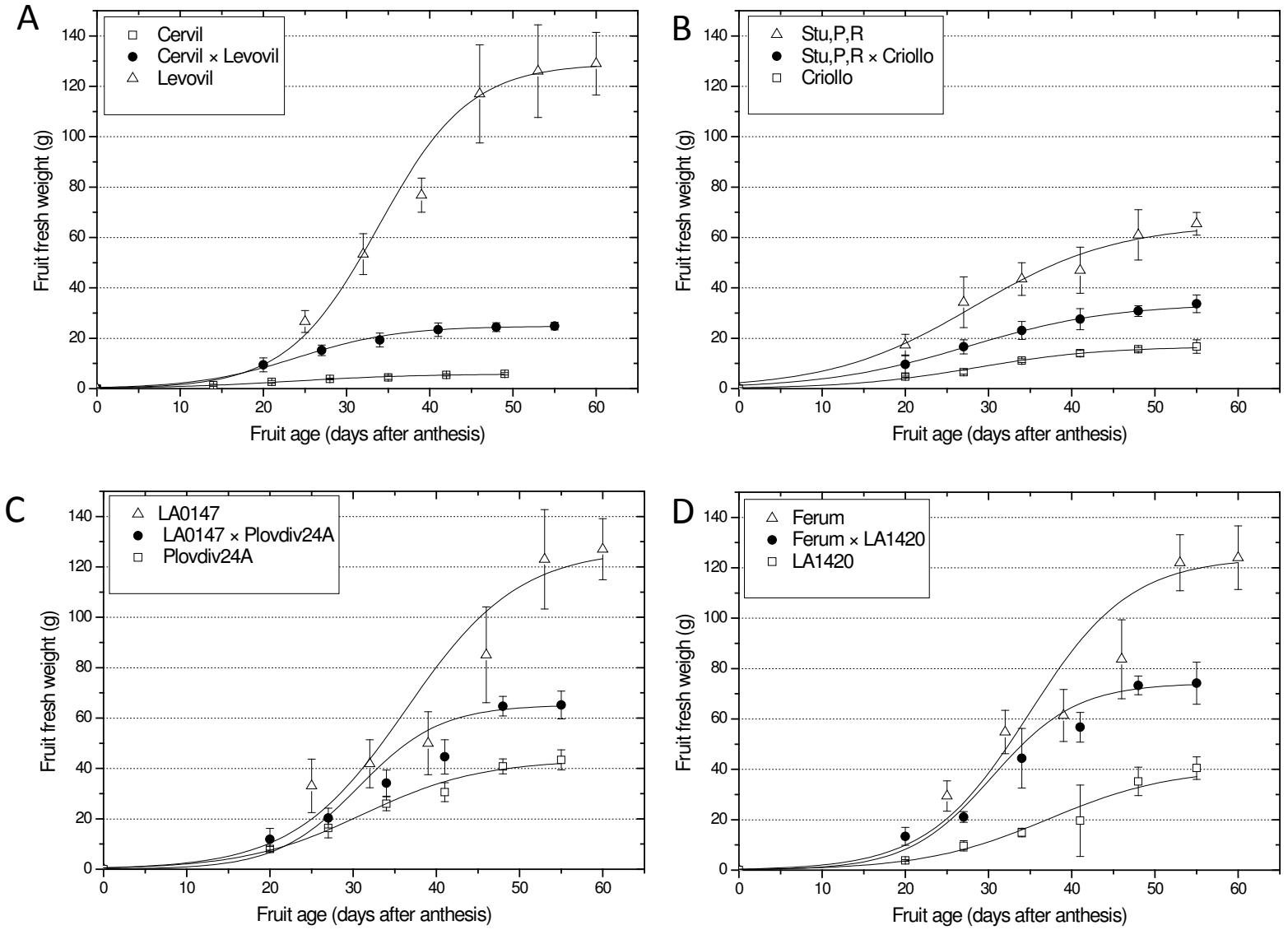
Supplemental table S7B Pairwise pearson coefficient R values and pvalues of 34 metabolites, 5 phenotypic traits, 26 enzyme activities and 188 significant protein spots at orange red stage *(too large to be included in the thesis)*



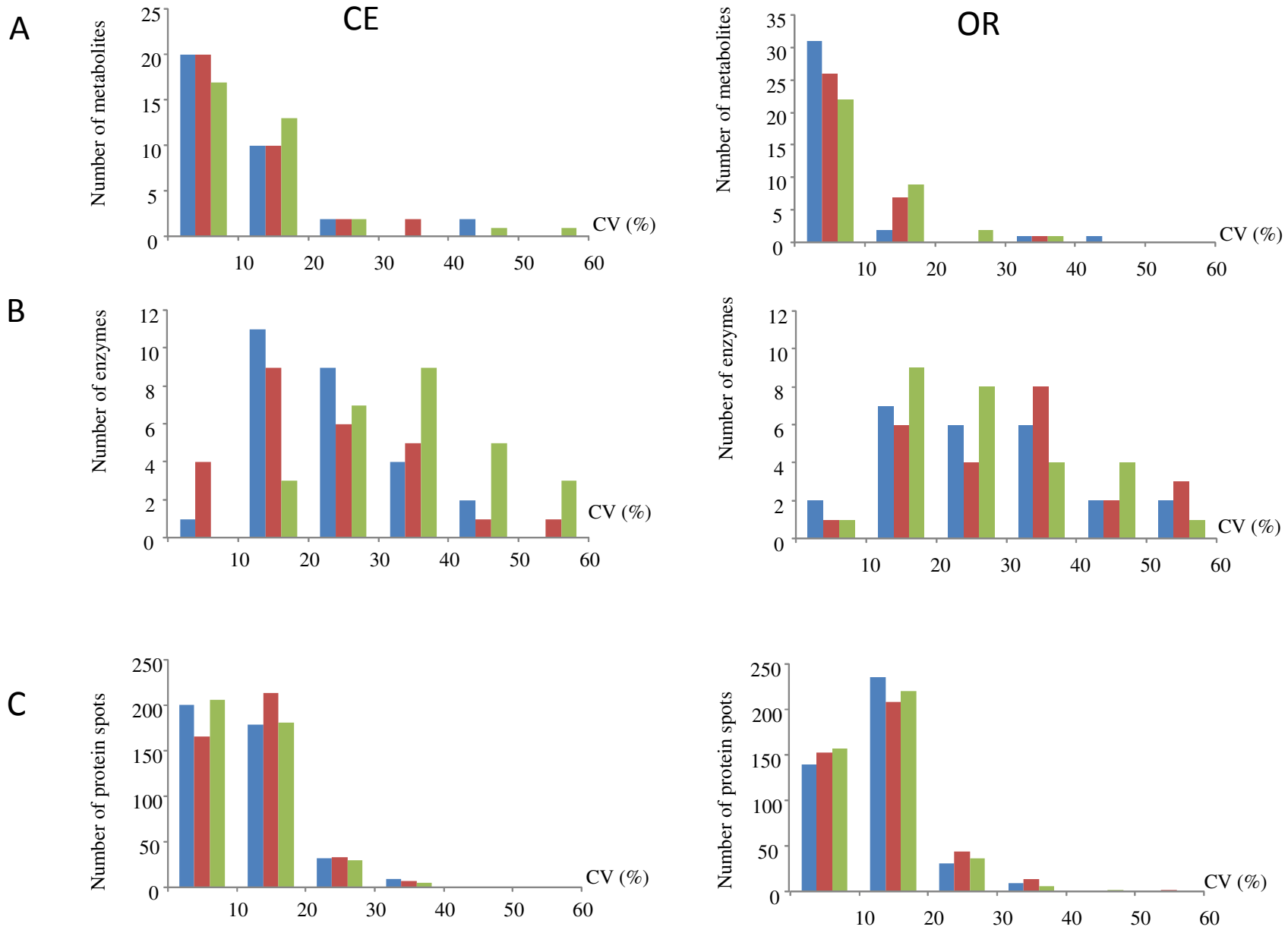
Supplemental Figure S1. Eight divergent lines and four F_1 used in this study. A: Four F_1 derived from the crosses of the eight lines; B: fruits from the eight lines.



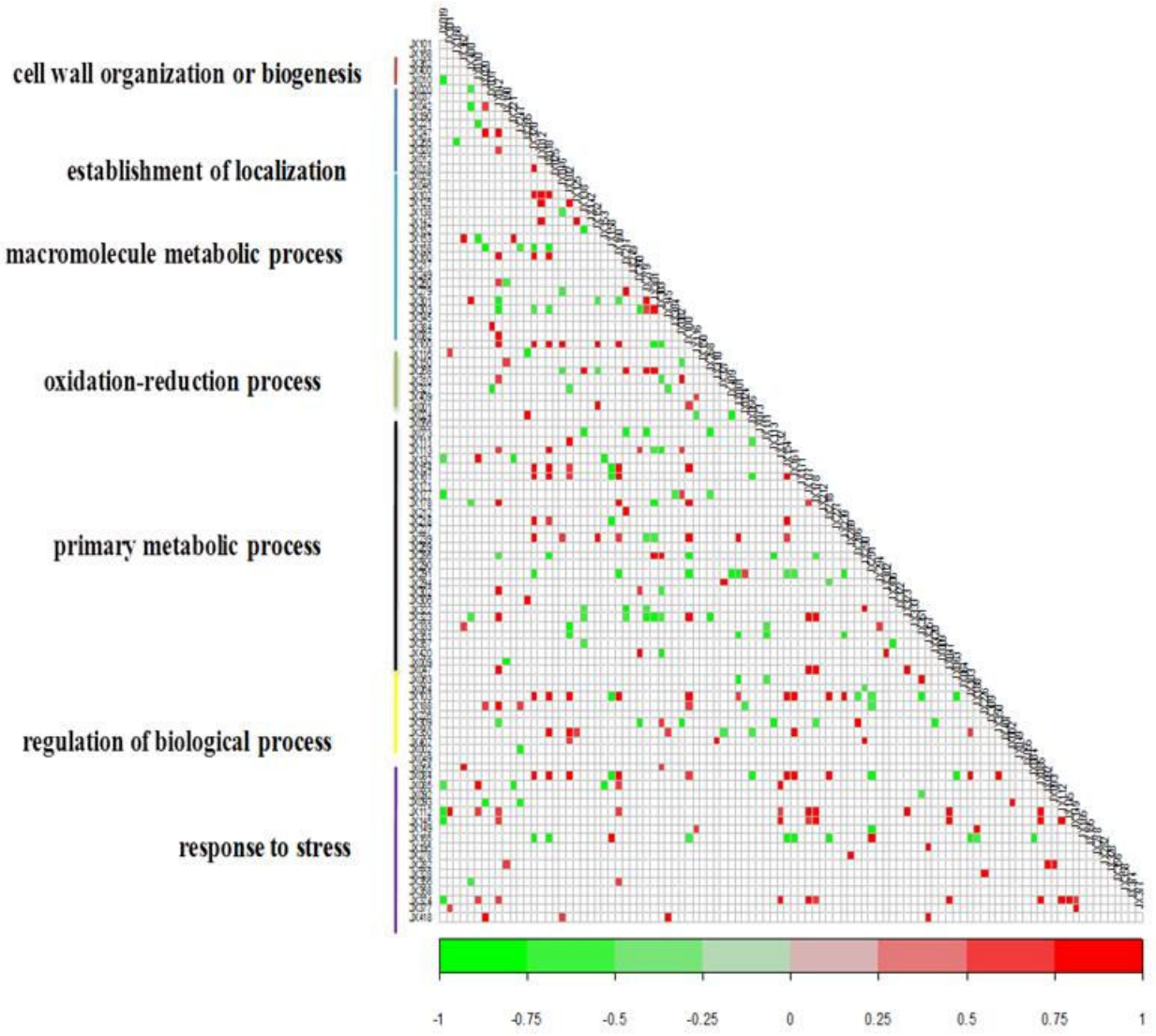
Supplemental Figure S2. Un-rooted neighbor joining tree of the eight parental lines and Heinz1706 based on 139 SNP markers



Supplemental figure S3 Fruit fresh weight increase of the four crosses during fruit ageing. Curves were fitted to three-parameter Sigmoid logistic function.

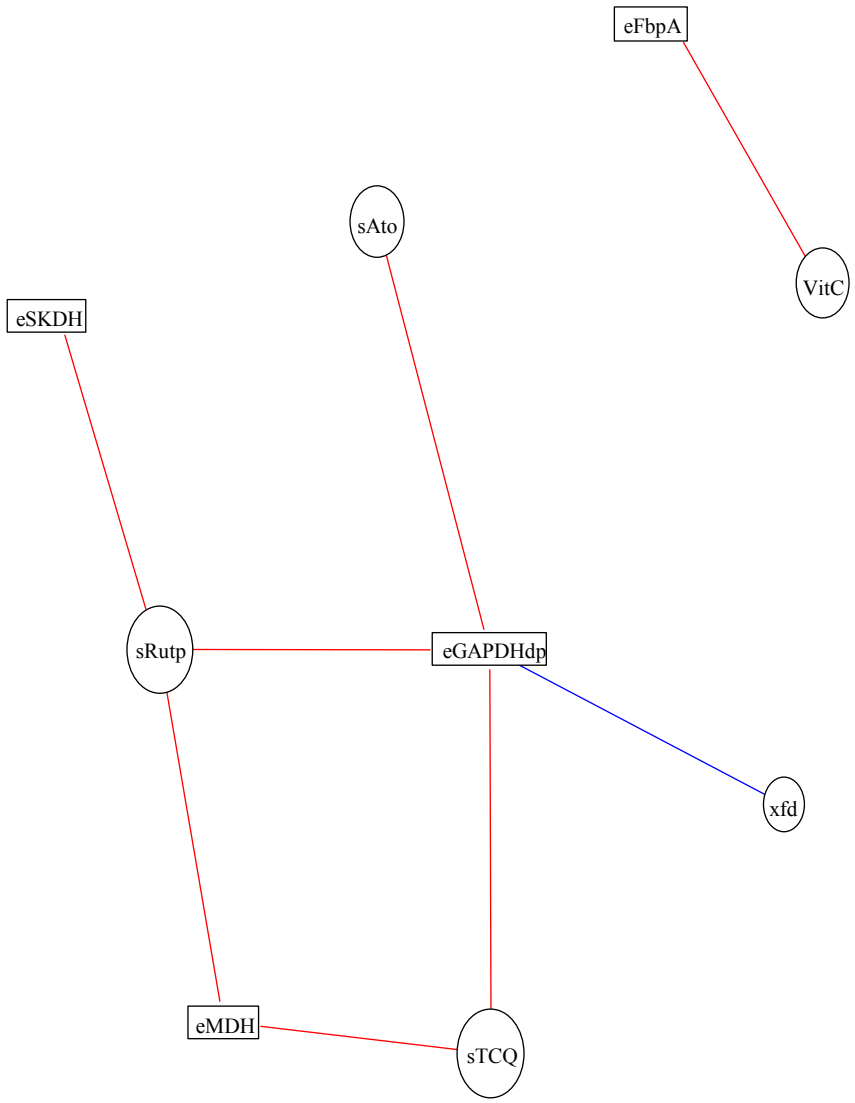


Supplemental Figure S4. Distribution of average coefficient of variation of 34 metabolites (A), 26 enzyme activities (B) and 424 protein spot amounts (C) in 4 *S. lycopersicum* (blue), 4 *S. l. cerasiforme* (dark red) and 4 hybrids (green) at cell expansion (CE) and orange red (OR) stage.

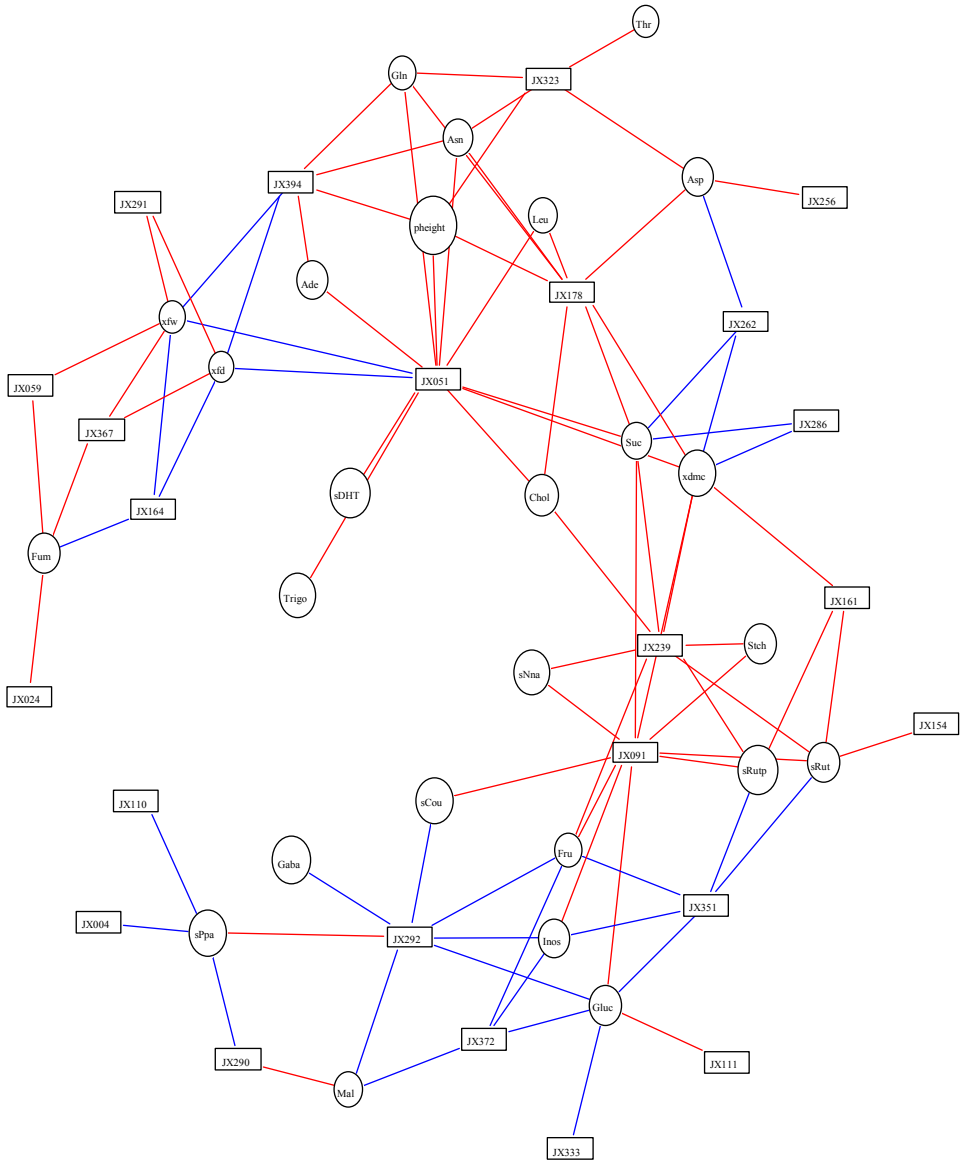


Supplemental Figure S5 Correlation among the amounts of 101 protein spots at orange red stage.

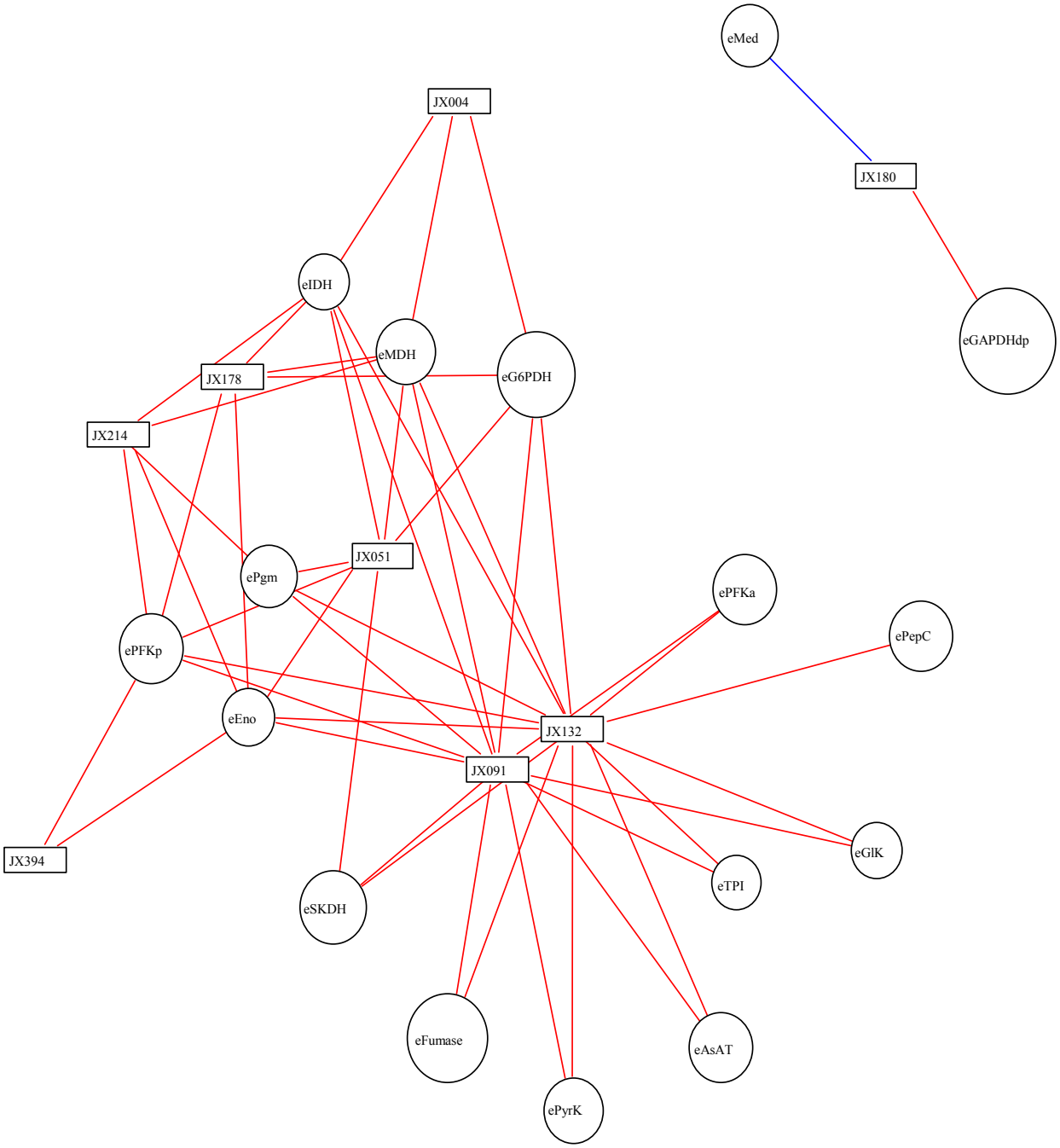
Based on variation analysis, a total of 117 protein spots were significantly different among genotypes at both cell expansion and orange-red stage. These protein spots were then grouped according to their functional classes. Seven groups (which have more than 5 protein spots) with a total of 101 protein spots were kept for representing the correlations among protein spots. Only significant correlations ($P<0.01$) are shown. Positive and negative correlations are presented in red and green, respectively.



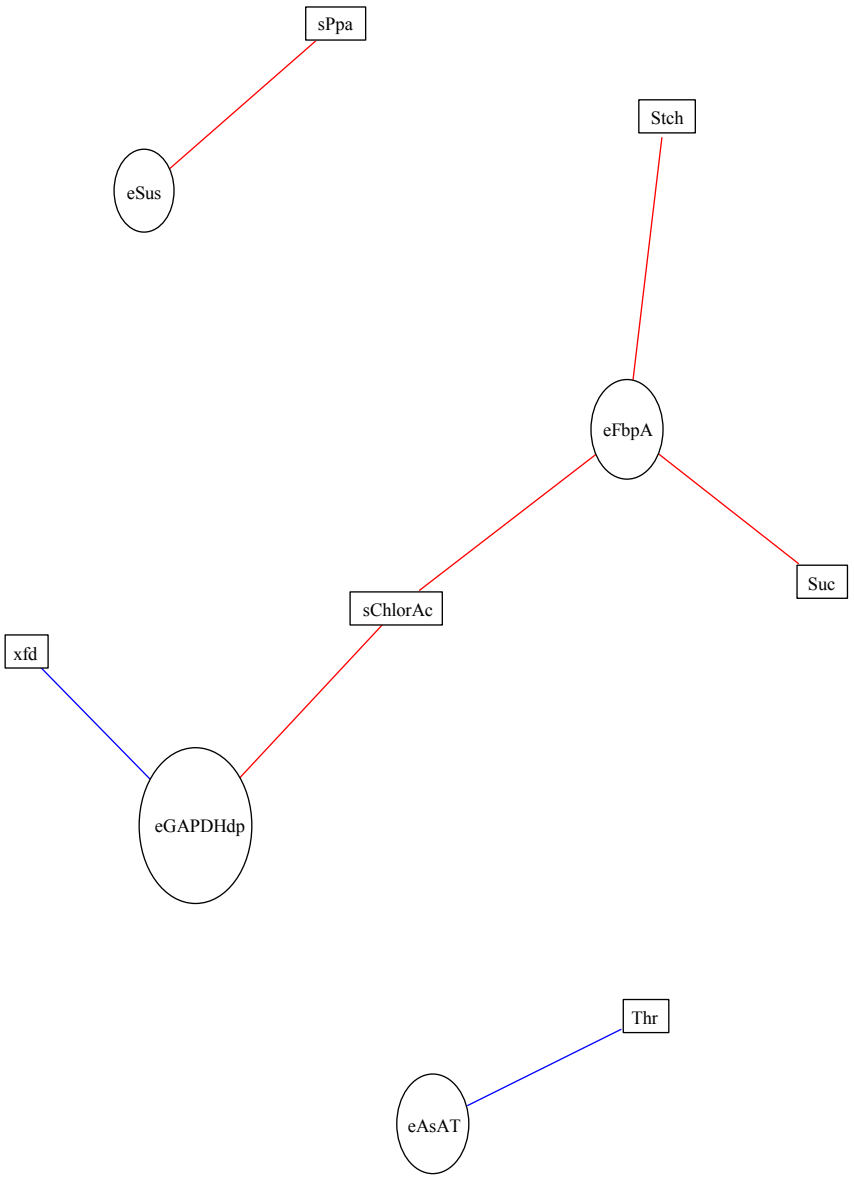
Supplemental figure S6. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and 26 enzyme activities at cell expansion stage. Metabolites, phenotypic traits, and enzymes were represented as different nodes (circle for metabolites and phenotypic traits, square for enzymes). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($|r|>0.71$, $p<0.01$).



Supplmental figure S7. Network analysis based on sPLS analysis of 34 metabolites, 5 phenotypic traits and the amounts of 60 protein spots at orange red stage. The 60 protein spots were significant among genotypes at cell expansion stage and were classified to primary metabolic (58 spots) and vitamin (2 spots) synthesis classes. Metabolites, phenotypic traits, and protein spots were represented as different nodes (circle for metabolites and phenotypic traits, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($|r|>0.71$; $p<0.01$). JX004: Malate dehydrogenase; JX024:Adenosine kinase; JX051: acid invertase; JX059: Succinyl-CoA ligase; JX091: Malic enzyme; JX110: Phosphoglucomutase; JX111: Inorganic pyrophosphatase; JX154: Malate dehydrogenase; JX161: Enolase; JX164: Inorganic pyrophosphatase family protein : JX178: Pyruvate dehydrogenase E1 component alpha subunit: JX239: Acid beta-fructofuranosidase ; JX256: Enolase; JX262: Alpha-mannosidase; JX286: Lipoxygenase homology domain-containing protein 1 JX290: LL-diaminopimelate aminotransferase; JX291: Lactoylglutathione lyase; JX292: 2-isopropylmalate synthase 1; JX323: Enolase; JX333: Ribulose-5-phosphate-3-epimerase; JX351: Ubiquilin-1; JX367: Phosphoribulokinase/uridine kinase; JX372: Diaminopimelate epimerase family protein; JX394: Adenylate kinase



Supplmental Figure S8. Network analysis based on sPLS analysis of 26 enzyme activities and the amounts of 60 protein spots at orange red stage. The 60 protein spots were significant among genotypes at orange red stage and were classified to primary metabolic (58 spots) and vitamin (2 spots) synthesis classes. Enzymes and protein spots were represented as different nodes (circle for enzyme, square for protein spots). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown (($|r|>0.71$; $p<0.01$). JX004: Malate dehydrogenase; JX051: Acid beta-fructofuranosidase; JX091: Malic enzyme ; JX132: Oxalyl-CoA decarboxylase; JX178: Pyruvate dehydrogenase E1 component alpha subunit; JX180: Pyridoxal biosynthesis lyase; JX214: Enolase; JX394: Adenylate kinase



Supplemental Figure S9. Network analysis based on sPLS analysis of 26 enzyme activities at cell expansion stage and 34 metabolites, 5 phenotypic traits and at orange red stage. Metabolites, phenotypic traits, and enzymes were represented as different nodes (circle for enzymes, square for metabolites and phenotypic traits). Correlations were represented as links that connect the nodes. Positive and negative correlations are presented in red and blue, respectively. Only significant correlations are shown ($(|r|>0.71, p<0.01)$).

Supplemental Table S1. List of 24 primary metabolites and 11 secondary metabolites analyzed, with abbreviations and assignment to their metabolic process

Metabolites	Abbreviation	Functional class
glucose	Gluc	Sugars
sucrose	Suc	Sugars
fructose	Fru	Sugars
inositol	Inos	Sugar alcohol
alanine	Ala	Amino acids
asparagine	Asn	Amino acids
aspartate	Asp	Amino acids
GABA	Gaba	Amino acids
glutamine and pyroglutamine	Gln	Amino acids
isoleucine	Ileu	Amino acids
leucine	Leu	Amino acids
phenylalanine	Phe	Amino acids
tyrosine	Tyr	Amino acids
valine	Val	Amino acids
threonine	Thr	Amino acids
ascorbic acid	VitC	Vitamins
citrate	Cit	Organic acids
malate	Mal	Organic acids
fumarate	Fum	Organic acids
starch	Stch	Starch
trigonelline	Trigo	others
adenosine_like	Ade	others
choline	Chol	others
Alpha-tomatine	sAto	secondary metabolites - glycoalkaloids
Dehydrotomatine	sDHT	secondary metabolites - glycoalkaloids
OH-lycoperoside A B C	sOHl	secondary metabolites - glycoalkaloids
Chlorogenic acid (5CQA)	sChl	secondary metabolites - phenolic compounds
Coumaric acid hexose/coumaroyl hexose	sCou	secondary metabolites - phenolic compounds
Crypto-chlorogenic acid (3CQA)	sCry	secondary metabolites - phenolic compounds
Naringenin chalcone	sNna	secondary metabolites - phenolic compounds
Rutin	sRut	secondary metabolites - phenolic compounds
Rutin-pentoside	sRutp	secondary metabolites - phenolic compounds
Tricaffeoylquinic acid (TCQ)	sTCQ	secondary metabolites - phenolic compounds
Pentothenic acid	sPpa	secondary metabolites -Vitamin B5

Supplemental Table S2. List of the enzymes analysed, with abbreviations and assignment to their respective pathways

Enzyme full names	Abbreviation	EC number	Pathway
Alanine aminotransferase	eAlAT	2.6.1.2	Amino acids metabolism
Aspartate aminotransferase	eAsAT	2.6.1.1	Amino acids metabolism
Shikimate dehydrogenase	eSKDH	1.1.1.25	Amino acids metabolism
phosphoglycerate kinase	ePGK	2.7.2.3	Calvin cycle
Triose Phosphate Isomerase	eTPI	5.3.1.1	Calvin cycle
Enolase	eEno	4.2.1.11	Glycolysis
Glucose-6-phosphate dehydrogenase	eG6PDH	1.1.1.49	Glycolysis
Glyceraldehyde-3-phosphate dehydrogenase	eGAPDHd	1.2.1.12	Glycolysis
Glyceraldehyde-3-phosphate dehydrogenase (NAD(P))	eGAPDHdp	1.2.1.12	Glycolysis
Isocitrate dehydrogenase (NADP+)	eIDH	1.1.1.41	Glycolysis
Phosphoenolpyruvate carboxylase	ePepC	4.1.1.31	Glycolysis
Phosphofructokinase (ATP)	ePFKa	2.7.1.11	Glycolysis
Phosphofructokinase (PPi)	ePFKp	2.7.1.11	Glycolysis
Pyruvate kinase	ePyrK	2.7.1.40	Glycolysis
Fructokinase	eFRK	2.7.1.4	Sucrose breakdown
Glucokinase	eGlK	2.7.1.2	Sucrose breakdown
Acid Invertase	eInvA	3.2.1.26	Sucrose breakdown
Neutral Invertase	eInvN	3.2.1.26	Sucrose breakdown
Sucrose Synthase	eSus	2.4.1.13	Sucrose breakdown
Fructose-bisphosphate aldolase	eFbpA	3.1.3.11	Sucrose synthesis
Phosphoglucomutase	ePgm	5.4.2.2	Sucrose synthesis
Aconitase	eAco	4.2.1.3	TCA cycle
Fumarase	eFumase	4.2.1.2	TCA cycle
Malate dehydrogenase (NAD)	eMDH	1.1.1.37	TCA cycle
Malic enzyme (NAD)	eMed	1.1.1.37	TCA cycle
Malic enzyme (NADP+)	eMedp	1.1.1.37	TCA cycle

Supplemental table S3. Polymorphism rate between eight parental lines and Heinz1706 with 139 SNP. Parental lines from the same cross are indicated in the same color

Parental line	Cervil	Levovil	Stupick	Criollo	LA0147	Plovdiv24A	Ferum	LA1420	Heniz1706
Cervil									
Levovil	82%								
Stupick	64%	21%							
Criollo	65%	27%	34%						
LA0147	78%	11%	16%	25%					
Plovdiv24A	45%	44%	39%	43%	40%				
Ferum	71%	28%	20%	37%	20%	42%			
LA1420	71%	20%	26%	24%	15%	40%	27%		
Heniz1706	83%	12%	22%	28%	6%	45%	22%	19%	

Supplemental table S4A Mean values of 34 metabolite content, phenotypic traits and enzymatic activities at cell expansion stage and orange red stage

Traits	LevCE	CerLevCE	CerCE	StuCE	StuCriCE	CriCE	LA0147CE	LA0147PloCE	PloCE	FerCE	FerLA147CE	LA1420CE
Phenotypic traits												
xfw	29,2277	9,2899	1,23136	17,9386	9,68974	4,72563	29,2079	11,4504	7,66992	29,4716	13,2734	3,78883
xfd	41,302	26,872	13,838	32,2667	28,764	22,66	41,806	27,84	24,028	39,85	30,522	21,25
xdmc	6,152	8,34465	10,469	6,71188	7,124	7,74868	7,49878	7,66589	8,69254	8,19792	6,8204	7,32517
pheight	82,5	107,5	88,5	90,63	112	108,5	74,5	89	80	72,5	81,5	63
pstem	13,48	15,32	13,31	12,75	14,75	12,5	14,64	14,94	11,8	16,02	16,8	14,45
Primary metabolites												
Gluc	11941,2	10445,7	6968,6	10719,6	11294,0	9607,8	13168,2	13158,4	13125,0	12655,8	11074,3	11616,8
Suc	469,9	836,4	873,6	636,7	621,2	497,3	619,0	704,9	724,6	961,6	751,9	603,6
Fru	12264,4	10227,4	6651,2	10897,3	10890,8	9415,1	13041,8	12472,8	11815,6	12150,7	10891,3	11491,5
Inos	233,3	320,9	496,7	335,7	294,5	293,8	338,7	365,8	516,4	501,8	373,8	329,2
Ala	35,5	71,6	137,0	39,1	36,2	49,3	27,5	23,3	39,9	41,9	35,6	21,1
Asn	83,9	191,3	248,7	104,3	95,8	133,6	77,7	62,6	107,5	107,0	86,9	55,2
Asp	38,2	76,5	129,1	66,7	45,9	56,7	46,3	48,2	68,3	76,7	53,2	42,5
Gaba	381,1	714,5	577,5	381,7	458,9	370,7	317,6	324,7	392,6	412,5	327,3	332,3
Gln	635,6	1939,1	2652,7	870,8	823,3	1024,5	725,1	532,0	1048,5	998,4	872,8	438,1
Ileu	24,9	62,1	87,4	31,7	25,3	43,6	12,7	18,8	64,3	35,1	24,9	13,4
Leu	27,8	61,3	93,7	33,0	29,9	47,4	23,0	26,9	55,9	45,1	35,5	23,4
Phe	125,4	194,5	246,3	167,9	90,8	136,6	90,7	192,8	438,8	184,3	140,3	111,8
Tyr	25,7	34,9	76,8	35,6	12,5	22,3	11,5	18,9	52,0	25,5	17,2	16,7
Val	35,3	82,4	101,4	42,4	37,6	43,9	21,9	26,3	65,0	46,9	29,5	15,2
Thr	25,3	62,8	67,5	35,1	20,7	31,9	13,0	14,0	30,7	35,7	22,2	9,2
VitC	8,4	13,5	22,6	11,0	11,5	11,6	10,0	13,5	11,7	14,2	13,1	9,7
Cit	1622,4	3118,3	4377,7	3382,5	3943,7	4840,3	2193,1	2571,9	3189,9	2964,0	3595,6	3713,7
Mal	1302,1	1351,5	1634,2	1794,0	965,7	1089,6	1695,8	1190,8	1451,2	1988,5	1230,8	1078,4
Fum	0,61	0,49	0,92	0,59	0,39	0,71	0,89	0,81	0,81	0,91	0,78	1,37
Stch	1,55	5,26	6,88	1,22	2,89	3,31	2,36	3,44	7,56	4,52	4,27	2,90
Trigo	20,8	40,4	60,3	22,7	32,1	67,6	16,5	21,0	23,8	21,6	32,2	48,3
Ade	6,71	8,71	10,27	6,21	6,65	6,75	6,79	7,53	7,88	9,09	8,81	7,05
Chol	57,2	85,8	74,5	55,2	61,4	64,0	44,5	56,5	75,4	50,1	41,1	47,5
Secondary metabolits												
sChlorAc	4106,3	8752,9	11257,3	4751,3	5075,7	5472,8	5345,0	5277,7	6065,5	4172,1	4036,5	5553,6
sTCQ	50,0	103,6	127,7	81,8	62,7	106,4	105,6	98,5	78,8	52,4	70,8	124,3
sAto	60665	320814	460157	174380	224930	318493	172926	222135	239711	115516	163918	192061
sCry	892,1	680,7	867,3	1372,5	1056,4	766,8	1569,5	1283,7	489,7	1009,0	1096,9	484,2
sDHT	12130	164198	267130	39932	46059	89555	39575	48047	49172	22434	33946	47547
sRutp	1459,0	3706,2	5146,7	3067,5	2486,4	2415,2	2927,9	4494,7	4660,9	1743,5	2983,1	4330,3
sPpa	784,6	546,6	673,4	649,4	434,9	786,2	803,9	679,5	663,6	778,8	641,4	621,9
sCou	1276,5	2199,4	3088,2	1770,5	1280,3	2057,4	2104,9	1593,7	2016,7	746,9	890,0	1913,2
sNna	35,5	131,9	189,0	69,1	53,5	165,6	127,2	104,8	83,8	49,7	153,5	62,2
sOHl	482,1	517,6	442,9	631,4	708,7	687,7	622,4	635,1	699,3	604,2	591,6	434,8
sRut	4227	10804	15056	10375	7462	8484	8898	15790	21767	5150	6840	9601
Enzymes												
eAlAT	1128,2	1826,2	843,9	1761,6	1859,1	861,8	1240,1	875,7	1376,2	2040,8	1255,7	1616,2
eAsAT	3355,1	3972,2	4149,5	2834,3	3378,5	4508,9	4282,1	5628,4	4696,7	2317,7	6548,2	6351,1
eSKDH	134,6	159,8	262,9	121,6	153,8	161,6	190,9	214,8	179,2	136,5	207,4	181,3
ePGK	6499,1	7319,1	7567,5	4607,4	3737,8	6355,7	7930,8	5966,1	6796,8	4079,3	7106,6	6591,4
eTPI	41801	79861	90550	64274	44648	71202	82357	96038	80870	70323	87216	78121
eEno	254,8	381,9	486,5	265,1	315,5	338,7	430,1	479,2	389,7	275,9	494,3	376,4
eG6PDH	148,5	185,6	213,2	182,9	147,0	183,6	212,1	213,8	188,4	142,8	222,6	185,2
eGAPDHd	1756,4	2796,0	2879,1	1968,5	1617,4	2622,9	3100,6	2856,8	2810,7	1780,6	2474,9	2775,5
eGAPDHdp	483,9	615,6	1126,6	328,0	534,6	673,2	621,9	806,8	630,5	349,1	593,4	697,7
eIDH	146,0	101,9	157,3	153,9	97,5	148,0	119,0	195,1	164,7	133,6	232,0	121,3
ePepC	387,0	548,8	479,1	652,6	430,3	539,6	624,9	789,4	519,8	461,0	756,4	540,6
ePFKa	84,0	98,1	93,1	87,7	93,0	103,9	102,8	111,9	123,0	96,9	114,8	101,6
ePFKp	455,4	947,3	915,9	949,1	611,1	919,8	949,0	978,4	921,1	786,9	985,0	704,0
ePyrK	314,6	354,9	478,1	418,3	370,6	427,2	488,4	383,9	316,6	392,2	548,0	353,9
eFRK	134,0	275,5	278,7	194,7	73,1	172,4	218,9	171,2	208,8	150,8	373,5	176,6
eGlK	63,5	96,0	144,2	88,8	69,7	92,9	112,8	121,0	101,6	80,5	142,5	97,2
eInvA	253,9	270,4	422,4	448,1	212,8	213,0	519,3	227,1	197,4	267,2	403,3	143,5
eInvN	84,6	380,3	269,2	224,3	140,4	157,7	228,5	155,9	130,6	210,4	244,4	71,5
eSus	410,0	437,8	187,6	345,6	335,3	463,7	362,7	432,1	622,8	584,0	541,0	951,0
eFbpA	1440,3	2653,5	3596,9	2347,7	2698,9	2527,1	2607,5	3094,8	2165,2	2079,4	2564,5	2282,9
ePgm	1283,0	2015,7	2220,7	1537,7	1336,3	1868,7	2007,8	2211,7	1630,1	1489,0	2335,3	1980,1
eAco	68,8	121,7	170,3	179,6	86,8	132,7	172,1	223,7	120,1	171,8	285,6	122,1
eFumase	254,5	314,7	385,4	278,9	522,6	278,4	251,7	474,1	239,2	366,5	413,1	180,6
eMDH	15664,2	20909,7	29662,6	20404,0	15130,0	20448,4	22217,1	29477,1	24711,6	17941,1	21370,5	23718,9
eMed	614,0	472,5	773,3	615,7	708,0	639,7	670,7	853,2	549,2	455,0	349,6	674,7
eMedp	236,9	276,6	354,2	275,5	357,6	278,0	205,0	253,9	247,2	116,3	392,3	476,0

geno	LevOR	CerLevOR	CerOR	StuOR	StuCriOR	CriOR	LA0147COR	LA0147POR	PloOR	FerOR	FerLA147COR	LA1420OR
Phenotypic traits												
xfw	119,877	24,813	5,42418	65,4114	31,0459	15,7285	124,926	64,4233	40,8354	119,83	73,3327	38,036
xfd	71,13	38,134	22,978	54,7533	41,546	35,1	66,284	50,664	40,788	64,082	55,68	45,803
xdmc	4,89	7,2694	8,56998	5,62717	6,098	5,9494	5,33697	6,12018	7,03107	6,28692	5,77109	5,1759
pheight	118	167,5	146,3	134,12	173,5	172,5	112,5	132,5	118	111	123,88	98,5
pstem	13,28	15,55	13,27	8,64	14,98	13,04	14,27	14,34	11,44	15,11	15,07	13,01
Primary metabolites												
Gluc	10900,9	13819,3	15515,0	10686,7	10654,0	9435,2	10613,4	14272,4	15772,1	13761,9	11447,6	9164,9
Suc	371,5	1244,8	2636,7	730,6	698,2	566,7	531,0	720,4	1016,7	988,2	564,0	395,3
Fru	11664,2	14141,4	16420,9	11482,6	11508,5	10310,2	10793,0	13513,1	13847,5	12940,2	12047,1	9755,4
Inos	132,6	198,7	308,1	187,1	175,8	128,0	160,6	198,1	289,5	301,7	182,5	110,6
Ala	24,1	17,0	23,1	14,4	14,2	13,7	12,8	10,5	18,0	18,5	19,0	18,8
Asn	146,4	171,6	288,0	165,7	237,2	273,3	104,2	73,5	161,0	102,0	120,6	126,0
Asp	105,0	181,8	316,9	186,0	195,2	157,9	110,6	101,1	126,5	226,3	165,8	81,0
Gaba	235,2	295,7	238,9	337,2	262,9	126,8	201,4	201,1	374,0	318,2	202,8	116,8
Gln	692,2	1149,0	1774,5	888,8	1466,3	1370,8	604,4	404,3	1066,9	635,2	742,5	601,1
Ileu	24,4	32,2	24,7	21,2	38,0	35,7	15,2	14,2	46,7	20,7	16,4	13,8
Leu	44,6	58,3	69,2	43,9	63,3	51,5	33,1	24,6	69,1	52,2	45,2	38,3
Phe	143,3	177,2	299,7	166,6	220,2	192,7	105,4	170,1	504,5	174,5	158,1	160,6
Tyr	19,2	21,4	32,6	21,2	21,3	16,6	12,8	13,0	44,6	16,0	12,5	9,6
Val	19,4	19,6	8,6	13,0	15,1	10,7	11,2	11,5	26,6	13,5	10,3	9,3
Thr	30,5	38,6	29,6	37,8	46,7	32,5	20,7	13,5	30,4	27,7	17,1	11,8
VitC	11,4	14,5	24,8	9,4	12,6	11,2	8,2	9,2	13,7	19,0	15,6	14,4
Cit	2421,8	5728,5	7660,6	5880,3	9101,5	7479,6	4054,6	4100,8	4243,0	4204,5	6982,6	5731,0
Mal	1228,9	1636,4	1932,9	1780,4	686,4	413,9	1149,7	1223,9	1866,8	1366,1	383,8	201,7
Fum	0,50	0,29	0,30	0,45	0,20	0,15	0,59	0,38	0,44	0,80	0,33	0,00
Stch	0,01	0,17	1,11	0,00	0,05	0,05	0,01	0,07	0,06	0,19	0,03	0,01
Trigo	13,4	34,0	61,6	16,1	28,6	45,6	13,0	13,3	14,9	13,7	23,7	27,6
Ade	5,24	6,57	7,26	4,99	7,39	8,19	3,89	5,25	7,06	4,12	6,78	4,65
Chol	42,9	70,7	76,0	54,4	62,9	52,1	35,8	49,3	63,2	34,9	43,9	45,6
Secondary metabolits												
sChlorAc	2721,1	3667,6	6449,1	2998,4	3175,6	3478,2	2941,1	3293,8	2985,3	3569,4	3249,9	3315,0
sTCQ	5144,8	2814,3	3102,2	5122,8	2479,0	2320,7	3324,1	1790,8	708,2	1754,6	2673,5	3410,6
sAto	2170	1618	3326	5469	3561	5885	2951	3244	5402	2122	996	880
sCry	480,3	900,0	675,4	666,3	533,8	773,8	316,8	588,1	561,3	641,1	570,0	505,5
sDHT	1050	1109	7599	2546	1765	2532	1516	1305	2424	887	418	1594
sRutp	1234,7	3949,0	7445,1	3638,1	2942,7	2057,7	2967,7	3399,4	5133,1	3342,6	2115,5	3529,4
sPpa	316,2	102,2	73,6	109,9	165,3	383,3	188,6	180,0	239,0	203,5	304,0	607,5
sCou	1202,3	1357,2	3223,3	1460,8	1614,9	1194,3	1905,3	1595,1	2333,2	1782,0	1586,6	683,2
sNna	7847,0	13100,2	64355,1	22311,1	14790,9	8370,7	21322,9	25911,4	40233,4	7744,2	5485,0	144,4
sOHl	1199,4	819,5	1650,9	1341,7	1104,7	2794,9	1243,7	688,1	794,4	828,5	621,4	298,7
sRut	2673	5276	22223	5804	4154	3599	3829	7644	18039	2995	1669	1594
Enzymes												
eAlAT	711,5	815,9	1025,2	826,4	1507,5	911,0	964,6	239,2	3218,7	1717,8	1385,0	1383,8
eAsAT	2108,1	3259,8	3970,1	1968,4	2095,4	1490,2	1626,4	1581,4	1821,7	2582,0	3199,6	2055,6
eSKDH	62,6	112,5	208,0	87,1	75,4	85,4	80,2	89,8	69,0	37,6	71,0	82,2
ePGK	679,9	3709,5	3355,8	1667,4	4436,0	2262,1	1553,7	1817,8	1472,8	1210,7	1374,2	2684,4
eTPI	15737	18968	42118	30089	18159	15443	13542	16350	15502	14554	10801	16302
eEno	47,3	111,2	202,7	113,8	85,2	103,6	52,5	74,6	83,2	49,5	54,1	90,1
eG6PDH	57,8	85,5	140,5	85,5	64,8	69,7	50,0	53,7	55,9	49,9	53,6	66,0
eGAPDHd	1002,8	1799,9	1329,1	972,1	796,3	406,6	504,2	864,5	698,6	708,3	1168,7	943,7
eGAPDHdp	283,3	337,4	296,8	211,2	567,9	1293,8	65,6	345,4	185,9	347,2	255,6	274,5
eIDH	115,1	222,5	378,8	168,4	135,7	140,9	100,6	150,5	118,2	92,3	117,2	140,3
ePepC	150,6	260,5	274,2	237,3	153,5	144,9	94,6	131,0	207,1	133,8	154,7	223,8
ePFKa	47,0	51,2	121,8	62,6	56,4	62,0	53,8	61,3	42,3	45,5	63,1	47,1
ePFKp	113,1	309,7	599,0	309,1	183,5	266,3	120,8	183,7	147,7	121,6	206,8	201,7
ePyrK	129,9	161,5	454,8	221,4	151,6	187,6	121,8	139,0	131,7	125,4	227,3	109,2
eFRK	21,7	39,3	89,8	32,4	48,3	57,0	44,5	84,1	16,4	21,9	41,8	38,3
eGlK	9,2	32,0	38,6	21,0	15,9	16,5	11,3	18,6	14,4	19,1	25,5	19,5
eInvA	554,2	1244,8	1299,9	1351,7	962,8	522,3	943,7	703,9	488,1	183,5	783,2	926,8
eInvN	235,8	400,5	445,7	783,2	485,0	292,9	271,4	331,6	147,0	93,2	530,9	465,5
eSus	133,7	177,4	232,0	168,4	431,4	88,0	386,4	286,6	308,4	290,3	87,7	248,5
eFbpA	1030,7	1633,9	2257,3	1623,5	1268,8	1067,5	1053,3	1077,8	517,4	665,7	286,0	1466,3
ePgm	477,9	626,1	1322,6	570,0	601,2	493,3	425,9	545,9	415,5	541,9	536,2	694,3
eAco	56,5	122,0	112,9	97,5	49,0	29,1	42,9	60,2	44,5	100,5	67,2	7,9
eFumase	220,7	391,3	398,5	171,7	195,6	208,4	18,1	260,0	175,8	186,2	216,2	236,1
eMDH	5451,2	10086,1	17568,2	6259,8	7067,5	5416,8	4436,9	4859,9	4890,2	4554,1	6579,8	6080,4
eMed	462,8	336,5	380,2	453,5	197,2	222,1	337,1	284,1	342,6	273,0	304,5	222,2
eMedp	47,0	151,2	272,0	234,9	127,6	124,8	86,2	119,8	51,7	139,9	174,0	152,2

