Exploiting **wild relatives** of ***S. lycopersicum*** for quality traits

Ana Marcela Víquez Zamora

Exploiting wild relatives of *S. lycopersicum* for quality traits

Ana Marcela Víquez Zamora

Thesis committee

Promotor
Prof. Dr Richard G.F. Visser
Professor of Plant Breeding
Wageningen University

Co-promotors
Dr Adriaan W. (Sjaak) van Heusden
Senior Scientist, Wageningen UR Plant Breeding, Wageningen University & Research Centre

Dr Arnaud G. Bovy
Senior Scientist, Wageningen UR Plant Breeding, Wageningen University & Research Centre

Other members
Prof. Dr Bas J. Zwaan, Wageningen University
Prof. Dr Celistina Mariani, Radboud University, Nijmegen
Dr Wilco Ligterink, Wageningen University
Dr Gabino Sanchez Perez, Plant Research International,
Wageningen University & Research Centre

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Exploiting wild relatives of *S. lycopersicum* for quality traits

Ana Marcela Víquez Zamora

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 4 September 2015
at 4 p.m. in the Aula.

Ana Marcela Víquez Zamora
Exploiting wild relatives of *S. lycopersicum* for quality traits,
167 pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summary in English

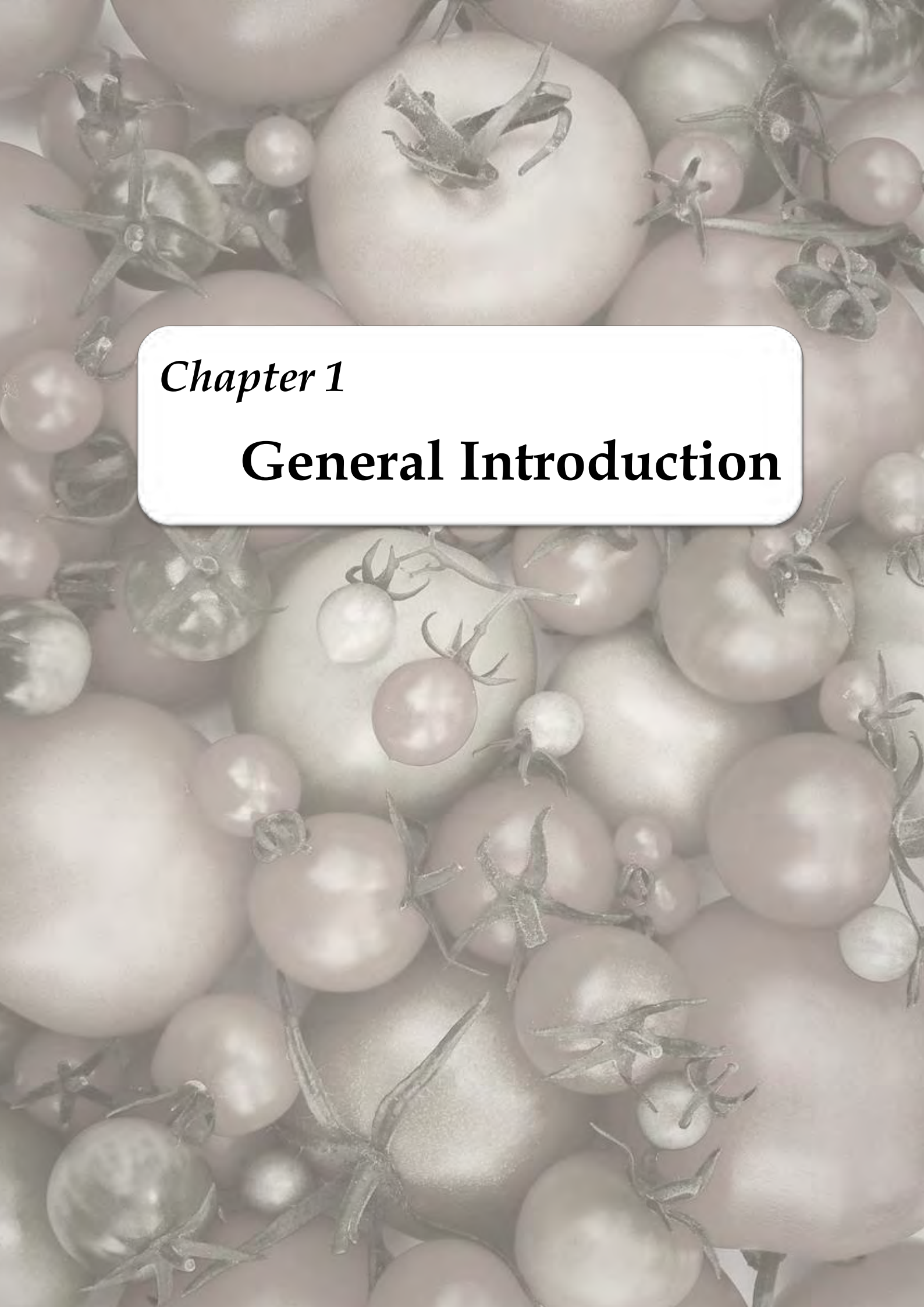
ISBN 978-94-6257-441-0

Contents

Preface

A good way to start this thesis would be to define quality. But, how can we define quality? *Quality* and *normal* are subjective terms; both concepts should meet your personal prejudices and expectations. Therefore, we can say that *quality* is the land of no one but at the same time of everybody.