Supplemental table S4B Mean values of the smounts of 256 significant protein spots among genotypes at cell expansion stage,

geno	CerCE	CerLe	LevCE	StuCE	StuCri	CriCE	LA01<	LA01<	PloCE	FerCE	FerLA	LA1420CE
JX006	1,75	0,78	0,85	0,57	0,76	0,77	0,67	0,70	0,75	0,71	0,77	0,84
JX019	3,52	5,53	5,18	5,71	5,12	4,38	5,48	4,76	4,78	4,94	4,09	6,29
JX101	0,62	1,92	1,48	1,81	1,61	1,51	1,67	1,84	1,77	1,65	1,61	1,45
JX104	0,76	0,58	0,81	0,88	0,89	0,97	0,93	0,96	1,16	1,13	1,08	1,00
JX168	1,02	1,34	1,29	1,42	1,19	1,62	1,11	1,09	1,15	1,04	1,42	1,34
JX243	1,71	1,42	1,96	1,94	1,65	1,80	1,70	1,93	2,13	1,58	1,87	2,00
JX248	0,22	0,73	0,56	0,58	0,58	0,54	0,59	0,56	0,44	0,67	0,53	0,48
JX359	2,44	1,86	1,72	1,78	1,73	1,76	1,80	2,15	2,29	1,87	1,79	1,81
JX362	0,89	1,71	1,78	2,29	2,16	2,32	1,96	1,87	1,99	1,94	1,90	1,96
JX400	0,67	0,70	0,83	0,60	0,74	0,56	0,81	0,61	1,19	0,81	0,59	0,68
JX166	3,80	3,05	3,55	3,72	3,25	2,69	3,44	3,19	3,07	2,99	3,09	2,90
JX175	2,56	2,51	2,53	2,81	2,72	3,03	2,86	3,33	4,01	2,81	3,39	3,04
JX225	0,52	0,64	0,48	0,75	0,96	0,52	0,60	0,58	0,41	0,35	0,52	0,40
JX289	6,95	10,26	11,82	12,43	8,47	4,82	14,57	10,43	5,46	12,22	8,72	6,21
JX030	2,36	2,56	3,04	2,79	2,71	2,64	2,96	3,07	2,48	2,64	3,14	3,25
JX038	0,90	1,18	0,33	0,45	0,51	0,35	0,40	0,36	0,31	0,41	0,39	0,42
JX040	1,32	0,60	0,35	0,44	0,37	0,46	0,35	0,39	0,38	0,37	0,36	0,38
JX244	0,21	0,69	0,55	0,68	0,60	0,34	0,71	0,76	0,46	0,98	0,63	0,38
JX010	0,24	0,17	0,11	0,12	0,11	0,10	0,14	0,12	0,09	0,13	0,12	0,11
JX020	1,28	1,04	0,94	0,93	0,97	0,84	1,01	1,01	0,99	0,55	0,62	0,87
JX037	0,68	0,28	0,33	0,22	0,25	0,15	0,22	0,29	0,33	0,26	0,30	0,40
JX042	1,97	1,83	1,26	1,07	1,39	1,38	1,26	1,59	1,64	0,53	0,81	0,97
JX062	0,07	0,11	0,11	0,10	0,10	0,10	0,10	0,08	0,07	0,13	0,12	0,11
JX069	0,87	0,63	0,60	0,58	0,51	0,57	0,46	0,56	0,45	0,45	0,51	0,59
JX189	0,17	0,15	0,14	0,24	0,17	0,22	0,16	0,18	0,15	0,21	0,18	0,21
JX190	1,36	2,14	2,10	1,94	2,19	2,35	2,04	1,81	1,79	1,95	2,16	2,09
JX221	0,10	0,15	0,22	0,24	0,21	0,23	0,23	0,21	0,23	0,28	0,24	0,23
JX247	4,94	4,58	4,63	3,86	4,36	3,74	4,12	4,13	4,49	1,63	3,07	3,87
JX254	0,36	0,53	0,42	0,43	0,41	0,28	0,34	0,44	0,31	0,41	0,38	0,64
JX265	1,42	2,03	2,40	1,87	1,80	1,50	2,16	2,24	2,00	2,69	2,46	2,31
JX320	0,48	0,33	0,33	0,33	0,34	0,31	0,43	0,35	0,37	0,34	0,27	0,30
JX331	1,87	2,07	2,85	2,57	2,50	2,67	2,09	2,38	2,05	2,22	2,67	3,10
JX364	1,27	0,96	0,88	1,06	0,98	1,08	0,87	1,11	1,11	0,80	0,89	1,05
JX386	0,84	0,58	0,59	0,61	0,61	0,63	0,63	0,55	0,57	0,75	0,55	0,63
JX389	3,16	3,46	4,04	3,46	3,45	3,15	4,00	3,71	3,63	3,88	3,87	3,81
JX423	0,96	1,08	1,31	1,24	1,29	1,08	1,02	1,03	0,98	0,82	1,21	1,17
JX021	1,58	1,61	1,68	2,02	1,69	1,92	2,15	2,02	1,81	1,68	2,18	2,28
JX023	1,71	2,30	1,85	2,31	1,83	1,30	1,98	1,42	0,89	1,44	2,26	2,46
JX224	1,35	1,77	1,51	1,97	1,75	1,46	2,13	2,26	1,92	2,12	2,42	2,13
JX270	5,41	5,83	5,70	6,85	6,21	6,64	7,10	6,85	6,95	6,68	8,24	7,38
JX373	2,84	2,82	3,89	3,71	2,94	3,00	3,89	3,69	3,72	3,90	4,33	3,79
JX012	0,53	0,38	0,25	0,31	0,34	0,33	0,27	0,34	0,32	0,33	0,30	0,29
JX014	14,37	7,73	4,93	4,09	11,15	15,08	5,00	7,59	8,57	4,90	4,10	4,28
JX018	0,14	0,13	0,08	0,11	0,10	0,11	0,07	0,09	0,10	0,07	0,11	0,10
JX025	0,16	0,31	0,36	0,30	0,35	0,26	0,26	0,25	0,26	0,30	0,24	0,23
JX031	0,55	0,49	0,38	0,41	0,52	0,51	0,31	0,41	0,53	0,34	0,35	0,41
JX046	0,83	0,92	0,59	1,03	0,92	0,81	0,76	0,74	0,65	0,80	0,75	0,64
JX054	0,16	0,25	0,22	0,26	0,23	0,24	0,26	0,23	0,22	0,27	0,30	0,26
JX061	0,13	0,19	0,19	0,25	0,22	0,20	0,27	0,20	0,21	0,26	0,29	0,26
JX095	0,59	0,84	0,83	0,98	0,75	0,90	0,88	0,69	0,82	0,80	0,67	0,34
JX102	1,11	0,89	0,74	0,73	0,82	0,91	0,86	1,07	1,06	1,11	1,03	1,01
JX107	1,51	1,10	1,09	1,17	1,15	1,56	1,24	1,40	1,38	1,22	1,39	1,37
JX114	1,49	1,63	1,94	1,62	1,75	1,31	1,50	2,21	2,51	2,01	1,75	1,46
JX120	1,66	1,26	0,95	1,16	1,28	1,31	1,10	1,35	1,34	1,29	1,24	1,30
JX125	1,37	1,33	1,30	1,35	1,35	1,64	1,55	1,71	1,62	1,68	1,74	1,59
JX137	0,40	0,29	0,21	0,18	0,21	0,27	0,22	0,23	0,28	0,26	0,23	0,19
JX138	1,21	1,09	1,11	1,24	0,80	0,97	0,93	0,95	0,95	1,08	0,95	1,11
JX139	0,76	0,66	0,46	0,42	0,45	0,48	0,43	0,65	0,87	0,51	0,42	0,48
JX142	1,12	0,96	0,94	1,04	0,94	1,01	1,17	1,19	1,13	1,25	1,26	1,13
JX148	0,30	0,38	0,38	0,34	0,38	0,93	0,40	0,43	0,26	0,30	0,38	0,64
JX152	0,87	0,72	0,57	0,65	0,79	0,84	0,64	0,84	0,76	0,61	0,79	0,72
JX153	0,16	0,26	0,38	0,39	0,42	0,50	0,36	0,37	0,40	0,36	0,41	0,37
JX158	0,50	0,63	0,75	0,88	0,86	0,89	0,91	0,66	0,34	0,92	1,03	0,83
JX160	0,70	0,66	0,49	0,77	0,69	0,58	0,58	0,80	0,94	0,70	0,58	0,77
JX172	0,83	0,57	0,50	0,55	0,59	0,54	0,55	0,69	0,55	0,55	0,59	0,59
JX174	0,64	0,64	0,59	0,58	0,59	0,65	0,65	0,80	0,85	0,63	0,61	0,68
JX181	0,59	0,43	0,38	0,56	0,43	0,46	0,52	0,53	0,48	0,48	0,44	0,51
JX182	0,55	0,76	0,80	0,83	0,91	0,88	0,88	0,91	0,93	0,90	0,93	0,81
JX196	0,69	0,83	0,81	0,89	0,86	0,94	0,80	0,88	1,00	0,90	1,07	0,84
JX201	3,18	2,70	2,75	3,05	3,21	3,11	3,00	3,35	3,38	3,47	3,56	3,45
JX202	2,42	2,52	2,70	2,80	2,64	2,15	2,46	2,62	3,05	3,28	3,10	3,20
JX211	0,20	0,34	0,39	0,44	0,32	0,46	0,43	0,26	0,31	0,28	0,27	0,49
JX217	0,70	0,83	0,90	0,87	0,76	0,47	0,69	1,04	0,92	0,65	0,67	0,54
JX219	1,12	0,88	0,76	0,78	0,77	0,78	0,82	0,87	1,04	1,06	0,92	0,78
JX231	0,66	0,97	0,70	0,96	1,02	0,96	0,96	0,92	0,70	0,79	1,05	0,81
JX232	0,51	0,62	0,46	0,59	0,55	0,64	0,41	0,41	0,56	0,54	0,80	0,52
JX249	0,95	1,07	1,06	1,37	0,88	0,70	0,70	0,84	0,73	0,69	0,74	0,77
JX258	0,73	0,66	0,65	0,83	0,86	0,70	0,72	0,65	0,64	0,74	0,79	0,78
JX260	0,37	0,47	0,32	0,45	0,36	0,43	0,31	0,27	0,31	0,36	0,36	0,41
JX279	0,46	0,40	0,96	0,46	0,37	0,59	0,48	0,46	0,45	0,71	0,42	0,61
JX287	1,60	2,07	2,57	2,11	1,82	1,50	2,21	2,14	1,62	1,68	2,27	2,12
JX301	2,34	2,79	3,37	2,27	2,46	2,23	2,71	2,27	1,90	2,78	2,05	2,52
JX303	0,86	1,05	1,25	0,82	0,98	0,97	1,22	0,87	1,06	1,10	0,81	1,16
JX336	1,58	1,49	1,72	1,52	1,59	1,57	1,63	1,82	1,80	1,84	1,62	1,50
JX345	0,60	0,57	0,43	0,57	0,58	0,57	0,58	0,95	1,58	0,48	0,61	0,62
JX360	0,63	0,76	0,80	0,90	1,04	1,17	0,91	0,67	0,64	0,81	1,19	1,18
JX365	0,78	0,57	0,57	0,70	0,72	0,95	0,91	0,59	0,43	0,60	0,70	0,66
JX382	0,42	0,53	0,64	0,53	0,52	0,49	0,50	0,50	0,44	0,54	0,58	0,41
JX383	0,87	0,80	0,63	0,72	0,75	0,75	0,65	0,69	0,62	0,63	0,74	0,68
JX384	0,41	0,30	0,29	0,36	0,28	0,24	0,25	0,26	0,28	0,26	0,29	0,26
JX385	0,94	0,87	0,76	0,76	0,77	0,61	0,78	0,96	0,83	0,88	0,88	0,85
JX390	2,16	1,77	1,56	1,79	1,97	2,00	1,69	1,98	1,88	1,83	1,99	2,09
JX396	1,38	1,29	1,02	0,93	1,23	1,15	1,57	1,30	1,20	1,48	1,20	1,03
JX408	0,80	1,02	1,08	1,12	1,17	1,30	1,24	1,08	1,22	1,30	1,24	1,04
JX417	1,76	1,39	1,32	1,22	1,83	2,97	1,35	1,44	1,42	1,18	1,87	2,04
JX274	1,02	0,96	1,06	1,12	1,02	1,12	1,17	1,20	1,21	1,32	1,53	1,21
JX146	19,50	11,11	4,97	4,19	6,79	10,33	4,75	12,85	21,11	4,35	7,39	10,63
JX398	0,80	0,91	1,09	0,78	0,94	0,71	0,91	1,02	0,77	0,90	0,98	0,94
JX082	0,22	0,30	0,29	0,28	0,24	0,16	0,26	0,29	0,31	0,23	0,26	0,23
JX098	1,19	0,82	0,56	0,66	0,67	0,58	0,87	0,77	0,65	0,94	0,73	0,75
JX100	2,32	1,66	1,62	1,58	1,99	2,17	1,61	1,61	1,59	1,43	1,49	1,72
JX116	1,30	1,14	0,86	1,11	1,31	1,62	1,04	1,27	1,64	1,16	1,21	1,36
JX150	0,44	0,53	0,49	0,31	0,67	1,17	0,45	0,31	0,31	0,53	0,34	0,35
JX162	1,19	0,84	0,91	0,59	1,23	1,69	0,75	0,78	0,83	0,76	1,09	1,22
JX187	0,92	1,10	1,01	1,14	1,18	1,12	1,00	1,18	1,05	1,04	1,31	1,19
JX222	1,07	0,50	0,34	0,47	1,26	1,54	0,54	0,84	0,67	0,53	1,49	1,37
JX266	3,40	3,99	4,70	4,26	4,15	5,67	3,82	4,81	5,87	4,15	4,71	4,33
JX268	0,42	0,51	0,63	0,48	0,36	0,19	0,49	0,74	0,41	0,60	0,46	0,38
JX296	2,41	3,23	2,96	2,73	3,09	2,11	2,83	2,38	2,55	4,02	2,85	2,45
JX310	1,03	1,62	1,59	1,60	1,44	0,86	1,56	1,69	1,71	1,46	1,48	1,55
JX326	2,49	2,57	2,74	2,54	2,78	2,94	2,76	2,84	2,77	2,73	2,85	2,40
JX327	3,59	3,24	2,77	3,03	3,23	3,06	3,22	3,03	2,41	3,39	3,10	3,13
JX361	0,52	0,76	0,86	0,87	0,80	0,62	0,94	0,76	0,59	0,94	0,92	0,83
JX391	2,51	1,44	1,07	1,24	1,46	0,89	1,75	1,97	1,86	1,48	1,35	0,87
JX409	1,75	1,69	1,66	1,65	1,63	1,65	1,62	1,71	1,94	1,76	1,56	1,53
JX001	3,71	2,25	0,88	0,73	2,30	3,85	0,61	2,36	3,55	4,51	2,68	0,75
JX008	1,45	1,35	1,16	0,98	1,18	1,08	1,20	1,21	1,21	1,15	1,07	0,98
JX015	5,53	4,45	3,38	3,39	3,83	4,18	3,91	5,26	4,68	3,99	4,53	4,08
JX024	1,36	1,42	1,58	1,19	1,14	0,79	1,41	1,28	0,90	1,13	1,42	1,24
JX026	2,00	2,17	2,78	2,40	2,11	2,22	2,74	2,61	2,73	2,68	2,56	2,41
JX027	0,26	0,33	0,35	0,27	0,34	0,32	0,33	0,28	0,46	0,35	0,36	0,30
JX029	0,49	0,53	0,45	0,49	0,45	0,43	0,51	0,50	0,54	0,55	0,65	0,57
JX032	0,80	1,02	0,78	0,94	1,03	1,03	0,91	0,96	0,99	1,00	1,12	1,09
JX033	1,19	1,37	1,21	1,16	1,32	1,38	1,41	1,35	1,37	1,17	1,21	1,08
JX053	2,25	2,23	1,99	2,86	2,56	1,97	2,44	1,90	2,69	2,26	2,24	2,46
JX056	0,73	0,43	0,33	0,35	0,31	0,37	0,31	0,37	0,60	0,38	0,40	0,41
JX073	3,76	4,69	4,43	3,93	3,82	3,78	4,28	4,15	4,26	4,14	4,37	3,97
JX087	0,63	0,81	0,33	0,51	0,58	0,49	0,35	0,33	0,42	0,23	0,35	0,41
JX106	0,96	1,00	1,23	1,74	1,57	1,82	0,85	1,24	1,74	1,21	1,54	1,38
JX111	1,00	1,18	1,52	1,01	1,18	0,86	1,21	1,27	1,35	1,51	1,21	1,21
JX113	2,78	1,63	1,72	1,78	1,80	1,92	1,83	2,16	2,02	1,60	1,71	1,80
JX132	0,12	0,14	0,10	0,09	0,08	0,06	0,10	0,07	0,09	0,08	0,10	0,13
JX144	0,96	1,06	1,42	0,97	1,30	1,07	1,07	0,97	1,07	1,02	1,07	1,28

JX151	1,30	1,37	1,49	1,61	1,49	1,64	1,77	1,59	1,53	1,80	1,98	1,74
JX154	2,36	1,61	1,41	1,42	1,67	1,82	1,72	1,91	1,90	1,70	1,68	1,69
JX157	1,80	1,53	1,23	1,28	1,46	1,76	1,46	1,38	1,66	1,19	1,36	1,30
JX161	2,06	1,62	1,58	1,61	1,70	1,72	1,69	1,72	1,44	1,39	1,62	1,40
JX171	0,50	0,58	0,41	0,48	0,47	0,64	0,38	0,25	0,33	0,55	0,63	0,48
JX177	4,34	4,23	3,68	3,28	3,35	3,21	3,63	4,21	3,73	4,16	3,77	3,29
JX178	0,82	0,59	0,59	0,60	0,73	0,68	0,67	0,69	0,67	0,73	0,73	0,59
JX179	1,14	1,07	1,04	1,19	1,45	1,73	1,12	1,24	1,10	1,18	1,43	1,66
JX205	1,98	2,33	1,99	1,66	2,12	1,82	1,79	2,00	2,04	1,84	1,78	1,77
JX206	1,83	1,45	1,89	1,44	1,85	1,90	1,74	1,97	1,75	1,63	1,98	2,32
JX212	0,28	0,27	0,37	0,32	0,24	0,31	0,38	0,33	0,45	0,29	0,32	0,53
JX218	5,84	4,74	4,00	3,46	4,46	3,92	4,26	4,26	4,83	3,81	3,91	4,24
JX220	0,28	0,46	0,40	0,38	0,45	0,42	0,46	0,42	0,44	0,50	0,46	0,42
JX227	0,77	0,83	0,46	0,30	0,44	0,49	0,87	0,64	0,35	0,75	0,83	0,50
JX228	0,57	0,83	0,74	0,90	0,75	1,05	0,75	0,45	0,48	1,06	1,18	0,89
JX230	0,15	0,22	0,22	0,21	0,21	0,18	0,26	0,22	0,20	0,25	0,26	0,23
JX233	0,71	0,68	0,74	0,86	0,87	0,86	0,75	0,66	0,55	0,64	0,81	0,93
JX239	2,42	2,77	2,66	1,92	2,58	2,47	2,05	1,63	1,53	1,92	2,06	1,76
JX255	0,22	0,23	0,33	0,27	0,31	0,27	0,28	0,26	0,35	0,34	0,30	0,28
JX264	2,11	1,76	1,82	1,80	2,06	1,78	1,90	1,94	2,00	1,85	1,77	1,55
JX269	0,41	0,96	1,48	0,29	0,31	0,33	0,34	0,37	0,34	0,44	0,33	0,41
JX275	1,09	1,68	1,77	1,61	1,65	1,68	1,58	1,40	1,49	1,80	2,09	1,60
JX286	1,24	2,02	2,07	0,85	1,71	2,09	1,93	1,75	0,92	1,51	1,89	1,97
JX290	1,40	1,31	1,36	1,32	0,94	0,43	1,29	1,40	1,46	1,19	0,95	0,80
JX291	0,97	3,62	4,34	4,52	3,66	1,93	4,12	3,75	1,53	4,52	5,00	4,69
JX294	0,25	0,28	0,35	0,38	0,40	0,30	0,16	0,17	0,23	0,35	0,37	0,23
JX302	0,46	0,52	0,73	0,72	0,56	0,61	0,78	0,68	0,73	0,74	0,83	0,64
JX305	0,22	0,29	0,38	0,34	0,45	0,36	0,27	0,24	0,26	0,28	0,29	0,36
JX306	0,47	0,54	0,64	0,57	0,38	0,28	0,53	0,52	0,63	0,50	0,56	0,59
JX311	1,02	0,57	0,43	0,41	0,46	0,48	0,41	0,57	0,80	0,40	0,52	0,42
JX313	0,12	0,17	0,25	0,22	0,30	0,26	0,21	0,22	0,24	0,22	0,28	0,22
JX322	0,58	0,82	0,83	0,68	0,65	0,64	0,65	0,67	0,75	0,64	0,61	0,52
JX323	4,17	4,00	3,59	3,46	3,61	3,33	3,91	4,70	4,22	4,60	3,89	3,79
JX329	0,41	0,37	0,45	0,45	0,49	0,58	0,44	0,43	0,50	0,50	0,55	0,44
JX333	0,45	0,50	0,48	0,61	0,53	0,61	0,49	0,42	0,44	0,48	0,52	0,56
JX342	0,49	0,58	0,66	0,60	0,59	0,54	0,54	0,64	0,57	0,69	0,68	0,68
JX351	0,38	0,67	0,88	0,87	0,93	0,83	0,90	0,64	0,29	0,34	0,55	0,83
JX357	1,44	1,41	1,66	1,52	1,49	1,53	1,57	1,63	1,81	1,38	1,32	1,08
JX363	0,44	0,86	1,10	1,00	1,06	0,93	1,03	1,13	1,07	0,98	1,10	0,91
JX397	0,45	0,58	0,79	0,65	0,86	1,05	0,75	0,65	0,66	0,80	0,83	0,82
JX402	0,39	0,45	0,42	0,45	0,52	0,65	0,43	0,47	0,46	0,44	0,50	0,59
JX406	0,25	0,30	0,23	0,24	0,31	0,24	0,19	0,19	0,18	0,29	0,21	0,22
JX411	0,47	0,56	0,49	0,49	0,61	0,51	0,38	0,43	0,38	0,47	0,48	0,42
JX412	0,68	0,75	0,77	0,92	0,89	0,73	0,73	0,91	0,86	0,84	0,88	0,75
JX413	3,53	3,80	3,84	3,47	3,03	3,30	4,07	3,81	3,90	3,88	4,16	3,49
JX420	0,63	0,86	0,77	0,84	0,79	0,76	0,61	0,68	0,51	0,64	0,70	0,85
JX422	0,80	0,91	1,09	0,91	0,98	0,91	0,91	1,11	1,15	1,14	1,28	1,18
JX009	2,78	1,96	2,01	2,23	2,18	2,03	2,05	2,11	1,58	2,01	1,85	1,79
JX028	0,49	0,67	0,86	0,76	0,68	0,75	0,81	0,70	0,54	0,70	0,96	0,78
JX047	0,53	0,93	0,31	0,66	0,80	0,43	0,46	0,37	0,39	0,38	0,33	0,37
JX063	0,48	0,65	0,85	0,97	0,74	0,67	0,87	0,62	0,70	0,82	0,70	0,80
JX064	3,19	1,73	1,81	1,71	1,79	2,91	2,16	1,56	2,29	1,68	1,77	2,23
JX094	4,62	3,71	3,48	4,28	3,93	3,57	4,02	3,78	3,40	3,59	3,33	3,81
JX096	1,26	0,70	0,81	0,64	0,90	1,45	0,71	0,49	0,40	0,68	0,73	0,70
JX103	0,45	0,44	0,27	0,35	0,35	0,29	0,41	0,36	0,44	0,29	0,34	0,28
JX108	1,90	2,48	3,04	2,94	3,14	2,81	2,86	2,68	2,22	3,57	3,78	2,55
JX188	3,96	4,23	3,84	3,92	4,35	4,24	3,96	4,56	4,51	3,60	4,62	4,03
JX203	1,74	1,35	1,54	1,30	1,30	1,18	1,40	1,51	1,41	1,41	1,37	1,20
JX209	0,27	0,23	0,30	0,18	0,24	0,42	0,18	0,26	0,31	0,20	0,26	0,30
JX226	0,19	0,18	0,22	0,24	0,23	0,24	0,23	0,17	0,08	0,29	0,30	0,23
JX309	0,86	1,24	1,65	1,15	0,97	1,19	1,63	1,36	1,49	1,56	1,36	1,26
JX350	1,46	1,38	1,35	2,99	1,42	0,98	1,49	3,00	2,61	1,18	1,31	1,29
JX407	8,54	8,45	6,74	6,40	5,21	5,22	6,98	7,89	8,46	6,50	6,05	5,49
JX002	0,76	0,33	0,26	0,40	0,54	0,73	0,41	0,38	0,36	0,53	0,54	0,57
JX003	0,55	0,26	0,18	0,35	0,24	0,37	0,19	0,30	0,43	0,20	0,33	0,28
JX016	1,34	1,57	1,22	1,42	1,65	1,56	1,59	1,80	1,65	1,68	1,76	1,64
JX048	7,56	7,32	7,70	7,06	5,31	5,63	6,84	8,19	11,01	3,87	5,57	7,11
JX049	9,83	8,28	5,75	6,82	6,52	6,36	6,38	6,01	6,47	5,20	6,00	6,23
JX052	1,00	1,06	0,90	1,17	1,16	0,94	0,79	0,91	0,89	0,91	0,98	1,18
JX055	0,83	0,90	1,03	0,94	0,96	0,92	0,99	0,79	1,06	1,10	0,87	0,92
JX074	1,19	0,94	0,76	1,08	1,10	1,21	0,78	0,90	1,17	0,75	1,03	1,06
JX084	0,87	1,05	1,12	0,85	0,83	0,70	0,85	1,27	1,78	0,82	1,04	0,95
JX085	0,59	0,46	0,21	0,27	0,33	0,21	0,23	0,23	0,29	0,20	0,18	0,24
JX090	1,21	0,63	0,92	0,83	0,60	0,71	0,76	0,84	1,39	0,48	0,59	0,70
JX092	2,40	2,61	1,91	1,82	1,78	1,32	1,47	1,56	2,28	1,45	1,34	1,68
JX093	1,74	1,69	1,80	1,14	1,53	0,87	1,21	1,13	1,61	1,52	1,20	0,89
JX097	0,58	0,66	0,43	0,70	0,79	0,90	0,46	0,54	0,56	0,43	0,53	0,61
JX112	1,70	1,75	1,36	1,29	1,48	1,43	1,27	1,40	1,38	1,44	1,19	1,14
JX140	0,48	0,18	0,18	0,23	0,27	0,43	0,18	0,20	0,34	0,16	0,22	0,30
JX141	0,92	0,55	0,55	0,66	0,74	1,00	0,59	0,65	0,83	0,46	0,67	0,96
JX145	2,83	0,95	0,96	0,96	0,93	0,67	1,12	1,08	0,75	1,72	0,92	0,74
JX149	0,75	0,28	0,30	0,41	0,57	0,79	0,20	0,32	0,55	0,29	0,36	0,31
JX165	0,77	0,89	1,21	1,33	1,30	1,26	1,18	0,79	0,54	1,09	1,34	1,45
JX194	2,17	2,02	2,00	2,43	2,39	2,46	2,06	2,32	2,30	2,26	2,59	2,85
JX195	5,13	5,53	5,90	5,23	6,20	5,96	5,42	5,70	6,49	5,64	4,97	4,84
JX200	2,32	2,20	2,26	2,57	2,43	2,17	2,32	2,50	2,01	2,39	2,69	2,60
JX215	0,96	1,03	1,10	1,06	1,13	1,20	1,09	1,28	1,46	1,18	1,18	1,13
JX259	0,61	0,69	0,50	0,58	0,56	0,41	1,24	0,86	0,59	0,61	1,00	1,65
JX278	0,51	0,59	0,55	0,21	0,34	0,35	0,32	0,24	0,34	0,30	0,20	0,70
JX282	3,37	4,58	3,25	2,79	2,91	1,47	1,95	2,08	3,11	1,34	1,60	2,23
JX295	0,18	0,19	0,12	0,15	0,24	0,16	0,11	0,15	0,16	0,12	0,11	0,12
JX314	0,64	0,61	0,77	0,80	0,85	0,53	0,83	0,75	0,60	0,77	0,72	0,96
JX328	0,29	0,43	0,41	0,33	0,34	0,40	0,43	0,38	0,35	0,40	0,49	0,48
JX332	0,45	0,40	0,36	0,45	0,50	0,37	0,31	0,40	0,38	0,31	0,44	0,48
JX346	0,84	1,05	1,19	0,94	0,77	0,89	1,12	1,05	1,49	1,12	1,21	1,07
JX354	9,34	12,01	12,63	13,64	12,54	7,23	12,48	11,67	7,91	13,40	14,12	13,06
JX356	0,58	0,56	0,45	0,62	0,52	0,45	0,58	0,57	0,53	0,51	0,45	0,45
JX358	0,57	0,57	0,46	0,61	0,61	0,56	0,48	0,53	0,52	0,62	0,93	0,47
JX374	3,47	3,47	2,36	2,12	3,12	3,17	2,46	2,90	2,69	2,45	2,40	2,85
JX377	6,87	4,89	3,81	4,75	4,35	4,10	4,28	3,90	3,97	4,83	3,37	4,07
JX404	3,09	2,72	2,95	3,23	3,64	3,62	3,44	3,83	4,13	3,00	2,93	3,46
JX418	0,56	0,48	0,32	0,37	0,44	0,61	0,28	0,49	0,43	0,26	0,29	0,30
JX043	0,25	0,17	0,16	0,17	0,27	0,23	0,27	0,20	0,19	0,19	0,26	0,18
JX242	0,41	0,79	0,89	0,66	0,75	0,72	0,76	0,59	0,77	0,79	0,83	0,67
JX312	1,90	1,50	1,06	0,99	1,37	1,41	1,23	1,93	1,25	1,39	1,38	1,18
JX348	0,46	0,43	0,61	0,69	0,79	0,73	0,65	0,52	0,41	0,52	0,70	0,68
JX393	0,40	0,43	0,48	0,60	0,70	1,30	0,54	0,46	0,36	0,50	0,63	0,53
JX133	1,63	1,06	0,43	0,42	1,07	1,31	1,55	1,35	1,47	0,55	0,39	0,90
JX369	2,30	2,32	2,35	2,08	1,83	1,17	2,58	2,49	2,39	2,63	2,51	2,18
JX395	4,32	3,46	2,72	3,07	3,26	2,70	3,19	3,57	4,07	3,38	3,10	3,18
JX197	0,73	0,88	0,87	1,17	0,82	0,89	1,04	0,95	0,87	1,01	1,18	1,14
JX273	1,39	2,01	2,41	2,08	1,58	0,86	2,37	2,22	2,65	2,16	2,24	2,00
JX325	10,83	8,79	6,91	7,61	7,88	7,77	7,83	7,94	8,00	7,26	6,29	7,56
JX368	1,28	1,94	2,36	2,21	2,02	1,43	1,68	2,04	2,00	2,08	2,56	2,48
JX075	0,27	0,34	0,32	0,36	0,31	0,26	0,36	0,39	0,41	0,36	0,34	0,34
JX185	3,88	3,84	3,01	3,76	3,87	3,53	3,49	3,11	3,37	3,76	3,20	3,41
JX223	0,27	0,55	0,73	0,82	0,76	0,72	0,84	1,09	0,79	0,99	0,78	0,82
JX007	5,07	3,71	3,22	3,40	2,52	1,99	3,10	2,91	2,01	2,92	3,37	3,10
JX017	1,94	2,20	2,07	2,23	2,84	2,14	2,02	1,73	1,27	1,72	1,68	1,86
JX034	10,02	8,04	8,48	8,03	6,74	4,42	8,68	7,35	4,47	8,53	8,41	8,14
JX173	0,41	0,50	0,55	0,57	0,51	0,47	0,59	0,50	0,52	0,51	0,65	0,64
JX335	0,91	1,01	0,97	1,00	1,00	0,86	0,91	0,99	0,82	0,90	1,21	1,05
JX349	1,48	2,54	1,47	2,26	3,28	2,75	1,59	1,66	1,80	1,72	2,03	1,36

Supplemental table S4C Mean values of the smounts of 188 significant protein spots among genotypes at orange red stage

geno	CerOR	CerLe	LevOI	StuOR	StuCri	CriOR	LA014	LA014	PloOR	FerOR	FerLA	LA1420OR
JX019	6,45	8,11	9,43	8,47	10,25	9,21	11,54	10,19	11,80	9,01	9,37	10,13
JX101	2,20	1,37	1,31	1,54	1,36	2,24	1,40	1,20	1,49	1,42	0,96	1,34
JX168	1,02	1,08	1,02	1,18	0,93	1,30	0,88	0,97	0,85	0,71	0,96	1,15
JX170	9,38	10,37	8,96	8,56	8,12	7,79	10,21	9,51	9,18	10,65	7,46	4,17
JX280	0,12	0,16	1,18	0,23	0,21	0,34	0,36	0,24	0,24	0,20	0,36	0,36
JX362	1,16	1,15	1,89	1,74	1,66	2,09	1,59	1,64	1,68	1,19	1,71	1,87
JX400	0,62	0,50	0,59	0,72	0,49	0,62	0,84	0,73	0,62	1,07	1,24	1,16
JX289	3,40	7,08	5,67	7,83	5,80	2,73	6,84	5,79	3,73	9,95	5,47	3,71
JX030	3,42	3,68	3,67	3,67	3,92	3,29	4,08	3,59	3,93	4,37	4,30	2,89
JX039	0,88	1,30	1,10	1,43	0,96	0,66	0,75	1,23	1,84	1,44	0,76	0,56
JX040	1,73	2,20	1,79	1,75	1,44	1,33	1,86	1,84	1,42	1,78	1,20	0,55
JX244	1,16	1,17	1,50	1,03	0,94	0,83	1,34	1,35	0,86	1,58	1,02	0,35
JX245	2,32	3,27	3,35	2,73	2,88	2,24	3,69	3,58	3,24	3,76	2,83	1,18
JX276	0,61	1,18	0,97	1,88	1,31	0,79	1,70	1,81	2,44	1,69	1,68	1,51
JX010	0,21	0,16	0,09	0,12	0,10	0,11	0,12	0,10	0,10	0,13	0,15	0,12
JX020	1,08	1,05	0,90	0,86	1,18	1,00	1,00	0,99	1,34	0,43	0,66	0,86
JX037	0,63	0,90	0,77	0,86	0,84	0,41	1,01	1,17	1,64	0,88	1,03	1,26
JX042	1,89	1,77	1,03	1,35	1,50	1,16	0,79	1,16	1,32	0,36	0,60	0,68
JX147	1,77	3,82	1,20	1,48	2,74	3,81	1,98	3,06	3,01	0,84	3,23	4,18
JX190	1,33	1,64	1,66	1,39	1,64	2,21	1,82	1,79	1,60	2,08	1,68	1,50
JX221	0,06	0,15	0,19	0,20	0,28	0,20	0,19	0,21	0,20	0,19	0,18	0,22
JX247	5,20	4,93	4,14	4,71	5,13	4,93	4,34	4,89	5,73	1,56	3,33	4,31
JX265	1,82	1,65	1,92	1,79	1,62	1,17	1,86	1,81	1,97	2,25	1,86	1,79
JX320	0,65	0,56	0,36	0,37	0,40	0,43	0,42	0,53	0,63	0,33	0,37	0,34
JX023	1,93	1,49	1,49	1,77	1,50	1,30	1,55	1,38	1,15	1,91	1,38	1,57
JX143	1,70	1,58	1,25	1,11	1,45	2,46	1,33	1,17	0,77	1,36	1,82	1,64
JX012	0,30	0,31	0,28	0,26	0,23	0,23	0,24	0,23	0,32	0,31	0,27	0,20
JX018	0,17	0,11	0,06	0,12	0,10	0,06	0,08	0,12	0,20	0,07	0,08	0,07
JX025	0,14	0,21	0,18	0,32	0,31	0,21	0,17	0,22	0,23	0,32	0,23	0,14
JX046	2,01	2,05	1,40	1,86	1,97	2,20	1,42	2,54	2,54	1,69	1,80	0,94
JX102	1,19	1,08	0,66	0,73	0,63	0,47	0,79	0,96	1,35	0,92	0,94	0,63
JX122	1,64	1,79	1,78	1,39	1,67	1,93	1,81	1,72	1,67	1,44	1,67	2,26
JX125	1,42	1,52	1,16	1,32	1,19	1,04	1,13	1,16	1,68	1,55	1,23	1,07
JX127	1,29	1,13	0,90	0,90	0,85	0,91	0,99	0,98	1,19	0,89	0,90	0,99
JX138	1,76	1,49	1,80	1,82	1,38	1,05	1,96	1,68	1,40	1,47	1,93	2,70
JX142	0,97	1,09	0,86	1,03	0,83	0,70	0,79	0,73	1,10	1,08	0,89	0,92
JX152	0,42	0,41	0,29	0,30	0,42	0,46	0,33	0,29	0,46	0,38	0,32	0,25
JX153	0,22	0,35	0,46	0,36	0,46	0,55	0,45	0,38	0,43	0,37	0,41	0,47
JX158	0,42	0,55	0,61	0,68	0,56	0,73	0,59	0,49	0,37	0,78	0,66	0,60
JX160	0,68	0,46	0,45	0,47	0,39	0,48	0,38	0,51	0,53	0,36	0,38	0,32
JX217	0,65	0,61	1,12	0,66	0,45	1,23	1,30	0,71	0,89	0,48	1,08	2,22
JX240	0,82	0,53	0,93	0,72	0,99	0,94	1,01	1,22	1,10	1,48	1,31	1,45
JX249	0,86	1,59	1,56	3,04	1,62	0,85	1,43	1,38	1,20	1,54	1,47	1,32
JX260	1,16	0,99	0,87	1,20	1,12	0,72	0,86	1,08	0,89	0,71	0,81	0,85
JX272	0,69	0,58	0,62	0,65	0,71	0,74	0,61	0,71	0,70	0,47	0,46	0,55
JX279	0,64	0,57	3,72	1,77	0,73	2,62	2,02	1,01	0,45	0,64	3,42	5,81
JX281	0,19	0,17	0,91	0,43	0,17	0,52	0,44	0,22	0,14	0,16	0,62	1,29
JX301	2,04	3,05	4,83	2,98	2,60	2,97	4,02	3,88	2,85	4,58	5,11	6,33
JX303	1,29	1,28	2,65	1,43	1,45	2,14	1,89	1,53	1,47	2,27	2,53	2,27
JX330	2,55	2,89	2,69	2,81	3,30	3,88	2,82	2,99	2,97	2,76	2,98	3,38
JX338	0,63	0,57	0,58	0,43	0,53	0,68	0,56	0,54	0,66	0,62	0,57	0,70
JX345	0,40	0,32	0,52	0,47	0,54	0,64	0,41	0,72	1,18	0,45	0,44	0,48
JX384	0,27	0,27	0,27	0,26	0,28	0,25	0,40	0,35	0,39	0,29	0,33	0,42
JX392	6,92	3,74	4,70	3,44	4,86	7,06	2,91	3,45	4,84	3,62	3,22	5,13
JX414	5,41	5,55	6,49	6,37	7,08	8,67	7,05	7,33	7,14	7,90	8,29	7,47
JX285	1,14	1,01	1,42	1,39	0,94	1,05	1,17	0,90	0,90	1,03	1,16	1,55
JX398	0,75	0,77	0,72	0,86	0,77	0,51	0,81	0,75	0,86	0,94	0,89	0,83
JX082	0,46	0,55	0,32	0,51	0,33	0,24	0,25	0,31	0,34	0,26	0,26	0,27
JX088	8,81	10,60	5,68	8,07	9,22	10,32	6,15	8,59	10,92	6,71	10,60	7,35
JX100	4,16	3,23	2,00	2,76	3,09	3,42	2,51	3,18	3,78	2,48	2,32	2,12
JX116	1,34	1,36	1,20	1,23	1,30	1,71	0,95	1,14	1,32	1,12	1,10	1,20
JX118	3,02	2,85	2,64	2,73	2,83	3,37	2,72	2,28	2,83	3,49	3,14	3,03
JX150	0,38	0,39	0,30	0,39	0,57	1,15	0,93	0,99	0,47	0,95	0,75	0,88
JX213	0,35	0,27	0,31	0,29	0,43	0,61	0,26	0,29	0,21	0,28	0,23	0,25
JX238	1,88	1,75	1,67	1,88	1,12	0,39	1,74	2,60	3,79	1,97	2,49	2,04
JX268	0,18	0,21	0,34	0,19	0,13	0,15	0,34	0,22	0,20	0,16	0,32	0,45
JX308	1,55	1,50	1,88	1,71	0,97	0,61	1,93	1,50	1,76	1,51	1,37	1,82
JX310	2,13	2,65	1,80	2,16	1,77	1,51	1,61	2,15	2,15	1,48	1,77	1,66
JX327	2,80	2,73	2,78	2,90	2,85	3,02	2,53	2,32	1,92	2,69	2,59	2,82
JX409	2,07	1,62	1,41	1,78	1,84	2,52	1,55	1,98	2,07	1,65	1,78	1,98
JX001	4,72	3,41	1,23	0,70	2,87	4,24	1,29	3,09	5,77	4,22	2,81	1,85
JX004	0,92	0,80	0,59	0,87	0,54	0,63	0,53	0,52	0,59	0,66	0,44	0,46
JX024	1,08	1,20	1,22	1,39	1,10	0,64	1,27	1,11	1,00	1,62	1,24	0,96
JX041	0,30	0,63	1,64	0,38	0,35	0,50	0,49	0,39	0,38	0,41	0,48	0,45
JX050	0,28	0,28	0,19	0,24	0,21	0,08	0,19	0,24	0,16	0,19	0,19	0,15
JX051	0,71	0,55	0,27	0,42	0,49	0,58	0,24	0,31	0,50	0,26	0,33	0,28
JX056	0,45	0,67	0,65	0,87	0,88	0,70	0,72	0,73	0,35	0,74	0,79	0,56
JX058	0,57	0,49	0,45	0,51	0,42	0,31	0,33	0,43	0,34	0,32	0,38	0,26
JX059	2,31	2,72	3,40	3,34	2,26	2,27	3,77	2,69	2,66	4,11	2,50	1,66
JX068	0,27	0,26	0,19	0,15	0,19	0,13	0,19	0,23	0,15	0,15	0,20	0,16
JX070	0,67	0,78	0,59	0,85	0,93	0,89	0,75	0,72	0,72	0,70	0,75	0,54
JX073	3,60	4,12	3,39	3,17	3,99	3,66	3,32	3,30	3,42	4,17	3,18	2,51
JX091	0,70	0,43	0,15	0,22	0,21	0,10	0,17	0,34	0,31	0,19	0,21	0,14
JX110	0,62	0,56	0,54	0,60	0,46	0,41	0,40	0,47	0,46	0,43	0,40	0,33
JX111	1,50	1,17	1,27	0,85	0,99	0,93	1,16	1,19	1,70	1,47	1,32	1,13
JX113	1,90	1,74	1,55	2,10	1,70	1,59	1,43	1,76	1,99	1,52	1,41	1,42
JX115	8,51	8,29	7,17	8,75	9,24	9,94	9,66	9,29	9,07	9,29	8,68	7,85
JX132	0,40	0,29	0,17	0,15	0,13	0,08	0,15	0,17	0,11	0,15	0,15	0,17
JX154	2,83	1,79	0,82	1,00	0,99	1,25	1,46	2,73	2,52	0,94	1,05	0,86
JX161	1,34	1,07	0,95	0,94	0,82	0,84	0,91	1,00	1,46	0,77	0,71	0,94
JX163	0,44	0,38	0,53	0,49	0,44	0,31	0,50	0,39	0,27	0,36	0,36	0,39
JX164	1,54	1,59	1,12	1,37	1,44	1,69	1,34	1,18	1,48	1,13	1,39	1,63
JX169	3,57	3,24	3,64	3,81	3,01	2,61	2,54	2,80	2,50	2,38	2,52	2,09
JX171	0,54	0,53	0,50	0,56	0,62	0,40	0,47	0,46	0,36	0,49	0,44	0,48
JX177	2,93	3,08	1,98	2,69	2,42	2,56	1,88	2,22	2,08	2,54	1,90	1,84

JX178	0,92	0,69	0,56	0,70	0,74	0,67	0,54	0,51	0,65	0,51	0,46	0,40
JX191	0,58	0,55	0,50	0,53	0,60	0,58	0,57	0,51	0,42	0,50	0,44	0,41
JX212	0,67	0,50	0,63	0,55	0,43	0,79	0,82	0,58	0,55	0,47	0,41	0,98
JX214	3,83	3,38	2,24	2,85	3,13	3,10	2,79	2,82	2,64	3,36	2,89	2,53
JX218	8,51	5,41	2,25	2,17	2,56	3,07	6,46	9,14	7,56	2,29	2,51	2,15
JX227	0,89	0,78	0,81	0,33	0,30	0,55	0,37	0,68	0,34	0,75	0,74	0,30
JX234	0,81	0,63	0,60	0,61	0,50	0,75	0,46	0,68	0,72	0,55	0,48	0,41
JX237	1,02	1,04	0,81	0,75	1,01	1,53	0,65	1,28	1,39	1,03	1,24	0,44
JX239	10,71	8,48	3,29	2,92	5,99	7,54	4,65	6,21	9,15	6,08	5,35	1,32
JX246	0,54	0,84	1,40	0,54	0,73	0,71	1,16	1,43	0,86	0,34	0,95	1,30
JX256	1,20	1,19	0,81	1,11	1,07	0,92	0,97	0,94	1,01	0,92	0,80	0,84
JX262	0,37	0,33	0,91	0,63	0,52	0,64	0,75	0,63	0,58	0,54	0,68	0,95
JX267	0,22	0,29	0,14	0,24	0,19	0,20	0,12	0,15	0,16	0,19	0,12	0,13
JX269	0,50	0,77	1,44	0,43	0,45	0,43	0,38	0,50	0,63	0,46	0,54	0,48
JX284	1,25	1,78	1,08	1,46	2,29	2,15	0,86	2,11	1,60	1,30	2,16	0,40
JX286	0,90	2,58	5,33	0,92	2,47	3,98	4,74	2,82	0,95	3,93	4,33	4,70
JX290	1,04	0,92	1,08	1,04	0,77	0,46	0,97	0,90	0,98	0,80	0,75	0,28
JX291	1,51	3,49	4,57	4,81	4,14	1,70	4,58	3,24	1,34	5,03	4,51	4,05
JX292	0,45	0,66	0,72	0,59	0,79	0,80	0,67	0,65	0,63	0,51	0,78	0,93
JX294	0,49	0,46	0,47	0,56	0,56	0,72	0,18	0,17	0,22	0,41	0,59	0,62
JX302	1,53	1,51	1,10	1,15	1,46	0,95	1,14	1,37	1,18	1,01	1,18	0,90
JX306	0,67	0,68	0,67	0,60	0,41	0,21	0,59	0,62	0,52	0,67	0,67	0,62
JX322	0,88	0,91	0,96	0,94	0,94	0,87	0,87	0,98	0,89	0,86	0,80	0,33
JX323	3,06	3,01	2,47	2,94	2,89	3,00	2,50	2,74	2,72	2,70	2,37	2,19
JX333	0,38	0,39	0,49	0,49	0,52	0,58	0,41	0,37	0,36	0,36	0,45	0,44
JX344	4,15	3,65	3,14	3,61	3,70	3,77	3,41	3,66	3,26	4,37	3,92	3,19
JX351	0,23	0,49	0,61	0,67	0,51	0,54	0,58	0,42	0,25	0,24	0,43	0,61
JX357	1,10	1,31	1,23	1,86	2,02	2,33	1,01	1,45	2,09	1,08	0,90	0,77
JX367	0,76	0,87	1,10	0,92	0,92	0,59	1,02	0,83	0,87	1,15	0,97	0,78
JX372	0,26	0,28	0,35	0,31	0,34	0,45	0,35	0,30	0,31	0,29	0,36	0,38
JX394	3,66	2,33	1,73	1,58	2,03	3,51	1,02	1,35	1,13	1,22	1,37	1,25
JX419	2,76	3,88	3,54	2,39	3,92	4,77	1,71	4,67	6,43	5,21	4,60	1,03
JX420	1,78	1,88	1,44	2,02	2,14	1,41	1,57	2,01	1,57	1,43	1,39	1,35
JX009	1,58	1,31	1,24	1,50	1,25	1,15	1,32	1,06	1,16	0,94	1,13	1,58
JX022	3,03	3,19	2,64	2,69	3,52	4,74	2,34	2,56	2,49	3,47	2,60	3,97
JX047	0,78	0,66	0,24	0,63	0,75	0,51	0,31	0,35	0,26	0,31	0,28	0,36
JX063	0,48	0,85	0,70	0,89	0,82	0,77	1,03	0,63	0,46	0,65	0,83	0,90
JX064	2,31	3,07	2,14	2,09	2,44	3,13	2,17	2,35	2,21	1,73	2,60	3,36
JX099	0,96	0,63	0,54	1,25	0,80	0,65	0,66	0,61	0,58	0,83	0,99	0,93
JX103	0,55	0,33	0,25	0,27	0,26	0,25	0,22	0,39	0,50	0,29	0,22	0,21
JX188	3,41	3,36	2,65	2,64	3,21	3,26	2,78	3,00	3,08	2,43	2,64	3,10
JX226	0,19	0,18	0,19	0,13	0,17	0,16	0,13	0,11	0,13	0,18	0,19	0,19
JX309	0,63	0,87	1,32	0,17	0,87	1,60	1,25	1,22	1,15	1,51	1,46	1,79
JX317	0,41	0,43	0,60	0,59	0,46	0,43	0,76	0,53	0,44	0,64	0,58	0,46
JX350	1,04	0,83	0,71	0,87	0,72	0,63	0,67	0,77	1,69	0,79	0,80	0,74
JX370	8,26	9,00	6,94	9,37	8,73	8,30	9,05	6,90	11,24	6,89	8,89	8,37
JX407	6,86	7,40	6,71	7,32	5,36	4,58	6,32	7,28	7,77	5,99	5,92	5,29
JX002	0,13	0,12	0,17	0,11	0,17	0,16	0,13	0,13	0,12	0,26	0,16	0,18
JX049	16,97	15,34	11,85	12,69	15,66	18,69	12,62	12,88	15,50	16,75	19,09	18,02
JX055	1,33	1,05	2,18	1,58	1,40	1,91	1,56	1,26	1,35	1,39	1,56	1,70
JX083	31,04	23,81	23,53	24,13	30,27	38,61	23,97	23,21	25,96	31,76	33,54	35,35
JX084	1,25	0,77	0,59	0,48	0,42	0,47	0,52	0,89	1,26	0,41	0,41	0,56
JX085	4,38	2,08	0,25	0,31	0,36	0,42	0,41	0,39	0,40	0,39	0,55	0,53
JX086	3,73	1,84	1,99	0,98	2,10	5,48	1,29	1,35	1,42	2,45	2,35	2,60
JX092	6,78	5,56	4,38	5,58	5,06	6,80	4,15	5,91	6,29	8,29	6,22	3,75
JX093	2,29	2,22	2,64	3,07	2,22	3,34	2,17	2,56	2,70	4,70	3,12	3,49
JX112	1,89	1,37	1,03	1,33	1,18	1,36	0,95	0,99	0,96	1,05	0,98	0,95
JX134	6,88	3,65	6,55	2,61	4,36	6,96	2,34	3,65	4,61	6,75	4,51	4,30
JX135	3,46	3,09	2,00	3,24	3,30	3,33	2,24	2,95	3,85	2,17	2,91	3,15
JX145	2,74	2,16	1,83	1,59	1,89	1,86	1,31	1,28	1,29	1,44	1,40	1,34
JX149	0,56	0,49	0,26	0,31	0,34	0,53	0,28	0,33	0,41	0,31	0,39	0,40
JX165	0,53	0,76	1,31	1,37	1,05	1,14	1,23	0,67	0,55	1,50	1,05	1,04
JX176	0,19	0,18	0,15	0,15	0,22	0,27	0,16	0,15	0,16	0,21	0,22	0,16
JX195	6,27	7,12	6,13	8,13	7,47	7,83	6,38	6,38	8,13	7,29	6,55	6,29
JX204	5,79	4,96	4,20	4,50	5,28	6,11	4,75	4,42	5,02	4,88	5,31	5,47
JX210	12,13	7,02	5,11	5,55	9,72	15,57	4,98	4,80	2,50	6,89	7,93	10,33
JX278	0,85	2,41	5,62	0,36	0,58	0,63	0,42	0,61	0,50	0,54	0,44	0,51
JX282	8,72	7,49	8,96	8,07	7,18	10,85	8,31	9,27	9,41	12,71	9,20	7,70
JX293	0,32	0,51	0,94	1,12	0,83	1,15	0,92	0,63	0,26	1,31	0,91	0,80
JX328	0,40	0,45	0,46	0,27	0,40	0,29	0,33	0,32	0,28	0,35	0,38	0,44
JX356	0,41	0,35	0,45	0,36	0,34	0,35	0,27	0,34	0,45	0,26	0,27	0,26
JX358	0,42	0,41	0,43	0,49	0,49	0,36	0,51	0,39	0,43	0,55	1,05	0,49
JX374	3,53	3,18	2,17	2,45	2,55	2,67	1,70	2,13	1,70	1,99	2,14	2,22
JX377	6,44	5,23	4,11	4,14	4,58	5,99	4,50	3,41	4,47	4,85	4,79	5,20
JX388	0,91	1,09	1,23	0,95	0,92	0,84	1,12	1,03	0,99	1,09	1,11	1,37
JX418	0,54	0,66	0,34	0,53	1,04	0,64	0,25	0,67	1,40	0,33	0,28	0,27
JX424	0,60	1,43	1,54	0,77	2,15	1,66	0,69	1,39	0,55	0,48	0,94	0,77
JX043	0,56	0,70	0,51	0,69	0,89	0,83	0,72	0,48	0,39	0,60	0,32	0,36
JX312	1,17	1,05	0,98	1,01	1,19	1,44	1,01	1,09	0,85	0,99	1,08	1,11
JX005	0,67	1,18	0,62	1,45	1,09	0,80	1,57	1,59	2,18	1,49	1,74	1,66
JX133	2,11	1,14	0,42	0,50	1,20	1,97	1,87	1,92	2,33	0,48	0,53	0,71
JX369	1,74	2,15	1,61	1,39	1,41	1,14	1,68	1,61	1,42	1,85	1,71	2,23
JX273	1,23	1,62	1,55	1,82	1,54	0,77	1,98	1,87	2,17	1,85	2,27	2,01
JX321	0,38	0,32	0,54	0,50	0,46	0,53	0,54	0,09	0,45	0,28	0,45	0,46
JX368	1,17	1,59	1,70	2,12	2,01	1,28	2,11	2,10	2,44	2,25	2,31	1,84
JX035	1,43	0,67	0,48	0,45	0,52	0,75	0,48	0,59	0,89	0,57	0,50	0,41
JX075	0,29	0,36	0,33	0,47	0,35	0,36	0,39	0,33	0,48	0,49	0,39	0,39
JX089	3,80	1,12	1,36	1,70	1,68	1,03	2,47	2,36	1,88	3,09	2,36	1,28
JX184	0,29	0,36	0,46	0,28	0,34	0,36	0,46	0,35	0,27	0,29	0,26	0,53
JX185	3,54	3,36	2,75	3,10	3,14	3,44	2,68	1,94	2,60	3,87	3,91	3,51
JX223	0,19	0,37	0,49	0,49	0,44	0,48	0,46	0,36	0,41	0,48	0,45	0,47
JX034	5,46	5,85	5,62	6,12	4,68	2,53	5,54	4,78	2,76	7,16	5,86	4,38
JX180	0,53	0,45	0,45	0,33	0,57	0,72	0,45	0,48	0,51	0,46	0,44	0,50

Supplemental table S5 Function description and variation analysis of 424 protein spots among 12 genotypes at cell expansion stage (CE) and orange red (OR) stage. Significant spots are highlight in yellow.

Spot ID	ITAG accession numbe	Position on genome (bp)	Protein function description	Protein classification	y ANOVA_genoty	ay ANOVA_stage	genotype and s	pe effect_Cy	pe effect_OR stage
JX001	Solyc02g080630.2.1	39392539-39397791	Lactoylglutathione lyase	primary metabolic process	7,06E-26	1,24E-08	5,34E-04	2,12E-12	1,93E-12
JX002	Solyc09g007150.2.1	775043-778675	Glutathione S-transferase	response to stress	1,42E-10	4,44E-28	7,09E-09	3,84E-07	1,53E-04
JX003	Solyc12g010020.1.1	3152019-3154797	Leucyl aminopeptidase	response to stress	3,11E-02	6,06E-06	7,64E-03	1,30E-04	7,33E-01
JX004	Solyc07g062650.2.1	62561877-62565142	Malate dehydrogenase	primary metabolic process	1,96E-04	1,41E-23	4,07E-01	1,56E-01	3,34E-04
JX005	Solyc01g057830.2.1	56934925-56939761	30S ribosomal protein S1	translation	2,51E-08	5,43E-02	2,32E-04	7,54E-01	3,44E-08
JX006	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis	6,83E-10	2,46E-15	5,88E-09	5,07E-09	4,73E-01
JX007	Solyc05g054760.2.1	63753656-63758079	Dehydroascorbate reductase	vitamin synthesis	2,10E-06	9,87E-06	9,49E-03	9,69E-06	1,44E-02
JX008	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process	1,81E-05	2,86E-22	6,52E-02	1,98E-03	8,52E-03
JX009	Solyc02g065400.2.1	31151784-31154764	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process	3,78E-06	6,73E-21	8,00E-03	3,14E-03	4,24E-05
JX010	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization	2,28E-12	4,12E-01	6,93E-01	1,66E-08	6,12E-04
JX011	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process	6,00E-02	4,31E-09	1,18E-01	1,53E-01	9,08E-02
JX012	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process	1,01E-08	5,94E-10	1,74E-05	1,12E-06	3,14E-03
JX013	Solyc08g065900.2.1	51450546-51456933	Charged multivesicular body protein 4b	establishment of localization	2,98E-01	2,96E-09	5,83E-01	1,98E-01	6,70E-01
JX014	Solyc07g064880.2.1	64080741-64083209	Small ubiquitin-related modifier	macromolecule metabolic process	5,25E-13	3,04E-14	1,04E-10	3,08E-09	4,14E-01
JX015	Solyc06g073190.2.1	41483219-41485991	Fructokinase-like	primary metabolic process	4,20E-03	6,69E-30	4,30E-03	7,08E-03	8,67E-01
JX016	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress	6,53E-02	4,65E-04	1,36E-01	4,37E-02	2,52E-01
JX017	Solyc07g064160.2.1	63644226-63646402	Thiazole biosynthetic enzyme	vitamin synthesis	5,25E-02	3,93E-09	4,83E-02	1,09E-06	2,89E-01
JX018	Solyc05g018700.2.1	22827490-22832816	Protein disulfide isomerase	macromolecule metabolic process	5,70E-09	3,66E-01	6,74E-05	2,35E-02	3,96E-07
JX019	Solyc02g079500.2.1	38618009-38620468	Peroxidase	cell wall organisation or biogenesis	2,35E-06	8,74E-26	1,69E-03	1,25E-02	1,52E-04
JX020	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization	4,72E-11	4,17E-01	6,75E-02	9,24E-06	3,20E-05
JX021	Solyc12g099000.1.1	64658262-64659443	S-adenosylmethionine synthase	hormone metabolic process	3,85E-02	1,32E-16	9,07E-02	4,17E-04	3,50E-01
JX022	Solyc01g111300.2.1	89348859-89349828	Cold shock protein-1	regulation of biological process	2,87E-02	5,71E-16	6,87E-04	5,70E-02	1,98E-04
JX023	Solyc01g101060.2.1	82678819-82681422	S-adenosylmethionine synthase	hormone metabolic process	9,21E-08	3,83E-05	7,34E-05	1,68E-05	3,11E-03
JX024	Solyc09g007940.2.1	1440058-1444221	Adenosine kinase	primary metabolic process	8,98E-08	3,97E-02	4,72E-03	3,65E-05	8,00E-04
JX025	Solyc08g067160.2.1	53325912-53329825	Acyl-protein thioesterase 2	macromolecule metabolic process	2,01E-06	4,86E-05	6,40E-02	1,81E-02	5,18E-05
JX026	Solyc09g082060.2.1	63298831-63303187	Cysteine synthase	primary metabolic process	2,60E-03	3,46E-13	2,26E-01	2,85E-02	7,55E-02
JX027	Solyc08g076990.2.1	58097110-58102792	Acetylornithine deacetylase	primary metabolic process	3,05E-01	1,70E-09	5,88E-01	3,76E-03	9,36E-01
JX028	Solyc02g080420.2.1	39222034-39228232	RNA Binding Protein 45	regulation of biological process	2,05E-02	2,19E-01	3,31E-01	4,16E-02	2,33E-01
JX029	Solyc08g075160.2.1	56470989-56477178	Bifunctional purine biosynthesis protein purH	primary metabolic process	1,71E-02	6,62E-14	1,13E-02	7,69E-03	2,81E-01
JX030	Solyc03g078400.2.1	44381441-44383329	Actin	cytoskeleton organization and biogenesis	5,95E-03	3,88E-15	1,17E-03	2,78E-02	4,19E-03
JX031	Solyc08g076970.2.1	58078222-58080829	Acetylornithine deacetylase or succinyl-diaminopimelate desucci	macromolecule metabolic process	1,50E-03	2,47E-01	6,72E-01	3,71E-02	1,13E-01
JX032	Solyc01g094200.2.1	77521221-77531982	NAD-dependent malic enzyme 2	primary metabolic process	1,82E-02	7,54E-14	7,63E-05	3,24E-03	6,36E-03
JX033	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process	3,87E-02	2,18E-05	2,31E-02	8,70E-03	1,10E-01
JX034	Solyc05g054760.2.1	63753656-63758079	Dehydroascorbate reductase	vitamin synthesis	4,95E-12	8,06E-16	1,58E-01	4,37E-05	2,82E-07
JX035	Solyc02g078540.2.1	37780626-37784558	Unknown Protein	Unknown	4,72E-08	4,61E-03	4,06E-08	6,31E-01	3,53E-11
JX036	Solyc07g049530.2.1	57154220-57156391	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation	1,21E-03	7,60E-16	1,59E-03	2,12E-01	5,35E-03
JX037	Solyc07g053310.2.1	59102795-59108445	Adaptin ear-binding coat-associated protein 1	establishment of localization	4,81E-08	6,65E-24	7,30E-08	1,33E-02	1,90E-07
JX038	Solyc02g036350.2.1	21279767-21281954	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation	3,12E-02	5,32E-06	5,96E-02	5,30E-09	1,13E-01
JX039	Solyc07g049530.2.1	57154220-57156391	1-aminocyclopropane-1-carboxylate oxidase	developmental maturation	3,08E-08	5,71E-23	9,77E-08	8,65E-02	2,61E-06
JX040	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation	2,85E-10	4,70E-28	1,22E-07	2,11E-04	4,66E-08
JX041	Solyc09g013080.2.1	5484876-5491347	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpl	primary metabolic process	5,15E-05	1,63E-01	4,76E-01	1,22E-01	2,15E-03
JX042	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization	8,22E-16	1,22E-03	1,69E-01	2,66E-08	6,09E-07
JX043	Solyc11g013110.1.1	5961440-5965680	Anthocyanidin synthase	secondary metabolic process	4,41E-05	4,70E-19	5,68E-04	3,99E-02	9,07E-04
JX044	Solyc02g079500.2.1	38618009-38620468	Peroxidase	cell wall organisation or biogenesis	3,64E-03	1,08E-26	2,11E-03	6,72E-02	5,61E-03
JX045	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	3,90E-02	4,22E-11	5,40E-02	5,57E-01	6,72E-02
JX046	Solyc10g080210.1.1	60883700-60890335	Polygalacturonase A	macromolecule metabolic process	1,89E-07	3,50E-25	5,16E-06	4,46E-03	2,53E-05
JX047	Solyc03g097270.2.1	53034151-53040369	Cysteine proteinase inhibitor	regulation of biological process	6,40E-08	2,60E-01	5,22E-01	6,70E-05	4,08E-03
JX048	Solyc04g082200.2.1	63550865-63552237	Dehydrin	response to stress	4,68E-04	9,33E-14	1,02E-02	3,16E-03	1,18E-02
JX049	Solyc06g076570.1.1	43954001-43954465	class I heat shock protein	response to stress	1,82E-04	1,49E-26	8,42E-04	1,35E-05	4,56E-03
JX050	Solyc02g081160.2.1	39802521-39809198	Diphosphate-fructose-6-phosphate 1-phosphotransferase	primary metabolic process	2,69E-03	1,67E-02	1,97E-02	2,75E-01	1,03E-04
JX051	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	1,23E-04	5,98E-05	1,30E-04	1,17E-01	3,16E-04
JX052	Solyc04g011440.2.1	3894918-3898067	heat shock protein	response to stress	1,85E-02	9,57E-30	4,46E-02	3,88E-03	8,39E-02
JX053	Solyc01g090700.2.1	76099361-76107876	Enoyl-CoA-hydratase	primary metabolic process	4,69E-02	9,45E-08	1,43E-01	7,60E-03	1,23E-01
JX054	Solyc05g056310.2.1	64792047-64799309	T-complex protein 1 subunit gamma	macromolecule metabolic process	2,52E-01	3,70E-08	3,81E-02	3,56E-02	1,88E-01
JX055	Solyc02g079930.2.1	38895029-38898734	Phosphosulfolactate synthase	response to stress	7,75E-06	8,72E-19	2,09E-04	1,12E-02	3,00E-04
JX056	Solyc01g099190.2.1	81251679-81256014	Lipoxygenase	primary metabolic process	3,30E-01	8,40E-14	1,22E-08	1,92E-04	4,79E-04
JX057	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process	1,41E-01	6,95E-16	7,25E-03	8,85E-02	1,47E-02
JX058	Solyc01g005560.2.1	394402-399248	Isocitrate dehydrogenase	primary metabolic process	4,31E-02	1,19E-06	1,39E-04	1,47E-01	4,59E-04
JX059	Solyc06g083790.2.1	45398741-45407021	Succinyl-CoA ligase	primary metabolic process	4,45E-05	9,99E-01	7,79E-01	7,44E-02	3,15E-03
JX060	Solyc02g078360.2.1	37652619-37657447	Thioredoxin family protein	oxidation-reduction process	8,90E-03	6,45E-04	1,86E-01	7,28E-02	6,14E-02
JX061	Solyc05g013990.2.1	7492470-7499923	T-complex protein 1 subunit epsilon	macromolecule metabolic process	1,83E-04	3,38E-01	8,63E-02	3,40E-05	3,13E-01
JX062	Solyc10g005650.2.1	519710-532795	Peroxisomal targeting signal 1 receptor	establishment of localization	4,48E-02	2,13E-08	7,39E-03	6,60E-03	9,26E-02
JX063	Solyc02g091840.2.1	47648877-47652967	Receptor like kinase. RLK	regulation of biological process	6,75E-08	6,47E-01	2,82E-02	1,53E-02	1,17E-06
JX064	Solyc03g025950.2.1	7764256-7767163	Membrane-associated progesterone receptor component 1	regulation of biological process	4,79E-05	4,22E-04	7,27E-03	1,17E-02	2,14E-04
JX065	Solyc01g111760.2.1	89701090-89706839	V-type ATP synthase beta chain	establishment of localization	1,61E-01	8,16E-02	6,94E-03	6,21E-02	4,27E-02
JX066	Solyc12g042060.1.1	42533149-42537576	ATP-dependent clp protease ATP-binding subunit	establishment of localization	7,05E-03	6,87E-01	3,68E-02	6,65E-02	1,30E-02
JX067	Solyc07g049450.2.1	57040844-57048137	Thioredoxin/protein disulfide isomerase	macromolecule metabolic process	2,49E-02	5,50E-02	1,63E-01	2,60E-01	2,11E-02
JX068	Solyc02g081160.2.1	39802521-39809198	Diphosphate-fructose-6-phosphate 1-phosphotransferase	primary metabolic process	9,12E-03	7,31E-03	7,41E-02	5,27E-01	3,94E-03
JX069	Solyc01g111760.2.1	89701090-89706839	V-type ATP synthase beta chain	establishment of localization	5,70E-03	6,86E-03	6,02E-04	1,46E-04	3,89E-02
JX070	Solyc02g083590.2.1	41515715-41520837	Dehydroquinate synthase	primary metabolic process	6,90E-04	1,03E-13	3,94E-04	3,62E-01	5,38E-06
JX071	Solyc05g008450.2.1	2800495-2805716	Oxidoreductase FAD/NAD	oxidation-reduction process	3,30E-02	1,55E-02	7,21E-01	2,41E-01	2,13E-01
JX072	Solyc11g066100.1.1	48856641-48858939	heat shock protein	response to stress	2,20E-02	1,24E-06	5,46E-01	6,51E-02	3,63E-01
JX073	Solyc09g090140.2.1	65031288-65034684	Malate dehydrogenase	primary metabolic process	7,71E-05	5,47E-11	6,27E-04	3,26E-03	1,08E-03
JX074	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress	2,81E-05	5,95E-01	9,44E-02	1,70E-04	4,65E-02
JX075	Solyc04g076820.1.1	59289748-59291172	Octicosapeptide/Phox/Bem1p domain-containing protein	Unknown	5,59E-06	2,93E-04	1,36E-01	4,62E-02	2,17E-04
JX076	Solyc12g044600.2.1	45079934-45087825	NADP-dependent malic enzyme. chloroplastic	primary metabolic process	4,27E-02	4,53E-01	8,15E-02	1,12E-01	6,01E-02
JX077	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress	1,24E-03	4,01E-03	2,61E-01	7,88E-02	8,27E-03
JX078	Solyc12g008630.1.1	2007849-2015328	Mitochondrial processing peptidase alpha subunit	macromolecule metabolic process	4,06E-01	1,29E-05	4,75E-01	4,43E-01	4,81E-01
JX079	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress	2,37E-01	3,44E-04	4,93E-01	1,23E-01	6,59E-01
JX080	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress	2,30E-02	3,27E-02	5,38E-01	2,54E-01	1,03E-01
JX081	Solyc10g081240.1.1	61667783-61677144	Protein grpE	primary metabolic process	7,50E-01	4,83E-10	2,94E-01	8,12E-01	4,51E-01
JX082	Solyc09g082720.2.1	63816287-63819690	Aldo/keto reductase family protein	oxidation-reduction process	1,31E-10	1,76E-10	1,48E-06	5,37E-03	9,23E-08
JX083	Solyc06g076560.1.1	43949614-43950078	class I heat shock protein	response to stress	7,75E-05	4,80E-32	5,01E-05	4,05E-01	2,39E-04
JX084	Solyc10g078930.1.1	59896781-59900058	Activator of heat shock protein ATPase homolog 1	response to stress	1,75E-13	2,29E-12	4,33E-04	1,09E-04	3,45E-12
JX085	Solyc09g015000.2.1	7427223-7428264	class I heat shock protein	response to stress	8,68E-25	2,12E-16	3,14E-21	8,07E-04	9,46E-15
JX086	Solyc03g113930.1.1	58031590-58032156	class IV heat shock protein	response to stress	8,42E-13	6,58E-26	7,88E-12	3,23E-01	1,25E-08
JX087	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	1,41E-03	4,45E-05	6,84E-03	2,12E-02	9,43E-03
JX088	Solyc03g111720.2.1	56431489-56432545	Peptide methionine sulfoxide reductase msrA	oxidation-reduction process	4,69E-06	3,39E-35	1,03E-05	6,44E-01	9,81E-05
JX089	Solyc01g111040.2.1	89209383-89212670	EF-Hand containing protein-like	Unknown	5,57E-08	8,50E-21	2,61E-06	2,56E-01	8,44E-06
JX090	Solyc04g082200.2.1	63550865-63552237	Dehydrin	response to stress	9,10E-03	2,61E-11	1,50E-03	4,99E-03	1,21E-02
JX091	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process	5,97E-10	1,18E-04	4,21E-06	9,19E-02	1,79E-07
JX092	Solyc08g062340.2.1	48122173-48122945	Class II small heat shock protein Le-HSP17.6	response to stress	4,68E-07	5,63E-31	1,95E-07	1,99E-05	2,24E-05
JX093	Solyc08g078700.2.1	59635844-59637072	Heat shock protein 22	response to stress	4,42E-03	3,88E-17	6,94E-05	2,07E-03	2,93E-03
JX094	Solyc02g090030.2.1	46276521-46278207	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process	2,89E-04	1,54E-30	2,11E-01	4,00E-03	6,64E-02
JX095	Solyc12g088720.1.1	62418023-62421876	Polyadenylate-binding protein 2	macromolecule metabolic process	6,52E-04	8,44E-02	6,45E-01	1,65E-02	6,42E-02
JX096	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	regulation of biological process	2,78E-09	1,69E-01	9,57E-02	3,63E-09	2,06E-02
JX097	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress	1,85E-02	2,01E-29	2,70E-01	3,99E-03	3,64E-01
JX098	Solyc10g083650.1.1	62763114-62763800	Peroxiredoxin ahpC/TSA family	oxidation-reduction process	2,76E-04	1,51E-09	2,54E-02	7,61E-06	1,57E-02
JX099	Solyc02g077710.1.1	37172109-37173137	E6-2 protein kinase	regulation of biological process	1,44E-06	1,09E-15	2,44E-04	3,20E-01	2,08E-05
JX100	Solyc10g083650.1.1	62763114-62763800	Peroxiredoxin ahpC/TSA family	oxidation-reduction process	6,47E-08	4,24E-18	2,45E-03	8,81E-04	1,22E-04
JX101	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis	3,83E-03	8,56E-02	1,70E-10	1,64E-07	1,78E-04
JX102	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process	1,62E-14	7,07E-04	5,75E-07	7,62E-06	1,10E-08
JX103	Solyc08g075210.1.1	56497821-56499095	Acyltransferase-like protein	regulation of biological process	6,06E-08	8,60E-03	2,19E-02	1,77E-02	1,55E-05
JX104	Solyc01g104950.2.1	85031292-85035396	Alpha-L-arabinofuranosidase/beta-D-xylosidase	cell wall organisation or biogenesis	1,42E-03	1,55E-10	3,41E-03	1,10E-03	2,73E-02
JX105	Solyc02g065400.2.1	31151784-31154764	Oxygen-evolving enhancer protein 1 of photosystem II	regulation of biological process	1,68E-02	3,48E-17	3,10E-01	1,50E-01	2,64E-02
JX106	Solyc01g110450.2.1	88910585-88913966	NADP dependent sorbitol 6-phosphate dehydrogenase	primary metabolic process	9,04E-05	3,82E-04	2,71E-01	1,61E-05	2,65E-01
JX107	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process	1,94E-02	1,35E-17	6,79E-05	6,12E-04	2,55E-01
JX108	Solyc05g055160.2.1	64069121-64073511	DNAJ chaperone	regulation of biological process	2,21E-03	8,84E-13	7,06E-04	3,65E-05	7,99E-01
JX109	Solyc09g011030.2.1	4368272-4373182	Hsp70 nucleotide exchange factor fes1	regulation of biological process	3,15E-01	4,06E-09	1,05E-01	2,89E-01	2,10E-01
JX110	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process	5,20E-04	8,95E-06	1,11E-01	1,85E-01	3,88E-03
JX111	Solyc12g055830.1.1	47184732-47187107	Inorganic pyrophosphatase	primary metabolic process	2,60E-09	6,88E-01	2,11E-03	1,98E-03	3,90E-07
JX112	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress	8,29E-11	1,64E-08	1,09E-02	2,54E-03	1,57E-07
JX113	Solyc07g066600.2.1	65205704-65208684	Phosphoglycerate kinase	primary metabolic process	7,93E-07	8,92E-05	4,83E-03	2,34E-04	1,96E-03
JX114	Solyc02g081700.1.1	40134968-40135714	Proteasome subunit alpha type	macromolecule metabolic process	1,27E-02	1,96E-02	2,46E-02	1,50E-02	5,06E-01
JX115	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process	1,53E-03	2,70E-17	4,92E-02	5,73E-02	1,93E-03
JX116	Solyc02g086880.2.1	44063088-44067604	Formate dehydrogenase	oxidation-reduction process	1,59E-06	9,55E-01	1,55E-01	1,21E-03	3,98E-03
JX117	Solyc07g064800.2.1	64009800-64015982	Dihydrolipoyllysine-residue succinyltransferase component of 2-	macromolecule metabolic process	7,94E-01	1,15E-11	3,55E-01	5,56E-01	5,76E-01

JX118	Solyc06g005150.2.1	170218-173338	Ascorbate peroxidase	oxidation-reduction process	8,03E-03	2,21E-13	2,34E-02	1,41E-01	4,76E-03
JX119	Solyc06g083190.2.1	45004847-45008910	Peptidyl-prolyl cis-trans isomerase	macromolecule metabolic process	6,88E-02	5,15E-13	6,23E-02	6,79E-01	1,96E-02
JX120	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process	8,82E-05	3,38E-06	7,02E-04	3,50E-05	1,47E-01
JX121	Solyc10g005100.2.1	92148-94188	Salt stress root protein RS1	response to stress	2,93E-01	1,84E-04	7,55E-01	4,14E-01	5,55E-01
JX122	Solyc12g005080.1.1	38442-41171	Dihydrolipoyllysine-residue succinyltransferase component of 2-o	macromolecule metabolic process	4,39E-02	1,25E-11	1,22E-03	2,23E-01	2,36E-03
JX123	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process	4,18E-01	1,11E-23	4,39E-01	5,70E-01	2,75E-01
JX124	Solyc05g014470.2.1	8322381-8324881	Glyceraldehyde 3-phosphate dehydrogenase	primary metabolic process	7,49E-02	4,71E-14	9,99E-01	4,17E-01	5,49E-01
JX125	Solyc09g098150.2.1	67309562-67312625	Metacaspase 7	macromolecule metabolic process	4,22E-03	2,28E-06	1,15E-03	3,55E-02	1,51E-03
JX126	Solyc01g104170.2.1	84393496-84398307	Ankyrin repeat domain-containing protein 2	regulation of biological process	6,35E-02	2,36E-02	1,96E-01	6,39E-02	8,31E-01
JX127	Solyc10g084050.1.1	63048965-63053841	26S protease regulatory subunit 6B homolog	macromolecule metabolic process	3,34E-02	4,44E-04	3,54E-02	6,78E-01	5,20E-04
JX128	Solyc01g111120.2.1	89253979-89260194	Triosephosphate isomerase	primary metabolic process	3,28E-02	6,75E-08	4,79E-01	9,05E-01	1,69E-02
JX129	Solyc01g099760.2.1	81668247-81673265	26S protease regulatory subunit 6A homolog	macromolecule metabolic process	4,73E-01	2,02E-06	7,43E-01	8,66E-01	2,67E-01
JX130	Solyc05g005700.2.1	514498-518886	Aldehyde dehydrogenase 1	primary metabolic process	6,08E-01	2,74E-03	8,58E-01	7,65E-01	7,14E-01
JX131	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process	7,23E-01	5,68E-01	4,12E-01	2,22E-01	8,52E-01
JX132	Solyc02g091100.2.1	47093169-47096847	Oxalyl-CoA decarboxylase	primary metabolic process	1,33E-09	4,64E-11	2,90E-05	2,94E-02	2,63E-06
JX133	Solyc03g115650.2.1	59340664-59343595	Eukaryotic translation initiation factor 5A	translation	1,49E-24	4,70E-07	9,25E-05	2,50E-08	6,10E-15
JX134	Solyc01g102960.2.1	83371681-83372560	class IV heat shock protein	response to stress	1,40E-04	1,23E-18	1,51E-03	2,22E-01	2,28E-03
JX135	Solyc01g057000.2.1	50798585-50800986	Universal stress protein family protein	response to stress	2,22E-04	2,82E-26	3,18E-02	3,88E-01	3,28E-03
JX136	Solyc02g081170.2.1	39810752-39812645	Plastid-lipid-associated protein. chloroplastic	response to stress	5,72E-03	1,31E-22	2,08E-01	4,20E-01	3,46E-02
JX137	Solyc06g005940.2.1	921114-925482	Protein disulfide isomerase	macromolecule metabolic process	1,06E-02	4,17E-01	2,55E-01	3,10E-05	2,07E-01
JX138	Solyc02g092670.1.1	48256556-48258820	Subtilisin-like protease	macromolecule metabolic process	4,86E-10	1,29E-21	2,01E-07	4,96E-02	6,08E-08
JX139	Solyc05g056230.2.1	64732387-64738654	Calreticulin 2 calcium-binding protein	macromolecule metabolic process	6,28E-07	1,13E-03	9,96E-01	6,18E-05	3,73E-02
JX140	Solyc03g082920.2.1	46341372-46345339	Heat shock protein	response to stress	2,86E-05	6,08E-04	1,42E-01	4,24E-04	5,86E-02
JX141	Solyc03g082920.2.1	46341372-46345339	Heat shock protein	response to stress	3,64E-04	7,80E-06	1,88E-01	1,86E-02	2,06E-02
JX142	Solyc09g098150.2.1	67309562-67312625	Metacaspase 7	macromolecule metabolic process	3,70E-04	1,65E-08	8,77E-04	3,90E-03	2,40E-03
JX143	Solyc04g040180.2.1	31093421-31095617	S-adenosylmethionine-dependent methyltransferase	hormone metabolic process	6,60E-02	8,50E-14	1,85E-04	4,80E-01	6,72E-04
JX144	Solyc01g108540.2.1	87591739-87592913	Acetyl esterase	primary metabolic process	4,93E-01	5,09E-14	5,50E-01	5,64E-03	5,48E-01
JX145	Solyc02g031950.2.1	17912940-17915049	Pathogenesis-related protein-like protein	response to stress	1,11E-10	5,57E-09	2,53E-03	6,67E-05	1,76E-07
JX146	no identification				1,10E-13	8,82E-09	2,15E-04	2,28E-11	6,61E-03
JX147	Solyc10g081030.1.1	61534917-61538284	Nascent polypeptide-associated complex alpha subunit-like protei	establishment of localization	6,67E-04	1,22E-05	6,91E-02	1,36E-01	1,96E-04
JX148	Solyc00g187050.2.1	18478269-18481034	Leucyl aminopeptidase	macromolecule metabolic process	7,96E-09	9,65E-14	2,77E-06	2,15E-07	3,02E-01
JX149	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress	5,03E-10	6,28E-02	1,94E-03	2,29E-06	3,00E-03
JX150	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process	2,38E-06	6,57E-04	5,79E-03	3,36E-03	2,55E-04
JX151	Solyc10g085550.1.1	63996015-64000623	Enolase	primary metabolic process	1,42E-02	5,61E-16	2,60E-03	1,57E-02	1,56E-02
JX152	Solyc06g075010.2.1	42927697-42932331	chaperonin	macromolecule metabolic process	7,96E-08	1,12E-29	1,11E-03	2,21E-04	6,77E-05
JX153	Solyc03g121640.2.1	63788759-63796170	chaperonin	macromolecule metabolic process	1,14E-09	3,93E-03	6,99E-01	4,18E-05	5,29E-04
JX154	Solyc03g115990.1.1	59594180-59595418	Malate dehydrogenase	primary metabolic process	1,11E-09	1,26E-02	3,94E-04	7,05E-03	6,55E-06
JX155	Solyc03g114500.2.1	58538109-58542525	Enolase	primary metabolic process	5,84E-03	4,65E-20	4,50E-01	9,45E-02	7,87E-02
JX156	Solyc01g097340.2.1	79999429-80002633	NAD-dependent epimerase/dehydratase family protein-like protei	cell wall organisation or biogenesis	4,95E-03	7,82E-20	4,42E-01	2,24E-01	5,14E-03
JX157	Solyc08g082430.2.1	62404261-62407246	Nucleoside diphosphate kinase	primary metabolic process	3,44E-07	2,37E-20	2,74E-01	2,31E-05	1,60E-02
JX158	Solyc12g098420.1.1	64267844-64273580	Ubiquitin carboxyl-terminal hydrolase	macromolecule metabolic process	4,45E-12	6,15E-10	4,83E-02	1,11E-06	1,89E-04
JX159	Solyc12g096190.1.1	63553620-63556459	Tryptophan synthase beta chain	primary metabolic process	2,01E-01	9,61E-08	6,03E-01	2,10E-01	7,25E-01
JX160	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process	4,86E-04	2,92E-12	2,14E-02	1,43E-02	4,78E-03
JX161	Solyc03g114500.2.1	58538109-58542525	Enolase	primary metabolic process	1,00E-06	1,73E-23	1,82E-04	1,12E-02	2,43E-07
JX162	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	5,06E-05	1,78E-03	8,34E-03	1,55E-05	2,49E-01
JX163	Solyc04g082630.2.1	63844019-63847151	Glyceraldehyde-3-phosphate dehydrogenase B	primary metabolic process	6,58E-02	1,41E-13	6,26E-01	8,11E-01	1,22E-04
JX164	Solyc01g106430.2.1	86088310-86092110	Inorganic pyrophosphatase family protein	primary metabolic process	1,61E-02	3,20E-27	2,13E-02	1,35E-01	8,24E-04
JX165	Solyc10g078930.1.1	59896781-59900058	Activator of heat shock protein ATPase homolog 1	response to stress	1,67E-15	2,43E-02	1,70E-03	7,57E-07	2,05E-08
JX166	Solyc10g086580.1.1	64688483-64690961	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process	4,07E-02	5,17E-25	5,79E-03	2,80E-02	3,94E-02
JX167	Solyc11g072450.1.1	52705683-52710449	Mitochondrial F0 ATP synthase D chain	establishment of localization	3,11E-01	2,30E-20	6,58E-02	4,10E-01	2,20E-02
JX168	Solyc01g104950.2.1	85031292-85035396	Alpha-L-arabinofuranosidase/beta-D-xylosidase	cell wall organisation or biogenesis	1,56E-06	1,93E-08	6,26E-01	5,02E-03	2,30E-03
JX169	Solyc05g014470.2.1	8322381-8324881	Glyceraldehyde 3-phosphate dehydrogenase	primary metabolic process	5,58E-03	3,50E-23	3,09E-02	3,72E-01	9,70E-04
JX170	Solyc04g005340.2.1	248903-251735	Alpha-1 4-glucan protein synthase	cell wall organisation or biogenesis	5,31E-06	5,07E-25	5,75E-06	3,33E-01	4,54E-05
JX171	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process	1,31E-04	5,13E-01	7,42E-04	1,38E-03	4,41E-03
JX172	Solyc06g060290.2.1	34649921-34655100	Protein disulfide isomerase	macromolecule metabolic process	7,83E-04	5,83E-10	4,06E-02	1,35E-03	1,07E-01
JX173	Solyc09g082990.2.1	64083974-64088590	GDP-D-mannose-3â??.5â??-epimerase 2	vitamin synthesis	5,84E-02	5,91E-12	1,07E-02	1,89E-03	3,77E-01
JX174	Solyc12g006870.1.1	1315304-1322819	Acyl-protein thioesterase 2	macromolecule metabolic process	2,68E-02	1,95E-10	1,10E-05	1,24E-06	2,19E-01
JX175	Solyc09g064370.2.1	57150099-57154745	Alcohol dehydrogenase	cellular metabolic process	4,20E-04	1,61E-01	1,60E-02	6,47E-03	1,04E-02
JX176	Solyc07g041490.1.1	50632738-50633541	Stress responsive alpha-beta barrel domain protein	response to stress	1,78E-04	6,16E-14	8,87E-01	1,46E-01	2,56E-04
JX177	Solyc09g090140.2.1	65031288-65034684	Malate dehydrogenase	primary metabolic process	1,05E-06	4,08E-23	3,78E-03	8,19E-03	2,13E-06
JX178	Solyc12g009400.1.1	2682210-2685916	Pyruvate dehydrogenase E1 component alpha subunit	primary metabolic process	4,61E-05	1,17E-02	1,41E-02	1,43E-02	2,19E-03
JX179	Solyc08g014340.2.1	4063851-4071314	Cysteine synthase	primary metabolic process	7,39E-04	4,83E-12	1,64E-03	2,48E-05	5,10E-01
JX180	Solyc03g120090.1.1	62641034-62641951	Pyridoxal biosynthesis lyase pdxS	vitamin synthesis	4,34E-03	2,21E-01	2,65E-02	7,50E-01	7,57E-05
JX181	Solyc11g069790.1.1	51502451-51507359	chaperonin	macromolecule metabolic process	2,03E-02	9,83E-10	1,94E-01	2,19E-02	1,86E-01
JX182	Solyc04g079200.2.1	61339502-61344701	26S proteasome regulatory subunit	macromolecule metabolic process	1,24E-01	7,65E-04	2,36E-01	4,26E-02	3,52E-01
JX183	Solyc03g113800.2.1	57894787-57902443	Betaine aldehyde dehydrogenase	primary metabolic process	8,79E-02	6,12E-01	1,72E-03	5,32E-02	1,01E-02
JX184	Solyc09g005740.1.1	508767-509831	Chloroplast lumen common family protein	Unknown	1,81E-04	6,54E-01	1,60E-02	7,55E-02	1,80E-03
JX185	Solyc07g061790.2.1	61964267-61965352	Heme-binding protein 2	Unknown	1,16E-06	8,93E-05	3,84E-03	7,65E-03	1,87E-04
JX186	Solyc02g078120.1.1	37474003-37475736	Eukaryotic translation initiation factor 3 subunit 7	translation	1,31E-02	1,48E-05	1,35E-01	1,49E-01	3,50E-02
JX187	Solyc11g011470.1.1	4520459-4526103	NADH-ubiquinone oxidoreductase subunit	oxidation-reduction process	9,19E-03	1,17E-14	1,91E-02	1,27E-03	2,00E-01
JX188	Solyc05g008460.2.1	2805741-2809779	ATP synthase subunit beta	regulation of biological process	6,28E-04	7,95E-20	8,50E-02	4,92E-02	4,05E-03
JX189	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization	4,30E-03	3,24E-03	2,88E-02	2,13E-02	2,88E-02
JX190	Solyc04g055170.2.1	52933311-52937146	Annexin 2	establishment of localization	3,52E-08	1,48E-08	9,19E-03	7,21E-04	3,27E-05
JX191	Solyc12g043020.1.1	44067590-44080483	Dihydroxy-acid dehydratase	primary metabolic process	4,46E-04	3,68E-08	8,52E-02	6,61E-02	2,83E-03
JX192	Solyc11g039980.1.1	28095562-28096080	ATP synthase subunit alpha	establishment of localization	1,33E-01	2,85E-08	2,31E-01	4,33E-01	1,00E-01
JX193	Solyc01g066480.2.1	66874829-66881864	Fumarylacetoacetate hydrolase domain-containing protein 1	primary metabolic process	3,36E-02	4,80E-03	7,97E-02	2,27E-01	3,05E-02
JX194	Solyc11g066060.1.1	48824058-48826931	heat shock protein	response to stress	3,91E-03	2,28E-13	4,37E-01	4,88E-02	1,03E-01
JX195	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress	1,79E-06	3,41E-15	2,10E-02	1,79E-03	9,24E-04
JX196	Solyc11g069000.1.1	50652125-50657452	T-complex protein 1 subunit beta	macromolecule metabolic process	5,95E-02	1,39E-06	3,33E-02	4,90E-02	1,25E-01
JX197	Solyc06g071790.2.1	40601620-40603444	Elongation factor Tu	translation	1,00E-03	7,04E-01	7,34E-02	4,27E-03	5,27E-02
JX198	Solyc05g012480.2.1	5715224-5720976	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process	3,46E-02	4,65E-06	3,26E-01	7,94E-02	3,14E-01
JX199	Solyc03g031720.2.1	8453484-8459499	RNA Binding Protein 45	regulation of biological process	3,48E-01	4,25E-08	1,51E-01	2,45E-01	2,79E-01
JX200	Solyc09g010630.2.1	3965253-3968837	heat shock protein	response to stress	1,79E-01	1,64E-01	3,17E-02	2,97E-02	1,69E-01
JX201	Solyc01g028810.2.1	33327242-33332361	chaperonin	macromolecule metabolic process	6,16E-02	3,23E-08	2,35E-03	3,66E-02	1,73E-02
JX202	Solyc12g010040.1.1	3180908-3187436	Leucyl aminopeptidase	macromolecule metabolic process	1,69E-02	1,91E-01	2,01E-01	3,43E-03	8,48E-01
JX203	Solyc03g083390.2.1	46783589-46788273	Nuclear movement protein nudc	regulation of biological process	5,65E-02	2,30E-02	6,87E-03	1,72E-03	2,18E-01
JX204	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress	1,00E-03	2,83E-03	3,04E-02	1,40E-01	4,32E-03
JX205	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process	1,60E-02	7,09E-01	2,13E-02	4,54E-02	2,66E-02
JX206	Solyc05g017760.2.1	18312063-18317108	Acetyl-CoA C-acetyltransferase	primary metabolic process	1,06E-01	9,92E-01	4,64E-01	3,26E-02	4,77E-01
JX207	Solyc07g008720.2.1	3694316-3696017	Nascent polypeptide-associated complex subunit beta	establishment of localization	3,35E-01	3,64E-02	2,99E-01	9,96E-01	1,08E-02
JX208	Solyc01g106210.2.1	85918285-85922197	Chaperone DnaK	response to stress	3,37E-02	9,48E-02	5,47E-01	5,65E-01	5,71E-02
JX209	Solyc02g081140.2.1	39782470-39786000	UBX domain-containing protein	regulation of biological process	2,50E-01	1,51E-17	6,08E-01	8,11E-03	6,31E-01
JX210	Solyc03g082420.2.1	45898742-45899828	Heat shock protein	response to stress	1,60E-14	5,33E-32	1,05E-15	7,08E-01	2,78E-12
JX211	Solyc09g072560.2.1	60593208-60595291	Legumin 11S-globulin	macromolecule metabolic process	1,18E-03	1,15E-04	4,78E-03	2,47E-02	7,63E-03
JX212	Solyc09g082860.2.1	63941988-63946072	Sulfate adenylyltransferase	primary metabolic process	2,50E-08	5,49E-16	5,43E-03	3,16E-04	1,20E-04
JX213	Solyc02g062460.2.1	28653823-28656374	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	1,35E-05	3,61E-03	7,29E-03	1,16E-01	5,38E-04
JX214	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process	2,04E-05	3,72E-18	5,04E-02	6,51E-02	8,25E-06
JX215	Solyc01g010750.2.1	5783031-5788195	Stress responsive protein	response to stress	3,45E-05	6,15E-02	2,32E-01	4,05E-03	1,24E-02
JX216	Solyc06g073280.2.1	41539950-41545003	LL-diaminopimelate aminotransferase	primary metabolic process	6,61E-03	2,00E-04	7,88E-01	1,21E-01	1,41E-01
JX217	Solyc02g082920.2.1	41128832-41129990	Endochitinase	macromolecule metabolic process	3,92E-08	3,57E-05	2,39E-12	5,46E-08	3,25E-07
JX218	Solyc06g073190.2.1	41483219-41485991	Fructokinase-like	primary metabolic process	2,54E-18	2,29E-01	3,82E-13	1,80E-02	6,99E-13
JX219	Solyc02g082800.2.1	41021911-41031456	Ubiquilin-1	macromolecule metabolic process	3,66E-01	5,37E-02	3,35E-01	1,20E-02	4,49E-01
JX220	Solyc03g121720.2.1	63848472-63854092	2-hydroxy-3-oxopropionate reductase	primary metabolic process	2,30E-02	8,80E-19	1,17E-01	4,01E-02	1,77E-01
JX221	Solyc06g050550.2.1	29742430-29749862	Sorting nexin 1	establishment of localization	4,48E-13	1,45E-04	3,14E-03	5,99E-06	4,02E-07
JX222	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	5,47E-06	1,42E-17	2,45E-09	1,53E-06	3,97E-01
JX223	Solyc01g106320.2.1	86005942-86007808	Octicosapeptide/Phox/Bem1p domain-containing protein	Unknown	2,78E-06	8,91E-15	2,04E-02	2,07E-03	3,34E-05
JX224	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process	1,36E-02	4,41E-19	2,44E-02	1,20E-04	3,10E-01
JX225	Solyc09g011080.2.1	4413327-4416467	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process	3,31E-03	2,80E-12	2,45E-04	2,42E-03	6,91E-02
JX226	Solyc08g068310.2.1	54615972-54621953	RNA-binding La domain protein	regulation of biological process	1,39E-11	4,73E-12	9,58E-07	2,69E-09	3,54E-03
JX227	Solyc01g110450.2.1	88910585-88913966	NADP dependent sorbitol 6-phosphate dehydrogenase	primary metabolic process	1,46E-08	3,67E-01	1,12E-02	5,95E-05	2,60E-04
JX228	Solyc05g050120.2.1	59250938-59255250	Malic enzyme	primary metabolic process	1,01E-03	7,17E-03	3,14E-02	9,47E-03	2,85E-02
JX229	Solyc06g073090.2.1	41417554-41421363	Ribosomal subunit interface protein	macromolecule metabolic process	1,17E-02	1,54E-06	1,84E-02	6,42E-01	5,18E-03
JX230	Solyc08g080370.2.1	60840952-60844640	Acetylornithine aminotransferase	primary metabolic process	7,03E-01	2,97E-11	2,63E-03	1,68E-02	3,32E-01
JX231	Solyc08g074410.2.1	55693755-55701940	Tryptophanyl-tRNA synthetase	macromolecule metabolic process	3,57E-02	2,41E-11	2,02E-01	2,39E-02	5,05E-01
JX232	Solyc05g056310.2.1	64792047-64799309	T-complex protein 1 subunit gamma	macromolecule metabolic process	2,65E-01	5,23E-04	2,02E-04	1,50E-02	1,99E-02
JX233	Solyc01g006980.2.1	1552093-1555934	Malonyl CoA-acyl carrier protein transacylase containing protein	primary metabolic process	2,03E-03	4,30E-03	1,03E-01	6,77E-03	5,82E-02
JX234	Solyc05g053300.2.1	62572547-62577624	Dihydrolipoyl dehydrogenase	primary metabolic process	1,08E-03	1,77E-09	5,50E-03	2,03E-01	3,46E-03
JX235	Solyc08g022210.2.1	14105526-14113435	Methylthioribose-1-phosphate isomerase	primary metabolic process	7,36E-02	2,04E-02	5,83E-01	2,86E-01	2,00E-01
JX236	Solyc06g081980.1.1	44239265-44240194	Pyridoxal biosynthesis lyase pdxS	vitamin synthesis	3,13E-03	8,54E-10	2,68E-01	5,42E-02	4,94E-02

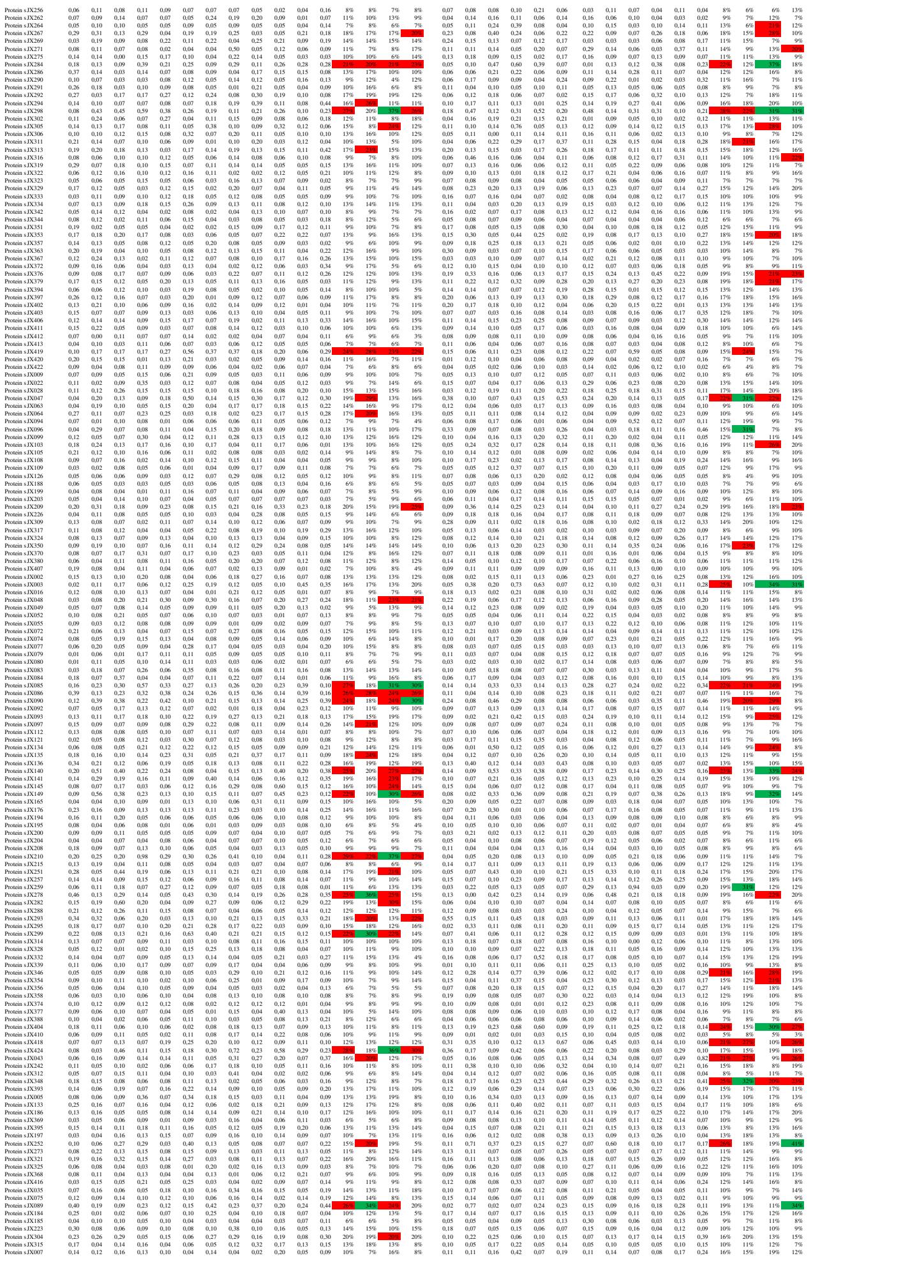
JX237	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	2,95E-08	3,89E-15	3,12E-04	1,41E-01	1,22E-06
JX238	Solyc02g062500.2.1	28750335-28754482	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	3,08E-07	4,04E-21	7,00E-07	3,52E-01	2,03E-05
JX239	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	1,95E-12	1,56E-23	1,74E-11	3,48E-02	9,89E-09
JX240	Solyc07g062970.2.1	62819985-62823101	Serine/threonine phosphatase family protein	macromolecule metabolic process	2,32E-04	2,22E-23	2,71E-04	2,34E-01	1,11E-03
JX241	Solyc09g009260.2.1	2643600-2645801	Fructose-bisphosphate aldolase	primary metabolic process	2,65E-02	8,90E-08	2,31E-03	1,15E-01	5,84E-03
JX242	Solyc09g064940.2.1	58086206-58088531	Phenazine biosynthesis protein PhzF family	secondary metabolic process	1,07E-05	8,85E-06	4,67E-01	3,83E-04	1,45E-02
JX243	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis	1,53E-02	1,78E-08	4,80E-01	2,45E-02	4,09E-01
JX244	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation	1,31E-07	2,15E-13	2,14E-02	6,79E-03	6,67E-05
JX245	Solyc09g089580.2.1	64646856-64649117	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation	2,86E-08	8,38E-32	8,62E-09	3,17E-01	4,53E-09
JX246	Solyc12g005860.1.1	490745-498964	3-isopropylmalate dehydratase large subunit	primary metabolic process	7,33E-07	8,02E-05	7,62E-02	6,06E-02	1,05E-04
JX247	Solyc04g073990.2.1	57610227-57612650	Annexin	establishment of localization	1,14E-17	1,95E-05	4,56E-02	1,49E-09	7,85E-08
JX248	Solyc04g005340.2.1	248903-251735	Alpha-1 4-glucan protein synthase	cell wall organisation or biogenesis	3,03E-02	3,96E-08	2,56E-01	3,17E-02	2,40E-01
JX249	Solyc11g069790.1.1	51502451-51507359	chaperonin	macromolecule metabolic process	2,56E-15	1,52E-18	1,79E-07	3,99E-06	1,16E-08
JX250	Solyc05g056490.2.1	64907109-64913076	3+apos	primary metabolic process	2,48E-01	1,03E-10	5,02E-01	2,37E-01	3,89E-01
JX251	Solyc04g076200.2.1	58724666-58728561	Universal stress protein family protein	response to stress	1,33E-02	7,92E-09	8,21E-01	3,39E-01	1,42E-01
JX252	Solyc06g071790.2.1	40601620-40603444	Elongation factor Tu	translation	1,14E-01	3,85E-07	6,86E-01	1,91E-01	4,58E-01
JX253	Solyc10g005890.2.1	683370-690823	DNA-damage inducible protein DDI1-like	macromolecule metabolic process	1,25E-01	7,96E-09	2,23E-01	2,53E-01	1,99E-01
JX254	Solyc01g079680.2.1	71306007-71308978	Ran GTPase activating protein	establishment of localization	1,33E-02	1,68E-08	3,13E-02	4,24E-02	2,79E-02
JX255	Solyc01g099240.2.1	81297091-81298341	3-hydroxyisobutyrate dehydrogenase	primary metabolic process	4,05E-02	2,30E-01	7,51E-02	1,64E-02	1,04E-01
JX256	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process	2,67E-02	2,53E-26	6,48E-02	3,44E-01	2,00E-03
JX257	Solyc07g063680.2.1	63306131-63309625	CHP-rich zinc finger protein-like	response to stress	8,87E-02	9,80E-08	7,44E-02	9,08E-02	1,34E-01
JX258	Solyc07g055320.2.1	60717008-60721284	ATP-dependent Zn protease cell division protein FtsH homolog	macromolecule metabolic process	2,00E-01	1,13E-01	1,24E-01	2,27E-02	4,22E-01
JX259	Solyc08g082820.2.1	62655311-62659585	Heat shock protein	response to stress	1,56E-03	5,88E-01	1,38E-01	8,66E-13	3,09E-01
JX260	Solyc02g088610.2.1	45222192-45229259	ATP-dependent chaperone ClpB	macromolecule metabolic process	2,65E-05	1,60E-28	8,48E-05	4,01E-02	1,69E-04
JX261	Solyc04g016470.2.1	7303341-7305043	Beta-1 3-glucanase	cell wall organisation or biogenesis	7,61E-02	3,18E-06	6,78E-02	1,01E-01	1,01E-01
JX262	Solyc06g068860.2.1	39073381-39083550	Alpha-mannosidase	primary metabolic process	1,02E-09	3,06E-24	8,50E-10	5,01E-02	6,82E-08
JX263	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis	1,25E-01	1,67E-29	6,33E-01	2,99E-01	3,85E-01
JX264	Solyc10g047950.1.1	38620947-38627409	Inorganic pyrophosphatase family protein	primary metabolic process	3,94E-03	8,94E-15	2,84E-03	3,73E-02	7,49E-03
JX265	Solyc06g082120.2.1	44333171-44336113	Ran GTPase binding protein	establishment of localization	2,99E-12	3,87E-08	1,41E-03	1,06E-08	7,38E-04
JX266	Solyc02g086880.2.1	44063088-44067604	Formate dehydrogenase	oxidation-reduction process	3,57E-04	5,74E-05	8,11E-01	2,63E-03	1,49E-01
JX267	Solyc10g085550.1.1	63996015-64000623	Enolase	primary metabolic process	3,72E-01	4,07E-18	4,52E-01	7,78E-01	1,47E-03
JX268	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process	5,98E-04	3,78E-11	5,79E-03	2,54E-02	5,07E-06
JX269	Solyc05g050800.2.1	60131934-60136497	Phosphoglycerate mutase family protein	primary metabolic process	9,32E-24	1,21E-03	3,96E-02	3,68E-12	3,15E-10
JX270	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process	1,24E-01	6,79E-32	4,45E-03	2,97E-02	2,65E-01
JX271	Solyc08g080140.2.1	60644754-60646888	dTDP-4-dehydrorhamnose reductase	primary metabolic process	1,90E-01	4,24E-12	3,09E-01	2,96E-01	1,85E-01
JX272	Solyc05g012480.2.1	5715224-5720976	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process	5,13E-01	6,32E-11	4,55E-01	6,29E-01	4,30E-03
JX273	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	translation	8,87E-13	6,79E-05	2,42E-01	1,04E-05	1,21E-06
JX274	Solyc04g009200.2.1	2695320-2699163	Glutamate-1-semialdehyde-2 1-aminomutase	nitrogen compound metabolic process	1,11E-03	1,47E-21	9,43E-03	1,79E-03	1,03E-01
JX275	Solyc04g016360.2.1	7158486-7165657	S-formylglutathione hydrolase	primary metabolic process	9,36E-05	3,05E-15	6,00E-01	5,77E-03	3,23E-02
JX276	Solyc03g095900.2.1	51028612-51030467	1-aminocyclopropane-1-carboxylate oxidase-like protein	developmental maturation	1,39E-12	5,54E-27	1,09E-08	8,16E-02	5,46E-09
JX277	Solyc10g080940.1.1	61453818-61456401	Tubulin beta chain	cytoskeleton organization and biogenesis	1,19E-01	2,01E-22	7,54E-02	1,25E-01	1,37E-01
JX278	Solyc09g015020.1.1	7440133-7440597	class I heat shock protein 3	response to stress	1,96E-09	4,81E-06	3,83E-08	9,55E-03	2,05E-06
JX279	Solyc08g079870.1.1	60466441-60468678	Subtilisin-like protease	macromolecule metabolic process	3,84E-11	1,11E-13	7,62E-10	1,22E-04	2,28E-07
JX280	Solyc01g059980.2.1	62201815-62226187	Beta-glucanase	cell wall organisation or biogenesis	5,41E-12	3,00E-11	1,59E-11	2,51E-01	1,06E-08
JX281	Solyc08g079870.1.1	60466441-60468678	Subtilisin-like protease	macromolecule metabolic process	4,09E-09	4,37E-11	9,03E-09	1,30E-01	1,42E-06
JX282	Solyc08g062450.1.1	48318166-48318642	class II heat shock protein	response to stress	6,20E-03	2,76E-33	6,09E-09	3,79E-03	5,97E-06
JX283	Solyc10g055810.1.1	52891942-52893092	Endochitinase	macromolecule metabolic process	4,81E-02	2,42E-06	4,00E-02	4,61E-01	6,59E-02
JX284	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	5,20E-04	1,29E-16	3,12E-03	6,48E-01	4,15E-03
JX285	Solyc05g005490.2.1	355897-361496	Carbonic anhydrase	nitrogen compound metabolic process	1,24E-03	3,41E-28	3,19E-04	4,86E-01	3,86E-05
JX286	Solyc04g054980.2.1	52664826-52665968	Lipoxygenase homology domain-containing protein 1	primary metabolic process	1,27E-18	1,69E-19	4,15E-11	8,46E-06	8,60E-11
JX287	Solyc07g041900.2.1	51828739-51831068	Cathepsin L-like cysteine proteinase	macromolecule metabolic process	3,49E-03	1,33E-21	7,62E-03	2,79E-03	1,50E-02
JX288	Solyc12g010320.1.1	3377663-3380422	Outer membrane lipoprotein blc	response to stress	2,09E-01	2,46E-31	4,65E-02	3,07E-01	1,21E-01
JX289	Solyc01g079220.2.1	70815571-70820929	NifU like protein	cellular metabolic process	2,37E-17	2,60E-18	6,93E-04	1,51E-08	1,97E-08
JX290	Solyc06g073280.2.1	41539950-41545003	LL-diaminopimelate aminotransferase	primary metabolic process	2,06E-16	4,00E-14	3,27E-02	3,93E-08	1,81E-07
JX291	Solyc02g080630.2.1	39392539-39397791	Lactoylglutathione lyase	primary metabolic process	3,89E-26	7,63E-01	4,64E-02	7,03E-12	4,79E-13
JX292	Solyc08g014130.2.1	3734998-3744536	2-isopropylmalate synthase 1	primary metabolic process	1,13E-04	1,95E-26	4,24E-03	1,12E-01	1,55E-03
JX293	Solyc10g084400.1.1	63281886-63284225	Glutathione S-transferase	response to stress	1,05E-07	5,65E-08	3,86E-03	6,87E-02	3,27E-05
JX294	Solyc03g116110.2.1	59675160-59679638	Alpha/beta hydrolase fold protein	primary metabolic process	1,23E-10	5,70E-11	1,61E-04	1,80E-04	3,26E-06
JX295	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress	1,71E-03	1,36E-20	1,67E-01	2,43E-03	6,28E-02
JX296	Solyc09g007270.2.1	865197-869322	Ascorbate peroxidase	oxidation-reduction process	1,67E-04	1,64E-34	4,95E-02	4,53E-03	1,26E-02
JX297	Solyc10g006650.2.1	1157432-1161954	Flavoprotein wrbA	oxidation-reduction process	7,09E-02	1,54E-20	8,82E-01	5,71E-01	2,97E-02
JX298	Solyc12g096190.1.1	63553620-63556459	Tryptophan synthase beta chain	primary metabolic process	8,33E-04	1,02E-02	5,16E-01	1,45E-01	1,51E-02
JX299	Solyc11g020040.1.1	10015582-10019521	Chaperone DnaK	response to stress	9,02E-01	1,97E-12	3,31E-01	4,12E-01	8,75E-01
JX300	Solyc01g100520.2.1	82287668-82292399	ATP-dependent Clp protease proteolytic subunit	establishment of localization	2,23E-02	2,87E-06	1,03E-02	6,25E-01	1,57E-02
JX301	Solyc07g051850.2.1	57713505-57718748	Aspartic proteinase	macromolecule metabolic process	3,13E-08	2,34E-12	3,79E-06	2,05E-02	4,27E-06
JX302	Solyc01g005560.2.1	394402-399248	Isocitrate dehydrogenase	primary metabolic process	6,13E-02	2,33E-19	1,85E-05	6,56E-03	3,47E-03
JX303	Solyc07g051850.2.1	57713505-57718748	Aspartic proteinase	macromolecule metabolic process	4,44E-07	6,00E-18	1,01E-04	5,87E-03	8,20E-05
JX304	Solyc03g025850.2.1	7627656-7630095	Remorin 1	Unknown	5,73E-02	4,40E-09	2,15E-03	1,41E-01	9,57E-03
JX305	Solyc04g011510.2.1	3944572-3949105	Triosephosphate isomerase	primary metabolic process	2,01E-02	4,61E-01	4,58E-01	7,84E-03	2,17E-01
JX306	Solyc07g064590.2.1	63891191-63895627	Ubiquitin thioesterase OTU1	primary metabolic process	3,77E-11	1,68E-03	2,96E-02	5,94E-04	4,65E-07
JX307	Solyc04g072400.2.1	57015439-57020648	Endoribonuclease E-like protein	macromolecule metabolic process	1,00E-02	5,62E-19	6,24E-03	4,84E-01	1,04E-02
JX308	Solyc02g062460.2.1	28653823-28656374	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	4,03E-03	2,33E-09	2,74E-03	4,28E-01	3,95E-03
JX309	Solyc03g097270.2.1	53034151-53040369	Cysteine proteinase inhibitor	regulation of biological process	1,02E-09	5,72E-03	1,16E-04	1,76E-05	1,59E-05
JX310	Solyc09g082720.2.1	63816287-63819690	Aldo/keto reductase family protein	oxidation-reduction process	3,65E-12	8,91E-16	2,30E-07	4,15E-05	5,83E-09
JX311	Solyc12g044740.1.1	45436676-45448004	Ubiquitin carboxyl-terminal hydrolase	primary metabolic process	5,75E-09	8,47E-01	1,83E-01	2,67E-06	1,50E-02
JX312	Solyc01g107700.2.1	86915524-86922302	Kynurenine formamidase	secondary metabolic process	6,01E-04	1,03E-06	2,48E-02	1,77E-02	6,02E-04
JX313	Solyc06g071000.2.1	40003245-40009711	N-succinylglutamate 5-semialdehyde dehydrogenase	primary metabolic process	1,54E-05	3,68E-01	9,34E-01	1,97E-02	8,12E-03
JX314	Solyc09g090330.2.1	65188896-65193572	Harpin binding protein 1	response to stress	3,05E-04	1,30E-13	2,98E-02	8,02E-04	3,26E-02
JX315	Solyc03g025850.2.1	7627656-7630095	Remorin 1	Unknown	1,58E-02	1,74E-16	1,71E-01	1,18E-01	6,59E-02
JX316	Solyc01g097460.2.1	80067637-80074843	Ribose-5-phosphate isomerase	primary metabolic process	1,43E-02	2,90E-20	2,98E-01	1,68E-01	5,87E-02
JX317	Solyc01g104170.2.1	84393496-84398307	Ankyrin repeat domain-containing protein 2	regulation of biological process	1,34E-05	1,22E-11	3,12E-03	3,07E-01	2,22E-05
JX318	Solyc01g067740.2.1	69074003-69078931	Superoxide dismutase	oxidation-reduction process	5,91E-01	1,86E-14	1,96E-01	5,76E-02	4,89E-01
JX319	Solyc06g063090.2.1	36234795-36239218	Alanine aminotransferase	primary metabolic process	1,28E-01	3,50E-19	3,13E-02	1,11E-01	8,25E-02
JX320	Solyc12g009060.1.1	2350521-2354721	Charged multivesicular body protein 2a	establishment of localization	5,62E-07	2,66E-07	1,46E-02	9,54E-03	2,36E-04
JX321	Solyc03g112150.1.1	56698575-56700008	Elongation factor Tu	translation	3,00E-06	9,58E-01	7,68E-04	2,85E-01	6,56E-07
JX322	Solyc01g112280.2.1	90105382-90109100	Succinyl-diaminopimelate desuccinylase	primary metabolic process	4,45E-06	3,93E-08	5,88E-02	1,98E-02	8,87E-04
JX323	Solyc09g009020.2.1	2370061-2375045	Enolase	primary metabolic process	3,90E-03	2,33E-19	2,49E-03	1,80E-02	1,18E-03
JX324	Solyc07g008170.2.1	2882629-2893614	Methyl binding domain protein	regulation of biological process	7,85E-02	3,75E-10	5,51E-01	5,09E-02	3,69E-01
JX325	Solyc01g011000.2.1	6848050-6851207	Eukaryotic translation initiation factor 5A	translation	9,86E-06	1,00E-04	1,49E-01	1,27E-03	8,71E-03
JX326	Solyc06g005160.2.1	182627-185280	Ascorbate peroxidase	oxidation-reduction process	2,98E-03	3,84E-24	8,63E-02	4,83E-02	2,31E-02
JX327	Solyc05g005480.2.1	352211-355615	Oxidoreductase zinc-binding dehydrogenase	oxidation-reduction process	5,72E-09	7,34E-12	9,46E-03	1,23E-05	4,14E-04
JX328	Solyc03g007670.2.1	2200919-2205536	SGT1	response to stress	8,31E-05	3,67E-02	2,68E-02	1,41E-02	2,16E-03
JX329	Solyc02g093830.2.1	49113784-49119691	Glucose-6-phosphate 1-dehydrogenase	primary metabolic process	1,85E-01	3,59E-02	4,17E-03	2,52E-02	6,40E-02
JX330	Solyc12g009250.1.1	2529842-2532163	chaperonin	macromolecule metabolic process	9,74E-03	7,63E-12	2,50E-02	7,68E-01	4,88E-04
JX331	Solyc12g055800.1.1	47132031-47139706	V-type ATP synthase alpha chain	establishment of localization	4,06E-04	2,43E-04	6,41E-03	1,59E-03	1,89E-02
JX332	Solyc06g076020.2.1	43582389-43585486	heat shock protein	response to stress	2,56E-01	8,35E-02	4,02E-02	1,92E-02	2,72E-01
JX333	Solyc03g115820.2.1	59463075-59469502	Ribulose-5-phosphate-3-epimerase	primary metabolic process	1,99E-06	8,42E-06	5,93E-01	1,53E-02	8,33E-04
JX334	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process	3,01E-02	8,03E-02	2,40E-01	2,47E-01	3,49E-02
JX335	Solyc09g009390.2.1	2835367-2840425	Monodehydroascorbate reductase	vitamin synthesis	6,24E-01	1,26E-05	5,40E-01	3,61E-02	8,90E-01
JX336	Solyc12g009140.1.1	2453127-2457231	Proteasome subunit alpha type	macromolecule metabolic process	2,65E-02	3,74E-01	5,10E-04	5,42E-03	1,38E-02
JX337	Solyc05g012070.2.1	5295089-5298245	Alpha-1 4-glucan-protein synthase	cell wall organisation or biogenesis	7,56E-01	1,29E-07	3,50E-01	1,19E-01	6,55E-01
JX338	Solyc02g088700.2.1	45268178-45273518	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process	1,85E-03	2,24E-03	1,98E-03	1,35E-01	9,13E-04
JX339	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress	9,98E-03	1,87E-01	1,13E-02	1,49E-01	6,04E-03
JX340	Solyc01g007320.2.1	1849252-1851757	ATP synthase subunit beta chloroplastic	establishment of localization	2,45E-01	3,04E-09	2,91E-01	6,39E-01	1,40E-02
JX341	Solyc06g082090.2.1	44317047-44320702	Methionine aminopeptidase	macromolecule metabolic process	3,21E-02	1,09E-04	2,06E-01	2,60E-01	1,32E-01
JX342	Solyc07g064810.2.1	64017771-64025505	Imidazole glycerol phosphate synthase subunit hisF	primary metabolic process	1,87E-02	1,26E-01	2,76E-02	1,42E-02	8,56E-02
JX343	Solyc02g087300.1.1	44325520-44326428	Protein transport SEC13-like protein	establishment of localization	1,62E-03	4,52E-01	4,77E-01	8,35E-02	4,22E-02
JX344	Solyc11g011960.1.1	4912805-4919067	UTP-glucose 1 phosphate uridylyltransferase	primary metabolic process	1,06E-01	9,53E-04	4,65E-03	2,71E-01	6,06E-04
JX345	Solyc02g081400.2.1	39962544-39967049	Intracellular protease PfpI family protein expressed	macromolecule metabolic process	6,61E-16	9,47E-05	1,13E-01	7,20E-08	4,93E-07
JX346	Solyc09g011670.2.1	4913443-4917118	Universal stress protein family protein	response to stress	7,71E-02	1,07E-14	7,11E-02	4,11E-03	4,09E-01
JX347	Solyc10g086580.1.1	64688483-64690961	Ribulose-1 5-bisphosphate carboxylase/oxygenase activase 1	cellular metabolic process	1,95E-01	3,23E-03	4,28E-01	9,28E-02	3,33E-01
JX348	Solyc09g011140.2.1	4499977-4502723	Tropinone reductase I	secondary metabolic process	9,52E-08	5,56E-20	3,70E-01	8,34E-06	2,19E-02
JX349	Solyc07g064160.2.1	63644226-63646402	Thiazole biosynthetic enzyme	vitamin synthesis	4,25E-09	9,05E-21	3,24E-07	2,55E-11	1,13E-01
JX350	Solyc03g098700.1.1	54427774-54428433	Cysteine protease inhibitor 8	regulation of biological process	2,86E-10	2,90E-15	3,02E-06	3,66E-07	4,20E-03
JX351	Solyc02g082800.2.1	41021911-41031456	Ubiquilin-1	primary metabolic process	1,94E-20	4,19E-16	5,62E-03	4,24E-12	4,68E-07
JX352	Solyc03g120230.2.1	62741263-62746412	MAR-binding filament-like protein 1	cytoskeleton organization and biogenesis	4,22E-01	9,62E-08	7,84E-01	6,67E-01	5,60E-01
JX353	Solyc10g008740.2.1	2822974-2826287	Magnesium chelatase ATPase subunit I	primary metabolic process	2,03E-03	1,45E-17	4,15E-01	8,13E-02	2,46E-02
JX354	Solyc01g109660.2.1	88321513-88322795	Glycine-rich RNA-binding protein	response to stress	1,64E-04	3,77E-06	2,49E-02	4,29E-04	3,58E-02
JX355	Solyc06g007200.2.1	1270929-1276535	Methionine aminopeptidase	macromolecule metabolic process	1,45E-03	1,70E-07	2,45E-02	1,48E-01	9,37E-03