Even so, another aspect within *quality* is that it implies the need of a reference parameter. This is favourable, because the possibility to compare facilitates research. In the work reported in my thesis our aim was to explore and to compare. We explored tomato germplasm from a genetic and metabolic point of view. Then, we compared the data in order to make hypotheses for the underlying mechanisms of traits, their genetics or even for the use of different technologies that could help breeders to exploit wild relatives to comply with different market expectations on *quality*.



Chapter 1

General Introduction

Tomato history & etymology

Etymologically, the word tomato was derived from the Náhuatl *tomātl*. The native Mexicans formed the word out of a combination of the verb '*tomaua*' that means to gain weight and '*atl*' that means water. Moreover, the Aztecs made the differentiation between *tomātl* and *xitomātl* referring *xitomātl* to a 'plump thing with navel'. With the navel distinction, it is likely that they were describing the rounded knotty depression formed when the peduncle is detached from the fruit, as in the fruits we consume today (Long, 1995).

In 1694, Tournefort assigned to tomatoes the botanical name *Lycopersicon esculentum*, which literally means "edible wolf peach". This, due to the association of tomato plants to deadly nightshades that according to German myths were used by sorcerers to summon or even become werewolves. Later, in 1753, Linnaeus began to consistently use Latin binomials in *Species Plantarum*. He classified tomatoes in the genus *Solanum* and assigned to the cultivated tomato the name of *S. lycopersicum* and *S. peruvianum* to more wild relatives. Though, the following year (1754), Miller kept on using Tournefort's botanical name to distinguish the genus from the nightshades (*Solanaceae*) (Peralta *et al.*, 2006).

The genus continued to be referred as *Lycopersicon esculentum* until almost the 21st century when taxonomists started to agree on the similarities of tomatoes with the nightshades. Spooner *et al.* (1993; 2005) gave phylogenetic, distribution and morphologic evidence that tomatoes belong to the genus *Solanum*. The use of *Solanum lycopersicum* has now become fully integrated in Solanaceae germplasm worldwide.

Crop and importance

Tomatoes are part of the Solanaceae family which comprises over 3000 species including potatoes, aubergines, peppers, petunias and *Physalis*. Approximately half of the Solanaceae family is represented by the genus *Solanum* which comprises potato as well (Bergougnoux, 2014). According to the available information about the production of the top 50 commodities in the world, tomato is on the ninth place. The most produced commodities are sugar cane, maize, rice (paddy), wheat, potatoes, sugar beet, cassava and soybeans. But according to product value, tomato is the fourth commodity in the world after rice, wheat and soybeans (FAOStat, 2015: http://faostat.fao.org/).

Today, it would be difficult to imagine the Mediterranean or even the Northern European cuisine without tomato or potato. But neither Europe nor Asia knew tomatoes before the sixteenth century. Meaning that within the past 400 years China, India, Turkey and Egypt incorporated tomatoes in their cultures and became four of the top five tomatoes' producers in the world. The production worldwide accounts for more than 161 million tonnes in 2012,

from which China alone produces more than 50 million tonnes (FAO, 2014). The other country in the top five is the United States of America. Heinz ketchup and Campbell's canned tomato soup pushed the market and consumption of the fruits. Tomatoes are known and produced globally. Although the Netherlands, Belgium, Iceland, United Kingdom and Ireland are the countries producing the highest yields per hectare, Asia is by far the largest tomato producing continent (Fig. 1). This figure demonstrates that tomatoes adapted, evolved and transformed societies and economies around the world (Harvey *et al.*, 2002).

Botanically, tomatoes are fruits. They are formed from the plant flowers' ovaries and contain seeds. Still, according to culinary uses, tomatoes are more served in salads or as part of the main courses and not as desserts. Therefore, people tend to perceive these fruits as vegetables. In general, the species *S. lycopersicum* is self-compatible. However, some tomato wild relatives have self-incompatibility mechanisms. Some wild tomatoes can have flowers where the pistil extends out and the stamens remain within the closed corolla not allowing self-pollination and the formation of fruits.