JX356	Solyc01g005520.2.1	349916-352687	Tetratricopeptide TPR_2 repeat protein	response to stress	3,40E-05	7,08E-19	3,65E-04	4,69E-05	4,52E-03
JX357	Solyc02g082830.1.1	41056853-41058151	Phosphoserine aminotransferase	primary metabolic process	7,99E-10	3,34E-01	3,66E-05	1,01E-02	1,42E-06
JX358	Solyc01g103450.2.1	83821528-83826037	Chaperone DnaK	response to stress	2,20E-15	1,30E-04	2,88E-02	2,27E-07	2,00E-07
JX359	Solyc04g011400.2.1	3868800-3872796	UDP-glucose 4-epimerase	cell wall organisation or biogenesis	7,82E-02	2,60E-24	8,45E-03	3,35E-02	2,00E-01
JX360	Solyc05g056230.2.1	64732387-64738654	Calreticulin 2 calcium-binding protein	macromolecule metabolic process	1,49E-04	6,78E-05	7,06E-01	6,74E-04	6,89E-02
JX361	Solyc09g011240.2.1	4573233-4578442	Reductase 2	oxidation-reduction process	1,19E-04	2,52E-22	2,03E-04	1,28E-04	1,96E-01
JX362	Solyc09g010930.2.1	4264025-4269432	NAD-dependent epimerase/dehydratase	cell wall organisation or biogenesis	1,58E-08	1,12E-05	3,85E-02	9,24E-05	5,47E-04
JX363	Solyc02g091490.2.1	47344370-47349179	Fructokinase 3	primary metabolic process	6,62E-07	1,36E-04	1,65E-01	5,41E-04	1,14E-02
JX364	Solyc05g008460.2.1	2805741-2809779	ATP synthase subunit beta	establishment of localization	1,01E-03	1,23E-15	2,27E-01	2,44E-03	1,22E-01
JX365	Solyc01g009420.2.1	3621816-3626739	Bifunctional polymyxin resistance arnA protein	macromolecule metabolic process	2,81E-04	8,73E-03	2,02E-02	5,72E-04	2,48E-02
JX366	Solyc12g056830.1.1	48264644-48265396	ATP synthase delta subunit	establishment of localization	8,42E-02	6,98E-12	7,02E-02	3,37E-01	1,10E-02
JX367	Solyc08g076220.2.1	57405726-57410850	Phosphoribulokinase/uridine kinase	primary metabolic process	5,56E-05	3,70E-02	1,70E-01	1,18E-01	1,82E-04
JX368	Solyc11g072190.1.1	52501034-52504379	Elongation factor beta-1	translation	1,62E-09	1,36E-01	9,99E-03	1,00E-04	3,16E-05
JX369	Solyc12g042650.1.1	43448722-43449792	40S ribosomal protein S12	translation	4,44E-11	2,23E-15	2,66E-04	9,85E-07	5,19E-05
JX370	Solyc07g044860.2.1	55244985-55247025	Oxygen-evolving enhancer protein 2. chloroplastic	regulation of biological process	1,63E-01	1,42E-05	1,52E-01	7,16E-01	9,63E-04
JX371	Solyc09g008280.1.1	1749950-1751122	S-adenosylmethionine synthase	hormone metabolic process	1,66E-01	1,74E-11	4,24E-01	5,54E-01	1,22E-02
JX372	Solyc09g005700.2.1	488084-495891	Diaminopimelate epimerase family protein	primary metabolic process	1,18E-01	5,62E-13	2,71E-01	7,44E-01	1,53E-03
JX373	Solyc10g083970.1.1	62989745-62990917	S-adenosylmethionine synthase	hormone metabolic process	3,51E-02	7,25E-17	9,73E-04	2,51E-04	6,34E-01
JX374	Solyc06g009020.2.1	2965668-2967884	Glutathione S-transferase	response to stress	1,13E-10	1,55E-06	5,07E-02	3,66E-04	2,00E-06
JX375	Solyc10g008010.2.1	2177130-2182889	Proteasome subunit alpha type	macromolecule metabolic process	2,54E-01	9,76E-02	6,89E-02	1,08E-01	1,79E-01
JX376	Solyc07g062570.2.1	62472541-62478080	Ubiquitin-conjugating enzyme E2 N	primary metabolic process	2,59E-02	4,91E-01	3,37E-01	1,74E-01	1,23E-01
JX377	Solyc09g090980.2.1	65698002-65700023	Major allergen Mal d 1	response to stress	3,70E-07	2,50E-02	7,32E-02	3,52E-03	8,70E-05
JX378	Solyc05g052150.2.1	61597939-61598511	ATP synthase subunit delta+apos;. mitochondrial	establishment of localization	2,36E-02	2,88E-16	6,30E-01	8,13E-01	1,52E-02
JX379	Solyc05g053810.2.1	62993932-62998448	Serine hydroxymethyltransferase	primary metabolic process	3,00E-02	1,66E-08	2,24E-01	6,02E-02	1,62E-01
JX380	Solyc02g014150.2.1	6519093-6531932	Photosystem II stability/assembly factor Ycf48-like protein	regulation of biological process	1,65E-03	4,77E-05	1,42E-01	1,55E-01	1,10E-02
JX381	Solyc01g099760.2.1	81668247-81673265	26S protease regulatory subunit 6A homolog	macromolecule metabolic process	2,97E-01	6,60E-07	8,35E-01	5,36E-01	5,70E-01
JX382	Solyc06g062950.1.1	36119748-36122078	Subtilisin-like protease	macromolecule metabolic process	2,37E-02	1,42E-02	9,86E-03	5,84E-03	6,11E-02
JX383	Solyc01g028810.2.1	33327242-33332361	chaperonin	macromolecule metabolic process	1,50E-03	2,84E-05	3,44E-01	1,43E-03	1,17E-01
JX384	Solyc02g086910.2.1	44078232-44082017	Peptidyl-prolyl cis-trans isomerase cyclophilin-type	macromolecule metabolic process	1,35E-03	1,39E-02	7,73E-08	2,33E-04	1,18E-04
JX385	Solyc04g007120.2.1	832713-843272	UV excision repair protein RAD23	macromolecule metabolic process	1,23E-05	2,95E-01	7,81E-01	1,40E-02	8,44E-03
JX386	Solyc11g020300.1.1	10728385-10741180	Translocon Tic40	establishment of localization	3,17E-02	2,52E-01	6,26E-02	2,39E-02	1,50E-01
JX387	Solyc02g070510.2.1	34815271-34822336	Proteasome subunit alpha type	macromolecule metabolic process	6,89E-03	6,23E-07	9,01E-01	1,93E-01	8,31E-02
JX388	Solyc01g106260.2.1	85976768-85981140	Chaperone DnaK	response to stress	1,24E-03	4,65E-01	3,56E-03	5,23E-01	3,70E-05
JX389	Solyc08g062660.2.1	48903019-48905978	Ran GTPase binding protein	establishment of localization	8,06E-05	4,26E-12	4,18E-01	6,08E-03	4,15E-02
JX390	Solyc05g053470.2.1	62706122-62712174	chaperonin	macromolecule metabolic process	7,65E-05	4,09E-07	5,13E-02	3,14E-03	1,49E-02
JX391	Solyc07g043420.2.1	54503345-54505070	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	3,21E-04	2,82E-13	2,96E-04	8,94E-06	9,86E-01
JX392	Solyc01g105060.2.1	85094784-85097712		macromolecule metabolic process	2,41E-08	2,97E-19	1,51E-05	6,86E-01	3,70E-06
JX393	Solyc08g074620.1.1	55911248-55913011	Polyphenol oxidase	secondary metabolic process	9,15E-08	1,94E-06	1,00E-06	6,96E-07	7,30E-01
JX394	Solyc06g065270.2.1	37095259-37099538	Adenylate kinase	primary metabolic process	1,15E-12	1,70E-03	6,17E-10	2,03E-01	8,47E-09
JX395	Solyc02g086730.1.1	43937531-43938103	50S ribosomal protein L12-C	translation	3,31E-02	1,26E-20	3,75E-02	4,30E-02	1,64E-01
JX396	Solyc01g100030.2.1	81869800-81871185	Deoxyuridine 5+apos-triphosphate nucleotidohydrolase	macromolecule metabolic process	6,98E-02	8,57E-22	9,25E-02	4,02E-02	9,50E-01
JX397	Solyc08g079020.2.1	59830669-59836240	Adenine phosphoribosyltransferase-like protein	primary metabolic process	3,84E-06	8,57E-10	9,72E-02	2,02E-04	2,77E-02
JX398	no identification				4,26E-04	9,69E-05	1,16E-02	1,69E-02	2,55E-03
JX399	Solyc02g088700.2.1	45268178-45273518	Mitochondrial processing peptidase beta subunit	macromolecule metabolic process	3,05E-01	2,03E-07	8,75E-02	2,58E-01	5,46E-03
JX400	Solyc03g007520.2.1	2089415-2095197	Proline-rich cell wall protein-like	cell wall organisation or biogenesis	7,80E-04	4,03E-01	2,34E-05	5,35E-03	4,01E-04
JX401	Solyc02g063130.2.1	29750711-29757113	UV excision repair protein RAD23	macromolecule metabolic process	8,78E-02	2,02E-13	9,39E-01	2,90E-01	5,88E-01
JX402	Solyc08g014340.2.1	4063851-4071314	Cysteine synthase	primary metabolic process	1,57E-04	3,41E-09	2,18E-01	4,09E-03	6,72E-02
JX403	Solyc11g011960.1.1	4912805-4919067	UTP-glucose 1 phosphate uridylyltransferase	primary metabolic process	2,78E-03	2,69E-10	5,20E-02	1,79E-01	5,08E-03
JX404	Solyc12g056230.1.1	47552627-47555380	Glutathione peroxidase	response to stress	3,84E-01	2,40E-03	6,65E-01	3,34E-02	6,89E-01
JX405	Solyc01g097450.2.1	80067623-80073617	Thioredoxin-like protein 1	oxidation-reduction process	6,90E-03	9,68E-01	9,65E-01	4,23E-01	7,23E-02
JX406	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process	4,04E-03	1,91E-01	3,08E-01	1,76E-02	1,89E-01
JX407	Solyc01g099770.2.1	81675491-81677536	Translationally-controlled tumor protein homolog	regulation of biological process	4,89E-09	3,65E-02	5,62E-01	5,41E-05	1,80E-03
JX408	Solyc06g082630.2.1	44672378-44677754	26S protease regulatory subunit 6B	macromolecule metabolic process	5,25E-03	7,42E-02	2,06E-02	1,21E-03	1,49E-01
JX409	Solyc06g060260.2.1	34615474-34625952	Stromal ascorbate peroxidase 7	oxidation-reduction process	6,35E-06	5,24E-05	5,89E-05	3,96E-02	7,74E-05
JX410	Solyc08g079170.2.1	59970339-59976456	Stress-induced protein sti1-like protein	response to stress	9,81E-03	2,86E-06	4,01E-02	9,94E-02	5,67E-03
JX411	Solyc04g045340.2.1	31642460-31657602	Phosphoglucomutase	primary metabolic process	2,55E-04	1,49E-03	6,89E-01	1,19E-02	7,06E-02
JX412	Solyc05g051850.2.1	61409236-61413326	Inositol-3-phosphate synthase	primary metabolic process	2,12E-02	9,88E-11	3,78E-03	2,39E-03	1,43E-01
JX413	Solyc01g007860.2.1	2018307-2022494	Ubiquitin-conjugating enzyme family protein-like	primary metabolic process	7,59E-03	7,67E-15	5,95E-02	9,02E-03	2,38E-01
JX414	Solyc07g042250.2.1	52684617-52687267	chaperonin	macromolecule metabolic process	3,03E-02	1,18E-01	7,52E-02	8,92E-01	4,65E-03
JX415	Solyc03g120280.1.1	62781048-62782397	RAN binding protein 3	establishment of localization	7,84E-03	9,91E-01	8,03E-01	2,84E-01	1,18E-01
JX416	Solyc12g010060.1.1	3203778-3206878	Eukaryotic translation initiation factor 5A	translation	2,75E-02	1,62E-01	1,98E-01	8,19E-01	1,97E-02
JX417	Solyc02g086830.2.1	44017964-44022130	Protease Do-like	macromolecule metabolic process	5,02E-04	2,65E-14	4,88E-05	8,65E-05	6,51E-01
JX418	Solyc04g071620.2.1	56178656-56180338	ASR4	response to stress	7,67E-09	3,57E-05	3,35E-05	2,73E-05	2,46E-05
JX419	Solyc03g083910.2.1	47397595-47401871	Acid beta-fructofuranosidase	primary metabolic process	9,07E-04	1,42E-15	1,82E-03	7,87E-01	3,91E-03
JX420	Solyc01g080460.2.1	72206073-72214485	Pyruvate phosphate dikinase	primary metabolic process	7,34E-08	6,29E-32	3,21E-06	3,10E-02	1,59E-06
JX421	Solyc02g062500.2.1	28750335-28754482	2-oxoglutarate-dependent dioxygenase	oxidation-reduction process	7,65E-03	2,55E-11	1,99E-02	2,60E-01	1,19E-02
JX422	Solyc09g083410.2.1	64466144-64472295	Amidase hydantoinase/carbamoylase family protein expressed	primary metabolic process	9,25E-07	3,88E-07	1,07E-03	4,92E-06	3,80E-01
JX423	Solyc12g055800.1.1	47132031-47139706	V-type ATP synthase alpha chain	establishment of localization	4,03E-03	2,76E-04	2,98E-01	2,74E-03	2,52E-01
JX424	Solyc02g031950.2.1	17912940-17915049	Pathogenesis-related protein-like protein	response to stress	3,84E-05	1,01E-05	3,60E-05	5,08E-01	2,17E-08

Supplemental table S6 CV within each genotype and in average at cell expansion stage (CE) and orange red stage (OR). CVs in red (0.2<CV<0.3), in green (CV>0.3)

Class	Trait	CerCE	CerLev(	LevCE	StuCE	StuCriC	CriCE	LA0147(	LA0147]	PloCE	FerCE	FerLA1·	LA1420(	Average	4 cera	4 lyco	4hybrids	CerOR	CerLev(	LevOR	StuOR	StuCriO	CriOR	LA0147O	LA0147]	PloOR	FerOR	FerLA1·	LA1420(	Average	4 cera	4 lyco	4 hybrids
Metaboli	Gluc	0,01	0,01	0,03	0,02	0,01	0,03	0,04	0,02	0,02	0,01	0,02	0,10	3%	4%	3%	1%	0,01	0,03	0,01	0,02	0,04	0,04	0,02	0,03	0,01	0,01	0,04	0,03	2%	2%	2%	4%
Metaboli	Suc	0,00	0,05	0,13	0,03	0,05	0,05	0,05	0,06	0,02	0,02	0,07	0,05	5%	3%	6%	6%	0,04	0,01	0,08	0,04	0,02	0,02	0,05	0,00	0,06	0,08	0,02	0,01	4%	4%	6%	1%
Metaboli	Fru	0,01	0,01	0,02	0,02	0,01	0,03	0,04	0,02	0,01	0,02	0,02	0,09	2%	3%	2%	2%	0,02	0,03	0,02	0,01	0,04	0,03	0,02	0,03	0,01	0,02	0,06	0,04	3%	3%	2%	4%
Metaboli	Ala	0,08	0,10	0,27	0,05	0,08	0,19	0,16	0,17	0,08	0,02	0,08	0,25	13%	15%	13%	11%	0,11	0,13	0,09	0,07	0,02	0,10	0,08	0,10	0,03	0,06	0,06	0,07	8%	7%	7%	8%
Metaboli	Asn	0,08	0,10	0,16	0,07	0,09	0,14	0,18	0,21	0,05	0,05	0,10	0,25	12%	13%	11%	12%	0,06	0,10	0,05	0,09	0,03	0,02	0,07	0,17	0,03	0,06	0,06	0,02	6%	3%	7%	9%
Metaboli	Asp	0,05	0,08	0,01	0,01	0,06	0,06	0,08	0,06	0,06	0,05	0,07	0,10	6%	7%	4%	7%	0,07	0,09	0,05	0,03	0,08	0,06	0,08	0,10	0,05	0,01	0,11	0,02	6%	5%	4%	9%
Metaboli	Gaba	0,05	0,04	0,05	0,04	0,01	0,04	0,09	0,11	0,04	0,03	0,02	0,01	4%	3%	5%	5%	0,08	0,04	0,12	0,03	0,05	0,05	0,05	0,07	0,03	0,06	0,06	0,02	6%	4%	7%	6%
Metaboli	Gln	0,07	0,04	0,21	0,04	0,07	0,17	0,16	0,24	0,04	0,07	0,06	0,36	13%	16%	12%	10%	0,04	0,13	0,10	0,05	0,03	0,03	0,07	0,13	0,03	0,05	0,04	0,04	6%	3%	7%	8%
Metaboli	Ileu	0,08	0,09	0,29	0,08	0,08	0,19	0,26	0,33	0,11	0,08	0,05	0,49	18%	22%	18%	14%	0,11	0,09	0,10	0,16	0,07	0,05	0,07	0,09	0,01	0,07	0,06	0,09	8%	6%	10%	8%
Metaboli	Leu	0,05	0,08	0,23	0,05	0,04	0,19	0,07	0,18	0,13	0,05	0,05	0,34	12%	18%	10%	9%	0,07	0,09	0,09	0,08	0,04	0,03	0,04	0,12	0,02	0,09	0,12	0,08	7%	5%	8%	9%
Metaboli	Phe	0,01	0,03	0,17	0,07	0,08	0,14	0,13	0,17	0,05	0,06	0,04	0,26	10%	12%	11%	8%	0,04	0,11	0,14	0,10	0,03	0,03	0,04	0,10	0,01	0,03	0,06	0,07	6%	4%	8%	7%
Metaboli	Tyr	0,03	0,07	0,25	0,10	0,04	0,18	0,11	0,23	0,06	0,12	0,05	0,30	13%	14%	14%	10%	0,00	0,09	0,16	0,14	0,09	0,05	0,06	0,13	0,01	0,02	0,08	0,07	8%	3%	9%	10%
Metaboli	Val	0,08	0,09	0,18	0,05	0,05	0,18	0,14	0,21	0,05	0,06	0,03	0,21	11%	13%	11%	10%	0,17	0,10	0,15	0,11	0,08	0,05	0,04	0,08	0,04	0,05	0,09	0,03	8%	7%	9%	9%
Metaboli	Thr	0,09	0,10	0,27	0,09	0,04	0,20	0,20	0,30	0,06	0,11	0,08	0,31	15%	16%	17%	13%	0,13	0,10	0,14	0,18	0,09	0,02	0,12	0,28	0,06	0,09	0,12	0,08	12%	7%	13%	15%
Metaboli	Chloro	0,04	0,03	0,04	0,03	0,08	0,05	0,06	0,13	0,05	0,06	0,04	0,03	5%	4%	5%	7%	0,05	0,01	0,03	0,04	0,04	0,02	0,03	0,04	0,03	0,06	0,08	0,05	4%	4%	4%	4%
Metaboli	VitC	0,02	0,09	0,07	0,13	0,15	0,07	0,05	0,06	0,07	0,07	0,22	0,04	9%	5%	8%	13%	0,08	0,17	0,08	0,02	0,11	0,13	0,08	0,06	0,07	0,06	0,05	0,12	9%	10%	6%	10%
Metaboli	Cit	0,03	0,03	0,04	0,03	0,01	0,02	0,07	0,01	0,02	0,05	0,06	0,11	4%	4%	5%	3%	0,03	0,03	0,02	0,02	0,01	0,02	0,04	0,01	0,02	0,05	0,03	0,06	3%	3%	3%	2%
Metaboli	Mal	0,02	0,03	0,10	0,05	0,01	0,02	0,09	0,02	0,07	0,05	0,03	0,10	5%	5%	7%	2%	0,05	0,04	0,05	0,07	0,04	0,05	0,07	0,07	0,02	0,07	0,05	0,08	5%	5%	6%	5%
Metaboli	Fum	0,05	0,08	0,07	0,10	0,03	0,03	0,05	0,10	0,03	0,12	0,07	0,05	6%	4%	8%	7%	0,09	0,10	0,08	0,04	0,19	0,09	0,18	0,07	0,13	0,02	0,07	0,00	0%	8%	8%	11%
Metaboli	Stch	0,08	0,10	0,27	0,73	0,06	0,23	0,06	0,22	0,26	0,20	0,10	0,06	20%	16%	32%	12%	0,03	0,30	1,41	0,71	1,41	1,07	1,41	0,82	0,53	0,28	1,41	0,76	85%	60%	95%	99%
Metaboli	Trigo	0,02	0,08	0,19	0,06	0,01	0,02	0,07	0,06	0,03	0,03	0,07	0,16	7%	6%	9%	6%	0,04	0,16	0,09	0,01	0,04	0,02	0,13	0,10	0,03	0,02	0,04	0,14	7%	6%	7%	8%
Metaboli	Ade	0,09	0,03	0,05	0,03	0,05	0,02	0,10	0,02	0,01	0,04	0,08	0,09	5%	5%	5%	4%	0,05	0,07	0,06	0,04	0,10	0,04	0,05	0,09	0,02	0,03	0,08	0,07	6%	4%	4%	9%
Metaboli	Chol	0,05	0,02	0,07	0,02	0,02	0,02	0,02	0,07	0,02	0,02	0,02	0,06	3%	4%	3%	3%	0,02	0,00	0,05	0,03	0,00	0,03	0,03	0,04	0,03	0,01	0,04	0,04	3%	3%	3%	2%
Metaboli	Inos	0,05	0,01	0,05	0,03	0,07	0,03	0,03	0,04	0,03	0,03	0,04	0,04	4%	4%	3%	4%	0,10	0,03	0,12	0,10	0,04	0,03	0,08	0,07	0,05	0,06	0,04	0,05	6%	6%	9%	4%
Metaboli	sChl	0,09	0,19	0,11	0,11	0,04	0,04	0,12	0,12	0,06	0,03	0,07	0,07	9%	6%	9%	11%	0,08	0,09	0,03	0,11	0,08	0,09	0,13	0,07	0,11	0,11	0,07	0,02	8%	7%	10%	8%
Metaboli	sTCQ	0,53	0,54	0,18	0,38	0,43	0,06	0,29	0,38	0,49	0,31	0,46	0,24	36%	33%	29%	45%	0,07	0,11	0,11	0,07	0,12	0,01	0,03	0,28	0,05	0,07	0,14	0,07	9%	5%	7%	16%
Metaboli	sAto	0,02	0,27	0,12	0,05	0,13	0,01	0,08	0,05	0,03	0,05	0,12	0,05	8%	3%	8%	14%	0,05	0,09	0,15	0,13	0,16	0,03	0,10	0,24	0,04	0,15	0,18	0,20	13%	8%	13%	17%
Metaboli	sCry	0,08	0,17	0,50	0,24	0,22	0,35	0,34	0,21	0,59	0,53	0,10	0,37	31%	35%	40%	18%	0,43	0,23	0,44	0,29	0,30	0,38	0,22	0,25	0,57	0,30	0,43	0,39	35%	44%	31%	30%
Metaboli	sDHT	0,04	0,45	0,16	0,10	0,19	0,07	0,16	0,06	0,05	0,05	0,15	0,12	13%	7%	12%	21%	0,15	0,32	0,05	0,15	0,13	0,04	0,08	0,25	0,04	0,36	0,29	0,10	16%	8%	16%	25%
Metaboli	sRutp	0,11	0,18	0,07	0,06	0,16	0,02	0,14	0,11	0,08	0,08	0,10	0,15	10%	9%	9%	14%	0,05	0,20	0,09	0,02	0,08	0,11	0,06	0,19	0,08	0,24	0,19	0,04	11%	7%	10%	16%
Metaboli	sPpa	0,05	0,06	0,03	0,15	0,12	0,05	0,11	0,04	0,03	0,08	0,15	0,01	7%	4%	9%	9%	0,04	0,07	0,02	0,05	0,10	0,05	0,10	0,15	0,06	0,13	0,17	0,05	8%	5%	8%	12%
Metaboli	sCou	0,15	0,16	0,13	0,19	0,21	0,17	0,35	0,19	0,19	0,25	0,48	0,22	22%	18%	23%	26%	0,12	0,07	0,15	0,04	0,11	0,07	0,11	0,06	0,13	0,10	0,21	0,30	12%	15%	10%	11%
Metaboli	sNna	0,25	0,47	0,48	0,77	1,14	0,03	0,45	0,19	0,29	0,12	0,43	0,44	42%	25%	45%	56%	0,13	0,32	0,06	0,11	0,10	0,02	0,05	0,26	0,09	0,27	0,35	0,50	19%	19%	12%	26%
Metaboli	sOHl	0,26	0,30	0,14	0,05	0,11	0,03	0,02	0,09	0,04	0,03	0,07	0,10	10%	11%	6%	14%	0,05	0,21	0,07	0,10	0,12	0,02	0,07	0,23	0,05	0,04	0,07	0,12	10%	6%	7%	16%
Metaboli	sRut	0,09	0,18	0,01	0,03	0,16	0,02	0,13	0,09	0,05	0,05	0,10	0,11	9%	7%	6%	14%	0,05	0,09	0,08	0,04	0,06	0,02	0,04	0,18	0,08	0,26	0,16	0,10	10%	6%	11%	12%
Enzyme	eAlAT	0,79	0,45	0,11	0,63	0,18	0,40	0,13	0,40	0,28	0,00	0,86	0,45	39%	48%	22%	47%	0,56	0,32	0,70	0,88	0,59	0,53	0,30	0,94	0,15	0,93	0,00	0,40	52%	41%	70%	46%
Enzyme	eAsAT	0,30	0,13	0,16	0,19	0,48	0,46	0,14	0,36	0,20	0,61	0,12	0,26	28%	31%	27%	27%	0,17	0,17	0,59	0,63	0,35	0,70	0,45	0,15	0,72	0,19	0,19	0,10	37%	42%	47%	21%
Enzyme	eSKDH	0,15	0,10	0,03	0,27	0,04	0,07	0,07	0,10	0,11	0,06	0,40	0,11	13%	11%	11%	16%	0,13	0,03	0,11	0,08	0,05	0,09	0,06	0,13	0,14	0,19	0,04	0,03	9%	10%	11%	6%
Enzyme	ePGK	0,11	0,28	0,27	0,34	0,52	0,33	0,17	0,25	0,32	0,24	0,35	0,27	29%	25%	25%	35%	0,67	0,38	0,10	0,42	0,79	0,69	0,14	0,23	0,73	0,01	0,24	0,80	43%	72%	17%	41%
Enzyme	eTPI	0,14	0,18	0,34	0,22	0,77	0,16	0,21	0,18	0,19	0,23	0,31	0,45	28%	24%	25%	36%	0,23	0,20	0,36	0,74	0,21	0,28	0,50	0,57	0,61	0,46	0,47	0,33	41%	36%	51%	36%
Enzyme	eCiS	0,31	0,04	0,45	0,00	0,14	0,03	0,00	0,72	0,00	0,18	0,68	0,39	25%	18%	16%	40%	0,36	0,55	0,61	0,80	0,00	0,00	#DIV/0!	0,00	0,26	0,00	0,33	0,15	#DIV/0!	19%	#DIV/0!	22%
Enzyme	eEno	0,10	0,23	0,26	0,04	0,25	0,24	0,04	0,15	0,03	0,05	0,27	0,14	15%	13%	10%	22%	0,10	0,24	0,13	0,26	0,39	0,25	0,46	0,04	0,24	0,11	0,11	0,20	21%	20%	24%	19%
Enzyme	eG6PDH	0,09	0,26	0,10	0,13	0,80	0,28	0,07	0,11	0,16	0,05	0,32	0,07	20%	15%	9%	37%	0,16	0,13	0,17	0,14	0,08	0,04	0,11	0,12	0,25	0,01	0,25	0,14	13%	15%	11%	15%
Enzyme	eGAPDHd	0,15	0,31	0,13	0,34	0,76	0,18	0,33	0,42	0,18	0,21	0,30	0,13	29%	16%	25%	45%	0,38	0,41	0,14	0,32	0,19	0,27	0,01	0,66	0,52	0,32	0,14	0,32	31%	37%	20%	35%
Enzyme	eGAPDHdp	0,38	0,38	0,48	0,25	0,59	0,57	0,65	0,05	0,23	0,42	0,45	0,23	39%	35%	45%	37%	0,62	0,41	0,24	0,94	0,96	0,29	0,70	0,42	0,59	0,34	0,07	0,68	52%	54%	55%	47%
Enzyme	eIDH	0,20	0,55	0,13	0,32	0,63	0,08	0,44	0,12	0,12	0,09	0,32	0,31	28%	17%	25%	41%	0,13	0,13	0,04	0,14	0,09	0,09	0,06	0,10	0,10	0,09	0,09	0,08	10%	10%	8%	10%
Enzyme	ePepC	0,19	0,15	0,08	0,05	0,66	0,32	0,07	0,08	0,08	0,12	0,33	0,05	18%	16%	8%	30%	0,28	0,06	0,41	0,27	0,34	0,07	0,59	0,18	0,06	0,31	0,44	0,22	27%	16%	39%	26%
Enzyme	ePFKa	0,15	0,26	0,10	0,19	0,04	0,19	0,09	0,14	0,30	0,07	0,25	0,06	15%	17%	11%	17%	0,16	0,38	0,23	0,16	0,03	0,05	0,18	0,23	0,29	0,15	0,16	0,16	18%	17%	18%	20%
Enzyme	ePFKp	0,23	0,33	0,48	0,13	0,21	0,32	0,04	0,12	0,29	0,11	0,22	0,25	23%	27%	19%	22%	0,11	0,04	0,36	0,15	0,04	0,10	0,23	0,25	0,33	0,09	0,21	0,19	17%	18%	21%	13%
Enzyme	ePyrK	0,05	0,21	0,06	0,05	0,02	0,40	0,16	0,58	0,32	0,17	0,35	0,24	22%	25%	11%	29%	0,33	0,50	0,27	0,24	0,19	0,13	0,35	0,12	0,14	0,08	0,27	0,09	23%	17%	23%	27%
Enzyme	eFRK	0,22	0,38	0,17	0,21	0,71	0,34	0,19	0,28	0,34	0,18	0,23	0,12	28%	25%	19%	40%	0,17	0,41	0,22	0,17	0,64	0,36	0,43	0,65	0,14	0,46	0,54	0,25	37%	23%	32%	56%
Enzyme	eGlK	0,20	0,09	0,17	0,10	0,70	0,33	0,05	0,18	0,12	0,08	0,14	0,15	19%	20%	10%	28%	0,19	0,17	0,80	0,31	0,14	0,16	0,64	0,51	0,25	0,15	0,41	0,22	33%	21%	47%	31%
Enzyme	eInvA	0,08	0,58	0,00	0,44	0,73	0,57	1,18	0,39	0,00	0,50	0,76	0,30	46%	24%	53%	62%	0,30	0,16	0,62	0,15	0,15	0,09	0,31	0,06	0,90	0,43	0,16	0,24	30%	38%	38%	13%
Enzyme	eInvN	0,23	0,72	0,88	0,09	0,72	0,36	0,21	0,24	0,74	0,40	0,60	0,18	45%	38%	39%	57%	0,16	0,25	0,47	0,22	0,18	0,24	0,62	0,01	0,46	0,81	0,24	0,23	33%	27%	53%	17%
Enzyme	eSus	0,54	0,35	0,06	0,66	0,51	0,37	0,47	0,23	0,35	0,06	0,61	0,47	39%	43%	31%	42%	0,56	0,45	0,66	0,59	0,00	0,00	0,66	0,06	0,00	0,54	0,23	0,64	37%	30%	61%	19%
Enzyme	eFbpA	0,25	0,16	0,11	0,13	0,12	0,25	0,15	0,04	0,23	0,12	0,15	0,26	16%	25%	13%	12%	0,02	0,34	0,00	0,22	0,22	0,30	0,45	0,24	0,83	0,57	0,36	0,09	30%	31%	31%	29%
Enzyme	ePgm	0,20	0,08	0,11	0,11	0,67	0,19	0,03	0,06	0,16	0,04	0,27	0,02	16%	14%	7%	27%	0,07	0,61	0,09	0,27	0,12	0,22	0,21	0,08	0,52	0,14	0,22	0,03	22%	21%	18%	26%
Enzyme	eAco	0,14	0,23	0,41	0,27	0,96	0,38	0,38	0,39	0,14	0,21	0,65	0,16	36%	21%	32%	56%	0,29	0,16	0,43	0,11	0,36	0,45	0,54	0,42	0,81	0,13	0,07	0,57	36%	53%	30%	25%
Enzyme	eFumase	0,20	0,78	0,25	0,03	0,12	0,16	0,74	0,29	0,42	0,19	0,56	0,12	32%	23%	30%	44%	0,21	0,22	0,70	0,38	0,64	0,59	0,00	0,38	0,15	0,18	0,56	0,31	36%	31%	32%	45%
Enzyme	eMDH	0,11	0,06	0,21	0,10	0,70	0,19	0,06	0,16	0,12	0,15	0,54	0,06	21%	12%	13%	36%	0,13	0,10	0,02	0,23	0,01	0,37	0,20	0,35	0,30	0,35	0,07	0,18	19%	24%	20%	13%
Enzyme	eMed	0,23	0,71	0,17	0,41	0,10	0,36	0,06	0,16	0,52	0,23	0,39	0,18	29%	32%	22%	34%	0,69	0,00	0,34	0,65	0,34	0,12	0,07	0,66	0,90	0,42	0,52	0,94	47%	66%	37%	38%
Enzyme	eMedp	0,12	0,28	0,31	0,09	0,18	0,07	0,28	0,24	0,11	0,61	0,40	0,08	23%	9%	32%	27%	0,24	0,50	0,46	0,12	0,09	0,10	0,59	0,26	0,71	0,31	0,26	0,29	33%	34%	37%	28%
Protein	sJX006	0,05	0,15	0,17	0,05	0,13	0,17	0,05	0,10	0,16	0,05	0,11	0,17	11%	14%	8%	12%	0,03	0,28	0,16	0,05	0,29	0,14	0,04	0,25	0,03	0,11	0,20	0,11	14%	8%	9%	26%
Protein	sJX019	0,04	0,03	0,16	0,06	0,06	0,11	0,12	0,08	0,12	0,16	0,08	0,21	10%	12%	12%	6%	0,10	0,18	0,05	0,07	0,08	0,02	0,08	0,13	0,12	0,04	0,06	0,04	8%	7%	6%	11%
Protein	sJX044	0,23	0,09	0,22	0,20	0,07	0,10	0,12	0,21	0,04	0,14	0,14	0,24	15%	15%	17%	13%	0,04	0,10	0,17	0,15	0,21	0,01	0,01	0,18	0,08	0,15	0,06	0,11	11%	6%	12%	14%
Protein	sJX101	0,21	0,06	0,04	0,08	0,04	0,08	0,06	0,17	0,03	0,04	0,07	0,09	8%	10%	5%	8%	0,03	0,02	0,12	0,13	0,03	0,25	0,12	0,09	0,05	0,16	0,23	0,16	12%	12%	13%	9%
Protein	sJX104	0,07	0,16	0,09	0,03	0,08	0,17	0,10	0,17	0,12	0,13	0,05	0,12	11%	12%	9%	11%	0,05	0,14	0,11	0,04	0,07	0,11	0,04	0,16	0,02	0,05	0,13	0,24	10%	11%	6%	12%
Protein	sJX156	0,11	0,18	0,03	0,14	0,02	0,10	0,16	0,21	0,22	0,07	0,06	0,11	12%	14%	10%	12%	0,16	0,08	0,20	0,25	0,06	0,12	0,16	0,22	0,11	0,09	0,05	0,12	13%	12%	18%	10%
Protein	sJX168	0,05	0,08	0,17	0,08	0,10	0,16	0,08	0,11	0,16	0,05	0,03	0,08	10%	11%	10%	8%	0,12	0,16	0,15	0,02	0,12	0,11	0,08	0,18	0,06	0,04	0,15	0,02	10%	8%	7%	15%
Protein	sJX170	0,12	0,10	0,06	0,07	0,06	0,16	0,09	0,06	0,02	0,04	0,05	0,13	8%	11%	6%	7%	0,04	0,14	0,05	0,04	0,19	0,00	0,04	0,03	0,06	0,20	0,06	0,11	8%	5%	8%	11%
Protein	sJX243	0,11	0,17	0,13	0,07	0,07	0,15	0,09	0,05	0,04	0,02	0,09	0,10	9%	10%	8%	10%	0,13	0,13	0,09	0,20	0,09	0,17	0,06	0,11	0,08	0,22	0,15	0,14	13%	13%	15%	12%
Protein	sJX248	0,28	0,34	0,12	0,13	0,26	0,06	0,08	0,15	0,22	0,25	0,16	0,12	18%	17%	14%	23%	0,17	0,05	0,23	0,24	0,02	0,17	0,47	0,10	0,03	0,15	0,12	0,48	19%	21%	27%	7%
Protein	sJX261	0,13	0,10	0,09	0,08	0,15	0,05	0,13	0,31	0,12	0,12	0,08	0,03	12%	8%	11%	16%	0,13	0,17	0,15	0,02	0,11	0,14	0,01	0,23	0,83	0,15	0,06	0,21	18%	33%	8%	14%
Protein	sJX263	0,02	0,08	0,05	0,09	0,08	0,09	0,05	0,05	0,04	0,04	0,02	0,15	6%	7%	6%	6%	0,06	0,07	0,14	0,10	0,11	0,15	0,07	0,08	0,11	0,05	0,15	0,18	11%	13%	9%	10%
Protein	sJX280	0,15	0,30	0,05	0,28	0,47	0,35	0,02	0,22	0,26	0,25	0,30	0,37	25%	28%	15%	32%	0,16	0,33	0,16	0,46	0,18	0,29	0,30	0,16	0,16	0,24	0,21	0,46	26%	27%	29%	22%
Protein	sJX337	0,03	0,02	0,05	0,06	0,04	0,04	0,02	0,10	0,09	0,07	0,02	0,04	5%	5%	5%	5%	0,05	0,11	0,05	0,17	0,11	0,11	0,43	0,08	0,08	0,10	0,06	0,11	12%	9%	19%	9%
Protein	sJX359	0,11	0,20	0,14	0,09	0,12	0,10	0,06	0,06	0,13	0,06	0,11	0,10	11%	11%	9%	12%	0,10	0,06	0,13	0,06	0,08	0,17	0,15	0,07	0,06	0,09	0,13	0,06	10%	10%	11%	9%
Protein	sJX362	0,11	0,18	0,02	0,03	0,09	0,13	0,12	0,15	0,08	0,06	0,04	0,16	10%	12%	6%	11%	0,15	0,19	0,10	0,10	0,16	0,06	0,05	0,09	0,12	0,04	0,02	0,21	11%	13%	7%	11%
Protein	sJX400	0,15	0,01	0,13	0,17	0,06	0,03	0,17	0,09	0,32	0,02	0,08	0,10	11%	15%	12%	6%	0,07	0,03	0,15	0,17	0,11	0,04	0,35	0,24	0,05	0,16	0,18	0,24	15%	10%	21%	14%
Protein	sJX166	0,05	0,02	0,09	0,20	0,02	0,05	0,02	0,14	0,04	0,06	0,10	0,06	7%	5%	9%	7%	0,15	0,08	0,03	0,18	0,14	0,08	0,12	0,06	0,04	0,11	0,13	0,14	11%	10%	11%	11%
Protein	sJX175	0,02	0,09	0,02	0,15	0,14	0,06	0,03	0,08	0,04	0,10	0,04	0,28	9%	10%	8%	9%	0,05	0,18	0,09	0,10	0,20	0,10	0,16	0,11	0,16	0,06	0,12	0,10	12%	10%	10%	15%
Protein	sJX225	0,18	0,06	0,22	0,07	0,14	0,28	0,16	0,15	0,19	0,48	0,15	0,57	22%	31%	23%	12%	0,15	0,18	0,20	0,06	0,13	0,20	0,19	0,11	0,15	0,14	0,19	0,19	16%	17%	15%	15%
Protein	sJX289	0,10	0,03	0,10	0,04	0,09	0,07	0,09	0,08	0,18	0,20	0,10	0,21	11%	14%	11%	7%	0,07	0,11	0,09	0,02	0,18	0,15	0,11	0,08	0,06	0,17	0,10	0,16	11%	11%	10%	12%
Protein	sJX347	0,07	0,11	0,23	0,16	0,05	0,04	0,02	0,12	0,07	0,09	0,07	0,10	10%	7%	13%	9%	0,11	0,21	0,11	0,26	0,22	0,19	0,16	0,07	1,00	0,13	0,24	0,24	24%	38%	17%	18%
Protein	sJX030	0,11	0,10	0,07	0,09	0,06	0,05	0,03	0,08	0,09	0,17	0,12	0,04	8%	7%	9%	9%	0,06	0,12	0,08	0,04	0,09	0,02	0,14	0,07	0,09	0,05	0,07	0,08	8%	6%	8%	9%
Protein	sJX277	0,06	0,09	0,19	0,16	0,09	0,39	0,15	0,22	0,03	0,21	0,03	0,14	15%	15%	18%	11%	0,14	0,10	0,10	0,19	0,19	0,06	0,27	0,13	0,11	0,27	0,23	0,32	18%	16%	21%	16%
Protein	sJX352	0,09	0,18	0,10	0,03	0,24	0,53	0,05	0,14	0,11	0,09	0,03	0,30	16%	26%	7%	15%	0,19	0,23	0,23	0,37	0,90	0,31	0,31	0,07	0,05	0,38	0,12	0,21	28%	19%	32%	33%
Protein	sJX036	0,19	0,13	0,24	0,28	0,13	0,01	0,04	0,31	0,14	0,13	0,16	0,33	17%	16%	17%	18%	0,08	0,03	0,07	0,07	0,09	0,03	0,18	0,03	0,52	0,06	0,08	0,04	11%	17%	9%	5%
Protein	sJX038	0,09	0,08	0,14	0,05	0,21	0,16	0,18	0,07	0,30	0,23	0,37	0,31	18%	22%	15%	18%	0,02	0,05	0,12	0,15	0,04	0,04	0,05	0,08	1,02	0,07	0,32	0,07	17%	29%	10%	12%
Protein	sJX039	0,03	0,12	0,09	0,24	0,15	0,18	0,11	0,40	0,10	0,13	0,37	0,26	18%	14%	14%	26%	0,12	0,07	0,12	0,28	0,08	0,08	0,12	0,06	0,19	0,08	0,17	0,15	13%	13%	15%	10%
Protein	sJX040	0,31	0,20	0,43	0,20	0,37	0,29	0,09	0,26	0,49	0,23	0,30	0,22	28%	33%	24%	28%	0,10	0,05	0,17	0,17	0,10	0,05	0,05	0,04	0,10	0,07	0,04	0,06	8%	8%	11%	5%
Protein	sJX244	0,29	0,55	0,26	0,23	0,16	0,28	0,25	0,13	0,20	0,12	0,13	0,50	26%	32%	22%	24%	0,13	0,15	0,10	0,13	0,12	0,04	0,30	0,15	0,13	0,15	0,13	0,48	17%	20%	17%	13%
Protein	sJX245	0,40	0,55	0,32	0,27	0,33	0,37	0,20	0,37	0,38	0,30	0,07	0,27	32%	36%	27%	33%	0,09	0,06	0,05	0,10	0,07	0,12	0,01	0,06	0,07	0,11	0,11	0,21	9%	12%	7%	8%
Protein	sJX276	0,25	0,17	0,14	0,36	0,04	0,12	0,47	0,10	0,09	0,31	0,13	0,13	19%	15%	32%	11%	0,07	0,09	0,20	0,08	0,16	0,16	0,02	0,09	0,08	0,16	0,12	0,13	11%	11%	11%	12%
Protein	sJX010	0,04	0,02	0,13	0,08	0,10	0,12	0,12	0,23	0,15	0,08	0,13	0,11	11%	10%	10%	12%	0,14	0,06	0,10	0,25	0,03	0,21	0,16	0,24	0,05	0,26	0,20	0,13	15%	13%	19%	13%
Protein	sJX013	0,13	0,05	0,05	0,07	0,08	0,14	0,14	0,12	0,02	0,11	0,03	0,03	8%	8%	9%	7%	0,17	0,13	0,16	0,16	0,13	0,19	0,10	0,16	0,12	0,17	0,08	0,20	15%	17%	15%	12%
Protein	sJX020	0,07	0,07	0,01	0,09	0,05	0,08	0,06	0,23	0,08	0,19	0,04	0,09	9%	8%	8%	10%	0,13	0,20	0,07	0,04	0,08	0,10	0,18	0,01	0,13	0,26	0,22	0,05	12%	10%	14%	13%
Protein	sJX037	0,47	0,15	0,24	0,19	0,34	0,11	0,08	0,16	0,04	0,42	0,14	0,23	22%	21%	24%	20%	0,04	0,07	0,05	0,08	0,20	0,12	0,02	0,19	0,09	0,24	0,13	0,07	11%	8%	10%	15%
Protein	sJX042	0,18	0,10	0,05	0,05	0,14	0,15	0,05	0,07	0,03	0,19	0,01	0,06	9%	11%	8%	8%	0,12	0,26	0,14	0,14	0,14	0,00	0,22	0,18	0,08	0,16	0,17	0,16	15%	9%	17%	19%
Protein	sJX062	0,06	0,13	0,10	0,15	0,07	0,11	0,07	0,19	0,19	0,10	0,09	0,27	13%	16%	10%	12%	0,06	0,15	0,09	0,08	0,22	0,20	0,16	0,03	0,20	0,14	0,09	0,08	13%	14%	12%	12%
Protein	sJX065	0,15	0,12	0,07	0,08	0,13	0,35	0,29	0,07	0,11	0,01	0,28	0,15	15%	19%	11%	15%	0,05	0,09	0,12	0,13	0,12	0,30	0,15	0,18	0,05	0,08	0,05	0,10	12%	12%	12%	11%
Protein	sJX066	0,34	0,28	0,17	0,12	0,12	0,08	0,25	0,12	0,27	0,14	0,09	0,21	18%	22%	17%	15%	0,06	0,10	0,19	0,20	0,10	0,22	0,21	0,07	0,14	0,22	0,23	0,11	16%	14%	21%	12%
Protein	sJX069	0,12	0,01	0,19	0,03	0,07	0,16	0,07	0,10	0,10	0,06	0,07	0,19	10%	14%	9%	6%	0,11	0,08	0,09	0,11	0,28	0,26	0,23	0,10	0,11	0,13	0,08	0,06	14%	13%	14%	14%
Protein	sJX147	0,54	0,10	0,47	0,85	0,64	0,26	0,68	0,52	0,45	0,34	0,36	1,18	53%	61%	59%	40%	0,49	0,13	0,75	0,64	0,14	0,10	0,36	0,20	0,04	0,35	0,26	0,18	30%	20%	53%	18%
Protein	sJX167	0,21	0,09	0,06	0,12	0,09	0,04	0,06	0,14	0,11	0,16	0,15	0,04	11%	10%	10%	12%	0,23	0,10	0,22	0,06	0,00	0,07	0,21	0,11	0,01	0,11	0,07	0,12	11%	11%	15%	7%
Protein	sJX189	0,10	0,08	0,22	0,15	0,09	0,10	0,21	0,12	0,29	0,13	0,04	0,09	13%	14%	18%	8%	0,16	0,18	0,13	0,07	0,05	0,12	0,10	0,11	0,10	0,21	0,06	0,23	13%	15%	13%	10%
Protein	sJX190	0,07	0,18	0,09	0,07	0,05	0,04	0,06	0,12	0,06	0,03	0,04	0,06	7%	6%	6%	10%	0,09	0,05	0,02	0,03	0,04	0,02	0,10	0,08	0,13	0,10	0,11	0,10	7%	8%	6%	7%
Protein	sJX192	0,10	0,19	0,22	0,07	0,13	0,04	0,10	0,06	0,04	0,07	0,02	0,11	10%	7%	11%	10%	0,09	0,11	0,18	0,13	0,05	0,15	0,11	0,20	0,15	0,16	0,04	0,25	13%	16%	15%	10%
Protein	sJX207	0,04	0,11	0,08	0,21	0,08	0,19	0,13	0,24	0,09	0,14	0,14	0,04	13%	9%	14%	14%	0,04	0,16	0,02	0,10	0,18	0,11	0,13	0,06	0,02	0,17	0,03	0,09	9%	7%	11%	11%
Protein	sJX221	0,21	0,08	0,04	0,17	0,09	0,15	0,03	0,10	0,01	0,09	0,10	0,04	9%	10%	8%	9%	0,15	0,04	0,10	0,09	0,20	0,01	0,05	0,03	0,13	0,06	0,02	0,10	8%	10%	8%	7%
Protein	sJX247	0,07	0,06	0,09	0,08	0,04	0,06	0,05	0,09	0,07	0,14	0,01	0,04	7%	6%	9%	5%	0,07	0,07	0,07	0,05	0,05	0,01	0,22	0,06	0,10	0,13	0,10	0,08	8%	7%	12%	7%
Protein	sJX254	0,17	0,19	0,09	0,14	0,31	0,22	0,15	0,19	0,30	0,31	0,31	0,16	21%	21%	17%	25%	0,08	0,16	0,20	0,23	0,03	0,10	0,07	0,21	0,26	0,22	0,11	0,25	16%	17%	18%	13%
Protein	sJX265	0,02	0,04	0,03	0,11	0,06	0,04	0,07	0,06	0,01	0,07	0,09	0,06	5%	3%	7%	6%	0,11	0,05	0,05	0,09	0,01	0,28	0,08	0,07	0,08	0,07	0,09	0,09	9%	14%	7%	5%
Protein	sJX300	0,14	0,12	0,17	0,13	0,09	0,51	0,14	0,08	0,29	0,09	0,12	0,10	17%	26%	13%	10%	0,13	0,06	0,10	0,19	0,05	0,18	0,25	0,24	0,32	0,33	0,18	0,57	22%	30%	22%	13%
Protein	sJX320	0,21	0,14	0,07	0,06	0,04	0,08	0,12	0,16	0,02	0,11	0,12	0,14	11%	11%	9%	11%	0,13	0,07	0,19	0,27	0,10	0,08	0,16	0,13	0,15	0,25	0,10	0,21	15%	14%	22%	10%
Protein	sJX331	0,18	0,11	0,12	0,03	0,07	0,06	0,07	0,13	0,05	0,17	0,01	0,13	9%	11%	10%	8%	0,03	0,11	0,18	0,06	0,07	0,20	0,13	0,04	0,07	0,13	0,04	0,06	9%	9%	12%	7%
Protein	sJX340	0,20	0,11	0,19	0,08	0,07	0,09	0,07	0,08	0,08	0,04	0,09	0,07	10%	11%	10%	8%	0,04	0,07	0,03	0,06	0,09	0,15	0,11	0,01	0,05	0,08	0,04	0,05	7%	7%	7%	5%
Protein	sJX343	0,13	0,06	0,10	0,05	0,04	0,14	0,04	0,16	0,04	0,04	0,11	0,01	8%	8%	6%	9%	0,03	0,14	0,09	0,08	0,16	0,04	0,09	0,17	0,07	0,08	0,10	0,23	11%	9%	9%	14%
Protein	sJX364	0,10	0,08	0,13	0,08	0,04	0,08	0,11	0,04	0,05	0,10	0,01	0,17	8%	10%	10%	4%	0,07	0,28	0,20	0,21	0,23	0,22	0,19	0,32	0,06	0,33	0,01	0,12	19%	12%	23%	21%
Protein	sJX366	0,13	0,05	0,08	0,12	0,09	0,03	0,31	0,10	0,19	0,14	0,12	0,08	12%	11%	16%	9%	0,10	0,24	0,05	0,07	0,16	0,19	0,04	0,11	0,10	0,03	0,15	0,17	12%	14%	5%	16%
Protein	sJX378	0,06	0,06	0,08	0,13	0,07	0,06	0,03	0,24	0,10	0,04	0,08	0,10	9%	8%	7%	11%	0,12	0,09	0,10	0,12	0,10	0,19	0,10	0,18	0,07	0,04	0,22	0,09	12%	12%	9%	15%
Protein	sJX386	0,09	0,23	0,06	0,07	0,11	0,09	0,07	0,14	0,23	0,03	0,07	0,13	11%	13%	6%	14%	0,14	0,05	0,04	0,12	0,10	0,35	0,09	0,01	0,09	0,10	0,11	0,11	11%	17%	9%	7%
Protein	sJX389	0,03	0,06	0,00	0,05	0,04	0,11	0,06	0,11	0,05	0,09	0,05	0,05	6%	6%	5%	6%	0,08	0,08	0,05	0,07	0,04	0,15	0,08	0,01	0,03	0,08	0,09	0,08	7%	8%	7%	5%
Protein	sJX415	0,15	0,05	0,16	0,13	0,09	0,43	0,12	0,14	0,06	0,10	0,07	0,07	13%	17%	13%	9%	0,53	0,06	0,19	0,13	0,08	0,50	0,24	0,61	0,13	0,40	0,04	0,06	25%	31%	24%	20%
Protein	sJX423	0,06	0,03	0,10	0,07	0,12	0,16	0,06	0,01	0,10	0,15	0,01	0,14	8%	11%	9%	4%	0,07	0,10	0,15	0,20	0,13	0,17	0,19	0,08	0,04	0,06	0,06	0,13	11%	10%	15%	9%
Protein	sJX021	0,02	0,14	0,01	0,09	0,03	0,15	0,03	0,07	0,11	0,08	0,06	0,02	7%	8%	5%	8%	0,04	0,15	0,04	0,03	0,16	0,14	0,70	0,14	0,07	0,13	0,08	0,16	15%	10%	23%	14%
Protein	sJX023	0,08	0,08	0,33	0,15	0,13	0,17	0,11	0,06	0,07	0,03	0,04	0,03	11%	9%	15%	8%	0,03	0,11	0,11	0,02	0,15	0,16	0,14	0,02	0,13	0,07	0,19	0,13	11%	11%	8%	12%
Protein	sJX143	0,16	0,17	0,13	0,06	0,20	0,04	0,11	0,10	0,76	0,16	0,14	0,11	18%	27%	11%	15%	0,14	0,09	0,19	0,14	0,14	0,20	0,19	0,32	0,28	0,21	0,20	0,06	18%	17%	18%	19%
Protein	sJX224	0,03	0,18	0,09	0,05	0,08	0,03	0,10	0,17	0,07	0,05	0,04	0,12	8%	6%	7%	12%	0,23	0,11	0,08	0,73	0,07	0,09	0,17	0,08	0,20	0,08	0,04	0,20	17%	18%	27%	8%
Protein	sJX270	0,03	0,11	0,09	0,07	0,02	0,11	0,12	0,12	0,02	0,03	0,07	0,22	8%	9%	8%	8%	0,02	0,20	0,03	0,16	0,23	0,14	0,16	0,13	0,05	0,09	0,14	0,24	13%	11%	11%	18%
Protein	sJX371	0,08	0,05	0,38	0,13	0,07	0,12	0,10	0,22	0,03	0,07	0,02	0,15	12%	10%	17%	9%	0,13	0,17	0,09	0,07	0,09	0,05	0,14	0,20	0,09	0,02	0,09	0,13	11%	10%	8%	14%
Protein	sJX373	0,06	0,12	0,02	0,11	0,14	0,09	0,03	0,10	0,07	0,05	0,05	0,14	8%	9%	5%	10%	0,07	0,14	0,19	0,07	0,10	0,12	0,08	0,10	0,17	0,15	0,04	0,36	13%	18%	13%	10%
Protein	sJX011	0,09	0,07	0,06	0,10	0,05	0,20	0,13	0,10	0,08	0,14	0,18	0,14	11%	13%	11%	10%	0,17	0,23	0,08	0,08	0,09	0,20	0,25	0,16	0,19	0,13	0,26	0,05	16%	15%	13%	18%
Protein	sJX012	0,10	0,06	0,09	0,02	0,11	0,06	0,04	0,10	0,16	0,07	0,04	0,12	8%	11%	6%	8%	0,18	0,05	0,06	0,10	0,12	0,17	0,12	0,12	0,07	0,10	0,03	0,05	10%	12%	9%	8%
Protein	sJX014	0,12	0,16	0,10	0,19	0,03	0,18	0,15	0,25	0,07	0,25	0,13	0,13	15%	13%	17%	14%	0,12	0,09	0,21	0,13	0,26	0,05	0,29	0,19	0,06	0,14	0,10	0,07	14%	7%	19%	16%
Protein	sJX018	0,18	0,17	0,17	0,24	0,09	0,22	0,34	0,01	0,03	0,29	0,06	0,21	17%	16%	26%	8%	0,04	0,16	0,44	0,20	0,12	0,34	0,11	0,18	0,14	0,13	0,30	0,08	19%	15%	22%	19%
Protein	sJX025	0,13	0,25	0,19	0,02	0,13	0,10	0,04	0,06	0,33	0,16	0,06	0,07	13%	15%	10%	13%	0,14	0,03	0,05	0,08	0,13	0,07	0,23	0,20	0,24	0,09	0,24	0,15	14%	15%	11%	15%
Protein	sJX031	0,21	0,11	0,13	0,19	0,23	0,10	0,16	0,18	0,13	0,30	0,19	0,09	17%	13%	19%	18%	0,18	0,27	0,12	0,17	0,18	0,22	0,08	0,12	0,21	0,15	0,12	0,17	17%	20%	13%	17%
Protein	sJX046	0,12	0,08	0,09	0,18	0,11	0,10	0,03	0,12	0,20	0,07	0,10	0,08	11%	13%	9%	10%	0,02	0,13	0,02	0,09	0,11	0,27	0,05	0,07	0,06	0,13	0,14	0,23	11%	14%	7%	11%
Protein	sJX054	0,26	0,13	0,08	0,11	0,06	0,16	0,10	0,09	0,13	0,11	0,20	0,09	13%	16%	10%	12%	0,13	0,17	0,16	0,18	0,11	0,10	0,10	0,28	0,29	0,08	0,10	0,19	16%	18%	13%	17%
Protein	sJX057	0,03	0,14	0,13	0,06	0,05	0,34	0,13	0,15	0,09	0,07	0,07	0,22	12%	17%	10%	10%	0,15	0,04	0,15	0,13	0,08	0,08	0,10	0,14	0,15	0,16	0,15	0,18	13%	14%	13%	10%
Protein	sJX061	0,22	0,10	0,19	0,05	0,19	0,08	0,01	0,16	0,07	0,03	0,09	0,06	10%	11%	7%	14%	0,07	0,31	0,18	0,25	0,09	0,22	0,14	0,14	0,24	0,05	0,08	0,10	16%	16%	16%	15%
Protein	sJX067	0,12	0,32	0,39	0,14	0,15	0,34	0,08	0,05	0,18	0,27	0,13	0,21	20%	21%	22%	16%	0,18	0,20	0,26	0,25	0,16	0,12	0,14	0,12	0,20	0,11	0,15	0,15	17%	16%	19%	16%
Protein	sJX078	0,06	0,05	0,30	0,06	0,03	0,10	0,06	0,02	0,10	0,03	0,08	0,05	8%	8%	11%	4%	0,08	0,10	0,03	0,08	0,15	0,08	0,09	0,16	0,07	0,17	0,04	0,02	9%	6%	9%	11%
Protein	sJX095	0,15	0,10	0,15	0,05	0,16	0,25	0,04	0,35	0,29	0,19	0,13	0,11	16%	20%	11%	18%	0,05	0,20	0,11	0,54	0,29	0,26	0,41	0,18	0,30	0,07	0,23	0,17	23%	20%	28%	23%
Protein	sJX102	0,02	0,05	0,06	0,06	0,05	0,05	0,10	0,05	0,10	0,09	0,06	0,09	7%	6%	8%	5%	0,08	0,10	0,08	0,12	0,07	0,22	0,10	0,04	0,05	0,18	0,10	0,11	10%	11%	12%	8%
Protein	sJX107	0,08	0,03	0,06	0,05	0,06	0,16	0,03	0,05	0,05	0,04	0,03	0,09	6%	10%	4%	4%	0,02	0,01	0,08	0,17	0,07	0,06	0,07	0,05	0,06	0,13	0,05	0,02	7%	4%	11%	5%
Protein	sJX114	0,08	0,10	0,10	0,06	0,10	0,04	0,02	0,15	0,33	0,07	0,09	0,16	11%	15%	6%	11%	0,05	0,06	0,10	0,12	0,09	0,16	0,07	0,06	0,01	0,18	0,07	0,09	9%	8%	12%	7%
Protein	sJX117	0,10	0,05	0,02	0,11	0,04	0,09	0,08	0,17	0,05	0,05	0,09	0,18	9%	10%	6%	9%	0,06	0,15	0,17	0,22	0,28	0,07	0,10	0,21	0,18	0,16	0,12	0,11	15%	11%	16%	19%
Protein	sJX119	0,19	0,15	0,11	0,06	0,06	0,08	0,02	0,20	0,07	0,12	0,03	0,14	10%	12%	8%	11%	0,03	0,11	0,14	0,08	0,17	0,03	0,16	0,07	0,03	0,11	0,05	0,06	9%	4%	13%	10%
Protein	sJX120	0,02	0,03	0,02	0,02	0,07	0,02	0,02	0,08	0,10	0,09	0,05	0,15	6%	7%	4%	6%	0,08	0,07	0,06	0,00	0,15	0,15	0,06	0,03	0,06	0,12	0,03	0,05	7%	8%	6%	7%
Protein	sJX122	0,04	0,07	0,02	0,08	0,03	0,03	0,10	0,14	0,06	0,02	0,04	0,07	6%	5%	6%	7%	0,03	0,07	0,03	0,10	0,13	0,11	0,07	0,12	0,14	0,14	0,03	0,07	9%	9%	8%	9%
Protein	sJX125	0,01	0,06	0,10	0,02	0,08	0,19	0,10	0,09	0,07	0,06	0,04	0,15	8%	11%	7%	7%	0,06	0,16	0,06	0,22	0,13	0,14	0,05	0,06	0,04	0,06	0,11	0,05	10%	7%	10%	12%
Protein	sJX127	0,17	0,15	0,08	0,12	0,03	0,04	0,21	0,08	0,09	0,08	0,05	0,09	10%	10%	12%	8%	0,03	0,08	0,06	0,07	0,10	0,08	0,12	0,14	0,06	0,08	0,12	0,08	9%	6%	8%	11%
Protein	sJX129	0,10	0,07	0,10	0,10	0,08	0,07	0,11	0,09	0,04	0,12	0,07	0,04	8%	7%	11%	8%	0,03	0,12	0,07	0,02	0,09	0,02	0,11	0,03	0,09	0,11	0,06	0,07	7%	5%	8%	8%
Protein	sJX137	0,12	0,04	0,13	0,06	0,08	0,17	0,17	0,08	0,12	0,09	0,04	0,26	11%	17%	11%	6%	0,17	0,05	0,07	0,23	0,17	0,23	0,12	0,14	0,12	0,10	0,04	0,90	19%	36%	13%	10%
Protein	sJX138	0,12	0,01	0,17	0,14	0,05	0,14	0,08	0,19	0,08	0,04	0,09	0,14	10%	12%	11%	9%	0,08	0,16	0,06	0,05	0,05	0,06	0,12	0,02	0,06	0,19	0,04	0,06	8%	6%	11%	7%

Protein sJX139	0,14	0,12	0,15	0,05	0,11	0,27	0,05	0,06	0,01	0,33	0,20	0,04	13%	12%	15%	12%	0,38	0,30	0,12	0,08	0,29	0,32	0,07	0,28	0,21	0,21	0,28	0,20	23%	28%	12%	28%
Protein sJX142	0,07	0,05	0,11	0,02	0,02	0,03	0,04	0,12	0,09	0,06	0,06	0,13	7%	8%	6%	6%	0,06	0,13	0,06	0,21	0,05	0,19	0,09	0,07	0,06	0,06	0,06	0,08	9%	10%	11%	8%
Protein sJX148	0,23	0,08	0,11	0,03	0,10	0,20	0,08	0,03	0,05	0,16	0,14	0,20	12%	17%	10%	9%	0,20	0,38	0,17	0,22	0,22	0,10	0,13	0,04	0,24	0,44	0,29	0,31	23%	21%	24%	23%
Protein sJX152	0,11	0,05	0,13	0,07	0,00	0,04	0,17	0,02	0,06	0,04	0,06	0,10	7%	8%	10%	3%	0,06	0,14	0,17	0,14	0,11	0,03	0,10	0,03	0,07	0,17	0,08	0,16	11%	8%	15%	9%
Protein sJX153	0,42	0,17	0,07	0,09	0,05	0,19	0,07	0,08	0,07	0,07	0,14	0,13	13%	20%	7%	11%	0,08	0,10	0,08	0,25	0,14	0,07	0,11	0,19	0,13	0,06	0,08	0,13	12%	10%	13%	13%
Protein sJX158	0,13	0,08	0,07	0,11	0,05	0,01	0,17	0,22	0,10	0,05	0,12	0,11	10%	9%	10%	12%	0,14	0,08	0,05	0,03	0,05	0,14	0,08	0,06	0,07	0,21	0,08	0,16	10%	13%	9%	7%
Protein sJX160	0,11	0,04	0,11	0,08	0,16	0,17	0,16	0,19	0,04	0,05	0,12	0,29	13%	15%	10%	13%	0,05	0,09	0,12	0,23	0,14	0,15	0,15	0,28	0,02	0,15	0,21	0,24	15%	11%	16%	18%
Protein sJX172	0,11	0,06	0,10	0,05	0,06	0,22	0,10	0,07	0,07	0,05	0,04	0,13	9%	13%	8%	6%	0,13	0,18	0,02	0,12	0,12	0,33	0,10	0,09	0,05	0,26	0,18	0,12	14%	16%	13%	14%
Protein sJX174	0,03	0,02	0,08	0,10	0,06	0,03	0,04	0,10	0,02	0,03	0,04	0,04	5%	3%	6%	6%	0,09	0,07	0,06	0,07	0,09	0,09	0,19	0,05	0,11	0,06	0,18	0,09	10%	9%	9%	10%
Protein sJX181	0,16	0,02	0,21	0,06	0,11	0,10	0,08	0,11	0,13	0,08	0,03	0,01	9%	10%	11%	7%	0,09	0,22	0,08	0,07	0,03	0,24	0,38	0,25	0,11	0,18	0,06	0,06	15%	13%	18%	14%
Protein sJX182	0,28	0,05	0,20	0,10	0,11	0,08	0,06	0,15	0,05	0,10	0,10	0,09	11%	12%	11%	10%	0,42	0,16	0,14	0,04	0,15	0,05	0,19	0,19	0,15	0,11	0,06	0,08	15%	18%	12%	14%
Protein sJX196	0,08	0,11	0,03	0,06	0,07	0,13	0,13	0,06	0,16	0,05	0,14	0,13	9%	12%	7%	9%	0,10	0,16	0,10	0,08	0,05	0,07	0,04	0,18	0,14	0,09	0,06	0,04	9%	9%	8%	11%
Protein sJX198	0,23	0,06	0,05	0,10	0,01	0,09	0,08	0,13	0,08	0,05	0,09	0,07	9%	11%	7%	7%	0,10	0,18	0,06	0,07	0,07	0,08	0,07	0,10	0,08	0,06	0,10	0,09	9%	9%	6%	11%
Protein sJX201	0,09	0,07	0,06	0,02	0,10	0,03	0,06	0,05	0,07	0,17	0,06	0,02	7%	5%	8%	7%	0,07	0,09	0,01	0,04	0,14	0,08	0,14	0,02	0,09	0,09	0,06	0,05	7%	7%	7%	8%
Protein sJX202	0,15	0,12	0,05	0,05	0,05	0,06	0,03	0,11	0,15	0,12	0,02	0,07	8%	11%	6%	7%	0,06	0,03	0,13	0,07	0,07	0,04	0,16	0,11	0,15	0,06	0,06	0,18	9%	11%	10%	7%
Protein sJX211	0,10	0,22	0,24	0,21	0,05	0,32	0,29	0,15	0,01	0,21	0,30	0,03	18%	12%	24%	18%	0,46	0,42	0,44	0,18	0,40	0,38	0,13	0,41	0,13	0,06	0,50	0,57	34%	39%	20%	43%
Protein sJX217	0,09	0,06	0,02	0,09	0,09	0,06	0,06	0,02	0,09	0,13	0,11	0,13	8%	9%	8%	7%	0,04	0,10	0,02	0,38	0,20	0,18	0,16	0,21	0,25	0,15	0,20	0,20	17%	17%	18%	18%
Protein sJX219	0,03	0,10	0,09	0,10	0,13	0,06	0,23	0,19	0,12	0,08	0,12	0,09	11%	8%	12%	13%	0,14	0,06	0,08	0,15	0,18	0,26	0,07	0,06	0,07	0,09	0,06	0,90	18%	34%	10%	9%
Protein sJX229	0,02	0,04	0,11	0,14	0,08	0,06	0,03	0,31	0,08	0,01	0,01	0,12	8%	7%	7%	11%	0,18	0,06	0,06	0,09	0,10	0,05	0,47	0,16	0,02	0,09	0,08	0,11	12%	9%	18%	10%
Protein sJX231	0,15	0,11	0,08	0,05	0,04	0,15	0,21	0,16	0,20	0,17	0,05	0,17	13%	17%	13%	9%	0,24	0,12	0,07	0,42	0,10	0,30	0,08	0,06	0,31	0,03	0,15	0,16	17%	25%	15%	11%
Protein sJX232	0,23	0,13	0,23	0,14	0,19	0,18	0,16	0,12	0,10	0,17	0,13	0,16	16%	17%	17%	14%	0,17	0,05	0,12	0,17	0,30	0,12	0,03	0,11	0,25	0,13	0,08	0,26	15%	20%	11%	13%
Protein sJX240	0,16	0,24	0,19	0,12	0,42	0,19	0,27	0,09	0,20	0,11	0,16	0,05	18%	15%	17%	23%	0,13	0,51	0,26	0,43	0,15	0,14	0,13	0,11	0,18	0,14	0,15	0,08	20%	13%	24%	23%
Protein sJX249	0,11	0,07	0,08	0,09	0,12	0,12	0,05	0,04	0,23	0,02	0,11	0,17	10%	16%	6%	9%	0,10	0,10	0,10	0,06	0,11	0,19	0,21	0,24	0,13	0,17	0,04	0,07	13%	12%	13%	12%
Protein sJX253	0,06	0,11	0,08	0,03	0,11	0,08	0,21	0,15	0,10	0,13	0,08	0,07	10%	8%	11%	11%	0,25	0,13	0,07	0,15	0,25	0,13	0,17	0,38	0,38	0,29	0,31	0,48	25%	31%	17%	27%
Protein sJX258	0,15	0,06	0,07	0,05	0,03	0,09	0,10	0,11	0,02	0,03	0,05	0,16	8%	10%	6%	6%	0,09	0,25	0,06	0,12	0,08	0,20	0,30	0,17	0,03	0,07	0,04	0,04	12%	9%	14%	14%
Protein sJX260	0,19	0,17	0,14	0,05	0,12	0,08	0,07	0,22	0,15	0,13	0,08	0,28	14%	18%	10%	15%	0,05	0,13	0,12	0,02	0,12	0,21	0,08	0,11	0,15	0,01	0,04	0,13	10%	14%	6%	10%
Protein sJX272	0,18	0,08	0,19	0,03	0,05	0,06	0,08	0,53	0,04	0,28	0,05	0,07	14%	9%	15%	18%	0,11	0,13	0,09	0,13	0,11	0,07	0,03	0,07	0,14	0,25	0,09	0,07	11%	10%	12%	10%
Protein sJX279	0,13	0,45	0,03	0,33	0,13	0,14	0,11	0,31	0,07	0,14	0,03	0,16	17%	12%	15%	23%	0,07	0,26	0,31	0,75	0,20	0,21	0,19	0,26	0,02	0,11	0,38	0,05	23%	9%	34%	28%
Protein sJX281	0,13	0,22	0,00	0,45	0,20	0,30	0,18	0,24	0,22	0,11	0,01	0,29	20%	23%	19%	17%	0,12	0,09	0,20	0,97	0,17	0,11	0,23	0,32	0,31	0,09	0,32	0,15	26%	17%	37%	22%
Protein sJX283	0,18	0,15	0,07	0,38	0,11	0,08	0,33	0,06	0,35	0,17	0,12	0,12	18%	18%	24%	11%	0,39	0,10	0,19	1,21	0,02	0,30	0,44	0,33	0,20	0,24	0,24	0,15	32%	26%	52%	17%
Protein sJX287	0,06	0,18	0,07	0,19	0,10	0,05	0,07	0,08	0,14	0,16	0,13	0,04	11%	7%	12%	12%	0,02	0,06	0,05	0,01	0,05	0,01	0,40	0,08	0,17	0,08	0,04	0,04	8%	6%	14%	6%
Protein sJX301	0,32	0,13	0,07	0,11	0,03	0,06	0,19	0,07	0,19	0,04	0,08	0,13	12%	18%	10%	8%	0,19	0,18	0,07	0,12	0,11	0,10	0,25	0,24	0,16	0,06	0,15	0,11	15%	14%	12%	17%
Protein sJX303	0,16	0,04	0,04	0,30	0,10	0,15	0,09	0,05	0,07	0,04	0,12	0,12	11%	13%	12%	8%	0,12	0,14	0,05	0,09	0,12	0,04	0,22	0,16	0,16	0,02	0,09	0,31	13%	16%	10%	12%
Protein sJX307	0,21	0,12	0,06	0,05	0,04	0,06	0,19	0,11	0,12	0,22	0,05	0,13	11%	13%	13%	8%	0,08	0,05	0,03	0,15	0,32	0,10	0,09	0,05	0,13	0,04	0,05	0,06	9%	9%	7%	12%
Protein sJX330	0,07	0,06	0,03	0,10	0,07	0,13	0,05	0,03	0,08	0,03	0,10	0,08	7%	9%	5%	6%	0,04	0,07	0,08	0,09	0,13	0,08	0,03	0,10	0,04	0,04	0,08	0,02	7%	5%	6%	10%
Protein sJX336	0,05	0,08	0,06	0,08	0,03	0,04	0,03	0,05	0,08	0,05	0,04	0,07	6%	6%	6%	5%	0,05	0,05	0,06	0,08	0,12	0,11	0,03	0,32	0,01	0,10	0,02	0,06	8%	6%	7%	13%
Protein sJX338	0,12	0,08	0,05	0,07	0,08	0,07	0,07	0,04	0,06	0,03	0,05	0,07	7%	8%	6%	6%	0,08	0,16	0,06	0,09	0,09	0,04	0,04	0,07	0,09	0,04	0,07	0,10	8%	8%	6%	10%
Protein sJX341	0,08	0,05	0,04	0,05	0,06	0,13	0,10	0,14	0,02	0,04	0,15	0,30	10%	13%	6%	10%	0,05	0,13	0,13	0,02	0,15	0,07	0,09	0,08	0,17	0,14	0,11	0,20	11%	12%	10%	12%
Protein sJX345	0,09	0,16	0,10	0,05	0,02	0,10	0,06	0,19	0,19	0,03	0,14	0,18	11%	14%	6%	13%	0,13	0,07	0,04	0,38	0,05	0,18	0,16	0,20	0,12	0,14	0,18	0,04	14%	12%	18%	13%
Protein sJX355	0,19	0,04	0,16	0,05	0,12	0,05	0,16	0,11	0,17	0,15	0,20	0,13	13%	13%	13%	12%	0,03	0,30	0,03	0,13	0,20	0,21	0,26	0,09	0,28	0,07	0,18	0,40	18%	23%	12%	19%
Protein sJX360	0,17	0,27	0,17	0,10	0,15	0,13	0,14	0,05	0,29	0,07	0,09	0,16	15%	19%	12%	14%	0,11	0,42	0,11	0,51	0,17	0,07	0,07	0,21	0,23	0,03	0,07	0,19	18%	15%	18%	22%
Protein sJX365	0,10	0,18	0,01	0,08	0,29	0,11	0,16	0,05	0,02	0,03	0,03	0,14	10%	9%	7%	14%	0,15	0,24	0,07	0,24	0,18	0,20	0,22	0,14	0,15	0,16	0,18	0,15	17%	16%	17%	18%
Protein sJX375	0,08	0,20	0,05	0,15	0,04	0,11	0,01	0,12	0,04	0,12	0,12	0,04	9%	7%	8%	12%	0,05	0,05	0,19	0,10	0,10	0,05	0,48	0,07	0,04	0,06	0,11	0,27	13%	10%	21%	8%
Protein sJX381	0,07	0,21	0,04	0,05	0,09	0,13	0,16	0,21	0,09	0,10	0,07	0,13	11%	10%	9%	15%	0,11	0,20	0,11	0,06	0,13	0,35	0,12	0,11	0,24	0,04	0,09	0,09	14%	20%	8%	13%
Protein sJX382	0,09	0,08	0,08	0,08	0,05	0,19	0,05	0,07	0,15	0,05	0,03	0,18	9%	16%	7%	6%	0,03	0,11	0,09	0,07	0,11	0,09	0,05	0,13	0,02	0,06	0,06	0,30	9%	11%	7%	10%
Protein sJX383	0,11	0,13	0,03	0,02	0,06	0,06	0,06	0,02	0,04	0,10	0,05	0,02	6%	6%	5%	6%	0,10	0,13	0,05	0,06	0,17	0,33	0,10	0,16	0,08	0,09	0,03	0,12	12%	16%	8%	12%
Protein sJX384	0,06	0,07	0,10	0,17	0,09	0,04	0,13	0,11	0,05	0,11	0,06	0,12	9%	7%	13%	8%	0,05	0,11	0,20	0,21	0,11	0,17	0,05	0,03	0,07	0,09	0,15	0,06	11%	9%	14%	10%
Protein sJX385	0,04	0,10	0,11	0,09	0,12	0,12	0,03	0,09	0,16	0,05	0,05	0,13	9%	11%	7%	9%	0,11	0,06	0,00	0,07	0,10	0,34	0,09	0,09	0,14	0,06	0,10	0,21	12%	20%	6%	9%
Protein sJX387	0,07	0,06	0,11	0,14	0,07	0,03	0,12	0,12	0,11	0,16	0,04	0,12	10%	9%	13%	7%	0,03	0,03	0,07	0,02	0,14	0,17	0,10	0,03	0,02	0,09	0,04	0,10	7%	8%	7%	6%
Protein sJX390	0,11	0,05	0,04	0,02	0,03	0,01	0,09	0,09	0,03	0,03	0,10	0,05	5%	5%	4%	7%	0,01	0,10	0,05	0,10	0,10	0,06	0,05	0,10	0,07	0,09	0,03	0,04	7%	4%	7%	8%
Protein sJX392	0,21	0,08	0,22	0,18	0,11	0,09	0,12	0,21	0,13	0,06	0,13	0,16	14%	15%	14%	13%	0,18	0,07	0,20	0,12	0,25	0,07	0,06	0,14	0,04	0,09	0,12	0,04	11%	8%	12%	15%
Protein sJX396	0,09	0,14	0,16	0,26	0,04	0,07	0,07	0,21	0,12	0,04	0,10	0,31	13%	15%	13%	12%	0,05	0,46	0,15	0,10	0,10	0,13	0,16	0,15	0,14	0,07	0,22	0,10	15%	10%	12%	23%
Protein sJX399	0,06	0,11	0,04	0,08	0,10	0,14	0,09	0,52	0,01	0,14	0,13	0,06	12%	7%	9%	22%	0,07	0,12	0,06	0,07	0,16	0,09	0,10	0,09	0,08	0,16	0,09	0,07	10%	8%	10%	11%
Protein sJX401	0,03	0,10	0,08	0,06	0,05	0,12	0,11	0,17	0,11	0,05	0,11	0,07	9%	8%	8%	11%	0,07	0,16	0,05	0,04	0,12	0,26	0,09	0,16	0,16	0,04	0,10	0,08	11%	14%	6%	14%
Protein sJX408	0,08	0,06	0,05	0,13	0,07	0,08	0,09	0,09	0,12	0,03	0,10	0,06	8%	8%	7%	8%	0,13	0,11	0,09	0,07	0,06	0,05	0,22	0,13	0,08	0,10	0,02	0,11	10%	9%	12%	8%
Protein sJX414	0,05	0,07	0,14	0,12	0,03	0,20	0,05	0,19	0,00	0,08	0,06	0,07	9%	8%	10%	9%	0,07	0,09	0,03	0,08	0,12	0,03	0,05	0,12	0,11	0,12	0,03	0,22	9%	11%	7%	9%
Protein sJX417	0,09	0,05	0,32	0,30	0,08	0,15	0,11	0,14	0,12	0,28	0,08	0,20	16%	14%	25%	9%	0,15	0,05	0,11	0,29	0,24	0,32	0,32	0,24	0,05	0,08	0,23	0,30	20%	21%	20%	19%
Protein sJX274	0,23	0,14	0,03	0,06	0,06	0,13	0,06	0,10	0,08	0,02	0,06	0,03	8%	12%	4%	9%	0,12	0,07	0,06	0,26	0,12	0,16	0,06	0,08	0,13	0,13	0,03	0,17	12%	15%	13%	8%
Protein sJX285	0,07	0,15	0,23	0,08	0,12	0,21	0,65	0,30	0,30	0,34	0,22	0,27	24%	21%	33%	20%	0,07	0,14	0,12	0,06	0,15	0,05	0,12	0,12	0,08	0,03	0,02	0,12	9%	8%	8%	10%
Protein sJX146	0,15	0,21	0,08	0,11	0,02	0,12	0,11	0,05	0,12	0,20	0,20	0,04	12%	11%	13%	12%	0,18	0,17	0,13	0,07	0,17	0,03	0,09	0,02	0,05	0,42	0,07	0,64	17%	22%	18%	11%
Protein sJX398	0,11	0,11	0,12	0,10	0,11	0,00	0,01	0,08	0,08	0,07	0,11	0,18	9%	9%	7%	10%	0,14	0,12	0,04	0,18	0,05	0,15	0,08	0,07	0,05	0,11	0,01	0,10	9%	11%	10%	6%
Protein sJX060	0,25	0,15	0,16	0,03	0,05	0,17	0,26	0,18	0,15	0,11	0,08	0,11	14%	17%	14%	11%	0,16	0,12	0,12	0,05	0,13	0,39	0,11	0,12	0,17	0,10	0,16	0,20	15%	23%	10%	13%
Protein sJX071	0,16	0,01	0,25	0,22	0,11	0,12	0,04	0,13	0,11	0,10	0,02	0,17	12%	14%	15%	7%	0,09	0,14	0,43	0,20	0,12	0,04	0,18	0,20	0,06	0,13	0,08	0,02	14%	5%	24%	14%
Protein sJX082	0,28	0,10	0,18	0,11	0,15	0,14	0,08	0,05	0,05	0,12	0,16	0,05	12%	13%	12%	11%	0,14	0,06	0,09	0,19	0,09	0,11	0,04	0,08	0,09	0,05	0,08	0,07	9%	10%	9%	8%
Protein sJX088	0,25	0,15	0,11	0,11	0,10	0,12	0,09	0,48	0,30	0,17	0,09	0,31	19%	25%	12%	21%	0,13	0,11	0,02	0,07	0,05	0,21	0,16	0,12	0,13	0,11	0,02	0,08	10%	14%	9%	7%
Protein sJX098	0,12	0,14	0,14	0,04	0,02	0,05	0,04	0,17	0,13	0,12	0,07	0,10	10%	10%	9%	10%	0,15	0,15	0,03	0,16	0,67	0,31	0,44	0,33	0,04	0,12	0,16	0,23	23%	18%	19%	33%
Protein sJX100	0,11	0,07	0,13	0,07	0,04	0,03	0,13	0,13	0,08	0,16	0,17	0,10	10%	8%	12%	10%	0,10	0,15	0,09	0,06	0,09	0,11	0,18	0,25	0,06	0,13	0,15	0,03	12%	8%	11%	16%
Protein sJX116	0,12	0,09	0,06	0,08	0,19	0,16	0,13	0,05	0,08	0,07	0,13	0,11	11%	12%	8%	12%	0,07	0,07	0,11	0,02	0,15	0,13	0,14	0,06	0,17	0,06	0,11	0,11	10%	12%	8%	10%
Protein sJX118	0,07	0,06	0,06	0,04	0,12	0,09	0,04	0,05	0,11	0,08	0,03	0,06	7%	8%	5%	7%	0,02	0,07	0,06	0,02	0,03	0,12	0,08	0,16	0,07	0,06	0,14	0,06	7%	7%	6%	10%
Protein sJX150	0,32	0,25	0,30	0,24	0,16	0,45	0,28	0,10	0,13	0,10	0,18	0,18	22%	27%	23%	17%	0,39	0,15	0,22	0,55	0,26	0,14	0,26	0,04	0,16	0,28	0,38	0,24	26%	23%	33%	21%
Protein sJX162	0,08	0,15	0,25	0,17	0,04	0,23	0,15	0,05	0,09	0,12	0,11	0,06	13%	12%	17%	9%	0,18	0,16	0,16	0,24	0,31	0,48	0,21	0,14	0,03	0,12	0,10	0,15	19%	21%	18%	18%
Protein sJX187	0,04	0,07	0,08	0,01	0,03	0,09	0,04	0,03	0,06	0,09	0,06	0,11	6%	7%	5%	5%	0,17	0,07	0,19	0,12	0,09	0,07	0,11	0,12	0,05	0,13	0,02	0,12	10%	10%	14%	7%
Protein sJX213	0,15	0,15	0,17	0,19	0,11	0,03	0,06	0,05	0,07	0,10	0,06	0,11	10%	9%	13%	9%	0,05	0,12	0,10	0,14	0,17	0,27	0,33	0,21	0,42	0,13	0,15	0,12	18%	21%	17%	16%
Protein sJX222	0,08	0,11	0,13	0,23	0,11	0,37	0,04	0,12	0,16	0,16	0,15	0,07	14%	17%	14%	12%	0,12	0,03	0,15	0,63	0,13	0,13	0,42	0,19	0,07	0,36	0,29	0,24	23%	14%	39%	16%
Protein sJX238	0,11	0,28	0,06	0,03	0,21	0,13	0,07	0,16	0,38	0,15	0,12	0,21	16%	21%	8%	19%	0,10	0,25	0,23	0,19	0,04	0,18	0,39	0,16	0,18	0,09	0,18	0,27	19%	18%	22%	16%
Protein sJX266	0,08	0,05	0,16	0,11	0,24	0,09	0,17	0,04	0,09	0,11	0,03	0,05	10%	8%	14%	9%	0,10	0,10	0,15	0,07	0,15	0,25	0,17	0,08	0,12	0,04	0,14	0,15	13%	15%	11%	12%
Protein sJX268	0,06	0,10	0,04	0,08	0,04	0,08	0,38	0,50	0,10	0,03	0,15	0,13	14%	9%	13%	20%	0,27	0,04	0,11	0,18	0,19	0,28	0,15	0,30	0,26	0,24	0,18	0,15	20%	24%	17%	18%
Protein sJX296	0,10	0,10	0,05	0,12	0,07	0,08	0,13	0,08	0,14	0,21	0,07	0,20	11%	13%	13%	8%	0,08	0,05	0,08	0,08	0,11	0,07	0,06	0,01	0,11	0,07	0,04	0,04	7%	8%	7%	5%
Protein sJX297	0,07	0,17	0,08	0,03	0,04	0,49	0,14	0,59	0,26	0,23	0,12	0,17	20%	25%	12%	23%	0,01	0,03	0,03	0,08	0,06	0,09	0,07	0,05	0,09	0,07	0,01	0,13	6%	8%	6%	4%
Protein sJX308	0,16	0,21	0,32	0,09	0,07	0,08	0,08	0,02	0,08	0,04	0,15	0,11	12%	11%	13%	11%	0,07	0,07	0,07	0,06	0,10	0,22	0,04	0,20	0,36	0,03	0,44	0,12	15%	19%	5%	20%
Protein sJX310	0,09	0,16	0,11	0,12	0,02	0,03	0,05	0,06	0,03	0,00	0,04	0,16	7%	8%	7%	7%	0,02	0,05	0,11	0,07	0,05	0,09	0,05	0,05	0,05	0,10	0,02	0,03	6%	5%	8%	4%
Protein sJX318	0,15	0,02	0,12	0,22	0,00	0,17	0,15	0,18	0,04	0,02	0,10	0,13	11%	12%	13%	8%	0,07	0,25	0,22	0,16	0,14	0,07	0,28	0,25	0,19	0,06	0,18	0,18	17%	13%	18%	20%
Protein sJX326	0,02	0,10	0,04	0,05	0,05	0,04	0,05	0,06	0,05	0,03	0,06	0,10	5%	6%	4%	6%	0,06	0,10	0,06	0,09	0,08	0,07	0,14	0,05	0,13	0,11	0,12	0,09	9%	9%	10%	9%
Protein sJX327	0,04	0,04	0,03	0,07	0,05	0,03	0,03	0,04	0,08	0,06	0,04	0,05	5%	5%	5%	4%	0,04	0,06	0,09	0,09	0,10	0,09	0,05	0,03	0,06	0,05	0,04	0,07	6%	6%	7%	6%
Protein sJX361	0,11	0,06	0,17	0,15	0,08	0,08	0,07	0,05	0,11	0,13	0,02	0,07	9%	9%	13%	5%	0,17	0,18	0,17	0,11	0,11	0,18	0,15	0,13	0,18	0,02	0,09	0,07	13%	15%	11%	13%
Protein sJX391	0,03	0,09	0,08	0,22	0,17	0,27	0,02	0,22	0,20	0,16	0,18	0,08	14%	14%	12%	17%	0,10	0,05	0,25	0,49	0,29	0,19	0,32	0,34	0,23	0,44	0,18	0,14	25%	17%	38%	21%
Protein sJX405	0,71	0,06	0,11	0,18	0,05	0,13	0,14	0,18	0,13	0,11	0,17	0,09	17%	26%	13%	11%	0,85	0,14	0,07	0,08	0,06	0,13	0,14	0,12	0,06	0,04	0,37	0,09	18%	28%	8%	17%
Protein sJX409	0,13	0,04	0,09	0,06	0,04	0,06	0,04	0,04	0,01	0,04	0,01	0,02	5%	5%	6%	3%	0,05	0,09	0,12	0,02	0,05	0,05	0,10	0,09	0,07	0,06	0,06	0,19	8%	9%	8%	7%
Protein sJX421	0,25	0,19	0,08	0,16	0,02	0,07	0,07	0,25	0,14	0,11	0,15	0,14	14%	15%	11%	15%	0,18	0,16	0,27	0,17	0,06	0,24	0,23	0,03	0,10	0,04	0,04	0,17	14%	17%	18%	7%
Protein sJX001	0,13	0,20	0,46	0,32	0,05	0,12	0,19	0,12	0,06	0,05	0,10	0,28	17%	15%	26%	12%	0,04	0,01	0,06	0,08	0,12	0,11	0,22	0,06	0,06	0,06	0,05	0,49	11%	18%	11%	6%
Protein sJX004	0,02	0,12	0,14	0,10	0,07	0,09	0,02	0,13	0,28	0,14	0,03	0,08	10%	12%	10%	9%	0,11	0,14	0,07	0,09	0,23	0,18	0,12	0,22	0,08	0,15	0,12	0,32	15%	17%	11%	18%
Protein sJX008	0,06	0,06	0,13	0,06	0,10	0,06	0,04	0,15	0,04	0,10	0,07	0,01	7%	4%	8%	10%	0,13	0,08	0,06	0,13	0,21	0,12	0,17	0,22	0,04	0,17	0,20	0,15	14%	11%	13%	18%
Protein sJX015	0,07	0,06	0,11	0,07	0,05	0,05	0,06	0,28	0,05	0,02	0,14	0,01	8%	5%	7%	13%	0,13	0,07	0,13	0,06	0,08	0,06	0,12	0,13	0,23	0,13	0,16	0,19	12%	15%	11%	11%
Protein sJX024	0,11	0,14	0,01	0,11	0,11	0,06	0,06	0,06	0,22	0,06	0,07	0,10	9%	12%	6%	10%	0,08	0,09	0,14	0,14	0,09	0,18	0,02	0,12	0,11	0,23	0,04	0,15	11%	13%	13%	8%
Protein sJX026	0,15	0,10	0,06	0,09	0,02	0,08	0,08	0,03	0,13	0,12	0,05	0,17	9%	13%	9%	5%	0,13	0,04	0,06	0,08	0,14	0,17	0,17	0,10	0,07	0,09	0,08	0,23	11%	15%	10%	9%
Protein sJX027	0,16	0,07	0,04	0,23	0,05	0,17	0,10	0,10	0,14	0,06	0,06	0,07	10%	14%	11%	7%	0,13	0,20	0,04	0,17	0,07	0,08	0,18	0,14	0,14	0,08	0,02	0,42	14%	19%	12%	11%
Protein sJX029	0,17	0,08	0,03	0,10	0,04	0,15	0,06	0,14	0,07	0,08	0,06	0,03	9%	11%	7%	8%	0,11	0,21	0,08	0,11	0,03	0,08	0,07	0,09	0,10	0,02	0,08	0,04	9%	8%	7%	10%
Protein sJX032	0,02	0,10	0,11	0,09	0,10	0,09	0,08	0,06	0,08	0,08	0,05	0,02	7%	5%	9%	8%	0,04	0,14	0,05	0,08	0,11	0,13	0,09	0,06	0,16	0,14	0,08	0,10	10%	11%	9%	10%
Protein sJX033	0,05	0,05	0,01	0,07	0,08	0,09	0,06	0,08	0,04	0,04	0,07	0,12	6%	8%	5%	7%	0,05	0,02	0,04	0,05	0,07	0,19	0,25	0,21	0,02	0,05	0,06	0,17	10%	11%	10%	9%
Protein sJX041	0,24	0,11	0,05	0,23	0,13	0,30	0,54	0,14	0,74	0,53	0,55	0,41	33%	42%	34%	23%	0,16	0,16	0,50	0,11	0,10	0,39	0,22	0,28	0,25	0,25	0,21	0,14	23%	24%	27%	19%
Protein sJX045	0,03	0,09	0,11	0,09	0,04	0,13	0,01	0,27	0,22	0,04	0,23	0,10	11%	12%	6%	16%	0,11	0,05	0,32	0,68	0,48	0,17	0,23	0,23	0,04	0,13	0,11	0,12	22%	11%	34%	22%
Protein sJX050	0,27	0,36	0,25	0,22	0,25	0,20	0,02	0,14	0,16	0,11	0,05	0,35	20%	25%	15%	20%	0,14	0,10	0,16	0,10	0,13	0,24	0,01	0,23	0,20	0,20	0,10	0,24	15%	21%	12%	14%
Protein sJX051	0,06	0,22	0,11	0,10	0,17	0,19	0,02	0,29	0,19	0,10	0,12	0,07	14%	13%	8%	20%	0,09	0,28	0,20	0,37	0,13	0,25	0,16	0,04	0,25	0,15	0,17	0,15	19%	19%	22%	16%
Protein sJX053	0,07	0,10	0,08	0,04	0,07	0,01	0,17	0,02	0,08	0,21	0,05	0,07	8%	6%	12%	6%	0,21	0,09	0,08	0,20	0,30	0,14	0,24	0,39	0,43	0,27	0,29	0,48	26%	31%	20%	27%
Protein sJX056	0,26	0,14	0,09	0,21	0,13	0,18	0,14	0,04	0,04	0,16	0,03	0,20	13%	17%	15%	8%	0,04	0,20	0,17	0,13	0,06	0,13	0,10	0,21	0,06	0,12	0,06	0,29	13%	13%	13%	13%
Protein sJX058	0,15	0,11	0,18	0,13	0,21	0,09	0,19	0,20	0,03	0,12	0,14	0,38	16%	16%	16%	17%	0,13	0,12	0,10	0,27	0,10	0,03	0,11	0,11	0,07	0,05	0,11	0,16	11%	10%	13%	11%
Protein sJX059	0,05	0,18	0,23	0,17	0,13	0,10	0,14	0,06	0,02	0,36	0,17	0,29	16%	11%	22%	14%	0,12	0,09	0,04	0,23	0,13	0,28	0,25	0,07	0,06	0,20	0,11	0,30	16%	19%	18%	10%
Protein sJX068	0,12	0,18	0,09	0,13	0,27	0,10	0,02	0,12	0,18	0,16	0,08	0,32	15%	18%	10%	16%	0,16	0,01	0,28	0,47	0,12	0,25	0,05	0,13	0,11	0,23	0,09	0,14	17%	16%	26%	9%
Protein sJX070	0,06	0,06	0,05	0,22	0,03	0,17	0,06	0,12	0,03	0,09	0,26	0,08	10%	9%	11%	12%	0,02	0,04	0,08	0,03	0,02	0,05	0,04	0,06	0,13	0,05	0,10	0,19	7%	10%	5%	5%
Protein sJX073	0,08	0,04	0,06	0,08	0,03	0,00	0,03	0,06	0,02	0,02	0,03	0,08	5%	5%	5%	4%	0,07	0,07	0,06	0,12	0,03	0,16	0,10	0,11	0,02	0,07	0,12	0,04	8%	7%	9%	8%
Protein sJX076	0,17	0,07	0,24	0,16	0,08	0,19	0,25	0,21	0,15	0,21	0,27	0,20	18%	18%	21%	16%	0,05	0,11	0,21	0,02	0,12	0,09	0,07	0,05	0,32	0,34	0,15	0,09	13%	14%	16%	11%
Protein sJX081	0,10	0,44	0,19	0,31	0,38	0,51	0,06	0,23	0,24	0,64	0,38	0,20	31%	26%	30%	36%	0,30	0,56	0,83	0,44	0,16	0,32	0,31	0,07	0,28	0,26	0,08	0,34	33%	31%	46%	22%
Protein sJX087	0,38	0,24	0,13	0,52	0,10	0,07	0,34	0,18	0,39	0,08	0,17	0,26	24%	27%	27%	17%	0,05	0,12	0,13	0,11	0,25	0,16	0,18	0,10	0,76	0,15	0,17	0,03	19%	25%	14%	16%
Protein sJX091	0,37	0,14	0,17	0,11	0,27	0,36	0,13	0,21	0,26	0,24	0,20	0,38	24%	34%	16%	21%	0,09	0,27	0,10	0,09	0,21	0,34	0,11	0,29	0,42	0,11	0,32	0,09	20%	24%	10%	27%
Protein sJX106	0,23	0,07	0,20	0,08	0,10	0,09	0,14	0,02	0,09	0,14	0,07	0,20	12%	15%	14%	7%	0,36	0,09	0,13	0,10	0,10	0,14	0,29	0,08	0,11	0,33	0,11	0,17	17%	19%	21%	9%
Protein sJX110	0,09	0,05	0,10	0,07	0,03	0,09	0,13	0,18	0,11	0,04	0,11	0,08	9%	9%	8%	9%	0,08	0,18	0,18	0,09	0,16	0,27	0,01	0,08	0,04	0,06	0,11	0,24	12%	16%	8%	13%
Protein sJX111	0,06	0,04	0,10	0,13	0,02	0,01	0,06	0,17	0,08	0,17	0,10	0,17	9%	8%	11%	8%	0,01	0,06	0,16	0,07	0,12	0,06	0,11	0,05	0,05	0,10	0,02	0,08	8%	5%	11%	7%
Protein sJX113	0,12	0,09	0,00	0,08	0,15	0,04	0,11	0,05	0,12	0,13	0,12	0,04	9%	8%	8%	10%	0,08	0,03	0,03	0,12	0,15	0,01	0,08	0,04	0,11	0,12	0,12	0,12	8%	8%	9%	8%
Protein sJX115	0,03	0,09	0,01	0,13	0,07	0,05	0,07	0,07	0,07	0,06	0,08	0,10	7%	6%	7%	8%	0,06	0,09	0,10	0,12	0,02	0,02	0,04	0,04	0,01	0,08	0,03	0,06	5%	4%	9%	4%
Protein sJX123	0,06	0,04	0,05	0,09	0,10	0,06	0,09	0,02	0,03	0,03	0,04	0,07	6%	5%	6%	5%	0,03	0,03	0,05	0,07	0,08	0,03	0,12	0,05	0,04	0,09	0,08	0,03	6%	3%	8%	6%
Protein sJX124	0,10	0,09	0,03	0,14	0,05	0,08	0,07	0,01	0,04	0,06	0,03	0,08	7%	7%	8%	5%	0,03	0,11	0,11	0,11	0,11	0,05	0,12	0,01	0,06	0,07	0,12	0,21	9%	9%	10%	9%
Protein sJX128	0,06	0,09	0,06	0,12	0,07	0,10	0,03	0,15	0,05	0,02	0,04	0,06	7%	7%	6%	9%	0,09	0,06	0,06	0,07	0,12	0,06	0,05	0,17	0,02	0,10	0,04	0,11	8%	7%	7%	10%
Protein sJX130	0,07	0,06	0,10	0,13	0,19	0,02	0,08	0,22	0,18	0,15	0,10	0,14	12%	10%	11%	14%	0,11	0,10	0,11	0,13	0,11	0,25	0,15	0,09	0,14	0,13	0,14	0,28	14%	19%	13%	11%
Protein sJX131	0,08	0,04	0,06	0,12	0,05	0,07	0,23	0,15	0,16	0,10	0,11	0,08	10%	10%	13%	9%	0,12	0,21	0,16	0,10	0,20	0,24	0,09	0,09	0,03	0,03	0,11	0,30	14%	17%	10%	15%
Protein sJX132	0,06	0,32	0,26	0,17	0,13	0,37	0,17	0,35	0,02	0,05	0,06	0,23	18%	17%	16%	21%	0,11	0,35	0,16	0,19	0,05	0,28	0,08	0,20	0,02	0,18	0,20	0,27	17%	17%	15%	20%
Protein sJX144	0,20	0,07	0,07	0,12	0,08	0,05	0,15	0,05	0,13	0,10	0,12	0,01	10%	10%	11%	8%	0,17	0,15	0,15	0,16	0,41	0,28	0,37	0,35	0,15	0,51	0,27	0,34	27%	23%	30%	29%
Protein sJX151	0,05	0,13	0,07	0,15	0,03	0,08	0,15	0,12	0,11	0,04	0,08	0,08	9%	8%	10%	9%	0,17	0,16	0,17	0,46	0,06	0,16	0,05	0,19	0,13	0,21	0,05	0,12	16%	14%	22%	12%
Protein sJX154	0,15	0,12	0,17	0,12	0,07	0,02	0,07	0,12	0,08	0,16	0,06	0,01	10%	7%	13%	9%	0,06	0,08	0,06	0,08	0,05	0,09	0,13	0,26	0,39	0,06	0,13	0,07	12%	15%	8%	13%
Protein sJX155	0,10	0,07	0,16	0,09	0,10	0,05	0,05	0,10	0,06	0,06	0,14	0,08	9%	7%	9%	10%	0,10	0,11	0,07	0,24	0,20	0,25	0,17	0,15	0,16	0,26	0,14	0,11	16%	16%	18%	15%
Protein sJX157	0,05	0,09	0,08	0,01	0,13	0,03	0,05	0,08	0,09	0,06	0,07	0,02	6%	5%	5%	9%	0,13	0,20	0,09	0,23	0,12	0,26	0,07	0,15	0,15	0,02	0,14	0,07	14%	15%	10%	15%
Protein sJX159	0,09	0,33	0,32	0,16	0,21	0,18	0,07	0,10	0,09	0,26	0,19	0,26	19%	16%	20%	21%	0,23	0,27	0,29	0,29	0,15	0,27	0,30	0,19	0,38	0,29	0,34	0,52	29%	35%	29%	24%
Protein sJX161	0,08	0,15	0,14	0,08	0,08	0,08	0,08	0,09	0,03	0,02	0,05	0,09	8%	7%	8%	9%	0,13	0,02	0,12	0,12	0,08	0,07	0,06	0,09	0,02	0,08	0,06	0,13	8%	9%	9%	6%
Protein sJX163	0,16	0,22	0,28	0,13	0,16	0,04	0,10	0,03	0,25	0,28	0,05	0,17	16%	16%	20%	12%	0,14	0,06	0,09	0,04	0,12	0,08	0,02	0,13	0,10	0,12	0,04	0,23	10%	14%	7%	9%
Protein sJX164	0,04	0,03	0,11	0,13	0,06	0,12	0,08	0,14	0,02	0,08	0,09	0,01	8%	5%	10%	8%	0,07	0,05	0,04	0,02	0,01	0,19	0,07	0,01	0,06	0,13	0,08	0,06	7%	9%	7%	4%
Protein sJX169	0,10	0,17	0,05	0,08	0,13	0,08	0,05	0,06	0,07	0,07	0,08	0,04	8%	7%	6%	11%	0,12	0,22	0,09	0,13	0,09	0,11	0,09	0,06	0,10	0,05	0,14	0,21	12%	13%	9%	12%
Protein sJX171	0,03	0,02	0,17	0,19	0,15	0,20	0,08	0,19	0,05	0,29	0,18	0,13	14%	10%	19%	13%	0,07	0,03	0,01	0,11	0,08	0,13	0,07	0,09	0,20	0,13	0,05	0,21	10%	15%	8%	6%
Protein sJX177	0,10	0,09	0,09	0,09	0,07	0,05	0,01	0,16	0,04	0,10	0,03	0,06	7%	6%	7%	9%	0,07	0,05	0,03	0,12	0,06	0,08	0,14	0,04	0,06	0,08	0,08	0,13	8%	8%	9%	6%
Protein sJX178	0,06	0,04	0,08	0,08	0,04	0,06	0,06	0,16	0,09	0,12	0,07	0,12	8%	8%	8%	8%	0,26	0,13	0,16	0,07	0,09	0,16	0,10	0,13	0,17	0,01	0,11	0,18	13%	19%	9%	12%
Protein sJX179	0,05	0,11	0,10	0,14	0,06	0,14	0,11	0,09	0,01	0,06	0,09	0,06	8%	6%	10%	9%	0,15	0,12	0,21	0,17	0,11	0,06	0,10	0,15	0,13	0,29	0,05	0,18	14%	13%	19%	11%
Protein sJX183	0,34	0,31	0,22	0,09	0,25	0,15	0,10	0,02	0,14	0,16	0,13	0,18	17%	20%	14%	18%	0,08	0,22	0,11	0,02	0,04	0,34	0,35	0,16	0,12	0,16	0,14	0,07	15%	15%	16%	14%
Protein sJX191	0,06	0,07	0,08	0,03	0,08	0,08	0,08	0,11	0,09	0,06	0,16	0,10	8%	8%	6%	10%	0,06	0,10	0,05	0,08	0,05	0,09	0,11	0,09	0,13	0,13	0,03	0,14	9%	10%	9%	7%
Protein sJX193	0,16	0,09	0,21	0,18	0,09	0,08	0,08	0,06	0,23	0,07	0,21	0,17	14%	16%	13%	11%	0,15	0,10	0,17	0,05	0,10	0,08	0,29	0,15	0,05	0,15	0,11	0,10	13%	10%	17%	12%
Protein sJX205	0,02	0,03	0,15	0,14	0,06	0,07	0,08	0,07	0,09	0,06	0,13	0,13	8%	8%	10%	7%	0,04	0,12	0,01	0,03	0,12	0,06	0,08	0,06	0,15	0,07	0,07	0,05	7%	8%	5%	10%
Protein sJX206	0,15	0,15	0,03	0,16	0,14	0,15	0,11	0,14	0,09	0,10	0,09	0,09	12%	12%	10%	13%	0,03	0,31	0,11	0,32	0,19	0,16	0,20	0,14	0,15	0,10	0,20	0,37	19%	18%	18%	21%
Protein sJX212	0,08	0,28	0,19	0,26	0,06	0,11	0,04	0,13	0,07	0,14	0,17	0,13	14%	10%	16%	16%	0,13	0,19	0,05	0,08	0,16	0,10	0,06	0,46	0,16	0,08	0,22	0,02	14%	10%	7%	26%
Protein sJX214	0,05	0,11	0,13	0,09	0,06	0,03	0,12	0,10	0,05	0,05	0,03	0,14	8%	7%	10%	7%	0,02	0,02	0,08	0,07	0,10	0,00	0,10	0,05	0,10	0,08	0,02	0,14	6%	7%	8%	4%
Protein sJX216	0,40	0,31	0,12	0,48	0,10	0,11	0,35	0,39	0,19	0,39	0,20	0,05	26%	19%	33%	25%	0,05	0,56	0,15	0,06	0,17	0,13	0,48	0,60	0,29	0,09	0,19	0,11	24%	15%	19%	38%
Protein sJX218	0,16	0,14	0,26	0,11	0,02	0,12	0,07	0,14	0,04	0,06	0,04	0,06	10%	9%	13%	8%	0,11	0,06	0,05	0,05	0,03	0,07	0,09	0,05	0,20	0,18	0,16	0,14	10%	13%	9%	7%
Protein sJX220	0,09	0,10	0,11	0,05	0,04	0,14	0,12	0,06	0,11	0,10	0,10	0,26	11%	15%	10%	8%	0,05	0,16	0,15	0,16	0,23	0,12	0,15	0,17	0,18	0,38	0,13	0,23	17%	14%	21%	17%
Protein sJX227	0,17	0,17	0,15	0,09	0,09	0,04	0,14	0,14	0,05	0,28	0,14	0,31	15%	15%	16%	13%	0,34	0,25	0,21	0,14	0,20	0,14	0,09	0,23	0,14	0,19	0,12	0,17	19%	20%	16%	20%
Protein sJX228	0,23	0,07	0,32	0,12	0,16	0,20	0,24	0,05	0,26	0,38	0,17	0,10	19%	20%	27%	11%	0,11	0,18	0,10	0,18	0,09	0,22	0,01	0,08	0,16	0,17	0,00	0,31	14%	20%	12%	9%
Protein sJX230	0,27	0,15	0,10	0,09	0,04	0,07	0,15	0,12	0,08	0,09	0,13	0,13	12%	14%	11%	11%	0,05	0,31	0,16	0,12	0,15	0,05	0,27	0,26	0,17	0,04	0,10	0,29	16%	14%	15%	20%
Protein sJX233	0,03	0,13	0,04	0,16	0,14	0,06	0,07	0,22	0,17	0,06	0,13	0,05	10%	8%	8%	15%	0,17	0,10	0,08	0,13	0,12	0,14	0,11	0,13	0,13	0,24	0,47	0,06	16%	13%	14%	20%
Protein sJX234	0,21	0,05	0,15	0,06	0,03	0,20	0,09	0,06	0,07	0,00	0,12	0,09	9%	14%	8%	7%	0,07	0,19	0,17	0,02	0,17	0,17	0,15	0,22	0,11	0,07	0,11	0,25	14%	15%	10%	17%
Protein sJX235	0,02	0,03	0,16	0,06	0,03	0,14	0,05	0,48	0,09	0,01	0,11	0,12	11%	9%	7%	16%	0,07	0,05	0,05	0,20	0,26	0,12	0,14	0,06	0,12	0,14	0,14	0,12	12%	11%	13%	13%
Protein sJX237	0,14	0,12	0,07	0,13	0,10	0,05	0,22	0,44	0,15	0,11	0,15	0,00	14%	8%	13%	20%	0,11	0,05	0,04	0,04	0,10	0,07	0,04	0,25	0,03	0,11	0,22	0,04	9%	6%	6%	16%
Protein sJX239	0,11	0,07	0,37	0,31	0,10	0,05	0,07	0,04	0,09	0,05	0,17	0,15	13%	10%	20%	9%	0,07	0,06	0,05	0,04	0,14	0,07	0,58	0,05	0,10	0,11	0,06	0,11	12%	9%	19%	8%
Protein sJX241	0,11	0,04	0,18	0,07	0,12	0,16	0,14	0,45	0,11	0,09	0,10	0,16	14%	14%	12%	18%	0,04	0,23	0,07	0,22	0,15	0,23	0,11	0,17	0,24	0,17	0,06	0,09	15%	15%	14%	15%
Protein sJX246	0,10	0,17	0,39	0,03	0,05	0,11	0,25	0,10	0,12	0,63	0,23	0,01	18%	9%	32%	14%	0,11	0,27	0,21	0,15	0,22	0,09	0,14	0,10	0,23	0,16	0,20	0,34	19%	20%	16%	20%
Protein sJX250	0,30	0,08	0,02	0,07	0,05	0,03	0,14	0,04	0,16	0,22	0,16	0,05	11%	14%	11%	8%	0,28	0,08	0,13	0,13	0,38	0,14	0,21	0,36	0,32	0,26	0,30	0,47	25%	30%	18%	28%
Protein sJX255	0,22	0,12	0,15	0,16	0,07	0,09	0,03	0,01	0,15	0,14	0,06	0,04	10%	12%	12%	7%	0,15	0,36	0,20	0,03	0,01	0,11	0,23	0,12	0,09	0,01	0,08	0,49	16%	21%	12%	14%



Protein sJX017	0,09	0,13	0,01	0,08	0,08	0,04	0,07	0,03	0,13	0,14	0,08	0,05	8%	8%	8%	8%	0,12	0,15	0,11	0,12	0,09	0,03	0,80	0,03	0,11	0,19	0,04	0,03	15%	7%	30%	8%
Protein sJX034	0,12	0,09	0,10	0,09	0,20	0,37	0,12	0,08	0,08	0,09	0,01	0,01	11%	14%	10%	9%	0,03	0,06	0,12	0,15	0,14	0,14	0,06	0,10	0,07	0,05	0,13	0,16	10%	10%	10%	11%
Protein sJX173	0,06	0,02	0,11	0,01	0,13	0,14	0,05	0,17	0,07	0,05	0,08	0,09	8%	9%	5%	10%	0,11	0,18	0,13	0,26	0,05	0,10	0,11	0,21	0,11	0,14	0,15	0,22	15%	13%	16%	15%
Protein sJX180	0,06	0,04	0,11	0,23	0,04	0,11	0,08	0,04	0,13	0,11	0,12	0,30	11%	15%	13%	6%	0,16	0,11	0,08	0,16	0,14	0,05	0,01	0,15	0,07	0,02	0,14	0,05	9%	8%	7%	13%
Protein sJX236	0,15	0,18	0,17	0,04	0,12	0,08	0,02	0,14	0,04	0,13	0,12	0,16	11%	11%	9%	14%	0,14	0,18	0,10	0,06	0,08	0,09	0,07	0,08	0,15	0,14	0,20	0,09	12%	12%	9%	14%
Protein sJX335	0,18	0,04	0,17	0,06	0,14	0,04	0,04	0,08	0,05	0,02	0,02	0,10	8%	9%	7%	7%	0,14	0,21	0,09	0,09	0,11	0,31	0,05	0,09	0,18	0,19	0,07	0,24	15%	23%	11%	12%
PhenotyJX349	0,04	0,12	0,10	0,08	0,01	0,01	0,09	0,12	0,03	0,08	0,11	0,02	7%	3%	9%	9%	0,16	0,15	0,09	0,54	0,05	0,10	0,21	0,05	0,12	0,21	0,16	0,18	17%	14%	26%	10%
Phenotyxfw	0,00	0,14	0,10	0,28	0,11	0,04	0,18	0,12	0,06	0,10	0,33	0,10	13%	5%	16%	18%	0,08	0,04	0,05	0,10	0,07	0,02	0,05	0,04	0,06	0,04	0,05	0,12	6%	7%	6%	5%
Phenotyxfd	0,02	0,05	0,00	0,10	0,06	0,02	0,05	0,04	0,02	0,04	0,12	0,05	5%	2%	5%	7%	0,03	0,02	0,03	0,06	0,03	0,01	0,03	0,02	0,04	0,01	0,03	0,04	3%	3%	4%	3%
Phenotyxdmc	0,13	0,04	0,06	0,08	0,08	0,03	0,02	0,06	0,13	0,05	0,08	0,02	7%	8%	5%	7%	0,04	0,03	0,01	0,03	0,01	0,01	0,03	0,03	0,03	0,14	0,01	0,07	4%	4%	5%	2%
Phenotypheight	0,05	0,00	0,03	0,00	0,02	0,00	0,00	0,05	0,06	0,04	0,00	0,04	2%	4%	2%	2%	0,02	0,01	0,00	0,04	0,01	0,00	0,02	0,01	0,05	0,01	0,03	0,03	2%	3%	2%	2%
Phenotypstem	0,08	0,01	0,07	0,05	0,01	0,09	0,06	0,07	0,10	0,04	0,06	0,06	6%	8%	6%	4%	0,02	0,06	0,14	0,20	0,11	0,10	0,01	0,07	0,02	0,08	0,06	0,08	8%	5%	11%	7%

CHAPTER VII: DISCUSSION AND PERSPECTIVES

CHAPTER VII: DISCUSSION AND PERSPECTIVES

The objectives of this thesis were to characterize tomato genetic diversity at the molecular and proteome levels and to identify proteins and QTLs responsible for fruit quality traits in tomato. For this purpose, three independent studies were performed leading to (1) the discovery of new SNP markers, (2) their use for association studies and finally (3) the analysis of proteome variation in relation to phenotypes. In this chapter, the major results obtained in this study are discussed and a few perspectives are proposed.

From the re-sequencing of targeted regions in two lines to whole genome re-sequencing of hundreds lines

When this PhD started in 2009, the classic Sanger sequencing method which has been extensively used for DNA sequencing was still considered as the standard method in terms of both read length and sequencing accuracy (Bonetta 2006). However, Next Generation Sequencing (NGS) techniques emerged in the mid 2000's with several sequencing technologies including Roche 454, Illumina GA and ABI SOLID. They produce much more data at a much lower price. Roche 454 (GS FLX with Titanium-series kits) at that time delivered 1 million sequence reads with an average length of 350-400 nucleotides (i.e. 400 Mb by run). Five Gb of 35 bases sequences per run (using paired-end) were obtained from the Genome Analyzer GAII (Illumina). Therefore, there was much interest in applying NGS platforms for target sequencing of specific candidate genes (Barbazuk et al. 2007; Van Tassell et al. 2008).

Previous studies showed that the polymorphism rate in tomato is relatively low (Miller and Tanksley 1990). In the present study, efforts have been put towards the identification of polymorphisms in tomato by the integration of these later technologies (Illumina and 454 Roche sequencing platforms) to achieve the re-sequencing of targeted sequences. Our efforts aimed to cover about 0.2% of the tomato genome by the amplification at regions targeted by long-range PCR in two contrasted accessions, in order to identify polymorphisms. We identified more than 3000 SNPs from which we finally validated a set of 64 of these SNPs by developing CAPS markers. These markers were successfully used to assess the genetic

Table 7-1. Comparison of sequencing platforms, sorted by reads per run. Only main commercially available platforms are shown. Modified from (Glenn 2012).

Instrument	Generation of sequencing	Run time[a]	Millions of Reads/run	Bases / read[b]	Yield MB/run
3730xl (capillary)	First	2 h	0.000096	650	0.06
PacBio RS	Third	2 h	0.01	860 -1,500	5-10
454 GS Jr. Titanium	Second	10 h	0.1	400	50
454 FLX Titanium	Second	10 h	1	400	400
454 FLX+	Second	20 h	1	650	650
Illumina MiSeq *	Second	26 h	4	150+150	1200
Illumina GAIIx	Second	14 d	300	150+150	96
Illumina HiSeq 2500 – rapid *	Second	40 h	600	150+150	180
Illumina iScanSQ	Second	8.5 d	700	100+100	140
SOLiD – 5500xl *	Second	8 d	>1,410 [c]	75+35	155,1
Illumina HiSeq 1000	Second	8.5 d	≤1500	100+100	≤300,000
Illumina HiSeq 2000	Second	11.5 d	≤3000	100+100	≤600,000

[a] Instrument time for maximum read length

[b] Average length for high quality reads >200 bases (mode is higher); typical maximum for reads ≤150 bases (most reads reach this length)

[c] Mappable reads [number of raw high quality reads (as reported for all other platforms) is higher]

*Information based on company sources alone (independent data not yet available)

diversity in a core collection of 23 accessions and to map an F2 population derived from the two sequenced lines. The detected SNPs could be assayed into genotyping array and may be useful for association studies of fruit traits. The data and matching analysis tools produced in this study provide valuable genomic resources for tomato genetic diversity analysis and dissection of the genetic basis of quantitative traits. However, the technologies are evolving very fast. In the past three years, many sequencing platforms have been developed, with yields in numbers of reads and length still increasing (**Table 7-1)**. Many bioinformatics tools have been also developed to match the increasing instrument throughput and their specificity. The third generation sequencing technologies are now being developed (**Table 7-1**). The major advantages of third-generation sequencing systems (TGS) over NGS are: (i) the use of single DNA molecules, rather than clustered amplicons, as templates for sequencing. Therefore, it is supposed to remove the phase errors encountered by NGS; (ii) It can sequence longer reads than current NGS (Schadt et al. 2010). The long reads will allow scientists to assemble genomes and study large scale structural variations for which read length is particularly important. However, researchers have not been rapid to adopt TGS because of its still high error rate. In addition, TGS produce new type of sequence data, thus new bioinformatic tools need to be developed to handle the reads with high error rate, as currently available tools have focused mainly on high-throughput, high accuracy, short-read data (Weaver 2012). Thus, third generation sequencing will maybe coexist with NGS in the next few years.

Moreover, during this period and thanks to the use of NGS, the tomato genome has been fully sequenced (Sato et al. 2012). A high quality reference genome has been annotated. More than six millions of SNPs were discovered by sequencing a *S. pimpinellifollium* accession. Technology advances and availability of a high quality reference genome paved the way for whole genome re-sequencing of hundreds of individuals in tomato. In INRA, the 8 contrasted lines studied at the proteome level were re-sequenced and more than 4 millions of SNPs identified. Associated technologies such as SNP genotyping array were also developed. Hamilton et al. (2012) sequenced the transcriptome of six tomato accessions including fresh market, processing, cherry, and closely relatives *S. pimpinellifolium*. Using the transcriptome data coupled with the draft tomato genome sequence, they identified a large collection of putative SNPs to be used in a high-throughput genotyping array used today to genotype a large range of accessions. Finkers et al. (2012) plan to re-sequence the whole genome of 84 accessions, including 10 old varieties, 43 landraces and 30 wild accessions.

This set of accessions contains rich unexploited genetic variation and will provide new insight into the organization of the tomato genome. They will include DNA of an ancient herbarium accession collection from the XVIIth century. Huang et al. (2012) re-sequenced 120 red fruited tomato lines including large fruit tomatoes (*Solanum lycopersicum),* cherry tomatoes (*S. l.* var. *cerasiforme),* and wild red fruited tomatoes (*S. pimpinellifolium*) leading to the identification of more than 6 million SNPs. Further question is how to combine all the SNPs detected in these different experiments to obtain a catalogue of tomato SNPs as SNPs may be not be detected using the same software or using the same depth. Many software packages such as CLC bio, DNASTAR, MAQ have been developed for SNP detection as reviewed by Zhang et al. (2011). In our study, we compared two software packages (CLC bio and DNASTAR) to clean and map the reads onto the reference genome and search for polymorphisms. We obtained similar results using both GA2X and 454. Discrepancies in the percentage of unmapped reads and in the number of polymorphisms detected were observed between both software packages. Previous studies already identified such differences in the performance of assemblers (Feldmeyer et al. 2011). Thus common software and criteria should be needed for the selection of polymorphism site. A common database is also needed to store and manage the data. Presently only the SolCap SNPs are managed in the solgenomics database (http://solgenomics.net/), but it is planed that it will welcome the future sequences.

In this way, the large number of SNPs and still increasing information on tomato genome will allow us to revise the core collection constructed at INRA (Ranc et al. 2008) at the whole genome level. It will also help us to assess the genome wide pattern of allele frequency variation, identify recombination hot spots, and assess linkage disequilibrium in hundreds of individuals sampling wide range of genetic diversity. Moreover, understanding the domestication process is also a key to tomato breeding. It is a unique opportunity to studying rapid evolutionary processes on a short time scale thanks to massive sequencing technologies. Furthermore, these SNPs could be used for rapid marker development for whole genome association studies or quantitative trait QTL mapping and be exploited for marker assisted selection or genomic selection.

From mid-throughput SNPlex™ assay to high-throughput SNP-chip for association mapping

As a result from the increasing number of SNP available, there has been a high demand for genotyping technologies. In 2009, a multiplex technology SNPlex was commonly used genotyping assay (Berard et al. 2009). We developed four SNPlex each carrying 48 SNPs to genotype a collection of 188 accessions including 44 S. l. lycopersicum, 127 S. l. cerasiforme and 17 S. pimpinellifolium accessions. We showed the value of this genotyping assay in tomato. The percentage of SNPs successfully scored (73%) is consistent with the success rate reported by Pindo et al. (2008) and Berard et al. (2009). The results suggest that this assay is reliable, flexible and cost-effective for medium-throughput SNP detection. However, there were also some limits such as strong selection of SNPs to put in a SNPlex, high cost per data point and high failure rate (30%). This pioneering genotyping technology opened the way to new technologies more flexible or with higher throughput, like the one proposed by Fluidigm (Moonsamy et al. 2011) or Illumina VeraCodeTM and GoldenGateTM. Although SNPlex™ assay is no more used, the SNPs used in this assay may be adapted to other genotyping platforms to be used by tomato breeders.

A total of 121 informative SNPs were obtained and used to analyze LD decay and population structure comparing SSR and SNP markers. SSR and SNP markers revealed similar structure patterns with two main groups and many intermediates, but several accessions were clustered in different groups. We then compared the associations detected using different structure covariate and samples. SNPs better performed than SSR markers for the estimation of population structure leading to a lower discovery rate of false-positive associations. Although the same number of alleles was used, the SNPs better cover the whole genome. Several associations were overlapping with already known QTL regions, demonstrating the power of our approach. However, some new associations were also identified in new regions. It is noteworthy that population structure influences a lot the association results. As shown in our study, 121 SNP corrected better for the population structure than 20 SSR markers, and much less associations were found using the structure based on these SNPs, suggesting that many associations could be due to the population structure. Furthermore, 121 markers were not sufficient to fully assess the LD structure at the genome level. As more SNP markers are today available, they will allow association studies with a higher degree of accuracy and taking into account the population structure.

As previously reported in Ranc et al. (2012), at least 50,000 SNPs would be necessary for high-resolution association mapping in tomato. Benefiting from next-generation sequencing technologies, SNP-chip carrying thousands of SNPs have been developed and used for whole genome association studies in several species (Huang et al. 2010). These studies showed the power of large amount of SNP for the assessment of population structure. However, the current two SNP genotyping arrays available, one composed of 7,000 features and the other one containing 5,000 SNP markers,
were used to genotype about 300 accessions which were selected from genetic resource collection and breeding materials to cover the entire range of genetic diversity in tomato. These data are currently under investigation to study LD patterns at a medium density scale, to refine the genetic structure of the core collection and reveal new associations for traits of agronomical interest in this species.

New genotyping approaches such as genotyping-by-sequencing (GBS) have also been developed for germplasm characterization and trait mapping in diverse organisms (Elshire et al. 2011). This approach is simple, quick, extremely specific, highly reproducible and has been used to identify and map SNPs in rice (Huang et al. 2009; Huang et al. 2010). It may be potentially used in tomato and replace genotyping array.

In our study and the pilot association report of Ranc et al. (2012), we highlighted the greater efficiency of the mixed model (K+Q) in dealing with type I error rates for association mapping in tomato. This model was also successfully used for genome wide association studies in other species (Huang et al. 2010). Mixed model is based on single locus tests combined with the control of genomic background. It can handle the confounding effects of large numbers of loci of small effect. However, for quantitative traits controlled by several large-effect loci, this approach may not be appropriate (Yang et al. 2011). Taking into account multiple cofactors in the statistical model is thus needed for association mapping. A multi-locus mixed model for mapping complex traits in structured populations has been proposed by Segura et al. (2012). This method has been applied to human and Arabidopsis thaliana data, identifying new associations and evidence for allelic heterogeneity. It should be used also in tomato.

The muti-locus mixed model is based on a linear model and can easily be extended for Bayesian analysis (Stephens and Balding 2009) and allows for the integration of linkage

mapping study (Segura et al. 2012). This kind of combined approach has been applied to designed mapping families such as nested association mapping (NAM) in maize is based on crossing diverse lines to a reference parent, was reported by Yu et al. (2008), and showed the power to identify functional markers and provide information about the genetic architecture underlying complex traits (Guo et al. 2010).

Then following question is how to validate the detected associations and to identify the polymorphisms and genes responsible of trait variation. Commonly approaches to probe this question are: to increase the population size and/or develop independent populations to validate the associations. Once a candidate region is identified, combining linkage mapping between closely related individuals differing mainly at the candidate location may be useful. Finally, once a candidate gene is identified, its modification by transformation can be planed.

Combined approach can also be potentially used for multi-parent advanced generation intercross (MAGIC) population which is another design involving intercross of multiple parents, forming a single large population (Cavanagh et al. 2008). This population has been developed in tomato in INRA. The eight parents used to develop the MAGIC population and their 4 hybrids have been studied using system biology approaches.

High-throughput technologies opened the way for Systems biology approaches

The large scale high-throughput technologies are also revolutionizing biological approaches. They allow scientists to obtain large datasets from different levels of expression or different organs and to analyze a biological system as a whole. Systems biology is an integrative way to decipher the relationships and the interactions between genes, proteins and cell elements of a biological system and to study how they impact the function and behavior of that system. Recently, system biology approaches have been used to study the natural genetic variation in plant at different levels, such as metabolomics (Keurentjes 2009), proteomics (Stylianou et al. 2008) and transcriptomics (Keurentjes et al. 2008). In our work, this new analytical approach focused on proteome, metabolome and phenotypic levels was applied to study natural genetic variation of fruit quality traits in eight diverse accessions and their four corresponding F1 at cell expansion and orange red stages of fruit development.

We identified 424 variable protein spots by combining 2-DE and nano LC MS/MS and built the first comprehensive proteome reference map of the tomato fruit pericarp at two developmental stages from 12 genotypes. We have described the physiological function of the identified proteins. These data provide experimental evidence for tomato fruit proteins that had only been predicted by genome annotation and are valuable tools for comparative studies of protein expression. The proteomic data obtained in this study can also be used to improve genome annotation (Faurobert et al. 2007; Neilson et al. 2011).

Castellana et al. (2008) demonstrated that incorporating peptide MS/MS data into automatic annotation method can improve the genome of annotation of Arabidopsis. By using proteomics approach, they discovered 778 new genes, corrected the annotations of 695 genes, and determined that approximately 13% of the Arabidopsis proteome consisted of proteins, which had been incorrectly annotated. We will test this hypothesis in the future.

In this thesis, we used a classical two-dimensional gel electrophosis (2-DE) method combined with mass spectrometry for the protein identification. 2-DE is powerful to separate protein spots but it is not a perfect method due to the distortion of protein patterns caused by polymerisation and running procedure of gels (Rabilloud et al. 2002). However, many software packages such as SameSpot have been developed to align and compare the gel images. These software packages are now capable of multiple gel analysis, including filtering of 2-DE images, automatic spot detection, normalization of the volume of each protein spot, and statistical analysis.

In recent years, gel-free shotgun techniques such as multidimensional protein identification technology (MudPIT) were also developed. One main advantage of non-gel based methods over gel-based methods is that it can estimate protein abundance more accurately and can observe differential abundance of proteins over a larger dynamic range. This includes resolving protein isoforms, thus allowing the quantification of proteins that share the same sequence but have different post-translational modifications (Bindschedler and Cramer 2011). However, the presence of a wide diversity of secondary metabolites, carbonhydrates and lipids may prevent quantitative analysis of plant proteomes, as they interfere with the gel-free techniques currently used. Until recently, two-dimensional gel electrophosis combined with MS has been the method of choice for plant protein identification; however, non-gel-based shortgun LC-MS techniques are slowly taking over (Nogueira et al. 2012).

In parallel, we quantified 34 metabolites, 26 enzyme activities and measured five phenotypic traits. A large range of variability was detected for all the traits. Several inheritance modes were observed with a majority of additivity. Data integration was achieved through sPLS correlation networks. Complex networks were described and many significant associations were observed within and between levels of expression. This analysis provides better understanding of relationships among the elements (proteins, enzymes, metabolites and phenotypic traits) leading to tomato fruit quality. Clusters of interest can then be tested for enrichment of pathways, biological process, or any other kind of annotation. The data can also be mapped onto pathways already known to be involved in the process.

The transcriptome of the eight lines and four hybrids has also been assessed by digital gene expression analysis. It will therefore be interesting to compare transcriptome changes with proteome changes using highly accurate quantitative platforms to investigate the differential regulation of gene expression. Combination of genetics and genomics with quantitative proteomics has immense potential to unravel the influence of the genotype on the cellular phenotype (Cox and Mann 2011). In addition, the eight parental lines were also re-sequenced, leading to the discovery of more than 4 millions SNPs. In the near future, all these data will be integrated together to try to construct a global network and to decipher the genetic variation of quality traits in tomato. Polymorphisms in the genes for which proteins showed quantitative variations will be analysed. In a longer term, the data obtained in this study will be used for the characterization of QTL detected in a MAGIC population. Such a population derived from the eight lines used in this study has been constructed at INRA. It will be genotyped by a SNP-chip array carrying 1536 SNPs and phenotyped in two locations and QTL will be mapped. Once a QTL is mapped it will be interesting to come back to the characterization of the parental lines and to analyze the variations of proteins and genes in QTL locations. This population will be also latter useful for fine mapping of fruit quality traits in tomato.

Conclusion

In summary, this thesis focused on using genetic and genomic approaches for the improvement of tomato fruit quality. To achieve this goal, we performed a large-scale analysis involving genome, proteome, metabolome, enzyme activity and phenotypic levels. We effectively developed molecular markers (SNPs, CAPS) and characterized proteome variations (proteome map). We successfully applied new genetic approach (association mapping) and analytical approaches (system biology approaches) for the dissection of the genetic bases of fruit quality traits. The data produced in this study provide a platform for further genetic and genomics studies in tomato. All together, we can hope it will be useful to answer consumer's quest for fruit quality.

REFERENCES

References

A

Adhikari TB, Jackson EW, Gurung S, Hansen JM, Bonman JM (2011) Association mapping of quantitative resistance to phaeosphaeria nodorum in spring wheat landraces from the USDA national small grains collection. Phytopathology 101:1301-1310.

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. Mol Breed 19:341-356.

Al-Maskri AY, Sajjad M, Khan SH (2012) Association Mapping: A step forward to discovering new alleles for crop improvement. International Journal of Agriculture and Biology 14:153-160.

Alba R, Payton P, Fei ZJ, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni JJ (2005) Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. Plant Cell 17:2954-2965.

Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? . Plant Cell 21:1877-1896.

Alpert KB, Grandillo S, Tanksley SD (1995) *fw 2.2*: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. Theoretical and Applied Genetics 91:994-1000.

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus Lycopersicon. Theoretical and Applied Genetics 103:1283-1292.

Amarjeet K, Jatendar D, Sharma OP, Gupta GD, Vandna K (2011) Lycopene. International Journal of Pharmacy and Technology 3:Review 1605-1622.

Andersen JR, Zein I, Wenzel G, Krutzfeldt B, Eder J, Ouzunova M, Lubberstedt T (2007) High levels of linkage disequilibrium and associations with forage quality at a Phenylalanine Ammonia-Lyase locus in European maize (Zea mays L.) inbreds. Theoretical and Applied Genetics 114:307-319.

Aoki K, Yano K, Suzuki A, Kawamura S, Sakurai N, Suda K, Kurabayashi A, Suzuki T, Tsugane T, Watanabe M, Ooga K, Torii M, Narita T, Shin-i T, Kohara Y, Yamamoto N, Takahashi H, Watanabe Y, Egusa M, Kodama M, Ichinose Y, Kikuchi M, Fukushima S, Okabe A, Arie T, Sato Y, Yazawa K, Satoh S, Omura T, Ezura H, Shibata D (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. BMC Genomics 11.

Aranzana MJ, Kim S, Zhao KY, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang CL, Toomajian C, Traw B, Zheng HG, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:531-539.

Atares A, Moyano E, Morales B, Schleicher P, Osvaldo Garcia-Abellan J, Anton T, Garcia-Sogo B, Perez-Martin F, Lozano R, Borja Flores F, Moreno V, del Carmen Bolarin M, Pineda B (2011) An insertional mutagenesis programme with an enhancer trap for the identification and tagging of genes involved in abiotic stress tolerance in the tomato wild-related species Solanum pennellii. Plant Cell Reports 30:1865-1879.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng DZ, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B,

Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang CL, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:627-631.

Azanza F, Young TE, Kim D, Tanksley SD, Juvik JA (1994) Characterization of the effect of introgressed segments of chromosome -7 and chromosome-10 from lycopersicon chùielewskii on tomato soluble solids, pH, and yield. Theoretical and Applied Genetics 87:965-972.

B

Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, Zimmermann P, Grossniklaus U, Gruissem W, Baginsky S (2008) Genome-scale proteomics reveals Arabidopsis thaliana gene models and proteome dynamics. Science 320:938-941.

Baldwin EA, Nisperoscarriedo MO, Baker R, Scott JW (1991) Quantitative - analysis of flavor parameters in 6 Florida tomato cultivars (*Lycopersicon esculentum Mill*). Journal of Agricultural and Food Chemistry 39:1135-1140.

Baldwin EA, Scott JW, Einstein MA, Malundo TMM, Carr BT, Shewfelt RL, Tandon KS (1998) Relationship between sensory and instrumental analysis for tomato flavor. Journal of the American Society for Horticultural Science 123:906-915.

Barbazuk WB, Emrich S, Schnable PS (2007a) SNP Mining from Maize 454 EST Sequences. CSH protocols 2007:pdb.prot4786-pdb.prot4786.

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007b) SNP discovery via 454 transcriptome sequencing. Plant Journal 51:910-918.

Barone A, Chiusano ML, Ercolano MR, Giuliano G, Grandillo S, Frusciante L (2008) Structural and functional genomics of tomato. International journal of plant genomics 2008:820274-820274.

Barrero LS, Cong B, Wu F, Tanksley SD (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 49:991-1006.

Barrero LS, Tanksley SD (2004) Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. Theoretical and Applied Genetics 109:669-679.

Baxter CJ, Sabar M, Quick WP, Sweetlove LJ (2005) Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. J Exp Bot 56:1591-1604.

Benjamini Y, Hochberg Y (2000) On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. J Edu Behav Stat 25:60-83.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53-59.

Berard A, Le Paslier MC, Dardevet M, Exbrayat-Vinson F, Bonnin I, Cenci A, Haudry A, Brunel D, Ravel C (2009) High-throughput single nucleotide polymorphism genotyping in wheat (Triticum spp.). Plant Biotechnology Journal 7:364-374.

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. Nature Reviews Genetics 11:867-879.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley S (1998) Advanced backcross QTL analysis in tomato. I. Identification of QTLs for

traits of agronomic importance from Lycopersicon hirsutum. Theoretical and Applied Genetics 97:381-397.

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum* × *L-hirsutum*: Linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147:861-877.

Bindschedler LV, Cramer R (2011) Quantitative plant proteomics. Proteomics 11:756-775.

Bonetta L (2006) Genome sequencing in the fast lane. Nature Methods 3:141-147.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J, Roux F (2010) Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet 6:e1000940.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum L.*) cultivars. Genetics 172:1165-1177.

Brewer MT, Moyseenko JB, Monforte AJ, van der Knaap E (2007) Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. J Exp Bot 58:1339-1349.

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889-890.

Brummell DA, Harpster MH (2001) Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. Plant Molecular Biology 47:311-340.

Bucheli P, Voirol E, de la Torre R, Lopez J, Rytz A, Tanksley SD, Petiard V (1999) Definition of nonvolatile markers for flavor of tomato (*Lycopersicon esculentum Mill.*) as tools in selection and breeding. Journal of Agricultural and Food Chemistry 47:659-664.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS (2009) The genetic architecture of maize flowering time. Science (Washington) 325:714-718.

C

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. Proc Nat Acad Sci USA 101:15670-15675.

Carrari F, Baxter C, Usadel B, Urbanczyk-Wochniak E, Zanor M-I, Nunes-Nesi A, Nikiforova V, Centero D, Ratzka A, Pauly M, Sweetlove LJ, Fernie AR (2006) Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiol 142:1380-1396.

Castellana NE, Payne SH, Shen Z, Stanke M, Bafna V, Briggs SP (2008) Discovery and revision of Arabidopsis genes by proteogenomics. Proc Nat Acad Sci USA 105:21034-21038.

Causse M, Buret M, Robini K, Verschave P (2003) Inheritance of nutritional and sensory quality traits in fresh market tomato and relation to consumer preferences. Journal of food science 68:2342-2350.

Causse M, Chaib J, Lecomte L, Buret M, Hospital F (2007) Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. Theoretical and Applied Genetics 115:429-442.

Causse M, Damidaux R, Rousselle P (2006) Traditional and enhanced breeding for quality traits in tomato. Genetic improvement of Solanaceous crops, Vol II Enfield (USA): Science Publishers 12:153-192.

Causse M, Duffe P, Gomez MC, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. J Exp Bot 55:1671-1685.

Causse M, Friguet C, Coiret C, Lepicier M, Navez B, Lee M, Holthuysen N, Sinesio F, Moneta E, Grandillo S (2010) Consumer Preferences for Fresh Tomato at the European Scale: A Common Segmentation on Taste and Firmness. Journal of food science 75:S531-S541.

Causse M, Saliba-Colombani V, Lecomte L, Duffe P, Rousselle P, Buret M (2002) QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J Exp Bot 53:2089-2098.

Causse M, Saliba-Colombani V, Lesschaeve I, Buret M (2001) Genetic analysis of organoleptic quality in fresh market tomato. 2. Mapping QTLs for sensory attributes. Theoretical and Applied Genetics 102:273-283.

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology 11:215-221.

Chaib J, Devaux MF, Grotte MG, Robini K, Causse M, Lahaye M, Marty I (2007) Physiological relationships among physical, sensory, and morphological attributes of texture in tomato fruits. J Exp Bot 58:1915-1925.

Chapman NH, Bonnet J, Grivet L, Lynn J, Graham N, Smith R, Sun G, Walley PG, Poole M, Causse M, King GJ, Baxter C, Seymour GB (2012) High-resolution mapping of a fruit firmness-related quantitative trait locus in tomato reveals epistatic interactions associated with a complex combinatorial locus. Plant Physiology 159:1644-1657.

Chen FQ, Foolad MR, Hyman J, St Clair DA, Beelaman RB (1999) Mapping of QTLs for lycopene and other fruit traits in a *Lycopersicon esculentum* × *L-pimpinellifolium* cross and comparison of QTLs across tomato species. Mol Breed 5:283-299.

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

Chen KY, Tanksley SD (2004) High-resolution mapping and functional analysis of se2.1: A major stigma exsertion quantitative trait locus associated with the evolution from allogamy to autogamy in the genus lycopersicon. Genetics 168:1563-1573.

Churchill G, Airey DC, Allayee H, Angel JM, Attie AD, et al (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nature Genet 36:1133-1137.

Coaker GL, Meulia T, Kabelka EA, Jones AK, Francis DM (2002) A QTL controlling stem morphology and vascular development in *Lycopersicon esculentum* × *Lycopersicon hirsutum* (Solanaceae) crosses is located on chromosome 2. American Journal of Botany 89:1859-1866.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nature Genet 40:800-804.

Cox J, Mann M (2011) Quantitative, High-Resolution Proteomics for Data-Driven Systems Biology. In: Kornberg RD, Raetz CRH, Rothman JE, Thorner JW (eds) Annual

Review of Biochemistry, Vol 80, vol 80. Annual Review of Biochemistry. pp 273-299. doi:10.1146/annurev-biochem-061308-093216

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889-1913.

D

D'Hoop BB, Paulo MJ, Mank RA, van Eck HJ, van Eeuwijk FA (2008) Association mapping of quality traits in potato (Solanum tuberosum L.). Euphytica 161:47-60.

Dai N, German MA, Matsevitz T, Hanael R, Swartzberg D, Yeselson Y, Petreikov M, Schaffer AA, Granot D (2002) LeFRK2, the gene encoding the major fructokinase in tomato fruits, is not required for starch biosynthesis in developing fruits. Plant Science 162:423-430.

Darrigues A, Schwartz SJ, Francis DM (2008) Optimizing sampling of tomato fruit for carotenoid content with application to assessing the impact of ripening disorders. Journal of Agricultural and Food Chemistry 56:483-487.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics 12:499-510.

Davidovich-Rikanati R, Sitrit Y, Tadmor Y, Iijima Y, Bilenko N, Bar E, Carmona B, Fallik E, Dudai N, Simon JE, Pichersky E, Lewinsohn E (2007) Enrichment of tomato flavor by diversion of the early plastidial terpenoid pathway. Nature Biotechnology 25:899-901.

Davies JN, Hobson GE (1981) The constituents of tomato fruit - the influence of environment, nutrition, and genotype. Crit Rev Food Sci Nutr 15:205-280.

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Car(h)(t)aGene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703-1704.

De La Vega FA, Lazaruk KD, Rhodes MD, Wenz MH (2005) Assessment of two flexible and compatible SNP genotyping platforms: TaqMan (R) SNP genotyping assays and the SNPlex (TM) genotyping system. Mutat Res-Fundam Mol Mech Mutagen 573:111-135.

de Vienne D, Bost B, Fievet J, Zivy M, Dillmann C (2001) Genetic variability of proteome expression and metabolic control. Plant Physiology and Biochemistry 39:271-283.

Deborde C, Maucourt M, Baldet P, Bernillon S, Biais B, Talon G, Ferrand C, Jacob D, Ferry-Dumazet H, de Daruvar A, Rolin D, Moing A (2009) Proton NMR quantitative profiling for quality assessment of greenhouse-grown tomato fruit. Metabolomics 5:183-198.

Deschamps S, Campbell MA (2010) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breed 25:553-570.

Devicente MC, Tanksley SD (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theoretical and Applied Genetics 83:173-178.

Di Mascio P, Kaiser S, Sies H (1989) Lycopene as the most efficient biological carotenoid singlet oxygen quencher. Archives of Biochemistry and Biophysics 274:532-538.

Do PT, Prudent M, Sulpice R, Causse M, Fernie AR (2010) The influence of fruit load on the tomato pericarp metabolome in a Solanum chmielewskii introgression line population. Plant Physiology 154:1128-1142.

Dobrowolski MP, Forster JW (2007) Linkage disequilibrium-based association mapping in forage species. Association mapping in plants. doi:10.1007/978-0-387-36011-9_9

Doganlar S, Frary A, Ku H-M, Tanksley SD (2002) Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45:1189-1202.

Dorais M, Papadopoulos AP, Gosselin A (2001) Greenhouse tomato fruit quality. In: Horticultural Reviews, vol 26. pp 239-319

E

Edwards D, Henry RJ, Edwards KJ (2012) Preface: advances in DNA sequencing accelerating plant biotechnology. Plant Biotechnology Journal 10:621-622.

Egan AN, Schlueter J, Spooner DM (2012) Application of next-generation sequencing in plant biology. American Journal of Botany 99:175-185.

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate gene association mapping of Arabidopsis flowering time. Genetics 183:325-335.

Ehrenreich IM, Stafford PA, Purugganan MD (2007) The genetic architecture of shoot branching in Arabidopsis thaliana: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics 176:1223-1236.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. Plos One 6.

Enfissi EMA, Barneche F, Ahmed I, Lichtle C, Gerrish C, McQuinn RP, Giovannoni JJ, Lopez-Juez E, Bowler C, Bramley PM, Fraser PD (2010) Integrative Transcript and Metabolite Analysis of Nutritionally Enhanced *De-etiolated1* Downregulated Tomato Fruit. Plant Cell 22:1190-1215.

Eshed Y, Gera G, Zamir D (1996) A genome-wide search for wild-species alleles that increase horticultural yield of processing tomatoes. Theoretical and Applied Genetics 93:877-886.

Eshed Y, Zamir D (1995) An introgression line population of lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147-1162.

Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. Genetics 143:1807-1817.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620.

F

Faurobert M, Iijima Y, Aoki K (2012) Proteomics and Metabolomics. In Genetics, genomics and breeding of tomato.

Faurobert M, Mihr C, Bertin N, Pawlowski T, Negroni L, Sommerer N, Causse M (2007) Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiology 143:1327-1346.

Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M (2011) Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, *Basommatophora*, *Pulmonata*), and a comparison of assembler performance. BMC Genomics 12.

Ferry-Dumazet H, Gil L, Deborde C, Moing A, Bernillon S, Rolin D, Nikolski M, de Daruvar A, Jacob D (2011) MeRy-B: a web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. BMC Plant Biology 11.

Filzmoser P, Gschwandtner M, Todorov V (2012) Review of sparse methods in regression and classification with application to chemometrics. Journal of Chemometrics 26:42-51.

Finkers R, Sandra S, Sander P (2012) 150 tomato genome (re-) sequencing project. Paper presented at the SOL 2012: From the Bench to Innovative Applications, University of Neuchâtel, Switzerland.

Fischhoff DA, Bowdish KS, Perlak FJ, Marrone PG, McCormick SM, Niedermeyer JG, Dean DA, Kusanokretzmer K, Mayer EJ, Rochester DE, Rogers SG, Fraley RT (1987) Insect tolerant transgenic tomato plants. Bio-Technology 5:807-813.

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374.

Flint J, Mott R (2001) Finding the molecular basis of quantitative traits: Successes and pitfalls. Nature Reviews Genetics 2:437-445.

Frary A, Doganlar S (2003) Comparative Genetics of Crop Plant Domestication and Evolution. Turk J Agric For 27:59-69.

Frary A, Fulton TM, Zamir D, Tanksley SD (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theoretical and Applied Genetics 108:485-496.

Frary A, Nesbitt TC, Grandillo S, van der Knaap E, Cong B, Liu JP, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science 289:85-88.

Fraser PD, Enfissi EMA, Halket JM, Truesdale MR, Yu DM, Gerrish C, Bramley PM (2007) Manipulation of phytoene levels in tomato fruit: Effects on isoprenoids, plastids, and intermediary metabolism. Plant Cell 19:3194-3211.

Fridman E, Carrari F, Liu Y-S, Fernie AR, Zamir D (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. Science 305:1786-1789.

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Nat Acad Sci USA 97:4718-4723.

Fridman E, Zamir D (2003) Functional divergence of a syntenic invertase gene family in tomato, potato, and Arabidopsis. Plant Physiology 131:603-609.

Friedman M (2002) Tomato glycoalkaloids: Role in the plant and in the diet. Journal of Agricultural and Food Chemistry 50:5751-5780.

Frusciante L, Carli P, Ercolano MR, Pernice R, Di Matteo A, Fogliano V, Pellegrini N (2007) Antioxidant nutritional quality of tomato. Molecular Nutrition & Food Research 51:609-617.

Fulton TM, Beck-Bunn T, Emmatty D, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1997) QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTLs found in other wild species. TAG Theoretical and Applied Genetics V95:881-894.

Fulton TM, Bucheli P, Voirol E, LÃpez J, PÃtiard V, Tanksley SD (2002) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica V127:163-177.

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter 13:207-209.

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. Theoretical and Applied Genetics 100:1025-1042.

G

Gady ALF, Hermans FWK, Van de Wal M, van Loo EN, Visser RGF, Bachem CWB (2009) Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations. Plant Methods 5.

Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, Lyngsoe R, Schultheiss SJ, Osborne EJ, Sreedharan VT, Kahles A, Bohnert R, Jean G, Derwent P, Kersey P, Belfield EJ, Harberd NP, Kemen E, Toomajian C, Kover PX, Clark RM, Raetsch G, Mott R (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature 477:419-423.

Ganal MW, Lapitan NLV, Tanksley SD (1988) A molecular and cytogenetic survey of major repeated DNA-sequences in tomato (lycopersicon-esculentum). Acta Horticulturae 213:262-268.

Garcia V, Stevens R, Gil L, Gilbert L, Gest N, Petit J, Faurobert M, Maucourt M, Deborde C, Moing A, Poessel J-L, Jacob D, Bouchet J-P, Giraudel J-L, Gouble B, Page D, Alhagdow M, Massot C, Gautier H, Lemaire-Chamley M, de Daruvar A, Rolin D, Usadel B, Lahaye M, Causse M, Baldet P, Rothan C (2009) An integrative genomics approach for deciphering the complex interactions between ascorbate metabolism and fruit growth and composition in tomato. Comptes Rendus Biologies 332:1007-1021.

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schuler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13:93-102.

Georgiady MS, Whitkus RW, Lord EM (2002) Genetic analysis of traits distinguishing outcrossing and self-pollinating forms of currant tomato, Lycopersicon pimpinellifolium (Jusl.) Mill. Genetics 161:333-344.

Gibon Y, Blaesing OE, Hannemann J, Carillo P, Hohne M, Hendriks JH, Palacios N, Cross J, Selbig J, Stitt M (2004) A Robot-based platform to measure multiple enzyme activities in Arabidopsis using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. The Plant cell 16:3304-3325.

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

Giovannoni J (2001) Molecular biology of fruit maturation and ripening. Annual Review of Plant Physiology and Plant Molecular Biology 52:725-749.

Giovannoni JJ (2004) Genetic regulation of fruit development and ripening. Plant Cell 16:S170-180.

Giovannucci E (1999) Tomatoes, tomato-based products, lycopene, and cancer: Review of the epidemiologic literature. Journal of the National Cancer Institute 91:317-331.

Glenn TC (2012) Field guide to next-generation DNA sequencers. Molecular Ecology Resources 11:759-769.

Goff SA (2011) A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding. New Phytologist 189:923-937.

Goldman IL, Paran I, Zamir D (1995) Quantitative trait locus analysis of a recombinant inbred line population derived from a *Lycopersicon esculentum* × *Lycopersicon cheesmanii* cross. TAG Theoretical and Applied Genetics 90:925-932.

Gomez-Romero M, Segura-Carretero A, Fernandez-Gutierrez A (2010) Metabolite profiling and quantification of phenolic compounds in methanol extracts of tomato fruit. Phytochemistry 71:1848-1864.

Gomez L, Bancel D, Rubio E, Vercambre G (2007) The microplate reader: an efficient tool for the separate enzymatic analysis of sugars in plant tissues - validation of a micro-method. Journal of the Science of Food and Agriculture 87:1893-1905.

Gomez L, Rubio E, Auge M (2002) A new procedure for extraction and measurement of soluble sugars in ligneous plants. Journal of the Science of Food and Agriculture 82:360-369.

Gonzalo MJ, van der Knaap E (2008) A comparative analysis into the genetic bases of morphology in tomato varieties exhibiting elongated fruit shape. Theoretical and Applied Genetics 116:647-656.

Gorguet B, Eggink PM, Ocana J, Tiwari A, Schipper D, Finkers R, Visser RGF, van Heusden AW (2008) Mapping and characterization of novel parthenocarpy QTLs in tomato. Theoretical and Applied Genetics 116:755-767.

Grandilio S, Ku HM, Tanksley SD (1996) Characterization of fs8.1, a major QTL influencing fruit shape in tomato. Mol Breed 2:251-260.

Grandillo S, Chetelat. R, al. e (2011) Solanum sect. Lycopersicum wild crop relatives: Genomic and breeding resources. C Kole, Springer Berlin Heidelberg:129-215.

Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. Theoretical and Applied Genetics 99:978-987.

Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. TAG Theoretical and Applied Genetics 92:935-951.

Guillet-Claude C, Birolleau-Touchard C, Manicacci D, Fourmann M, Barraud S, Carret V, Martinant JP, Barriere Y (2004) Genetic diversity associated with variation in silage corn digestibility for three O-methyltransferase genes involved in lignin biosynthesis. Theoretical and Applied Genetics 110:126-135.

Guo B, Sleper DA, Beavis WD (2010) Nested Association Mapping for Identification of Functional Markers. Genetics 186:373-383.

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Molecular Biology 57:461-485.

Gur A, Zamir D (2004) Unused natural variation can lift yield barriers in plant breeding. Plos Biology 2:1610-1615.

Gutierez RA, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR, Coruzzi GM (2008) Systems approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene CCA1. Proc Nat Acad Sci USA 105:4939-4944.

H

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.

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. Plos One 2.

Hamilton JP, Sim SC, Stoffel K, Deynze Av, Buell CR, Francis DM (2012) Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. Plant Genome 5:17-29.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2:618-620.

Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, Yamazaki Y, Kanaya S, Nakamura Y, Kitayama M, Suzuki H, Sakurai N, Shibata D, Tokuhisa J, Reichelt M, Gershenzon J, Papenbrock J, Saito K (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. Journal of Biological Chemistry 280:25590-25595.

Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc Nat Acad Sci USA 104:6478-6483.

Ho LC (1996) The mechanism of assimilate partitioning and carbohydrate compartmentation in fruit in relation Ito the quality and yield of tomato. J Exp Bot 47:1239-1243.

Hobson GE, Bedford L (1989) The composition of cherry tomatoes and its relation to consumer acceptability. Journal of Horticultural Science 64:321-329.

Holtan HEE, Hake S (2003) Quantitative trait locus analysis of leaf dissection in tomato using Lycopersicon pennellii segmental introgression lines. Genetics 165:1541-1550.

Huang S, ZHANG J, ZHANG Z (2012) A map of genome variation in the red tomato cluster. Paper presented at the SOL 2012: From the Bench to Innovative Applications, University of Neuchâtel, Switzerland.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Research 19:1068-1076.

Huang X, Paulo M-J, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Nat Acad Sci USA 108:4488-4493.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, Han B (2010a) Genome-wide association studies of 14 agronomic traits in rice landraces. Nature Genet 42:961-U976.

Huang Y-F, Madur D, Combes V, Ky CL, Coubriche D, Jamin P, Jouanne S, Dumas F, Bouty E, Bertin P, Charcosset A, Moreau L (2010b) The genetic architecture of grain yield and related traits in *Zea maize L.* revealed by comparing intermated and conventional populations. Genetics 186:395-U612.

I

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: Systems biology. Annual Review of Genomics and Human Genetics 2:343-372.

Igartua E, Casas AM, Ciudad F, Montoya JL, Romagosa I (1999) RFLP markers associated with major genes controlling heading date evaluated in a barley germ plasm pool. Heredity 83:551-559.

Initiative AG (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796-815.

International Rice Genome Sequencing P (2005) The map-based sequence of the rice genome. Nature 436:793-800.

Ivandic V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in Hordeum spontaneum. Plant Breeding 122:300-304.

Iwahashi Y, Hosoda H (2000) Effect of heat stress on tomato fruit protein expression. Electrophoresis 21:1766-1771.

J

Jansen RC, Stam P (1994) High-resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447-1455.

Jiménez-Gómez JM, Alonso-Blanco C, Borja A, Anastasio G, Angosto T, Lozano R, Martínez-Zapater JM (2007) Quantitative genetic analysis of flowering time in tomato. Genome 50:303-315.

Jimenez-Gomez JM, Maloof JN (2009) Sequence diversity in three tomato species: SNPs, markers, and molecular evolution. BMC Plant Biology 9:85.

Jones R (1986) Breeding for improved post-harvest tomato quality: genetical appearances. Acta Hortic 190:77-87.

Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol 7:198-210.

Jun T-H, Van K, Kim MY, Lee S-H, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica 162:179-191.

K

Kabelka E, Yang WC, Francis DM (2004) Improved tomato fruit color within an inbred backcross line derived from *Lycopersicon esculentum* and *L-hirsutum* involves the interaction of loci. Journal of the American Society for Horticultural Science 129:250-257.

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723.

Kelkel M, Schumacher M, Dicato M, Diederich M (2011) Antioxidant and anti-proliferative properties of lycopene. Free Radic Res 45:925-940.

Keurentjes JJB (2009) Genetical metabolomics: closing in on phenotypes. Current Opinion in Plant Biology 12:223-230.

Keurentjes JJB, Koornneef M, Vreugdenhil D (2008a) Quantitative genetics in the age of omics. Current Opinion in Plant Biology 11:123-128.

Keurentjes JJB, Sulpice R, Gibon Y, Steinhauser M-C, Fu J, Koornneef M, Stitt M, Vreugdenhil D (2008b) Integrative analyses of genetic variation in enzyme activities of primary carbohydrate metabolism reveal distinct modes of regulation in Arabidopsis thaliana. Genome Biology 9.

King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ (2012) Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Research 22:1558-1566.

Klee HJ, Giovannoni JJ (2011) Genetics and control of tomato fruit ripening and quality attributes. in: bassler bl, lichten m, schupbach g (eds) annual review genetics, Vol 45, vol 45. Annual Review of Genetics. pp 41-59.

Klee HJ, Hayford MB, Kretzmer KA, Barry GF, Kishore GM (1991) Control of ethylene synthesis by expression of a bacterial enzyme in transgenic tomato plants. Plant Cell 3:1187-1193.

Kliebenstein DJ (2009a) Advancing genetic theory and application by metabolic quantitative trait loci analysis. Plant Cell 21:1637-1646.

Kliebenstein DJ (2009b) Quantification of variation in expression networks. Methods in molecular biology (Clifton, NJ) 553:227-245.

Kliebenstein DJ (2010a) Systems Biology Uncovers the Foundation of Natural Genetic Diversity. Plant Physiology 152:480-486.

Kliebenstein DJ (2010b) Using network biology to identify quantitative genetic variation altering signaling in both plant host and generalist pathogens. Phytopathology 100:S165-S165.

Knapp S, Bohs L, Nee M, Spooner DM (2004) Solanaceae - a model for linking genomics with biodiversity. Comparative and Functional Genomics 5:285-291.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5.

Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446.

Ku HM, Doganlar S, Chen KY, Tanksley SD (1999) The genetic basis of pear-shaped tomato fruit. Theoretical and Applied Genetics 99:844-850.

Ku HM, Grandillo S, Tanksley SD (2000) fs8.1, a major QTL, sets the pattern of tomato carpel shape well before anthesis. Theoretical and Applied Genetics 101:873-878.

L

Labate JA, Baldo AM (2005) Tomato SNP discovery by EST mining and resequencing. Mol Breed 16:343-349.

Labate JA, Robertson LD, Baldo AM (2009) Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum L.*). Heredity 103:257-267.

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185-199.

Langlois D, Etievant PX, Pierron P, Jorrot A (1996) Sensory and instrumental characterisation of commercial tomato varieties. Zeitschrift Fur Lebensmittel-Untersuchung Und-Forschung 203:534-540.

Laval G, SanCristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genetics Selection Evolution 34:481-507.

Le Cao K-A, Rossouw D, Robert-Granie C, Besse P (2008) A sparse PLS for variable selection when integrating omics data. Statistical applications in genetics and molecular biology 7:35-Article 35.

Lecomte L, Duffé P, Buret M, Servin B, Hospital F, Causse M (2004a) Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds. Theoretical and Applied Genetics V109:658-668.

Lecomte L, Saliba-Colombani V, Gautier A, Gomez-Jimenez MC, Duffe P, Buret M, Causse M (2004b) Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. Mol Breed 13:1-14.

Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C (2005) Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. Plant Physiology 139:750-769.

Lemaux PG (2008) Genetically engineered plants and foods: A scientist's analysis of the issues (Part I). In: Annu. Rev. Plant Biol., vol 59. Annual Review of Plant Biology. pp 771-812.

Li L, Lu K, Chen Z, Mu T, Hu Z, Li X (2008) Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. Genetics 180:1725-1742.

Li XY, Xing JP, Gianfagna TJ, Janes HW (2002) Sucrose regulation of ADP-glucose pyrophosphorylase subunit genes transcript levels in leaves and fruits. Plant Science 162:239-244.

Lindhout P, Vanheusden S, Pet G, Vanooijen JW, Sandbrink H, Verkerk R, Vrielink R, Zabel P (1994) Perspectives of molecular marker assisted breeding for earliness in tomato. Euphytica 79:279-286.

Linnaeus C (1753) Species Plantarum. 1st ed Holmiae, Stockholm 2:1200pp.

Lippman Z, Tanksley SD (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 158:413-422.

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends in Genetics 23:60-66.

Lisec J, Romisch-Margl L, Nikoloski Z, Piepho HP, Giavalisco P, Selbig J, Gierl A, Willmitzer L (2011) Corn hybrids display lower metabolite variability and complex metabolite inheritance patterns. Plant J 68:326-336.

Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Nat Acad Sci USA 99:13302-13306.

Liu YS, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnology Journal 1:195-207.

M

Maghuly F, Kogler S, Marzban G, Nöbauer K, Razzazi E, Laimer M (2011) Proteomics, a systems biology based approach to investigations of Jatropha curcas seeds. BMC Proceedings 5:P162.

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. Genetics 175:879-889.

Mamidi S, Chikara S, Goos RJ, Hyten DL, Annam D, Moghaddam SM, Lee RK, Cregan PB, McClean PE (2011) Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. Plant Genome 4:154-164.

Manaa A, Ben Ahmed H, Valot B, Bouchet JP, Aschi-Smiti S, Causse M, Faurobert M (2011) Salt and genotype impact on plant physiology and root proteome variations in tomato. J Exp Bot 62:2797-2813.

Manoj K, Uday D (2010) RAPD based fingerprinting of tomato genotypes for identification of mutant and wild cherry specific markers. Journal of Plant Sciences 5:273-281.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380.

Marjanovic M, Stikic R, Vucelic-Radovic B, Savic S, Jovanovic Z, Bertin N, Faurobert M (2012) Growth and proteomic analysis of tomato fruit under partial root-zone drying. Omics : a journal of integrative biology 16:343-356.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nature Reviews Genetics 12:671-682.

Martinez-Valverde I, Periago MJ, Provan G, Chesson A (2002) Phenolic compounds, lycopene and antioxidant activity in commercial varieties of tomato (*Lycopersicum esculentum*). Journal of the Science of Food and Agriculture 82:323-330.

Mathieu S, Cin VD, Fei Z, Li H, Bliss P, Taylor MG, Klee HJ, Tieman DM (2009) Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. J Exp Bot 60:325-337.

Matsukura C, Aoki K, Fukuda N, Mizoguchi T, Asamizu E, Saito T, Shibata D, Ezura H (2008) Comprehensive Resources for Tomato Functional Genomics Based on the Miniature Model Tomato Micro-Tom. Current Genomics 9:436-443.

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

McKhann HI, Camilleri C, Berard A, Bataillon T, David JL, Reboud X, Le Corre V, Caloustian C, Gut IG, Brunel D (2004) Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. Plant Journal 38:193-202.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, Brown P, Browne C, Eller M, Guill K, Harjes C, Kroon D, Lepak N, Mitchell SE, Peterson B, Pressoir G, Romero S, Rosas MO, Salvo S, Yates H, Hanson M, Jones E (2009) Genetic properties of the maize nested association mapping population. Science (Washington) 325:737-740.

Menda N, Semel Y, Peled D, Eshed Y, Zamir D (2004) In silico screening of a saturated mutation library of tomato. Plant Journal 38:861-872.

Meyer RC, Kusterer B, Lisec J, Steinfath M, Becher M, Scharr H, Melchinger AE, Selbig J, Schurr U, Willmitzer L, Altmann T (2010) QTL analysis of early stage heterosis for biomass in Arabidopsis. Theoretical and Applied Genetics 120:227-237.

Mezmouk S, Dubreuil P, Bosio M, Decousset L, Charcosset A, Praud S, Mangin B (2011) Effect of population structure corrections on the results of association mapping tests in complex maize diversity panels. Theoretical and Applied Genetics 122:1149-1160.

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic-relationships and genetic-variation in the genus lycopersicon. Theoretical and Applied Genetics 80:437-448.

Miller P (1754) The gardeners dictionary. Abridged 4th ed John and James Rivington, London.

Minoia S, Petrozza A, D'Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bendahmane A, Carriero F (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC research notes 3:69-69.

Moco S, Capanoglu E, Tikunov Y, Bino RJ, Boyacioglu D, Hall RD, Vervoort J, De Vos RCH (2007) Tissue specialization at the metabolite level is perceived during the development of tomato fruit. J Exp Bot 58:4131-4146.

Monforte AJ, Friedman E, Zamir D, Tanksley SD (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: Deductions about natural variation and implications for germplasm utilization. Theoretical and Applied Genetics 102:572-590.

Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from Lycopersicon hirsutum chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theoretical and Applied Genetics 100:471-479.

Moonsamy PV, Bonella PL, Williams TC, Holcomb CL, Turenchalk GS, Blake LA, Hoglund BN, Rastrou M, Daigle DA, Simen BB, Goodridge D, Hillman G, Hamilton A, May AP, Erlich HA (2011) Use of the Fluidigm (r) access array (tm) system provides simplified amplicon library preparation in next generation sequencing for high throughput HLA genotyping. Hum Immunol 72:S142-S142.

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proc Nat Acad Sci USA 97:12649-12654.

Mounet F, Moing A, Garcia V, Petit J, Maucourt M, Deborde C, Bernillon S, Le Gall G, Colquhoun I, Defernez M, Giraudel J-L, Rolin D, Rothan C, Lemaire-Chamley M (2009) Gene and Metabolite Regulatory Network Analysis of Early Developing Fruit Tissues Highlights New Candidate Genes for the Control of Tomato Fruit Composition and Development. Plant Physiology 149:1505-1528.

Moy M, Slutzky S, Yeselson Y, Petreikov M, Shen S, Ori N, A. Schaffer A (2012) Tandem duplications in the promoter of the soluble invertase gene tiv are causal to the evolution of hexose accumulation in solanum fruit. Paper presented at the SOL 2012: From the Bench to Innovative Applications, University of Neuchâtel, Switzerland,

Moyle LC (2007) Comparative genetics of potential prezygotic and postzygotic isolating barriers in a lycopersicon species cross. Journal of Heredity 98:123-135.

Moyle LC, Graham EB (2005) Genetics of hybrid incompatibility between Lycopersicon esculentum and L-hirsutum. Genetics 169:355-373.

Moyle LC, Nakazato T (2008) Comparative genetics of hybrid incompatibility: Sterility in two Solanum species crosses. Genetics 179:1437-1453.

Mueller LA, Lankhorst RK, Tanksley SD, Giovannoni JJ, White R, Vrebalov J, Fei ZJ, Eck Jv, Buels R, Mills AA, Menda N, Tecle IY, Bombarely A, Stack S, Royer SM, Chang SB, Shearer LA, Kim B, Jo S, Hur C, Choi D, Li C, Zhao J, Jiang H, Geng Y, Dai Y (2009) A snapshot of the emerging tomato genome sequence. Plant Genome 2:78-92.

Munos S, Ranc N, Botton E, Berard A, Rolland S, Duffe P, Carretero Y, Le Paslier MC, Delalande C, Bouzayen M, Brunel D, Causse M (2011) Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. Plant physiology 156:2244-2254.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. Plant Cell 21:2194-2202.

N

Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA (2011) Less label, more free: Approaches in label-free quantitative mass spectrometry. Proteomics 11:535-553.

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

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics 12:443-451.

Nogueira FCS, Palmisano G, Schwammle V, Campos FAP, Larsen MR, Domont GB, Roepstorff P (2012) Performance of isobaric and isotopic labeling in quantitative plant proteomics. Journal of Proteome Research 11:3046-3052.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat Genet 30:190-193.

Novaes E, Drost DR, Farmerie WG, Pappas GJ, Jr., Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genomics 9.

Nuez F, Prohens J, Blanca JM (2004) Relationships origin, and diversity of Galapagos tomatoes: Implications for the conservation of natural populations. American Journal of Botany 91:86-99.

O

Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD (2004) Linkage disequilibrium mapping of Arabidopsis CRY2 flowering time alleles. Genetics 167:1361-1369.

Orsini E, Krchov LM, Uphaus J, Melchinger AE (2012) Mapping of QTL for resistance to first and second generation of European corn borer using an integrated SNP and SSR linkage map. Euphytica 183:197-206.

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction-fragment-length-polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theoretical and Applied Genetics 73:350-356.

Osorio S, Alba R, Damasceno CMB, Lopez-Casado G, Lohse M, Zanor MI, Tohge T, Usadel B, Rose JKC, Fei ZJ, Giovannoni JJ, Fernie AR (2011) Systems Biology of Tomato Fruit Development: Combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. Plant Physiology 157:405-425.

Osuna D, Usadel B, Morcuende R, Gibon Y, Blaesing OE, Hoehne M, Guenter M, Kamlage B, Trethewey R, Scheible W-R, Stitt M (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. Plant Journal 49:463-491.

P

Palma JM, Corpas FJ, del Rio LA (2011) Proteomics as an approach to the understanding of the molecular physiology of fruit development and ripening. Journal of Proteomics 74:1230-1243.

Paran I, Goldman I, Zamir D (1997) QTL analysis of morphological traits in a tomato recombinant inbred line population. Genome 40:242-248.

Paran I, van der Knaap E (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. J Exp Bot 58:3841-3852.

Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. Journal of Applied Genetics 52:413-435.

Park YH, West MAL, St Clair DA (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum L.*). Genome 47:510-518.

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato - comparision across species, generations, and environments. Genetics 127:181-197.

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine Mapping of Quantitative Trait Loci Using Selected Overlapping Recombinant Chromosomes, in an Interspecies Cross of Tomato. Genetics 124:735-742.

Paterson AH, Lander ES, Hewitt JD, Paterson S, Lincoln SL, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Pattin KA, White BC, Barney N, Gui J, Nelson HH, Kelsey KT, Andrew AS, Karagas MR, Moore JH (2009) A Computationally Efficient Hypothesis Testing Method for Epistasis Analysis Using Multifactor Dimensionality Reduction. Genetic Epidemiology 33:87-94.

Paulus C, Köllner B, Jacobsen H-J (1993) Physiological and biochemical characterization of glyoxalase I, a general marker for cell proliferation, from a soybean cell suspension. Planta 189:561-566.

Peng J, Wang H, Haley SD, Peairs FB, Lapitan NLV (2007) Molecular mapping of the Russian wheat aphid resistance gene Dn2414 in wheat. Crop Science 47:2418-2429.

Peralta IE, Knapp SK, Spooner DM (2005) New species of wild tomatoes (Solanum section Lycopersicon: Solanaceae) from Northern Peru. Syst Bot 30:424-434.

Peralta IE, Spooner D (2007) History, origin and early cultivation of tomato (*Solanaceae*). In: Razdan MK, Mattoo AK (eds) Genetic improvement of Solanaceous crops, vol 2. Science Publisher, Enfield (NH), pp 1-24

Peralta IE, Spooner DM (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (Solanum L. section Lycopersicon Mill. Wettst. subsection Lycopersicon). American Journal of Botany 88:1888-1902.

Peralta IE, Spooner DM, Knapp S (2008) Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). Systematic Botany Monographs 84.

Peterson DG, Pearson WR, Stack SM (1998) Characterization of the tomato (*Lycopersicon esculentum*) genome using in vitro and in situ DNA reassociation. Genome 41:346-356.

Pindo M, Vezzulli S, Coppola G, Cartwright DA, Zharkikh A, Velasco R, Troggio M (2008) SNP high-throughput screening in grapevine using the SNPlex$^{(TM)}$ genotyping system. BMC Plant Biology 8.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genet 38:904-909.

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. American Journal of Human Genetics 67:170-181.

Prudent M, Bertin N, Genard M, Munos S, Rolland S, Garcia V, Petit J, Baldet P, Rothan C, Causse M (2010) Genotype-dependent response to carbon availability in growing tomato fruit. Plant Cell and Environment 33:1186-1204.

Prudent M, Causse M, Genard M, Tripodi P, Grandillo S, Bertin N (2009) Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. J Exp Bot 60:923-937.

R

Rabilloud T, Luche S, Chevallet M (2002) Gel electrophoresis techniques in proteomic analysis. Biofutur:11-19.

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5:94-100.

Ramagli LS, Rodriguez LV (1985) Quantitation of microgram amounts of protein in two-dimensional polyacrylamide-gel electrophoresis sample buffer. Electrophoresis 6:559-563.

Ranc N, Munos S, Santoni S, Causse M (2008) A clarified position for solanum lycopersicum var. cerasiforme in the evolutionary history of tomatoes. BMC Plant Biology 8.

Ranc N, Munos S, Xu J, Le Paslier M-C, Chauveau A, Bounon R, Rolland S, Bouchet J-P, Brunel D, Causse M (2012) Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. G3 (Bethesda, Md) 2:853-864.

Ravel C, Praud S, Murigneux A, Linossier L, Dardevet M, Balfourier F, Dufour P, Brunel D, Charmet G (2006) Identification of Glu-B1-1 as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum L.*) by means of an association study. Theoretical and Applied Genetics 112:738-743.

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V, Ebmeyer E, Bothe R, Pietsch C, Wurschum T (2011) Association mapping for quality traits in soft winter wheat. Theoretical and Applied Genetics 122:961-970.

Remington DL, Ungerer MC, Purugganan MD (2001) Map-based cloning of quantitative trait loci: progress and prospects. Genetics Research 78:213-218.

Rick CM (1995) Lycopersicum esculentum (Solanaceae). Evolution of Crop Plants, Second Edition Longman Scientific & Technical, Essex, UK:452-457.

Rick CM, Holle M (1990) Andean *Lycopersicum esculentum* var. *cerasiforme*: Genetic Varaition and Its Evolutionary Significance. *Economic Botany* 44:69-78.

Rigola D, van Oeveren J, Janssen A, Bonne A, Schneiders H, van der Poel HJA, van Orsouw NJ, Hogers RCJ, de Both MTJ, van Eijk MJT (2009) High-throughput detection of induced mutations and natural variation using keypoint (TM) technology. Plos One 4.

Robbins MD, Sim SC, Yang WC, Van Deynze A, van der Knaap E, Joobeur T, Francis DM (2011) Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. J Exp Bot 62:1831-1845.

Robertson. L, Larate J (2007) Genetic resources of tomato (*Lycopersicon esculentum Mill.*) and wild relatives. Genetic improvement of solanaceous crops Enfield (USA): Science Publisher 2:25-75.

Rocco M, D'Ambrosio C, Arena S, Faurobert M, Scaloni A, Marra M (2006) Proteomic analysis of tomato fruits from two ecotypes during ripening. Proteomics 6:3781-3791.

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nature Reviews Genetics 7:862-872.

Roemer S, Fraser PD, Kiano JW, Shipton CA, Misawa N, Schuch W, Bramley PM (2000) Elevation of the provitamin A content of transgenic tomato plants. Nature Biotechnology 18:666-669.

Roessner-Tunali U, Hegemann B, Lytovchenko A, Carrari F, Bruedigam C, Granot D, Fernie AR (2003) Metabolic profiling of transgenic tomato plants overexpressing hexokinase reveals that the influence of hexose phosphorylation diminishes during fruit development. Plant Physiology 133:84-99.

Rose JKC, Bashir S, Giovannoni JJ, Jahn MM, Saravanan RS (2004a) Tackling the plant proteome: practical approaches, hurdles and experimental tools. Plant Journal 39:715-733.

Rose JKC, Saladie M, Catala C (2004b) The plot thickens: new perspectives of primary cell wall modification. Current Opinion in Plant Biology 7:296-301.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137-138.

Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. Journal of Molecular Biology 323:785-793.

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theoretical and Applied Genetics 111:1396-1408.

Roy JK, Bandopadhyay R, Rustgi S, Balyan HS, Gupta PK (2006) Association analysis of agronomically important traits using SSR, SAMPL and AFLP markers in bread wheat. Current Science 90:683-689.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods in molecular biology (Clifton, NJ) 132:365-386.

S

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. Trends in Plant Science 13:36-43.

Saito K, Matsuda F (2010) Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. In: Merchant S, Briggs WR, Ort D (eds) Annual Review of Plant Biology, Vol 61, vol 61. Annual Review of Plant Biology. pp 463-489. doi:10.1146/annurev.arplant.043008.092035

Saladie M, Matas AJ, Isaacson T, Jenks MA, Goodwin SM, Niklas KJ, Ren X, Labavitch JM, Shackel KA, Fernie AR, Lytovchenko A, O'Neill MA, Watkins CB, Rose JKC (2007) A reevaluation of the key factors that influence tomato fruit softening and integrity. Plant Physiology 144:1012-1028.

Saliba-Colombani V, Causse M, Gervais L, Philouze J (2000) Efficiency of AFLP, RAPD, and RFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43:29-40.

Saliba-Colombani V, Causse M, Langlois D, Philouze J, Buret M (2001) Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. Theoretical and Applied Genetics 102:259-272.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. . Biotechnology (Reading, Mass) 24:104-108.

Saravanan RS, Rose JKC (2004) A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. Proteomics 4:2522-2532.

Sato S and the tomato genome consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635-641.

Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Human Molecular Genetics 19:R227-R240.

Schaffer AA, Levin I, Oguz I, Petreikov M, Cincarevsky F, Yeselson Y, Shen S, Gilboa N, Bar M (2000) ADPglucose pyrophosphorylase activity and starch accumulation in immature tomato fruit: the effect of a *Lycopersicon hirsutum*-derived introgression encoding for the large subunit. Plant Science 152:135-144.

Schaffer AA, Petreikov M, Miron D, Fogelman M, Spiegelman M, Bnei-Moshe Z, Shen S, Granot D, Hadas R, Dai N, Levin I, Bar M, Friedman M, Pilowsky M, Gilboa N, Chen L (1999) Modification of carbohydrate content in developing tomato fruit. Hortscience 34:1024-1027.

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20:509-523.

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology 24:447-454.

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

Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT (2009) ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. Methods 48:240-248.

Segura V, Vilhjalmsson BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genet 44:825-U144.

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Nat Acad Sci USA 103:12981-12986.

Sheoran IS, Olson DJ, Ross AR, Sawhney VK (2005) Proteome analysis of embryo and endosperm from germinating tomato seeds. Proteomics 5:3752-3764.

Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. Plos One 7:e40563-e40563.

Sim S-C, Robbins MD, Chilcott C, Zhu T, Francis DM (2009) Oligonucleotide array discovery of polymorphisms in cultivated tomato (*Solanum lycopersicum L.*) reveals patterns of SNP variation associated with breeding. BMC Genomics 10.

Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004) Linkage disequilibrium mapping of a Verticillium dahliae resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theoretical and Applied Genetics 108:217-224.

Skot L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP, Thomas ID (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in Lolium perenne (L.). Genetics 177:535-547.

Slimestad R, Verheul M (2009) Review of flavonoids and other phenolics from fruits of different tomato (Lycopersicon esculentum Mill.) cultivars. Journal of the Science of Food and Agriculture 89:1255-1270.

Soller M, Brody T, Genizi A (1976) Power of experimental designs for detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theoretical and Applied Genetics 47:35-39.

Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes Solanum L. section Lycopersicon (Mill.) Wettst. Taxon 54:43-61.

Springer NM, Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell 19:2391-2402.

Steinhauser M-C, Steinhauser D, Gibon Y, Bolger M, Arrivault S, Usadel B, Zamir D, Fernie AR, Stitt M (2011) Identification of Enzyme Activity Quantitative Trait Loci in a *Solanum lycopersicum* × *Solanum pennellii* Introgression Line Population. Plant Physiology 157:998-1014.

Steinhauser MC, Steinhauser D, Koehl K, Carrari F, Gibon Y, Fernie AR, Stitt M (2010) Enzyme Activity Profiles during Fruit Development in Tomato Cultivars and Solanum pennellii. Plant Physiology 153:80-98.

Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. Nature Reviews Genetics 10:681-690.

Stevens MA (1986) Inheritance of tomato fruit quality components. Plant Breeding Reviews 4:273-311.

Stevens MA, Kader AA, Albrightholton M, Algazi M (1977) Genotypic variation for flavor and composition in fresh market tomatoes. Journal of the American Society for Horticultural Science 102:680-689.

Stevens R, Buret M, Duffe P, Garchery C, Baldet P, Rothan C, Causse M (2007) Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations. Plant Physiology 143:1943-1953.

Stevens R, Page D, Gouble B, Garchery C, Zamir D, Causse M (2008) Tomato fruit ascorbic acid content is linked with monodehydroascorbate reductase activity and tolerance to chilling stress. Plant Cell and Environment 31:1086-1096.

Stuber CW (ed) (2010) Heterosis in plant breeding, vol 12. Plant Breeding Reviews. John Wiley & Sons, Inc.,, Oxford, UK

Stylianou IM, Affourtit JP, Shockley KR, Wilpan RY, Abdi FA, Bhardwaj S, Rollins J, Churchill GA, Paigen B (2008) Applying gene expression, proteomics and single-nucleotide polymorphism analysis for complex trait gene identification. Genetics 178:1795-1805.

Sulpice R, Trenkamp S, Steinfath M, Usadel B, Gibon Y, Witucka-Wall H, Pyl E-T, Tschoep H, Steinhauser MC, Guenther M, Hoehne M, Rohwer JM, Altmann T, Fernie AR, Stitt M (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. Plant Cell 22:2872-2893.

Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW, Holland JB (2007) QTL mapping with near-isogenic lines in maize. Theoretical and Applied Genetics 114:1211-1228.

T

Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E (2002) Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. Journal of Agricultural and Food Chemistry 50:2005-2009.

Tanksley SD (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. Plant Cell 16:S181-S189.

Tanksley SD, Ganal MW, Prince JP, de-Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141-1160.

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theo App Genet 92:213-224.

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for solids content in tomato-a re-examination. Theoretical and Applied Genetics 75:811-823.

Taylor IB (1986) Biosystematics of the tomato. Tomato Crop: a Scientific Basis for Improvement Chapman and Hall, London:1-34.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286 - 289.

Thumma BR, Nolan MR, Evans R, Moran GF (2005) Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in Eucalyptus spp. Genetics 171:1257-1265.

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genet 43:159-U113.

Tieman D, Bliss P, McIntyre LM, Blandon-Ubeda A, Bies D, Odabasi AZ, Rodriguez GR, van der Knaap E, Taylor MG, Goulet C, Mageroy MH, Snyder DJ, Colquhoun T, Moskowitz H, Clark DG, Sims C, Bartoshuk L, Klee HJ (2012) The Chemical Interactions Underlying Tomato Flavor Preferences. Current Biology 22:1035-1039.

Tikunov Y, Lommen A, de Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. Plant Physiology 139:1125-1137.

Tobler AR, Short S, Andersen MR, Paner TM, Briggs JC, Lambert SM, Wu PP, Wang Y, Spoonde AY, Koehler RT, Peyret N, Chen C, Broomer AJ, Ridzon DA, Zhou H, Hoo BS, Hayashibara KC, Leong LN, Ma CN, Rosenblum BB, Day JP, Ziegle JS, De La Vega FM, Rhodes MD, Hennessy KM, Wenz HM (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. Journal of biomolecular techniques : JBT 16:398-406.

U

Usadel B, Blaesing OE, Gibon Y, Retzlaff K, Hoehne M, Guenther M, Stitt M (2008) Global transcript levels respond to small changes of the carbon status during progressive exhaustion of carbohydrates in Arabidopsis rosettes. Plant Physiology 146:1834-1861.

V

van Berloo R, Zhu A, Ursem R, Verbakel H, Gort G, van Eeuwijk FA (2008) Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theoretical and Applied Genetics 117:89-101.

van der Knaap E, Lippman ZB, Tanksley SD (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. Theoretical and Applied Genetics 104:241-247.

van der Knaap E, Sanyal A, Jackson SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168:2127-2140.

van der Knaap E, Tanksley SD (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. Theoretical and Applied Genetics 103:353-358.

van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in yellow stuffer tomato. Theoretical and Applied Genetics 107:139-147.

Van Deynze A, Stoffel K, Buell CR, Kozik A, Liu J, van der Knaap E, Francis D (2007) Diversity in conserved genes in tomato. BMC Genomics 8.

Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods 5:247-252.

Venter JC (2001) The sequence of the human genome (vol 292, pg 1304, 2001). Science 292:1838-1838.

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity 93:77-78.

Vriezen WH, Feron R, Maretto F, Keijman J, Mariani C (2008) Changes in tomato ovary transcriptome demonstrate complex hormonal regulation of fruit set. New Phytologist 177:60-76.

W

Wang H, Schauer N, Usadel B, Frasse P, Zouine M, Hernould M, Latche A, Pech JC, Fernie AR, Bouzayen M (2009) Regulatory features underlying pollination-dependent and -independent tomato fruit set revealed by transcript and primary metabolite profiling. Plant Cell 21:1428-1452.

Wang JK, Wan XY, Crossa J, Crouch J, Weng JF, Zhai HQ, Wan JM (2006) QTL mapping of grain length in rice (Oryza sativa L.) using chromosome segment substitution lines. Genet Res 88:93-104.

Weaver (2012) Promise and pitfalls of Third-Generation Sequencing. http://wwwbiotechniquescom/news/Promise-and-Pitfalls-of-Third-Generation-Sequencing/biotechniques-333200html#UFTGNa6zHtl.

Weckwerth W (2008) Integration of metabolomics and proteomics in molecular plant physiology - coping with the complexity by data-dimensionality reduction. Physiologia Plantarum 132:176-189.

Weller JI, Soller M, Brody T (1988) Linkage analysis of quantitative traits in an interspecific cross of tomato (*lycopersicon-esculentum* × *lycopersicon-pimpinellifolium*) by means of genetic-markers. Genetics 118:329-339.

Wilcox PL, Echt CE, Burdon RD (2007) Gene-assisted selection: applications of association genetics for forest tree breeding. Association mapping in plants. doi:10.1007/978-0-387-36011-9_10

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993) Genetic-analysis using random amplified polymorphic DNA markers. Method Enzymol 218:704-740.

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719-2733.

X

Xiao M, Jiang N, Schaffner E, Stockinger EJ, Knaap Evd (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science (Washington) 319:1527-1530.

Y

Yalcin B, Flint J, Mott R (2005) Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. Genetics 171:673-681.

Yan WG, Li Y, Agrama HA, Luo D, Gao F, Lu X, Ren G (2009) Association mapping of stigma and spikelet characteristics in rice (Oryza sativa L.). Mol Breed 24:277-292.

Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'Connell JR, Mangino M, Maegi R, Madden PA, Heath AC, Nyholt DR, Martin NG, Montgomery GW, Frayling TM, Hirschhorn JN, McCarthy MI, Goddard ME, Visscher PM, Consortium G (2011) Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics 19:807-812.

Yang WC, Bai XD, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis D (2004) Discovery of single nucleotide polymorphisms in Lycopersicon esculentum by computer aided analysis of expressed sequence tags. Mol Breed 14:21-34.

Yates HE, Frary A, Doganlar S, Frampton A, Eannetta NT, Uhlig J, Tanksley SD (2004) Comparative fine mapping of fruit quality QTLs on chromosome 4 introgressions derived from two wild tomato species. Euphytica 135:283-296.

Yong-Sheng L, Amit G, Gil R, Mathilde C, René D, Michel B, Joseph H, Dani Z (2003) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnology Journal 1:195-207.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539-551.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203-208.

Z

Zamir D, Tanksley SD (1988) Tomato genome is comprised largely of fast-evolving, low copy-number sequences. Molecular & General Genetics 213:254-261.

Zeng Z-B, Kao C-H, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res, Camb 74:279-289.

Zeng ZB (1994) A composite interval mapping method for locating multiple QTLs. Proceedings, 5th World Congress on Genetics Applied to Livestock Production, University of Guelph, Guelph, Ontario, Canada, 7-12 August 1994 Volume 21 Gene mapping; polymorphisms; disease genetic markers; marker assisted selection; gene expression; transgenes; non-conventional animal products; conservation genetics; conservation of domestic animal genetic resources:37-40.

Zhang J, Chiodini R, Badr A, Zhang GF (2011a) The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics 38:95-109.

Zhang J, Hao C, Ren Q, Chang X, Liu G, Jing R (2011b) Association mapping of dynamic developmental plant height in common wheat. Planta 234:891-902.

Zhang N, Brewer MT, van der Knaap E (2012) Fine mapping of *fw3.2* controlling fruit weight in tomato. Theoretical and Applied Genetics 125:273-284.

Zhang SB, Chen C, Li L, Meng L, Singh J, Jiang N, Deng XW, He ZH, Lemaux PG (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. Plant Physiology 139:1107-1124.

Zhao H, Nettleton D, Soller M, Dekkers JCM (2005) Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. Genet Res 86:77-87.

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nature Communications 2.

Annexes

This annexe presents two articles to which I contributed :

1- Ranc N, Munos S, Xu J, Le Paslier M-C, Chauveau A, Bounon R, Rolland S, Bouchet J-P, Brunel D, Causse M (2012) Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. G3 2:853-864

2- Text of an oral presentation to the International Congress of the Society for Horticulture in Europe, held in Angers, 23-27 July 2012, to be published in Acta Horticulturae.

Annexe I

Genome-Wide Association Mapping in Tomato (*Solanum lycopersicum*) Is Possible Using Genome Admixture of *Solanum lycopersicum* var. *cerasiforme*

Nicolas Ranc,*,[1] Stephane Muños,*,[2] Jiaxin Xu,*,[†] Marie-Christine Le Paslier,[‡] Aurélie Chauveau,[‡] Rémi Bounon,[‡] Sophie Rolland,* Jean-Paul Bouchet,* Dominique Brunel,[‡] and Mathilde Causse*,[3]

*INRA, Institut National de la Recherche Agronomique, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes, Avignon, 84143, France, [†]Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, People's Republic of China, and [‡]INRA, UR1279, Unité Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, Evry, 91057, France

ABSTRACT Genome-wide association mapping is an efficient way to identify quantitative trait loci controlling the variation of phenotypes, but the approach suffers severe limitations when one is studying inbred crops like cultivated tomato (*Solanum lycopersicum*). Such crops exhibit low rates of molecular polymorphism and high linkage disequilibrium, which reduces mapping resolution. The cherry type tomato (*S. lycopersicum var. cerasiforme*) genome has been described as an admixture between the cultivated tomato and its wild ancestor, *S. pimpinellifolium*. We have thus taken advantage of the properties of this admixture to improve the resolution of association mapping in tomato. As a proof of concept, we sequenced 81 DNA fragments distributed on chromosome 2 at different distances in a core collection of 90 tomato accessions, including mostly cherry type tomato accessions. The 81 Sequence Tag Sites revealed 352 SNPs and indels. Molecular diversity was greatest for *S. pimpinellifolium* accessions, intermediate for *S. l. cerasiforme* accessions, and lowest for the cultivated group. We assessed the structure of molecular polymorphism and the extent of linkage disequilibrium over genetic and physical distances. Linkage disequilibrium decreased under $r^2 = 0.3$ within 1 cM, and minimal estimated value ($r^2 = 0.13$) was reached within 20 kb over the physical regions studied. Associations between polymorphisms and fruit weight, locule number, and soluble solid content were detected. Several candidate genes and quantitative trait loci previously identified were validated and new associations detected. This study shows the advantages of using a collection of *S. l. cerasiforme* accessions to overcome the low resolution of association mapping in tomato.

KEYWORDS
tomato (*Solanum lycopersicum*)
admixture
association mapping
linkage disequilibrium

doi: 10.1534/g3.112.002667
Manuscript received March 16, 2012; accepted for publication May 24, 2012

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.112.002667/-/DC1
Sequence data from this article have been deposited to the European Nucleotide Archive under accession nos. HE805129 to HE805210.
[1]Present address: Syngenta Seeds SAS, 12, chemin de l'Hobit, B.P. 27, 31790 Saint-Sauveur, France.
[2]Present address: INRA, Laboratoire des Interactions Plantes Micro-organismes, 31326 Castanet Tolosan Cedex, France.
[3]Corresponding author: INRA UR 1052, Unité de Génétique et Amélioration des Fruits et Légumes, Domaine St Maurice BP 94, Montfavet Cedex, 84143 France. E-mail: Mathilde.Causse@avignon.inra.fr

Linkage mapping has proved its usefulness in detecting important qualitative and quantitative loci in crops (Doebley *et al.* 1997; Frary *et al.* 2000). Linkage mapping strategies are limited in detecting loci underlying quantitative traits (QTL) because, commonly, only two extreme parents are used for generating the segregating population, and only a few recombination events are studied (Flint-Garcia *et al.* 2005). Furthermore, the discovery of new genes underlying the variation of phenotypic traits is limited to those having a large effect on phenotypic variation (Buckler *et al.* 2002). Genetic resources consist of a large number of accessions with different histories, mutations, and recombination events and may represent a large reservoir of phenotypic and molecular diversity. The association mapping strategy has been proposed to identify polymorphisms involved in phenotypic variations and may prove useful in identifying interesting alleles for breeding purpose.

Recently, the value of association mapping in genetic studies has been described (Gupta *et al.* 2005; Zhu *et al.* 2008). New statistical methods have been developed to analyze structured samples (Pritchard *et al.* 2000b; Price *et al.* 2006; Yu *et al.* 2006), and these methods have been efficiently applied to plants (Thornsberry *et al.* 2001; Flint-Garcia *et al.* 2005; Zhao *et al.* 2007). One of the most important parameters in association mapping is the intensity of linkage disequilibrium (LD) over the genome. LD is defined as nonrandom association of alleles, and its intensity determines the resolution of association mapping (Rafalski 2002). When LD extends within several hundreds of base-pairs (bp), a large number of markers is necessary to cover the whole genome, and alleles at selected candidate genes should be tested for association. If it extends over greater distances, the whole genome may be scanned with a lower density of markers to identify polymorphisms associated with phenotypic variation. The extent of LD over the genome is expected to vary according to the species, genome region, and population under study (Nordborg and Tavare 2002). LD is expected to be stronger in inbred than outbred species because recombination is less effective in selfing species, where individuals are more likely to be homozygous at a given locus, than in outcrossing species (Flint-Garcia *et al.* 2003). Moreover, reduction in population size (bottleneck) increases the drift effect and, consequently, LD within and between chromosomes. Thus, inbred crops are theoretically less suitable for high-resolution association mapping because of their low level of molecular diversity and high overall genomic LD.

The cultivated tomato (*Solanum lycopersicum* var. *esculentum*, formerly *Lycopersicon esculentum*) is a diploid plant that is predominantly selfing and highly inbred. The tomato was domesticated from its wild relative, *S. pimpinellifolium*, with the first domesticated form presumably represented by *S. lycopersicum* var. *cerasiforme* (*i.e.*, the cherry tomato). The modern cultivated tomato accessions exhibit a low level of genetic diversity compared with their wild relatives as the result of several bottlenecks that occurred during domestication, migration, and selection; this low level of genetic diversity is exacerbated by the autogamous nature of this species (Yang *et al.* 2004; Van Deynze *et al.* 2007). As expected, LD extends through long genetic distances in the cultivated accessions (van Berloo *et al.* 2008). Part of the *S. lycopersicum* var. *cerasiforme* (*S. l. cerasiforme*) accessions display a genetic admixture pattern between cultivated and wild tomato accessions (Ranc *et al.* 2008). Such an admixture population could be compared with advanced intercrossed lines (*i.e.*, populations derived from two inbred strains that were randomly intercrossed for several generations). As a consequence, cherry-type tomatoes have a greater level of genetic diversity than *S. l. esculentum* and a greater phenotypic diversity than *S. pimpinellifolium*, which offers interesting properties for association mapping.

Association mapping has rarely been used to identify the molecular bases of QTL in the tomato, with the exception of analysis of two regions encompassing map-based cloned genes. Recently, association mapping was shown to be relevant in identifying quantitative trait nucleotides (QTN) responsible for locule number (LCN) differences between *S. l. cerasiforme* and *S. l. esculentum* (Muños *et al.* 2011). A sequence of 1800 bp containing the QTL *lcn2.1* was identified by map-based cloning. LD mapping detected two SNPs within this sequence that show highly significant associations with phenotypic variation. Previously, Nesbitt and Tanksley (2002) failed to find any association between fruit size and genomic sequence of the *fw2.2* region, which carries a QTL for fruit size; however, they studied only 39 cherry tomato accessions.

The objectives of the present study was to define the optimal conditions for whole-genome association in the tomato by using cherry tomato accessions and to assess the marker density needed to perform association mapping in this crop. This pilot study focused on chromosome 2 because several clusters of QTL for fruit morphology and quality traits have been mapped on this chromosome (Causse *et al.* 2002). Four genes underlying these QTL have been cloned: *fw2.2*, which is responsible for fruit weight (FW) variation (Frary *et al.* 2000); *Ovate*, which causes pear-shaped tomato fruit (Liu *et al.* 2002); *Cnr*, which causes nonripening fruit (Manning *et al.* 2006); and *lcn2.1*, responsible for LCN (Muños *et al.* 2011).

We genotyped a core collection of 90 accessions mainly composed of *S. l. cerasiforme* accessions by Sanger sequencing of DNA fragments. We sequenced 81 fragments mapped on chromosome 2 and spread over three different mapping densities: (1) a whole chromosome density (1 fragment/5 cM); (2) a fine mapping density (1 fragment/cM) and (iii) a physical mapping density (1 fragment/100 kb). For physical mapping density, we focused on regions in which QTL were previously fine mapped (Lecomte *et al.* 2004). In this study, we describe the level of molecular polymorphism detected. The extent of LD was assessed over the entire chromosome and over physical distances. Finally, association tests regarding FW, LCN, and soluble solid content (SSC) phenotypes were performed.

MATERIALS AND METHODS

Plant material

The accessions were sampled in a germplasm collection that is maintained and characterized at the Institut National de la Recherche Agronomique (INRA) in Avignon, France. These accessions are part of a core collection drawn from 380 accessions that maximizes both genetic and phenotypic diversity (Ranc *et al.* 2008). A set of 90 tomato accessions (supporting information, Table S1) was used for sequence analysis. This sample was composed of 63 cherry type tomato accessions (*i.e.*, *S. lycopersicum* var. *cerasiforme*, hereafter named *S. l. cerasiforme*), 17 large fruited accessions (*S. lycopersicum* var. *esculentum*, hereafter named *S. l. esculentum*), and 10 *S. pimpinellifolium* accessions. Accessions were derived from French researchers' prospecting, breeders' collections, the Tomato Genetics Resource Center (Davis, CA), the Centre for Genetic Resources (Wageningen, The Netherlands), the North Central Regional Plant Introduction Station (Ames, IA), and the N.I. Vavilov Research Institute of Plant Industry (St. Petersburg, Russia).

Tomato phenotyping

The accessions were grown during 2007 and 2008 summers in Avignon. Four plants per accession were bred in plastic greenhouse. Three harvests of 10 ripe fruits were done for each accession and were used as repetition in the phenotypic analysis. The 10 fruits were phenotyped for FW, LCN, and SSC. Year and accession effects were assessed by two-factor analysis of variance with [R] software (R Development Core Team 2005). Heritability estimations were calculated as following: $h_F^2 = \boldsymbol{\sigma}_g^2/(\boldsymbol{\sigma}_g^2 + \boldsymbol{\sigma}_e^2)$ with σ_g^2 and σ_e^2 the genetic and residual variance, respectively. σ_g^2 and σ_e^2were estimated by (MSc-MSe)/89 and MSe, respectively. MSc and MSe represent the mean squares of cultivar and residual effects, respectively. Because genetic effect over the two years was much significantly greater than year effect, we calculated associations by using accession adjusted means over years. FW and LCN were log transformed (Table S1). Pearson correlations were assessed among traits.

DNA fragments sequenced

The positions of the sequence tag sites (STS) along the chromosome 2 are shown in Figure 1. We used Primer3 (Rozen and Skaletsky 2000) to design pairs of primers for each STS based on sequence data of genes and markers mapped on chromosome 2 (http://solgenomics.net/).

These fragments were chosen to cover the entire chromosome with three different densities: (1) fragments every 5 cM chosen to cover the whole chromosome (2); fragments every cM chosen to cover the middle of the chromosome; and (3) fragments every 100 Kb chosen to cover five physical contigs representing candidate regions for fruit quality QTL: Contig1 (SL2.40ch02: 41129698.. 41563558), mapped around a sugar content QTL (sugs2.1); Contig2 (SL2.40ch02: 41752714.. 42140082), mapped in an LCN QTL (lcn2.1); Contig3 (SL2.40ch02: 42664935.. 43230501), mapped around a SSC QTL (ssc2.2); Contig4 (SL2.40ch02: 46744832.. 46893523), mapped in around FW QTL (fw2.2); and Contig5 (SL2.40ch02: 47342796.. 47472243), mapped around a sugar content QTL (sugs 2.2). Because of a low level of polymorphism previously described in *S. lycopersicum*, we targeted fragments for sequencing on intronic or intergenic regions. For a specific unigene, intron localization was predicted with tblastx on *Arabidopsis thaliana* genomic sequence and primers were designed on exonic sequence surrounding introns. The characteristics of the STS are presented in Table S2.

Fragment sequencing and analysis

Genomic DNA was isolated from 100 mg of frozen leaves using the DNeasy Plant Mini Kit (QIAGEN, Valencia, CA) according the manufacturer's recommendations. Amplification reactions were performed in a final volume of 5 μL in a reaction mix composed of 2.5 ng of template DNA, 0.4 pmol of each primer, 0.05 mM concentration of each deoxynucleotide, 2 mM $MgSO_4$, 1X *Taq* polymerase buffer P, and 0.03 units of Platinum *Taq* HiFi (Invitrogen, Carlsbad, CA). After 5 min of denaturation at 94°, 30 cycles were performed of 20 s at 94°, initial denaturation during 20 s at 55°, annealing during 2 min at 68°, followed by a final extension step of 5 min at 68°. Pairs of primers revealing single-band polymerase chain reaction (PCR) product were chosen for sequencing. PCR products were purified using the ExoSAP method with Exonuclease I (NEB, Beverly, MA) and Shrimp Alkaline Phosphatase (USB, Cleveland, OH). Fragments were sequenced with SP6 universal primer in an adapted 5-μL reaction volume method using BigDye terminator V3.1 and analyzed on an ABI 3730 xl sequencer (Applied Biosystems, Foster City, CA). Sequence alignment and SNP detection were performed using Genalys software available at http://software.cng.fr/ (Takahashi *et al.* 2003). Sequences of *lcn2.1*, previously obtained for this core collection (Muños *et al.* 2011), were added in this study (embl accession number JF284938 and JF284939). Genotype data are provided in Table S3.

Linkage disequilibrium

The molecular diversity was estimated by Watterson's θ. The LD parameter r^2 was estimated among loci with TASSEL (Bradbury *et al.* 2007), and the comparison-wise significance was computed by 1000 permutations. We compared different strategies for analyzing LD decay over genetic distances. We examined pairwise LD values, analyzing all polymorphisms with minor allele frequency (MAF) greater than 5% or only one polymorphism by fragment with the greatest heterozygosity index. We also compared pairwise LD decay between polymorphisms assessed in the whole population (N = 90) or only in the *cerasiforme* subset (n = 63). Pairwise r^2 were plotted according to genetic distance between two loci, and nonlinear regression fitted the decay of LD over genetic or physical distance. The decrease of LD over genetic distance was fitted by the equation: $y = a + be^{-c/x}$ using nonlinear regression, where y represents r^2 and x represents the genetic or physical distance in cM or kb (Tenesa *et al.* 2004).

Association analysis

An association study was performed with the set of 90 accessions. Several statistical models were tested: (1) the Simple general linear model (GLM); (2) the structured association model (Q model), taking into account only the structure of the collection; and (3) the mixed linear model (K+Q or MLM model), taking into account both kinship and structure, as described by Yu *et al.* (2006). The significance of associations between traits and markers was estimated with TASSEL. Population assignment of individuals was inferred by Structure 2.1 software (Pritchard *et al.* 2000a) based either on 20 simple sequence repeat (SSR) markers spread throughout the genome (Ranc *et al.* 2008) or on the genotypes of all the STS markers or just a subset of these markers. For inferring the most likely number of population, the Evanno *et al.* (2005) transformation method was used. The Ritland's matrix of relative kinship coefficients (Ritland 1996), implemented in the mixed linear model, was estimated using SPAGeDi (Hardy and Vekemans 2002) based on the set of SSR markers. According to Yu *et al.* (2006), the diagonal of the matrix was set to 2.0, and the negative values were set to 0. To deal with multiple testing, we computed adjusted *P* values using Benjamini and Hochberg (2000) procedures to control for the false discovery rate. Associations with an adjusted *P* value less than 0.005 were declared significant. For markers that were significantly associated with a trait, a general linear model with all fixed-effect terms was used to estimate R^2, the amount of phenotypic variation explained by each marker. The standardized effect of each marker was also calculated by dividing the difference of average values of the two homozygous classes by the phenotypic standard deviation for the trait (Weber *et al.* 2008). The accession used for tomato genome sequencing, Heinz 1706, was used as a reference for allele effect calculation.

RESULTS

Identification of polymorphisms on chromosome 2

Eighty-six pairs of primers, corresponding to 86 loci on chromosome 2, revealed a unique PCR band and were chosen for forward sense sequencing of the 90 tomato accessions. Five fragments were not readable because of heterozygous signals, probably due to the amplification of paralogous sequences. The 81 remaining fragments (Table S4) had an average size of 542 bp. Noncoding regions represented almost 69% (30,396 bp) of the total length sequenced (44,223 bp). Eleven fragments (13%) were monomorphic among the 90 accessions. Figure 1 shows the location and polymorphism content of the 70 polymorphic STS. A total of 300 single-nucleotide polymorphisms (SNPs) and 52 insertion-deletions (indels) were detected among 90 accessions. Only polymorphisms with MAF values greater than 5% were taken into account in the following description. Polymorphisms were analyzed according to species membership of accessions (Table 1). SNPs and indels were more frequent in noncoding regions, with an average of 8.7 polymorphisms per 1000 bp, than in the exonic parts of genes (average of 5.4 polymorphisms per 1000 bp). The molecular diversity decreased from wild to cultivated groups, whereas the number of polymorphisms dropped only for *S. l. esculentum* (Figure 2). *S. l. cerasiforme* shared polymorphisms with both cultivated and wild accessions. *S. l. cerasiforme* had only five specific polymorphisms, and 344 polymorphisms shared with one of the two other species (187 with *S. pimpinellifolium*, 11 with *S. lycopersicum*, and 146 with both species). Fifty-four percent of overall polymorphisms identified in *S. l. esculentum* corresponded to singletons within this group. Most of these polymorphisms were carried by two accessions (LA0409 and Stupicke Polni Rane).

The ratio of polymorphisms in noncoding regions to coding regions is similar in *S. pimpinellifolium* and *S. l. cerasiforme* but is strikingly higher in *S. l. esculentum* (Table 1). *S. l. esculentum* also

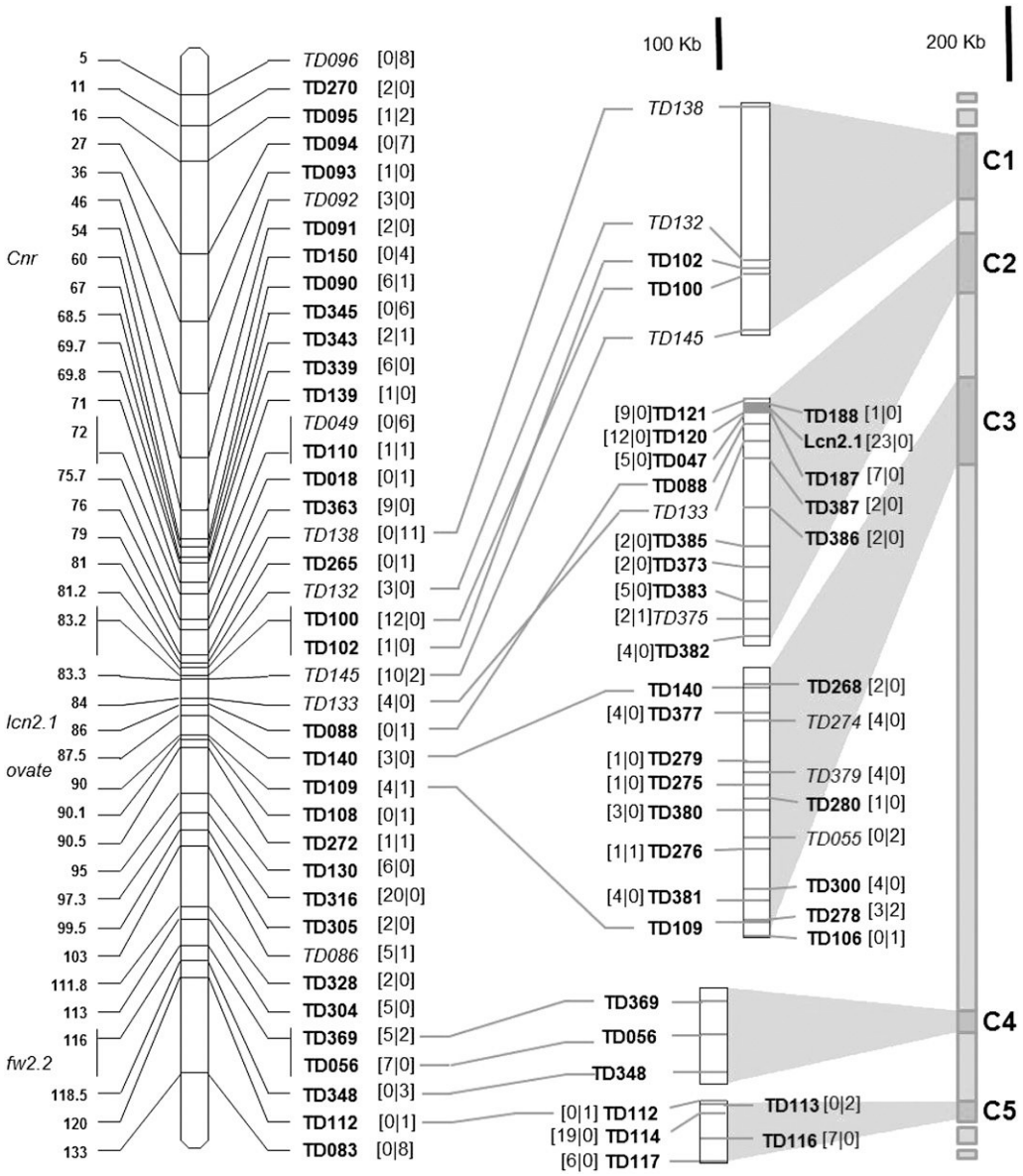


Figure 1 Genetic and physical location of the polymorphic fragments sequenced on chromosome 2. Genetic distances on the EXPEN2000 reference map are indicated on the left of the chromosome. Physical contigs are drawn on the right of the scheme. Cloned QTL are indicated on the left of the chromosome. Gray shaded area indicates homology of contigs on chromosome 2 pseudo-molecule. Numbers of polymorphisms (SNPs and indels) found in non-coding and coding regions are indicated within bracket in the first and second position, respectively. Markers in italics show high LD when compared together.

showed an excess of low frequency polymorphisms, as did *S. l. cerasiforme*, although to a lesser extent (Figure 3).

Linkage disequilibrium

We compared LD decay over genetic distances in different samples. LD decreased over shorter genetic distances when all polymorphisms per sequence were taken into account than when using a single polymorphism per fragment. Minimal difference was observed when only the *cerasiforme* subset was analyzed (Figure S1). LD was likely overestimated in the whole sample because of the genetic structure with both cultivated and wild accessions added to the *cerasiforme* subset. For further LD analysis, we focused on the 63 *cerasiforme* accessions. Based on the regression of LD over distances, LD decay reached $r^2 = 0.3$ for a genetic distance of 1 cM, and the minimal value

Table 1 Distribution and frequencies of polymorphisms (SNP and indel) across species and ratio of polymorphism in coding and noncoding region

	Number of Access.	Number of Total Polymorphic Sites	Number of Shared Polymorph.[a]			Polymorph. Frequency for 1000 bp		Noncoding/Coding Polymorphisms Ratio
			esc	*cera*	*pimpi*	coding	noncoding	
esc	17	157	0			1.66	4.27	2.57
cera	63	349	11	5		5.42	8.61	1.59
pimpi	10	336	0	187	3	5.27	8.25	1.57

All fragments (81) are taken into account.
[a] Numbers in diagonal indicate species specific polymorphisms.

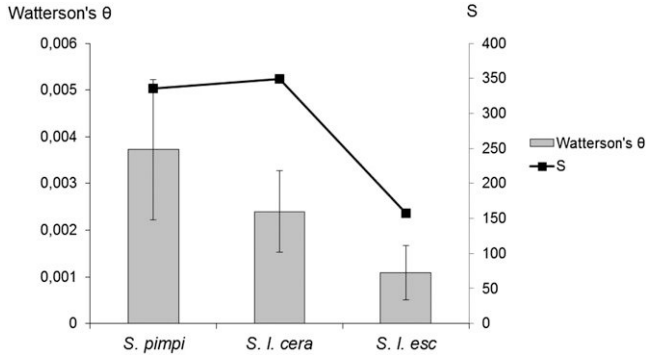


Figure 2 Molecular diversity of the three groups of tomato based on 352 polymorphisms. Molecular diversity was estimated by Watterson's θ and compared with the total number of polymorphisms (S) for S. pimpinellifolium, *S. l. cerasiforme*, and *S. l. esculentum*.

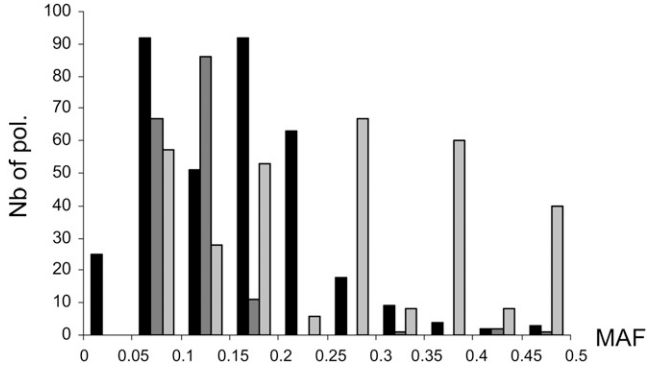


Figure 3 Distribution of polymorphism MAFs among tomato species. *S. l. cerasiforme* (n = 63) is represented in black, *S. l. esculentum* (n = 17) in dark gray, and *S. pimpinellifolium* (n = 10) in light gray. Polymorphisms with overall species MAF lower than 0.05 were previously discarded (see *Materials and Methods*).

of $r^2 = 0.09$ was obtained for distances of 13 cM (Figure 4A). Nevertheless, high r^2 (reaching the maximum value $r^2 = 1$) remained even within a distance of 60 cM, but only 28 sites of 340 (corresponding to 12 STS spread over chromosome 2) were responsible for these high pairwise LD values.

We assessed the extent of LD over physical distances within the five physical contigs covering a total of 1.86 Mb (Figure 4B). The minimal estimated r^2 fitted value of 0.13 was obtained within 20 kb, but high pairwise LD persisted within 400 kb. Figure 4C shows the matrix of LD between polymorphic sites of the physical contigs. The pattern of LD intensity over physical distances was heterogeneous. In Contig1, polymorphisms within STS formed blocks with high LD. In Contig2 and the first part of Contig3, STS did not form LD blocks. High LD between and within STS was interrupted by polymorphisms showing low LD with other polymorphisms. A striking break in the LD pattern over physical distance appeared in the middle of Contig3, where strong intrafragment blocks of LD but low LD between fragments were observed. To check whether this region corresponds to a hotspot of recombination, we used the tomato genome sequence to assess the physical positions of STS and the reference genetic map (EXPEN2000, http://solgenomics.net), and we calculated the ratio of physical to genetic distances among STS. The genetic *vs.* physical distance ratios in Contig3 were unevenly distributed with 136 kb/cM between TD140 and TD055 and 20 kb/cM between TD109 and TD106, suggesting the presence of a hotspot of recombination. The difference in LD behavior between and within contig clearly appears on graphical haplotypes (Figure S2).

Association mapping

The genetic structure of 90 tomato accessions was first estimated using 20 SSR markers spread over the genome. The most probable number of subpopulations in the sample was two (Figure S3). A subdivision in four populations was also detected, as previously shown with 318 accessions (Ranc *et al.* 2008). Twenty-six cherry tomato accessions were not clustered with high probability ($Q > 0.8$) within one structure group and were thus classified as an admixture between the two major groups (Table S1). The same trend of structure with only two populations was observed when estimating the structure with all the STS markers on chromosome 2.

FW and LCN were log-transformed to fit a normal distribution graphically, but LCN fitted a Poisson distribution. The three traits were correlated together (Figure S4). Broad-sense heritabilities were high: 0.94, 0.96, and 0.95 for SSC, FW and LCN, respectively. Genetic structure assessed by SSR markers had a significant effect on FW and SSC with R^2 values of 0.24 and 0.12, respectively, whereas population structure accounted only for 5% of the LCN variation. For association mapping, the mixed model taking into account both genetic structure assessed with all STS (Q_{STS}) and coancestry matrices (K+Q model) resulted in the best approximation of the expected cumulative distribution of *P* values, followed by the K+Q model with Q assessed with SSR markers (Q_{SSR}), then the structured association model (Q model) and the simple model (GLM; Figure 5).

We also tested alternative models to take the structure into account. Taking in the MLM model the four main coordinates of significant axes of principal components analysis provided almost similar results to the naïve model (Figure 5) as well as using k = 4 structure model (data not shown). When we used Q_{STS} in the MLM model, the probability plot was much closer to the diagonal (Figure 5), suggesting that the correction for the structure was much better, and thus we present the associations obtained with this model with corrected *P* values less than 0.05 (Table 2). With this model, we detected 14, 3, and 3 associations with FW, LCN, and SSC, respectively (Table 2). Using just a subset of 265, we found that STS avoiding the loci involved in the main regions where significant associations were detected provided the same results (data not shown).

Because the correction for structure when using Q_{SSR} was not fully satisfying with this model and many associations appeared significant, we only retained the most significant associations with adjusted *P* value lower than 0.005, although some association may still be false positive. With this model, we detected 37, 3, and 14 associations with FW, LCN, and SSC, respectively. Finally, taking into account both STS and SSR markers for the structure analysis gave the same results as the STS alone. For FW, LCN, and SSC, polymorphisms with the greatest *P* values explained a large part of the trait variation (22%, 44%, and 21%, respectively). As reference alleles are based on Heinz 1706, a genotype with large fruit and low SSC, allele effects were almost all positive for FW, whereas allele effects for SSC were all negative.

Figure 6 shows the significant associations between the polymorphisms and the traits along the chromosome with both MLM models. Most of the polymorphisms found in association with one of the traits were part of a dense chromosome region. For FW, the two strongest associations involved TD380-526 (fragment TD380 polymorphic site at position 526) on Contig3 and TD387-452 on Contig2. The r^2 value for LD estimation between these two SNPs is 0.41 in the whole accession sample (Figure S5). Because other sites revealed similar level of

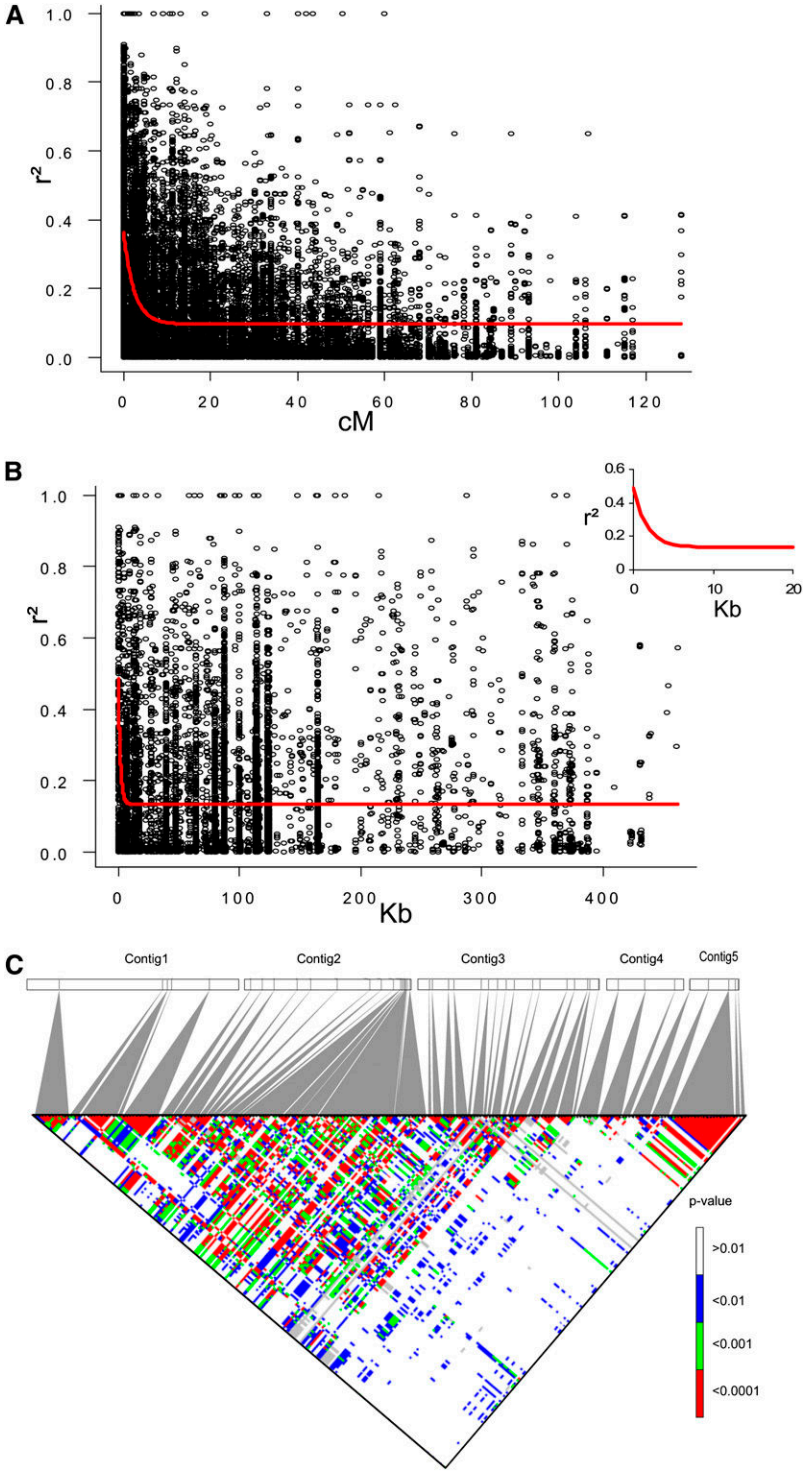


Figure 4 Estimates of LD (r^2) *vs.* genetic and physical distance on chromosome 2 for the 63 *S. l. cerasiforme* accessions. Only polymorphic sites having MAF greater than 5% are indicated (see *Materials and Methods*). (A) Decay of r^2 over genetic distance on chromosome 2. Plot of r^2 over distance was fitted by nonlinear regression (red curve). (B) Decay of r^2 over physical distance on the five major contigs. Plot of r^2 over distance is fitted by nonlinear regression (red curve). The inset shows a more detailed view of the LD decay curve for markers located less than 20 Kb apart. (C) Matrix of pairwise LD *P* value between and within physical contigs. *P* values were calculated with 1000 permutations.

LD and did not result in significant association, these two associations could correspond to two linked QTL on adjacent contigs. When only the 63 *S. l. cerasiforme* accessions were used for association analysis, TD387-452 was not associated with FW, but association with TD380-526 remained significant. A significant association for FW was detected with TD056-134, which corresponds to the 5′ region of the *fw2.2* QTL previously cloned by positional cloning (Nesbitt and Tanksley 2002). In addition, TD049-528 was associated with FW and colocated with FW2.1, a QTL for FW variation fine mapped in a biparental *S. l. esculentum* × *S. l. cerasiforme* progeny (Lecomte *et al.* 2004). Finally, we detected significant associations for FW with coding polymorphisms in the TD055 fragment corresponding to the *Ovate* gene.

For LCN, only three associations were significant (Table 2, Figure 6B). The greatest associations involved two SNPs that have been identified through map-based cloning as responsible for the LCN variation (Muños *et al.* 2011). LD between these two SNPs was extreme ($r^2 = 0.95$). The other significant association implicated TD373-391 on the same contig. TD373-391 showed the greatest r^2 with the

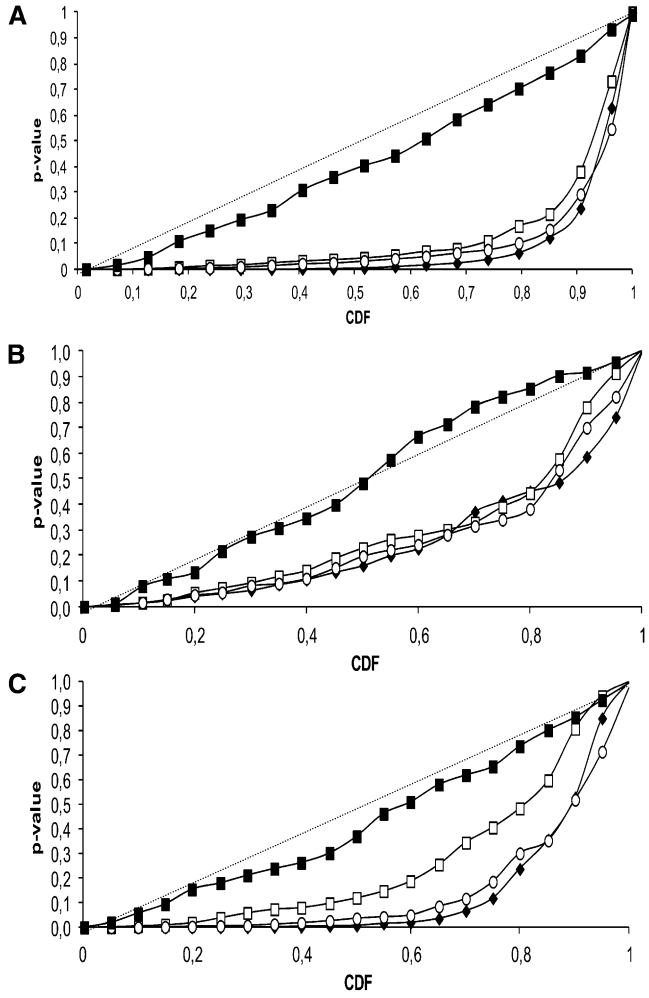


Figure 5 Cumulative density functions (CDF) using several alternative models of association. Model comparisons are performed for FW (A), LCN (B), and SSC (C). Associations are tested for all polymorphic sites with MAF >5% on 90 individuals. Naive GLM (black diamond) and K+Q models, with structure based on SSR markers (white squares), on 4 PCA axis (white circles) and on all STS markers (black squares) were tested. The diagonal indicates uniform distribution of *P* values under the expectation that random SNPs are unlinked to the polymorphisms controlling these traits (H0: no SNP effect).

lcn2.1 SNPs (r^2 = 0.47). This association may thus result from the LD with the functional *lcn*2.1 SNPs (Figure S5).

For SSC, the strongest associations were found with TD380-526 and TD387-452 loci, which were also significantly associated with FW. These results could be a consequence of the high negative correlation between FW and SSC (r = −0.66). Several of the other polymorphisms showing associations when using QSSR were in significant LD (Figure S5).

When we screened for associations using only the 63 cherry tomato accessions, a group of accession chosen to limit the population structure (Ranc *et al.* 2008), the MLM model was very close to the naïve GLM model (Figure S5). Many associations were no more significant whatever the model. For FW, significant associations were detected with loci that were also detected in the whole collection, TD380-526, TD056-134 (fw2.2), TD116-707, and TD117-219. A new association was detected with TD138-61. For LCN, the two SNPs in the *lcn2.1* locus remained significant. For SSC, the main association with TD380-526 was significant, as well as two with TD120 markers (Table S5).

DISCUSSION

To assess the genetic diversity among tomato accessions and analyze the extent of LD, we sequenced 81 DNA fragments, covering 44 kb, in 90 accessions of wild and cultivated tomatoes. We detected 352 polymorphic loci (SNP or indel). The extent of LD varied according to the regions, scales, and associations between phenotypes and polymorphisms that were successfully detected.

Power of *S. l. cerasiforme* for polymorphism discovery

The 63 *S. l. cerasiforme* accessions were previously sampled to maximize both genetic and phenotypic diversity. This sample captures 98% of SSR alleles identified in a larger sample of 144 cherry type accessions (Ranc *et al.* 2008). These accessions represent a large level of molecular variability that is almost identical to that of their wild progenitor, *S. pimpinellifolium*. In tomato, several studies were aimed at discovering SNPs and indels. Nesbitt and Tanksley (2002) searched for molecular polymorphisms in the *fw2.2* region within a collection of *S. l. esculentum* (N = 4) *and S. l. cerasiforme* (N = 39) accessions. They found only one SNP per 7 kbp within *S. l. esculentum* accessions and one SNP per 340 bp within the *S. l. cerasiforme* sample. Mining or resequencing ESTs is another strategy to discover SNPs. Using this method, Yang *et al.* (Yang *et al.* 2004) detected one SNP every 8500 bp in coding regions. Jimenez-Gomez and Maloof (2009) found more than 15,000 intraspecific polymorphisms in a set of 223,000 ESTs in *S. lycopersicum*. However, most of these polymorphisms have low allelic frequency in cultivated tomato. (Labate and Baldo 2005) reported a greater amount of polymorphic ESTs, but the studied accessions were described as highly variable compared with other *S. lycopersicum* accessions because of introgressions from wild relatives. Among the 1487 SNPs detected by Labate *et al.* (2009), only 162 were polymorphic in *S. lycopersicum* breeding germplasm, and most of them had minor allele frequency below 10%.

Van Deynze *et al.* (2007) increased the frequency of SNPs and indels compared with previous studies by focusing on gene introns. In the present study, the use of *S. l. cerasiforme* allowed us to detect 352 polymorphisms (SNP and indels) in 81 sequenced fragments. Four of the eleven monomorphic fragments (TD085, TD098, TD111, and TD384) contained only coding regions, which are less polymorphic. The difference in the polymorphism rate between species for noncoding regions may be a consequence of either (1) hitch-hiking of the region surrounding a selected polymorphism or (2) a demographic bottleneck during domestication associated with a reduction of the population effective size. *S. l. cerasiforme* suffered a decrease of its population effective size during domestication from *S. pimpinellifolium* (Bai and Lindhout 2007). The lack of diversity differences between *S. pimpinellifolium* and *S. l. cerasiforme* could be due to the greater number of accessions sequenced for the latter. The theta diversity statistic corrected for the unbalance in sample size and highlighted a higher molecular diversity in the wild sample. Molecular polymorphism is linked to the population effective size by the Watterson's estimate of the scaled mutation rate (per site) $\theta = 4Ne\mu$, where *Ne* is the population size and μ is the mutation rate. The transfer of the tomato from Mexico to Europe during the 16th century greatly reduced the effective population size of the tomato and subsequently decreased the amount of molecular diversity in *S. lycopersicum*. A selection pressure that targeted coding regions could explain the higher ratio between noncoding and coding polymorphisms for *S. lycopersicum*. The reduction of diversity could arise on the fragment targeted by selection but also on the region suffering genetic hitchhiking or background selection (Innan and Stephan 2003). Thus, a less

Table 2 Significant associations for fruit weight (FW), locule number (LCN), and soluble solid content (SSC) estimated with K+Q models on 90 accessions

Trait	Locus	Location[a]	Model A				MAF[e]	Model B
			P Value	Corrected *P* Value[b]	R^{2c}	a[d]		Corrected *P* Value[b]
log(FW)	TD091-415	54cM	0.0012	0.004	0.10	10.0	0.18	ns
log(FW)	TD091-607	54cM	8.12×10^{-04}	0.003	0.10	9.2	0.24	ns
log(FW)	TD049-528	72cM	6.04×10^{-04}	0.002	0.11	9.5	0.48	ns
log(FW)	TD363-213	76cM	0.0019	0.005	0.07	9.6	0.39	ns
log(FW)	TD383-419	84cM-c2.13	7.56×10^{-04}	0.003	0.12	12.1	0.11	ns
log(FW)	TD383-558	84cM-c2.13	6.36×10^{-04}	0.002	0.13	11.3	0.13	ns
log(FW)	TD383-60	84cM-c2.13	6.36×10^{-04}	0.002	0.13	11.3	0.13	ns
log(FW)	TD375-573	84cM-c2.14	0.0011	0.003	0.10	9.0	0.25	ns
log(FW)	TD133-115	84cM-c2.8	3.34×10^{-04}	0.002	0.09	7.2	0.33	ns
log(FW)	TD133-395	84cM-c2.8	5.57×10^{-04}	0.002	0.09	7.3	0.33	ns
log(FW)	TD387-452	84cM-c2.9	9.40×10^{-07}	4.14×10^{-05}	0.19	11.6	0.27	0.025
log(FW)	lcn2.1-686	86cM-c2.3	2.86×10^{-05}	0.001	0.12	−11.7	0.38	ns
log(FW)	lcn2.1-692	86cM-c2.3	8.95×10^{-06}	2.63×10^{-04}	0.15	−12.7	0.37	ns
log(FW)	TD274-17	87.5cM-c3.13	9.32×10^{-04}	0.003	0.08	8.9	0.26	ns
log(FW)	TD274-325	87.5cM-c3.13	4.76×10^{-04}	0.002	0.10	9.8	0.23	ns
log(FW)	TD377-96	87.5cM-c3.14	0.0014	0.004	0.09	8.3	0.17	ns
log(FW)	TD377-97	87.5cM-c3.14	0.0023	0.005	0.08	8.5	0.16	ns
log(FW)	TD377-98	87.5cM-c3.14	0.0014	0.004	0.09	8.3	0.17	ns
log(FW)	TD377-91	87.5cM-c3.14	0.0013	0.004	0.09	8.2	0.17	ns
log(FW)	TD379-326	88cM-c3.11	4.42×10^{-04}	0.002	0.12	14.4	0.15	0.001
log(FW)	TD380-256	89cM-c3.8	3.04×10^{-04}	0.002	0.11	9.5	0.21	ns
log(FW)	TD380-526	89cM-c3.8	6.13×10^{-08}	5.39×10^{-06}	0.22	13.2	0.36	0.002
log(FW)	TD280-328	89cM-c3.9	4.54×10^{-04}	0.002	0.10	10.5	0.48	ns
log(FW)	TD055-469	89.5cM-c3.7	9.46×10^{-05}	0.001	0.13	8.3	0.26	ns
log(FW)	TD278-267	90cM-c3.3	1.73×10^{-04}	0.002	0.11	12.0	0.21	0.023
log(FW)	TD278-21	90cM-c3.3	0.003 ns	0.02 ns	—	—	—	0.048
log(FW)	TD278-39	90cM-c3.3	5.23×10^{-04}	0.002	0.10	15.0	0.15	0.030
log(FW)	TD278-444	90cM-c3.3	2.30×10^{-04}	0.002	0.12	12.4	0.22	0.025
log(FW)	TD278-524	90cM-c3.3	3.81×10^{-04}	0.002	0.12	11.9	0.20	0.030
log(FW)	TD300-257	90cM-c3.5	1.95×10^{-04}	0.002	0.12	11.6	0.20	ns
log(FW)	TD300-41	90cM-c3.5	0.0011	0.003	0.11	9.2	0.33	ns
log(FW)	TD108-347	90.1cM	8.29×10^{-04}	0.003	0.10	7.4	0.27	ns
log(FW)	TD056-134	116cM-c4.7	3.49×10^{-04}	0.002	0.12	10.8	0.35	ns
log(FW)	TD369-493	116cM-c4.8	0.0025	0.005	0.09	11.1	0.26	ns
log(FW)	TD116-707	120cM-c4.3	4.90×10^{-05}	0.001	0.16	8.1	0.45	0.023
log(FW)	TD117-164	120cM-c4.4	1.16×10^{-04}	0.001	0.15	10.1	0.33	ns
log(FW)	TD117-176	120cM-c4.4	1.16×10^{-04}	0.001	0.15	10.1	0.33	0.029
log(FW)	TD083-246	133cM	0.0013	0.004	0.09	10.3	0.48	0.033
log(LCN)	TD373-391	86cM-c2.12	2.14×10^{-05}	0.002	0.21	−0.68	0.49	0.037
log(LCN)	lcn2.1-692	86cM-c2.3	5.93×10^{-13}	1.85×10^{-10}	0.44	−1.16	0.37	4.57×10^{-09}
log(LCN)	lcn2.1-686	86cM-c2.3	5.32×10^{-12}	8.30×10^{-10}	0.44	−1.21	0.38	1.34×10^{-08}
SSC	TD133-115	84cM-c2.8	1.87×10^{-05}	7.12×10^{-04}	0.16	−0.63	0.33	ns
SSC	TD133-395	84cM-c2.8	4.90×10^{-05}	0.002	0.15	−0.58	0.33	ns
SSC	TD387-452	84cM-c2.9	3.88×10^{-07}	5.89×10^{-05}	0.24	−0.86	0.27	0.018
SSC	TD047-274	86cM-c2.5	3.96×10^{-06}	2.01×10^{-04}	0.19	−1.00	0.12	ns
SSC	TD120-212	86cM-c2.6	3.10×10^{-04}	0.004	0.13	−0.58	0.33	ns
SSC	TD120-88	86cM-c2.6	2.22×10^{-04}	0.003	0.13	−0.59	0.32	ns
SSC	TD140-180	87.5cM-c3.15	1.90×10^{-04}	0.003	0.14	−0.73	0.21	ns
SSC	TD379-326	88cM-c3.11	0.008 ns	0.04 ns	—	—	—	0.045
SSC	TD380-256	89cM-c3.8	2.57×10^{-04}	0.003	0.13	−0.65	0.21	ns
SSC	TD380-526	89cM-c3.8	1.27×10^{-06}	9.68×10^{-05}	0.21	−0.70	0.36	0.022
SSC	TD280-328	89cM-c3.9	1.64×10^{-04}	0.003	0.14	−0.55	0.48	ns
SSC	TD055-469	89.5cM-c3.7	8.93×10^{-05}	0.002	0.15	−0.67	0.26	ns
SSC	TD117-164	120cM-c4.4	1.52×10^{-04}	0.003	0.14	−0.70	0.33	ns
SSC	TD117-176	120cM-c4.4	1.52×10^{-04}	0.003	0.14	−0.70	0.33	ns

Model A: MLM model, with structure based on 20 SSR (only *P* values less than 0.005 are shown with indication on allele effect); model B: MLM model with structure based on all STS loci on chromosome 2 (*P* values less than 0.05 are shown). MAF, minimal allele frequencies; ns, nonsignificant.

[a] Nomenclature for the location is as follows: "genetic distance on expen2000 reference map"-"the number of contig"."the fragment number on this contig".

[b] *P* values are corrected following the Benjamini & Hochberg (2000) procedure (see *Materials and Methods*).

[c] R^2 were calculated using Q model.

[d] Allele effects are indicated in grams for FW, mean number of locule for LCN, and °brix for SSC.

[e] MAFs are shown for each polymorphism.

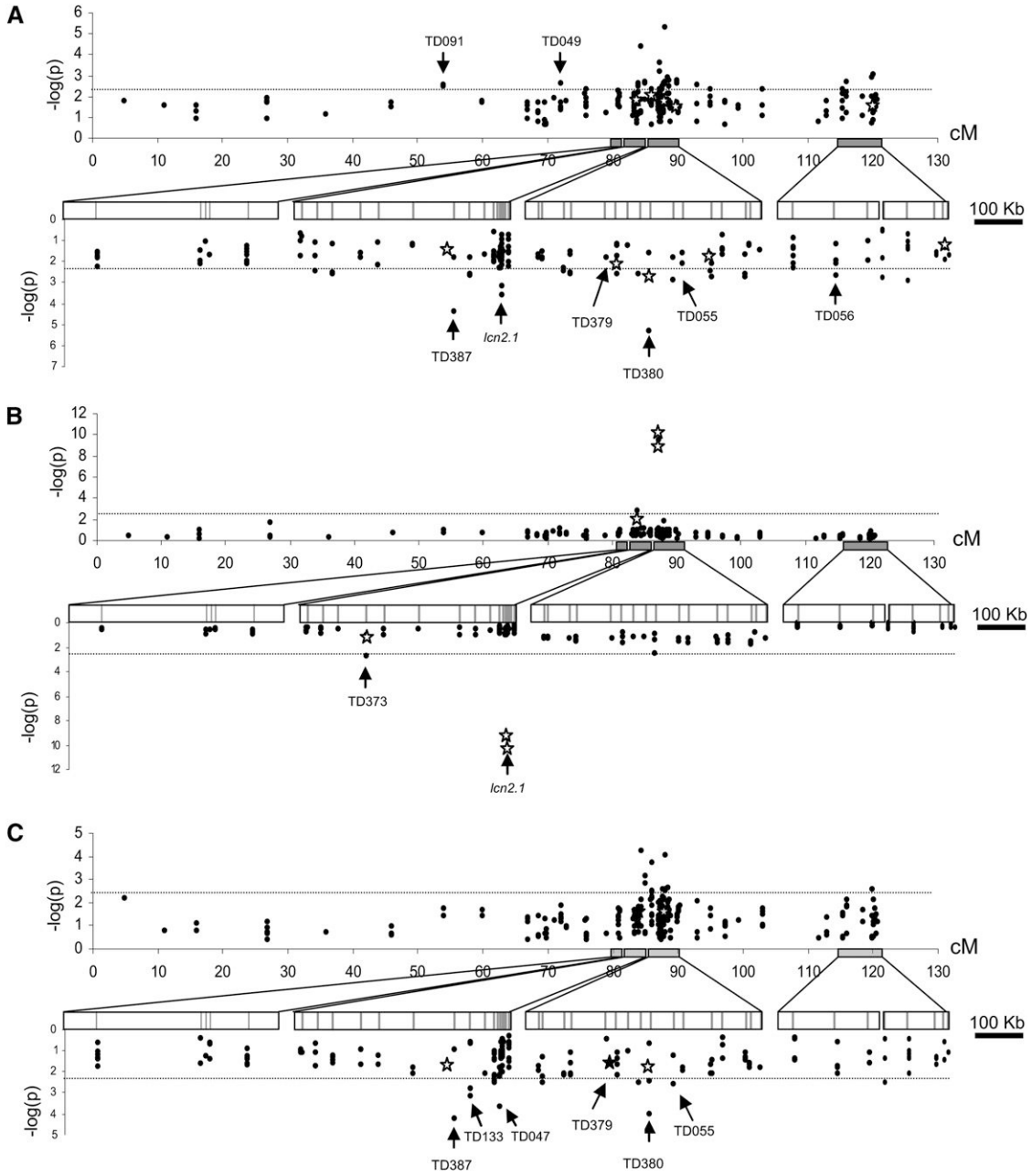


Figure 6 Plot of association *P* values over the chromosome 2. Associations are estimated for 90 accessions. K+Q model was used to screen for association between polymorphisms and (A) FW, (B) LCN, and (C) SSC. Stars indicate the associations detected with the structure assessed with all STS, and black dots the associations detected with 20 SSR markers. The upper part of each graph represents associations along genetic distance over the entire chromosome 2. The lower part shows associations for each physical contig. Arrows indicate the marker name of the most significant associations. Adjusted *P* values for multiple testing (see *Materials and Methods*) are shown.

drastic reduction in population size and continuous inter-mating with *S. pimpinellifolium* shaped a higher level of molecular diversity for *S. l. cerasiforme*.

LD decay over genetic and physical distances

An ancient admixture increased the polymorphism level of cherry tomatoes and limited their overall LD. We reached minimal LD values ($r^2 < 0.09$) with distances greater than 13 cM, but extreme LD values were still found over 60 cM for a few marker pairs. Our results support those from van Berloo *et al.* (2008), who described an LD extent ranging from 15 to 20 cM using AFLP markers in a cherry tomato sample (N = 18). Nesbitt and Tanksley (2002) showed that LD in *S. l. cerasiforme* could be broken within 150 kb around *fw2.2*. With an average ratio of 750 kb/cM on the whole tomato genome (Tanksley *et al.* 1992), the results of LD decay over physical and genetic distances are not consistent. In our *S. l. cerasiforme* sample, some r^2 values were still extreme over hundreds of kb, but the drop estimated by nonlinear regression indicated that minimal LD is reached over 20 kb. *Arabidopsis thaliana* also showed a large extent of LD over the *FRI* locus. LD extends to 200 kb, corresponding to one cM in this species (Nordborg *et al.* 2002). This estimate is locus-specific and when studies are performed on the whole genome, LD decays within 10 kb on average

(Kim *et al.* 2007). Nordborg (2000) estimated from simulations that LD should vanish over a scale of 10 kb for inbred species. Our results in the cherry tomato support these simulations. The results of LD decay over genetic distances in the tomato are similar to the LD pattern assessed in barley, another highly inbred crop (Zhang *et al.* 2009). In barley, large differences are observed in the LD decay pattern among cultivated accessions, landraces, and wild accessions (Caldwell *et al.* 2006). The greater LD extent for the crop compared with the wild ancestor or to *A. thaliana* could be due to a major bottleneck that fixed large haplotypes during domestication.

The LD pattern observed in the physical contigs is similar to haplotype blocks described in soybean landraces, which is also an inbred crop (Hyten *et al.* 2007). It is also similar to haplotype blocks in *A. thaliana* (Kim *et al.* 2007) and in humans (Daly *et al.* 2001). For the tomato, this LD pattern could be due to recent mutations with low frequencies (more than 50% of polymorphisms had MAF < 0.2). These polymorphism patterns may have evolved by lineage effects rather than by recombination and thus may decay in LD in a small region that is not correlated to distance. The high LD pattern described in the first part of Contig3 and Contig2 could have been shaped by selection. Clusters of QTL have been mapped in this region regarding LCN, fruit shape, FW, soluble solids, and sugar content (Lecomte *et al.* 2004). The selection of new advantageous mutations during domestication should have increased LD in domesticated accessions (Nordborg and Tavare 2002). In *A. thaliana*, LD blocks surrounding selected polymorphisms are significantly longer than blocks surrounding nonselected alleles (Kim *et al.* 2007). Finally, a recombination hotspot is likely responsible for the break in the LD pattern observed in Contig3. Mapping data offered direct confirmation of uneven distribution of recombination over Contig3, but the high density of polymorphisms detected in this study should be mapped on a large F2 population to confirm the presence of such a recombination hotspot (Drouaud 2006).

Candidate genes are validated by association mapping

Our approach using a core collection was efficient in detecting association in several candidate gene regions. Recently, *lcn2.1* was identified by the map-based cloning approach as a QTN responsible for variation in the tomato LCN (Muños *et al.* 2011). We used information on *lcn2.1* to highlight any possible effect of these two SNPs on FW and SSC. A significant association was found between these two SNPs and FW. Muños *et al.* (2011) highlighted the role of this locus in tomato domestication and further FW increase. This association was the only one with negative allelic effect. The reference genotype, Heinz 1706, has large fruits with only two locules, whereas almost all other two-locule genotypes carry small fruit. The large number of these small-fruit accessions in the reference group induced a negative effect for FW. Nesbitt and Tanksley (2002) could not detect any association in a *S. l. cerasiforme* sample between FW and polymorphisms in the *fw2.2* region cloned previously. These authors concluded that genes other than *fw2.2* are responsible for the variation of FW in cherry tomatoes. The number of accessions (39 *S. l. cerasiforme*, 4 *S. l. esculentum* and 3 *S. pimpinellifolium*) was the principal limitation of the study. Using 90 accessions selected to represent the diversity of a larger collection, we found a significant association with a polymorphic site located in the promoter of the gene. This polymorphism could be responsible for the phenotype variation or could be in LD with the responsible one. The entire cloned region should be sequenced and tested for association before concluding.

Association mapping for the discovery of new QTL and candidate genes

Many QTL related to fruit traits map to chromosome 2 (Causse *et al.* 2002, Labate *et al.*, 2007). These QTL and the QTL that were fine-mapped in the mapping population *S. l. cerasiforme* × S. *l. esculentum* for FW, LCN and SSC (Lecomte *et al.* 2004) were also identified by association mapping. The screening of polymorphisms on chromosome 2 with high-density markers allowed the detection of many new associations and identification of several putative new candidate genes. The number of significant associations found with FW can result from LD caused by strong selection on this phenotype (Bai and Lindhout 2007). TD380-526 showed the most significant association with FW. This fragment STS matched a predicted gene, Solyc02g085390.1.1, which is homologous to *A. thaliana*'s SNF2-like protein (AT5G66750). This gene has been characterized as an ATP-dependent helicase with chromatin remodeling activity. Chromatin remodeling proteins reconfigure protein–DNA interactions that accompany or induce changes in genome activity, such as gene expression (Kaya *et al.* 2001; Verbsky and Richards 2001). The other highly significantly associated fragment, TD387, has homology with a *S. lycopersicum* unigene (SGN-U596069) and matches the *S. lycopersicum* annotation Solyc02g084070.1.1. This gene has no homology with any gene of known function.

Another association was detected for FW with TD049-528. TD049 was tagged in the 3′ region of a gene coding for glyoxalase I (Solyc02g080630.1.1). This gene colocalizes with a QTL for FW variation in a mapping population derived from a *S. l. esculentum* x *S. l. cerasiforme* cross (Saliba-Colombani *et al.* 2001). Because of the putative impact of glyoxalase I protein on plant cell proliferation (Paulus *et al.* 1993), this gene represents a good candidate gene for FW variation. The two polymorphisms most significantly associated with FW were also associated with SSC. This could be due to the dilution effects of soluble sugars and acids according to fruit size (Prudent *et al.*, 2010). The two polymorphisms were no longer statistically associated with SSC when we added the FW effect as a covariate in the K+Q-model. We observed the same result for TD117 (Solyc02g091640.1.1, which codes for an Endoribonuclease E-like protein), which is genetically close to the *fw2.2* gene. The two other strongest associations, TD047 (promoter of Solyc02g083950.1.1, which codes for the WUSCHEL transcription factor) and TD133 (Solyc02g084030, which codes for a methionine sulfoxide reductase), are both located in the same region as TD120. Because TD047 and TD133 are separated by a distance of 2 cM, this region must be enriched in SNPs to locate precisely one or more responsible polymorphisms. TD055 mapped in a SSC QTL (brix2.2) described in the mapping population involving cherry tomato (Saliba-Colombani *et al.* 2001; Lecomte *et al.* 2004). TD055 was designed in the *Ovate* gene and showed association with SSC. *Ovate* is implicated in the modification of fruit shape, but no effect on SSC has yet been reported. This polymorphism could thus be in LD with the responsible polymorphism. SSC also showed significant association with TD140 (Solyc02g085100 0.1.1), which was identified as an aldose-1-epimerase. This enzyme catalyzes the transformation of alpha-D-glucose into beta-D-glucose and participates in glycolysis and gluconeogenesis. The aldose-1-epimerase thus represents a new candidate for SSC variation.

Optimal conditions for genome-wide association studies in tomato

We highlighted the greater efficiency of the K+Q-model in dealing with type I error rates for association mapping in the tomato. Information on the estimated familial relatedness in our sample did not influence

the results for association with FW because most of the false positives are also corrected with genetic structure information. The K+Q-model may prove its power in a sample of increased size as well as broader allelic diversity (Yu *et al.* 2006). A greater number of markers to detect structure may also reveal a more subtle structure. Taking in the MLM model the structure in 4 subgroups did not change the associations, neither using the coordinates of the first four axes of principal components analysis. The departure from the distribution of *P* values under the expectation that random SNPs are not linked to the polymorphisms controlling FW, SSC, and LCN indicates that our analysis did not succeed in correcting for the whole genetic structure. However, the number of polymorphisms tested was too small and nonrandomly spread over the genome. We then decided to focus only on highly significant associations to reduce the acceptance of false-positive associations. When we used in the MLM model the structure based on the STS detected on chromosome 2 (all or a subset excluding the positions with the main effects), the correction was much better (Figure 5), and many associations were no more significant (Table 2), confirming that many associations could be due to the structure. Nevertheless structure based on STS on chromosome 2 may capture a large part of the LD on that chromosome, and thus exclude interesting associations. Furthermore, the traits studied here have strongly evolved from wild to domesticated forms (as shown by the large part of variation explained by the structure). Correcting for the structure may thus hamper the discovery of relevant loci involved in domestication. For these reasons, we presented results of both models.

The core collection may be efficient in detecting polymorphisms with large effects on trait variation, but it will suffer a decrease of statistical power when dealing with low effect variants. A larger collection is necessary to map such genes with low effect. A higher power may be achieved by increasing the sample size rather than by increasing the number of polymorphisms (Long and Langley 1999). The density of markers needed for association analysis is estimated by LD decay over genetic or physical distance (Rafalski 2002). An r^2 value of 0.3 indicates a sufficiently strong LD to be useful for association mapping in human studies (Ardlie *et al.* 2002). In *S. l. cerasiforme* accessions, LD estimated values decayed below this value within 1 cM. One SNP per cM could thus be valuable for medium resolution genome-wide association. Nevertheless, many associations may not be detected with such low number of markers as for physical distances, even if extreme LD is still found over hundreds of kb, an estimate of LD decay indicates that LD is minimal after 20 kb. With a genome size of 950 Mb, a minimum set of 48,000 markers would thus be necessary to have a physical resolution for genome-wide association in tomato. This is a minimal number, which should be probably doubled to tag common polymorphisms in all regions. To validate these estimations based on LD, we looked at the number of significant associations for different marker densities. As expected, the number of SNPs in association with traits increased with densification of polymorphisms. Significant associations ($p < 0.005$) were found using a large mapping strategy (1 marker per 5 cM) for FW, but no association was found for LCN or SSC. The density of markers necessary for analysis will thus depend on the trait, the locus targeted, and the population studied. For example, it would not have been possible to physically map the *lcn*2.1 QTN using only LD because these two SNPs are in complete equilibrium with surrounding polymorphisms, except with TD373, which is located on the physical region of *lcn*2.1 (Muños *et al.* 2011).

Our results suggest that genome admixture of *S. l. cerasiforme* provides an interesting source of molecular diversity for the domesticated tomato. The design of our core collection was efficient enough to detect associations in all the candidate regions where QTL have been previously mapped. We highlighted the greater efficiency of the K+Q model in dealing with type I error rate even in a relatively small sample. Association mapping validated the polymorphisms discovered by positional cloning (*lcn2.1* and *fw2.2*) or fine mapping (*fw2.1*). The screening of polymorphisms along chromosome 2 with a high marker density allowed the detection of many new associations that were confirmed in a larger sample. We identified several putative new candidate genes. If we extrapolate our results to the whole genome, at least 50,000 SNPs will be necessary for high-resolution mapping in such a collection and the double would be more realistic to avoid SNP with low MAF. Due to the recent advances in next-generation sequencing technologies, the development of genomic tools (*i.e.* SNP-chip) of high to very high density will allow screening of the whole tomato genome for association with traits of interest.

ACKNOWLEDGMENTS

We thank Hélène Burck for characterizing and maintaining the INRA tomato Genetic Resources collection. We are grateful to Yolande Carretero, Esther Pelpoir, and Laure David for their help with growing and phenotyping cherry tomato accessions. We thank Aurélie Chauveau and Rémi Bounon for plant DNA sequencing. French INRA AIP BioRessources supported this work. The thesis of Nicolas Ranc, Ph.D., was financially supported by the EUSOL European project PL016214-2. Jiaxin Xu, Ph.D., was financially supported by the Chinese Scientific Council.

LITERATURE CITED

Ardlie, K. G., L. Kruglyak, and M. Seielstad, 2002 Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. 3: 299–309.

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

Benjamini, Y., and Y. Hochberg, 2000 On the adaptive control of the false discovery rate in multiple testing with independent statistics. J. Educ. Behav. Stat. 25: 60–83.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss *et al.*, 2007 TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

Buckler, I., S. Edward, and J. M. Thornsberry, 2002 Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5: 107–111.

Caldwell, K. S., J. Russell, P. Langridge, and W. Powell, 2006 Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172: 557–567.

Causse, M., V. Saliba-Colombani, L. Lecomte, P. Duffe, P. Rousselle *et al.*, 2002 QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. J. Exp. Bot. 53: 2089–2098.

Daly, M. J., J. D. Rioux, S. F. Schaffner, T. J. Hudson, and E. S. Lander, 2001 High-resolution haplotype structure in the human genome. Nat. Genet. 29: 229–232.

Doebley, J., A. Stec, and L. Hubbard, 1997 The evolution of apical dominance in maize. Nature 386: 485–488.

Drouaud, J., 2006 Variation in crossing-over rates across chromosome4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination hot spots. Genome Res. 16: 106–114.

Evanno, G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14: 2611–2620.

Flint-Garcia, S. A., J. M. Thornsberry, S. Edward, and I. V. Buckler, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54: 357–374.

Flint-Garcia, S. A., A.-C. Thuillet, J. Yu, G. Pressoir, S. M. Romero *et al.*, 2005 Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44: 1054–1064.

Frary, A., T. C. Nesbitt, A. Frary, S. Grandillo, E. D. Knaap *et al.*, 2000 *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85–88.

Gupta, P. K., S. Rustgi, and P. L. Kulwal, 2005 Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 57: 461–485.

Hardy, O. J., and X. Vekemans, 2002 spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2: 618–620.

Hyten, D. L., I.-Y. Choi, Q. Song, R. C. Shoemaker, R. L. Nelson *et al.*, 2007 Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175: 1937–1944.

Innan, H., and W. Stephan, 2003 Distinguishing the hitchhiking and background selection models. Genetics 165: 2307–2312.

Jimenez-Gomez, J., and J. Maloof, 2009 Sequence diversity in three tomato species: SNPs, markers, and molecular evolution. BMC Plant Biol. 9: 85.

Kaya, H., K.-i. Shibahara, K.-i. Taoka, M. Iwabuchi, B. Stillman *et al.*, 2001 *FASCIATA* genes for chromatin assembly factor-1 in *Arabidopsis* maintain the cellular organization of apical meristems. Cell 104: 131–142.

Kim, S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark *et al.*, 2007 Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 39: 1151–1155.

Labate, J. A., and A. Baldo, 2005 Tomato SNP discovery by EST mining and resequencing. Mol. Breed. 16: 343–349.

Labate, J. A., S. Grandillo, T. Fulton, S. Muños, A. L. Caicedo, *et al.*, 2007 Tomato, pp. 11–135 in *Genome Mapping and Molecular Breeding in Plants*, Vol. 5, Vegetables, edited by C. Kole. Springer-Verlag, Berlin/Heidelberg.

Labate, J., L. Robertson, F. Wu, S. Tanksley, and A. Baldo, 2009 EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (*Solanum lycopersicum* L.). Theor. Appl. Genet. 118: 1005–1014.

Lecomte, L., V. Saliba-Colombani, A. Gautier, M. C. Gomez-Jimenez, P. Duffé *et al.*, 2004 Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. Mol. Breed. V13: 1–14.

Liu, J., J. Van Eck, B. Cong, and S. D. Tanksley, 2002 A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc. Natl. Acad. Sci. USA 99: 13302–13306.

Long, A. D., and C. H. Langley, 1999 The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. 9: 720–731.

Manning, K., M. Tor, M. Poole, Y. Hong, A. J. Thompson *et al.*, 2006 A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat. Genet. 38: 948–952.

Muños, S., N. Ranc, E. Botton, A. Bérard, S. Rolland *et al.*, 2011 Increase in tomato locule number is controlled by two key SNP located near Wuschel. Plant Physiol. 156: 2244–2254.

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

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 154: 923–929.

Nordborg, M., and S. Tavare, 2002 Linkage disequilibrium: what history has to tell us. Trends Genet. 18: 83–90.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 30: 190–193.

Paulus, C., B. Köllner, and H.-J. Jacobsen, 1993 Physiological and biochemical characterization of glyoxalase I, a general marker for cell proliferation, from a soybean cell suspension. Planta 189: 561–566.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick *et al.*, 2006 Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000a Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly, 2000b Association mapping in structured populations. Am. J. Hum. Genet. 67: 170–181.

Prudent, M., N. Bertin, M. Génard, S. Muños, S. Rolland, *et al.*, 2010 Genotype-dependent response to carbon availability in growing tomato fruit. Plant Cell Env. 33: 1186–1204.

R Development Core Team, 2005 *R: A Language and Environment for Statistical Computing*, reference index version 2.2.1., edited by F. F. S. Computing, Vienna, Austria.

Rafalski, A., 2002 Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. 5: 94–100.

Ranc, N., S. Muños, S. Santoni, and M. Causse, 2008 A clarified position for *solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). BMC Plant Biol. 8: 130.

Ritland, K., 1996 Estimators for pairwise relatedness and individual inbreeding coefficients. Genet. Res. 67: 175–185.

Rozen, S., and H. Skaletsky, 2000 Primer 3 on the WWW for general users and for biologist programmers, pp. 365–386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by S. Krawetz and S. Misener. Humana Press, Totowa, NJ.

Saliba-Colombani, V., M. Causse, D. Langlois, J. Philouze, and M. Buret, 2001 Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. Theor. Appl. Genet. V102: 259–272.

Takahashi, M., F. Matsuda, N. Margetic, and M. Lathrop, 2003 Automated identification of single nucleotide polymorphisms from sequencing data. J. Bioinform. Comput. Biol. 1: 253–265.

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de-Vicente, M. W. Bonierbale *et al.*, 1992 High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160.

Tenesa, A., A. F. Wright, S. A. Knott, A. D. Carothers, C. Hayward *et al.*, 2004 Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum. Mol. Genet. 13: 25–33.

Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen *et al.*, 2001 Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28: 286–289.

van Berloo, R., A. Zhu, R. Ursem, H. Verbakel, G. Gort *et al.*, 2008 Diversity and linkage disequilibrium analysis within a selected set of cultivated tomatoes. Theor. Appl. Genet. 117: 89–101.

Van Deynze, A., K. Stoffel, C. R. Buell, A. Kozik, J. Liu *et al.*, 2007 Diversity in conserved genes in tomato. BMC Genomics 8: 465.

Verbsky, M. L., and E. J. Richards, 2001 Chromatin remodeling in plants. Curr. Opin. Plant Biol. 4: 494–500.

Weber, A. L., W. H. Briggs, J. Rucker, B. M. Baltazar, J. de Jesus Sanchez-Gonzalez *et al.*, 2008 The genetic architecture of complex traits in teosinte (*Zea mays ssp. parviglumis*): new evidence from association mapping. Genetics 180: 1221–1232.

Yang, W., X. Bai, E. Kabelka, C. Eaton, S. Kamoun *et al.*, 2004 Discovery of single nucleotide polymorphisms in Lycopersicon esculentum by computer aided analysis of expressed sequence tags. Mol. Breed. V14: 21–34.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

Zhang, L., S. Marchand, N. Tinker, and F. Belzile, 2009 Population structure and linkage disequilibrium in barley assessed by DArT markers. Theor. Appl. Genet. 119: 43–52.

Zhao, K., M. J. Aranzana, S. Kim, C. Lister, C. Shindo *et al.*, 2007 An *Arabidopsis* example of association mapping in structured samples. PLoS Genet. 3: e4.

Zhu, C., M. Gore, E. S. Buckler, and J. Yu, 2008 Status and prospects of association mapping in plants. The Plant Genome 1: 5–20.

Communicating editor: D. Zamir

Annexe II

A multi-level Omic approach of tomato fruit quality

J. Xu (1,2), L. Pascual (2), N. Desplat (2), M. Faurobert (1), Y. Gibon (3), A. Moing (3), M. Maucourt (4), P. Ballias (3), C. Deborde (3), Y. Liang (2), J.P. Bouchet (2), D. Brunel (5), M.C. Lepaslier (5), M. Causse (2)

(1) Northwest A&F University, College of Horticulture, Yang Ling, Shaanxin, 712100, P. R. China

(2) INRA, UR1052, GAFL, BP94, F-84143 Montfavet, France (mathilde.causse@avignon.inra.fr)
(3) INRA, UMR1332 Biologie du Fruit et Pathologie, BP 81, F-33140 Villenave d'Ornon, France

(4) Université de Bordeaux, UMR1332 Biologie du Fruit et Pathologie, BP 81, F-33140 Villenave d'Ornon, France

(5) INRA, UR1279, Unité Etude du Polymorphisme des Génomes Végétaux, CEA-Institut de Génomique-CNG, Evry, 91057, France

Keywords: *Solanum lycopersicum,* genomic, proteome, metabolome, SNP, fruit quality

Abstract

Improvement of fruit quality traits is a major goal for tomato breeding. Deciphering the genetic diversity and inheritance of fruit quality components is thus necessary. For this purpose, we carried out a large multi-level omic experiment. Eight contrasted lines and 4 of their F1 hybrids were phenotyped for fruit development traits. Fruit pericarp samples were analysed at 2 stages (cell expansion and orange ripe) and different scales: (1) untargeted profiling of major polar metabolites, (2) activities of 28 enzymes involved in primary metabolism, (3) proteome profiles revealed by 2D-PAGE and identification of 470 protein spots showing quantitative variations and (4) gene expression analysis by Digital Gene Expression. In parallel, the 8 lines were resequenced and more than 3 million SNPs identified when aligned on the reference tomato genome. This experiment allowed us to assess and compare the range of variability and inheritance mode of the metabolic traits and expression data. Correlation networks were constructed within and between levels of analysis to identify regulatory networks. Diversity of candidate genes could thus be analysed, relating the polymorphisms at the sequence levels with their expression.

INTRODUCTION

Dissection of the genetic variation and inheritance of phenotypic trait is the first step for plant improvement. Most phenotypic traits in plants are quantitatively inherited and controlled by the joint effects of several quantitative trait loci (QTL). QTL and association mapping are

widely used for the localization of polymorphisms responsible for phenotypic variation in plants (Morrell et al., 2012). Positional map-based cloning of QTL is the most straight-forward approach to identify genes involved in complex phenotypes, but cloning the causative genes is much more difficult than determining their positions. Few experiments of QTL map-based cloning have been reported in tomato (Frary et al., 2000; Fridman et al., 2000; Munos et al., 2011). Furthermore, DNA sequence variation (SNP or Indel) may not affect the traits directly. There are several intermediate levels between DNA genotypes and the traits phenotypes. The cascade of effects from DNA variation to trait phenotype is organized in complicated biological networks (Kliebenstein, 2010; Sulpice et al., 2010). Intermediate molecular phenotypes such as transcript and protein abundance also genetically vary in populations and are themselves quantitative traits (Rockman, 2006). High-throughput approaches have opened new prospects for analyzing biological systems and their complex functions at different levels including genomic, transcriptomic, proteomic, and metabolomic levels. System biology approaches integrating 'omic' resources and technologies offer new strategies for discovering links between co-regulated genes and pathways and ultimately, for predicting gene function and identifying regulatory genes in plants (Saito et al., 2008). It should enable us to understand the biology inside the black box of quantitative genetics relating genotype and phenotype in terms of causal networks of interacting genes. System approaches have been applied in yeast (Ideker et al., 2001), in the model plant Arabidopsis (Hirai et al., 2007) and in tomato (Mounet et al., 2009), at several levels.

Tomato (*Solanum lycopersicum*) is a model organism for the fleshy–fruited plants. Its genome has been almost fully sequenced (http://solgenomics.net/). The International Tomato Annotation Group (ITAG) identified more than 30,000 genes (Tomato Genome Consortium, in press). These data constitute a powerful tool for accelerating tomato functional genomics. We carried out an extensive multi-level omic experiment in order to dissect fruit quality at several scales. Eight contrasted lines and four of their F1 hybrids were phenotyped for fruit development traits. Fruits were harvested and pericarp samples analysed at metabolomic, proteomic, transcriptomic and genomic levels. This experiment allowed us to assess and compare the range of variability and the inheritance mode of the metabolic traits and expression data. Correlation networks were constructed within and between levels of analysis to identify regulatory networks. Diversity of chosen candidate genes was then analysed, relating the polymorphisms at the sequence levels with their expression.

MATERIALS AND METHODS

Plant material

The study was carried out using four *Solanum lycopersicum* lines (Levovil, Stupicke Polni Rane, LA0147, Ferum) with large fruits and four cherry type tomato, *S. l.* var *cerasiforme* lines (Cervil, Criollo, Plovdiv 24A, LA1420), as well as four hybrids between lines of the two groups. Ten plants per genotype were grown from February to August on 2010 in greenhouse in Avignon (France). For omic analyses, 20, 30 and 60 fruits were harvested for large, medium and cherry fruited tomato, respectively. Fruits were collected at two stages of

development, cell expansion stage (25, 20 and 14 days after anthesis for large, medium and small fruited tomato, respectively) and orange-red stage, according to the fruit colour. For each genotype and stage, three pools of 7 to 20 fruits were made by mixing fruits from truss 2 to 6 of 10 plants, but avoiding the first and last fruit of the truss. Pericarps were collected from each pool, immediately frozen, ground in liquid nitrogen and stored at -80 ˚C before analysis.

Methods

The transcriptome analyses were conducted on two biological replicates from cell expansion stage. RNA was isolated and purified using Qiagen RNeasy plant mini kit. RNA samples preparation was done with Illumina's Digital Gene Expression Tag Profiling Kit. Then tags were sequenced with illumina HiSeq 2000. Proteome and metabolome analyses were carried out on three biological replicates from the two stages. Protein isolation, separation by two-dimensional electrophoresis (2-DE), image analysis and mass spectrometry were carried out as previously described by Faurobert et al. (2007). Metabolome analyses were carried out using quantitative ^{1}H-NMR profiling of polar extracts, as described in Deborde et al.(2009) with minor modifications, on an Avance III 500 MHz spectrometer equipped with an ATMA inverse 5 mm probe. Enzymatic activity profiling was performed as described in Steinhauser et al. (2010). Correlation networks were reconstructed using sparse partial least squares regression (sPLS) analysis with R software (MixOmic package).

RESULTS AND DISCUSSION

Metabolite variation

The pericarp tissues of fruits at two stages of four lines with small fruits, four with large fruits and four of their hybrids were analysed. ^{1}H-NMR profiles allowed the quantification of 32 major polar metabolites, including sugars and sugar alcohols (sucrose, glucose, fructose, inositol), organic acids (citrate, malate, fumarate), amino acids (alanine, asparagine, aspartate, GABA, glutamine, isoleucine, leucine, phenylalanine, tyrosine, valine, threonine), phenolic acids (chlorogenate) and other compounds (trigonelline, choline). A large range of variability was observed among genotypes and between parents and hybrids. The main differences were observed between stages but the genotype also revealed a large range of variability (Table 1). Activities of 28 enzymes from central carbon metabolism were measured in pericarp tissue at two stages and also exhibited a large range of variation as illustrated on Figure 1 and Figure 2.

Proteome profile

A total of 1230 spots were detected using Samespots software, among which the abundance of 566 spots was significantly different between genotypes or stages. 424 spots were sequenced by LC-MS/MS and 422 proteins were identified. Among them, 355 spots corresponded to one gene, 25 spots corresponded to more than one gene with the same function, six spots corresponded to duplicated genes. A total of 36 spots were a mix of proteins corresponding to more than one gene with different functions. Thus, they were removed from analysis. Finally,

386 spots corresponding to 293 unique genes were used. Spots were classified according to the GO term related with Biological process. Spots in unknown category were classified based on their description when possible. The distribution of spot functions is show in Table 2. This classification is consistent with the spot functions identified by Faurobert et al. (2007). Among the 424 spots, 310 showed significant differences in volume according to stage, while 251 showed significant differences in volume according to genotype and 154 showed significant interactions between stage and genotype ($P<0.01$).

Transcriptome

The analysis of gene expression at cell expansion stage in the eight parental lines and four F1 hybrids allowed us to characterize the level of expression of more than 23,000 genes. Gene expression was analyzed at two different levels. In a first approach, we identified all the genes differentially expressed among the parental lines, and obtained a collection of 3,919 genes. The function of most of these genes was related with metabolic pathways that may be linked to the differences detected in the metabolomic profiles. Besides, we were also able to identify more than 100 transcription factors that may play a key role in tomato fruit development. A second approach was employed to analyze the genes differentially expressed between each couple of parents and their hybrid. This number of differentially expressed genes in each cross varied between 889 and 442, being greater in the cross between the two most distant lines. When we integrated the data from both approaches, we detected more than 500 genes that were differentially expressed just in the F1 hybrids.

Inheritance analysis

For each quantitative trait (metabolite content, enzyme activity, spot volume or gene expression), we assessed additivity and dominance and then the dominance over additivity ratio. Figure 3 compares the mode of inheritance of the traits in each cross. The number of variable traits varied among crosses in relation to the genetic distance between the parents. In average, more than 60% of the traits showed an additive inheritance, but differences were shown among crosses. A significant number of traits exhibited an over-dominant or over-recessive mode of inheritance, but no specific trend towards an excess of dominance or recessivity could be observed. These results are close to those of Schauer et al. (2008), except that we find a slightly higher rate of over-dominant and over-recessive traits. Several modes of inheritance are often observed for a category of traits (metabolite content, enzyme activity or gene/protein expression) as illustrated on Figure 2, suggesting different genetic controls in the different crosses.

Integrative analysis

Correlation networks were then analysed within and between levels of analysis. Many significant correlations were observed. For instance, several amino acids or sugars (fructose and glucose) varied in a coordinated manner. A few significant correlations between metabolites and enzyme activities were detected. Finally some significant correlations were observed between protein or gene expression and the metabolites or fruit size as illustrated Figure 4.

Genome sequences

We have resequenced the 8 genomes via GAII, with a depth varying from 9 to 26X. After alignment to the reference genome, the coverage rate was higher than 90% in every line. More than 3 millions SNPs were detected (defined with a depth higher than 8X), with large variation among lines (from 82,000 SNPs for the line the closest to the reference genome to more than 1,500,000 for the most distant one). Strong differences were also observed in the distribution of SNPs among the chromosomes. In average 1 SNP was detected every 382 bp (varying from 130 to 1309 according to the lines). Although only 2 to 3% of the SNPs are in coding regions, 4,000 to 25,000 SNP correspond to a non synonymous mutation. We now have to relate these data with the protein and gene expression variations.

CONCLUSION

This experiment using large 'omic' datasets provides a better understanding of hidden networks of molecular elements (genes, transcripts, proteins and metabolites) in tomato fruit. A multiallelic genetic intercross population (Cavanagh et al. 2008) has been constructed with the 8 parental lines and will soon be phenotyped. The detailed description of the 8 lines and their F1 hybrids will help us to identify the genes under the QTL.

Acknowledgements

The research was funded by ANR MAGICTomSNP project. Thanks to Yolande Carretero and Esther Pelpoir for technical help.

LITTERATURE CITED

Cavanagh C., Morell M., Mackay I. and Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Current Opinion in Plant Biology 11: 215-221

Deborde C., Maucourt M., Baldet P., Bernillon S., Biais B., Talon G., Ferrand C., Jacob D., Ferry-Dumazet H., de Daruvar A., Rolin D., Moing A. 2009. Proton NMR quantitative profiling for quality assessment of greenhouse-grown tomato fruit. Metabolomics 5 : 183-198

Faurobert M., Mihr C., Bertin N., Pawlowski T., Negroni L., Sommerer N., Causse M. 2007. Major proteome variations associated with cherry tomato pericarp development and ripening. Plant Physiol. 143: 1327-1346

Frary A., Nesbitt T.C., Grandillo S., van der Knaap E., Cong B, Liu J., Meller J., Elber R., Alpert KB., Tanksley SD. 2000. fw2.2: A Quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88.

Fridman E., Pleban T. and Zamir D. 2000. A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA 97: 4718–4723

Hirai M., Sugiyama K., Sawada Y., Tohge T., Obayashi T., Suzuki A., Araki R., Sakurai N., Suzuki H., Aoki K., Goda H., Nishizawa O.I., Shibata D. and Saito K. 2007. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. Proc. Natl. Acad. Sci. USA 104: 6478-6483.

Ideker T., Galitski T., Hood L. 2001. A new approach to decoding life: Systems biology. Ann. Rev. of Genomics and Human Genetics 2: 343-372.

Kliebenstein DJ. 2010. Systems biology uncovers the foundation of natural genetic diversity. Plant Physiol. 152: 480–486.

Morrell PL., Buckler ES., Ross-Ibarra J. 2012. Crop genomics: advances and applications. Nature Reviews Genetics 13: 85-96.

Mounet F., Moing A., Garcia V., Petit J., Maucourt M., Deborde C., Bernillon S., Le Gall G., Colquhoun I., Defernez M, Jean-Luc Giraudel JL., Dominique Rolin D., Rothan C. and Lemaire-Chamley M. 2009. Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. Plant Physiol. 149: 1505–1528

Muños S., Ranc N., Botton E., Bérard A., Rolland S., Duffé P., Carretero Y., Le Paslier MC., Delalande C., Bouzayen M., Brunel D. and Causse M. 2011. Increase in tomato locule number is controlled by two key SNP located near Wuschel. Plant Physiol 4: 2244-2254

Saito K., Hirai MY., Yonekura-Sakakibara K. 2008. Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. Trends Plant Sci. 13: 36-43

Schauer N., Semel Y., Balbo I., Steinfath M., Repsilber D., Selbig J., Pleban T., Zamir D., Fernie AR. 2008. Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20:509-523.

Steinhauser MC., Steinhauser D., Koehl K., Carrari F., Gibon Y., Fernie AR., Stitt M. 2010. Enzyme Activity Profiles during Fruit Development in Tomato Cultivars and *Solanum pennellii*. Plant Physiol. 153: 80-98

Sulpice R., Trenkamp S., Steinfath M., Usadel B., Gibon Y., Witucka-Wall H., Pyl E.T., Tschoep H., Steinhauser MC., Guenther M., Hoehne M., Rohwer JM., Altmann T., Fernie A. and Stitt M. (2010) Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. Plant Cell 22: 2872–2893

TABLES

Table 1. Range of variation observed for primary metabolites in fruits at 2 stages, cell expansion (CE) and orange ripe (OR), in 12 genotypes,. Only the minimum and maximum values are shown

	CE			OR			CE/OR	
	min	Max	max/min	min	max	max/min	min	max
glucose	68996.3	193536.8	2.81	158574.1	233208.3	1.47	0.42	0.90
sucrose	6417.2	11856.6	1.85	7596.8	28049.8	3.69	0.31	1.08
fructose	65853.1	198775.3	3.02	173280.2	238532.6	1.38	0.38	0.85
inositol	3780.8	6188.0	1.64	2135.3	4796.5	2.25	1.29	2.08
citrate	26295.3	62463.3	2.38	49525.9	149942.5	3.03	0.37	0.57
malate	13562.9	25447.1	1.88	3894.1	30910.2	7.94	0.59	3.78
fumarate	5.5	18.7	3.38	0	12.7	NA	0.88	3.70
alanine	283.6	1356.3	4.78	171.0	492.6	2.88	0.78	5.51
asparagine	758.6	2461.9	3.25	1200.6	4593.9	3.83	0.31	0.95
aspartate	576.5	1278.4	2.22	1564.4	3598.3	2.30	0.20	0.42
GABA	4099.5	8485.6	2.07	681.1	2464.1	3.62	1.83	7.31
glutamine	2756.4	22886.1	8.30	6605.5	24155.8	3.66	0.24	1.21
isoleucine	166.2	865.5	5.21	232.6	687.6	2.96	0.57	3.30
leucine	309.5	927.6	3.00	401.5	1042.5	2.60	0.40	1.26
phenylalanine	1209.4	4905.7	4.06	1988.7	7429.7	3.74	0.35	0.94
tyrosine	156.2	760.6	4.87	185.9	656.8	3.53	0.50	2.20
valine	214.6	1004.1	4.68	91.7	396.7	4.33	1.19	10.95
threonine	125.2	745.3	5.95	221.4	768.6	3.47	0.38	2.12
chlorogenate	295.5	1484.4	5.02	378.4	949.3	2.51	0.61	1.56
choline	551.2	1018.9	1.85	555.1	1036.1	1.87	0.72	1.11
trigonelline	220.9	872.8	3.95	217.0	766.8	3.53	0.90	1.24

Table 2. Function classification of the 424 protein spots sequenced

Function classification	Spot number
biosynthetic process	1
cellular metabolic process	25
cellular response to stimulus	2
developmental maturation	7
establishment of localization	23
macromolecular complex subunit organization	1
macromolecule metabolic process	97
nitrogen compound metabolic process	4
organelle organization	3
organic substance metabolic process	2
oxidation-reduction process	3
primary metabolic process	109
regulation of biological process	28
response to chemical stimulus	10

response to stress	90
Unknown	17

Figures

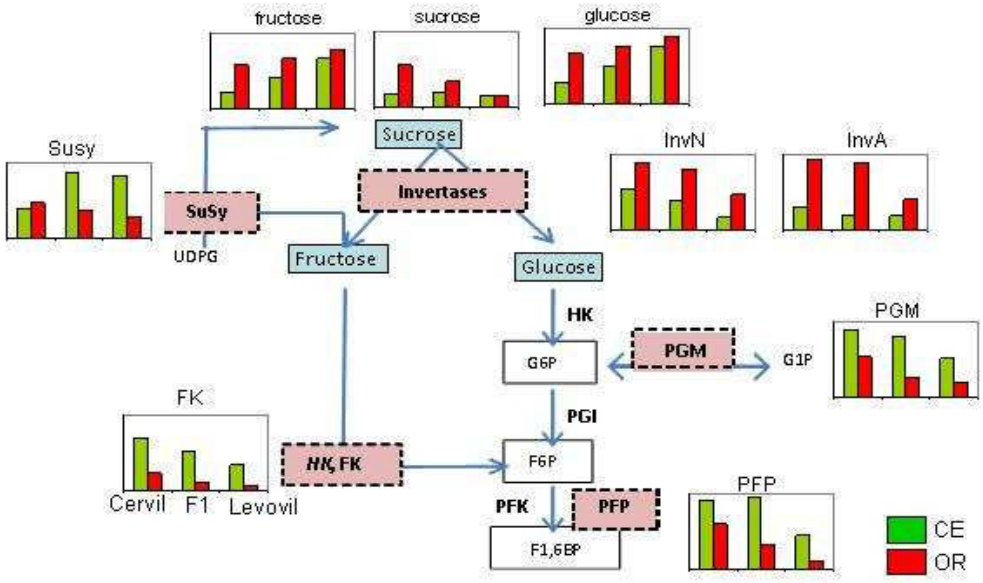


Figure 1: Genetic variation in sugar contents and related enzyme activities, sucrose synthase (SuSy), Neutral and acid inverstase (InvN, InvA), phosphogluco-mutase (PGM), fructokinase (FK) and phosphor-fructo phosphorylase (PFP), at cell expansion (CE) and orange ripe (OR) stages in 3 genotypes: a cherry tomato (Cervil), a large fruited line (Levovil) and their F1.

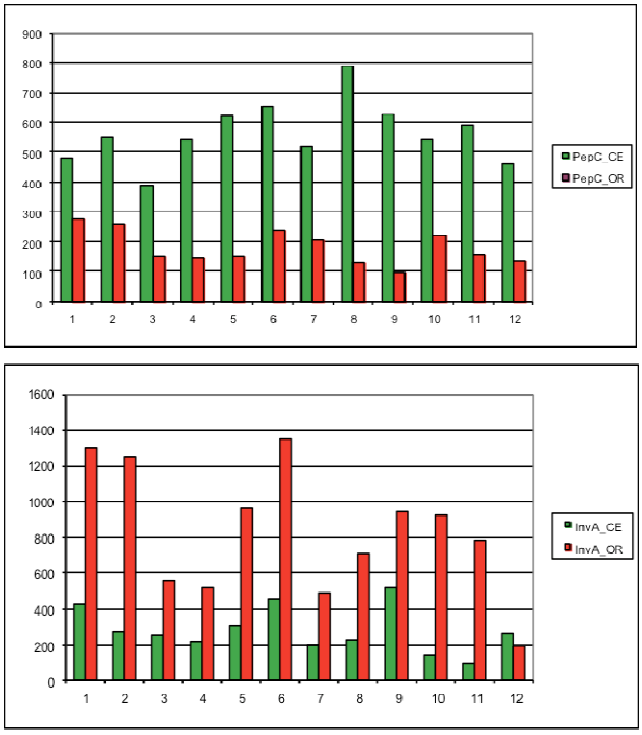


Figure 2: Variation in enzyme activity (Top: Phospho*enol*pyruvate carboxylase; bottom: Acid invertase) among the 12 genotypes at two stages, cell expansion (CE) and orange ripe (OR). The genotypes 2, 5, 8 and 11 correspond to hybrids between their two neighbours.

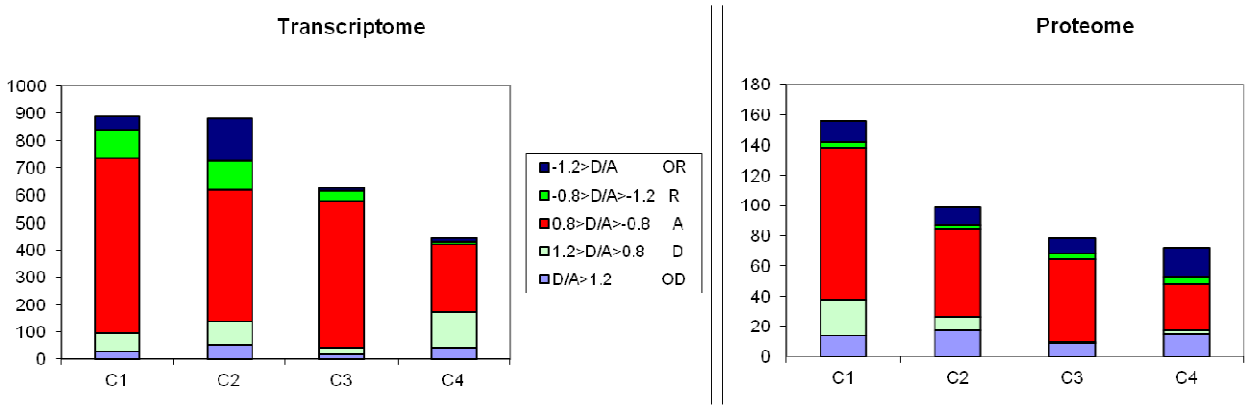


Figure 3: Mode of inheritance of the transcript and protein abundance. For each cross (C1 to C4), when the means of the 3 genotypes were significantly different, the F1 value was compared to its parental lines and the ratio of dominance to additivity (D/A) assessed.

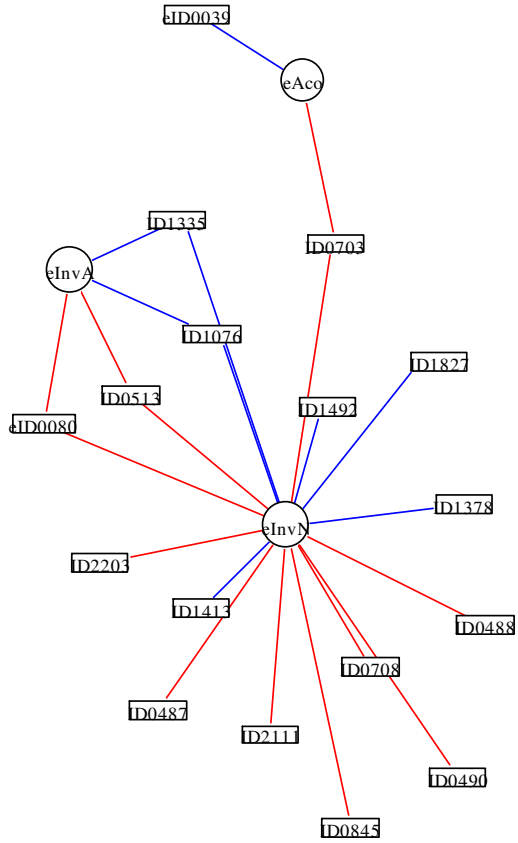


Figure 4: Example of a correlation network identified between enzyme activities and protein abundance at the cell expansion stage. The central role of neutral invertase is underlined. Circles correspond to enzyme activities and squares to protein spot volumes. The network was obtained by sPLS analysis. Only correlations with absolute values higher than 0.9 are represented.

Summary of thesis:

Fruit quality in tomato is highly dependent on genetic variation. Following domestication and modern breeding, molecular diversity of tomato has been strongly reduced, limiting the possibility to improve traits of interest. New molecular markers such as single nucleotide polymorphisms (SNP) constitute precious tools to saturate tomato genetic maps and identify quantitative trait loci (QTL) and associations in a poorly polymorphic species like tomato. The objectives of this study were to characterize tomato genetic diversity at the molecular levels and to try to identify QTLs, genes and proteins responsible for fruit quality traits in tomato. For this purpose, three independent studies were conducted leading to the discovery of new SNP markers, their use for association study and finally the analysis of proteome diversity in relation to physiological phenotypes. We first used two next-genration sequencing platforms (GA2 Illumina and 454 Roche) to re-sequence targeted sequences covering about 0.2% of the tomato genome from two contrasted accessions. More than 3000 SNPs were identified between the two accessions. We then validated 64 SNPs by developing CAPS markers. We thus showed the value of NGS for the discovery of SNPs in tomato and we produced low cost CAPS markers which could be used to characterize other tomato collections. A SNPlex™ array carrying 192 SNPs was then developed and used to genotype a broad collection of 188 accessions including cultivated, cherry type and wild tomato species and to associate these polymorphisms to ten fruit quality traits using association mapping approach. A total of 40 associations were detected and co-localized with previously mapped QTLs. Some other associations were identified in new regions. We showed the potential of using association genetics in tomato. Finally, a new analytical approach combining proteome, metabolome and phenotypic profiling were applied to study natural genetic variation of fruit quality traits in eight diverse accessions and their four corresponding F1s at cell expansion and orange-red stages. We identified 424 variable spots by combining 2-DE and nano LC MS/MS and built the first comprehensive proteome reference map of the tomato fruit pericarp at two developmental stages from the 12 genotypes. In parallel, we measured the variation of 34 metabolites, 26 enzyme activities and five phenotypic traits. A large range of variability and several inheritance modes were described in the four groups of traits. Data integration was achieved through sPLS and correlation networks. Many significant associations were detected within level and between levels of expression. This systems biology approach provides better understanding of networks of elements (proteins, enzymes, metabolites and phenotypic traits) in tomato fruits.

Keywords : Tomato ; Fruit quality ; Molecular markers ; Association mapping ; QTL ; Proteome ; Systems biology

Résumé de la thèse:

L'amélioration de la qualité du fruit de tomate dépend largement de la variation génétique. A la suite de la domestication et de la sélection moderne, la diversité moléculaire chez la tomate a été profondément réduite, limitant les possibilités d'amélioration. De nouveaux marqueurs moléculaires révélant les polymorphismes nucléotidiques (SNP) constituent des outils précieux pour saturer les cartes génétiques et identifier des QTL (*quantitative trait loci*) et des associations chez une espèce peu polymorphe comme la tomate. Les objectifs de cette étude étaient de caractériser la diversité génétique de la tomate au niveau moléculaire et de tenter d'identifier des QTL, des gènes et des protéines responsables de la variation de caractères de qualité du fruit. Pour cela, trois études indépendantes ont conduit à (1) la découverte de nouveaux marqueurs SNP, (2) leur utilisation en génétique d'association et (3) l'analyse de la diversité du protéome en relation avec des caractères physiologiques du fruit. Dans la première étude, nous avons comparé deux plateformes de reséquençage pour reséquencer des zones ciblées couvrant environ 0.2% du génome de deux accessions contrastées. Plus de 3000 SNPs ont été identifiés. Nous avons ensuite validé 64 SNPs en développant des marqueurs CAPS. Nous avons ainsi montré l'intérêt des techniques de reséquençage pour la découverte de SNP chez la tomate et produit des marqueurs simples qui peuvent être utiles pour caractériser de nouvelles ressources. Nous avons ensuite développé un ensemble de 192 SNPs et génotypé une collection de 188 accessions comportant des accessions cultivées, des type "cerise" et des formes sauvages apparentées et recherché des associations avec 10 caractères de qualité du fruit. Une quarantaine d'associations a été détectée dans des régions où des QTL avaient été préalablement identifiés. D'autres associations ont été identifiées dans de nouvelles régions. Nous avons ainsi confirmé le potentiel de la génétique d'association pour la découverte de QTL chez la tomate. Finalement une approche combinant l'analyse du protéome, du métabolome et de traits phénotypiques a été mise en œuvre pour étudier la variabilité naturelle de la qualité du fruit de huit lignées contrastées et de quatre de leurs hybrides, à deux stades de développement (expansion cellulaire et orange-rouge). Nous avons identifié 424 spots protéiques variables en combinant électrophorèse bidimensionnelle et nano LC MS/MS et construit une carte de référence du protéome de fruit de tomate. En parallèle, nous avons mesuré la variation de teneurs en 34 métabolites, les activités de 26 enzymes et cinq caractères phénotypiques. La variabilité génétique et les modes d'hérédité ont été décrits. L'intégration des données a été réalisée par construction de réseaux de corrélations et régression sPLS. Plusieurs associations ont été détectées intra et inter niveau d'expression, permettant une meilleure compréhension de la variation de la qualité des fruits de tomate.

Keywords : Tomate ; Qualité des fruits ; Marqueurs molécualaires ; QTL ; Génétique d'association ; Protéome ; Biologie des systèmes.