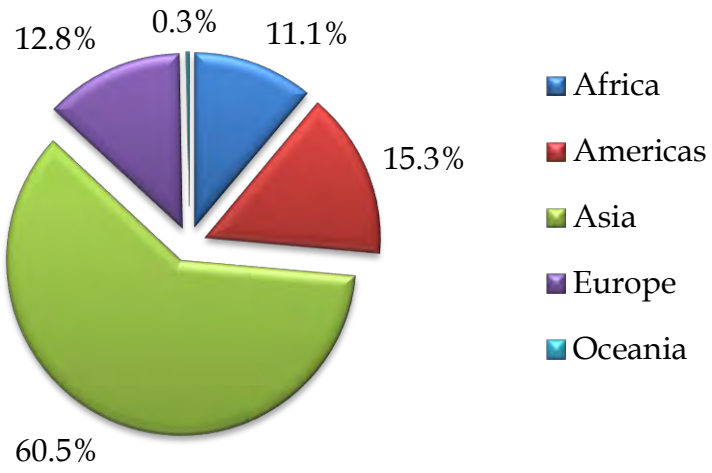


Figure 1. Production share by region. Average between 2012 – 2013. Source: FAOStat, 2014.

Among *S. lycopersicum* there are annual, herbaceous plants with upright, creeping or bushy habit according to the varieties. Tomatoes can be determinate or indeterminate plants. The flowers are small, yellow and star shaped. They get together on the same peduncle to form trusses of fruits that can vary in number. The berry type; fleshy fruit guard the seeds. The seeds are surrounded by a mucilaginous substance that provides seeds with some nutrients to germinate. The plants have composed leaves with a strong smell. Leaves, and even some stems, can have glandular, short or long multicellular trichomes. The complete cycle from seed to seed is between 90-120 days under optimal conditions. Tomatoes are noble fruits that

do not present thorns, hard cortex or bones. Once ripened, they are ready-to-eat from the vine.

Origin and wild relatives

Native tomato plants have their origin in the western part of the South American Andes (Rick 1973, Taylor 1986). They are distributed from Ecuador, including the Galápagos Islands, to the north part of Chile (Darwin *et al.*, 2003; Peralta and Spooner, 2005). Tomato wild relatives´ habitats comprise coasts, deserts, valleys, islands and even mountains of more than 3300 meters of altitude. The range of ecologically diverse places allowed the evolution of a wide diverse range of phenotypes (Blanca *et al.*, 2012).

Even though the place of origin is clear, the place of domestication of the cultivated tomatoes is somewhat unclear. It is likely that the first tomatoes were introduced in Europe by Hernán Cortés in 1523, after the invasion of Mexico. Towards Asia, Fernando de Magallanes in 1521 started trading Mexican goods with the Philippines and the neighbouring countries such as China, Japan and India. Commerce is a probable reason for the dissemination of tomatoes (Villareal, 1980).

José de Acosta reported for the first time in 1590, a tomato plant with fruits full of juice in his history book from the 'Indias'. Unluckily he didn't mention where he saw the plant, more towards Mexico or more towards Peru (Jenkins 1948). DeCandolle (1886) believed in a Peruvian hypothesis, while Jenkins (1948) developed the Mexican hypothesis. However, since there is a difficulty in identifying truly native *S. lycopersicum* or even true landraces from Mexico and Peru it is problematic to elucidate the factual first tomato place of domestication (Peralta *et al.*, 2006). However, it is clear that in general tomatoes went to other countries to stay.

Solanum section *Lycopersicon* includes thirteen species in four groups. The group with the same name 'Lycopersicon' includes the cultivated tomato *S. lycopersicum, S. pimpinellifolium, S. cheesmaniae* and *S. galapagense*. The Neolycopersicon group includes the species *S. pennellii.* The Eriopersicon group embraces *S. chilense, S. corneliomulleri, S. habrochaites, S. huaylasense,* and *S. peruvianum.* The group Arcanum includes *S. arcanum, S. chmielewskii* and *S. neorickii*. Furthermore, allied species are classified into two sections: *Lycopersicoides* (*S. lycopersicoides* and *S. sitiens*) and *Junglandifolia* (*S. jungladifolium* and *S. ochranthum*) (Grandillo *et al.*, 2011). Seventeen species closely related to, but with distinct characteristics from the tomatoes we eat today.

In this thesis we focus on exploring the variation between accessions of wild relatives of the species in the subsection Lycopersicon (Table 1). In general, they are herbaceous shrubs, with

scentless flowers with yellow petals and yellow fused anthers (Grandillo *et al.*, 2011). In literature, there is one variety that is considered as a wild cherry tomato referred as *Solanum lycopersicum* var. cerasiforme. However, as Peralta *et al.* (2008) stated, this wild cherry tomato might be a feral plant or a mixture between domesticated tomatoes and wild relatives such as *S. pimpinellifolium* (Nesbitt & Tanksley 2002; Ranc *et al.*, 2008; Xu *et al.*, 2013).

Since the species from the group *Lycopersicon* are bilaterally compatible with *S. lycopersicum*, they represent a good source to explore variation for quality traits that can be incorporated into the cultivated tomatoes.

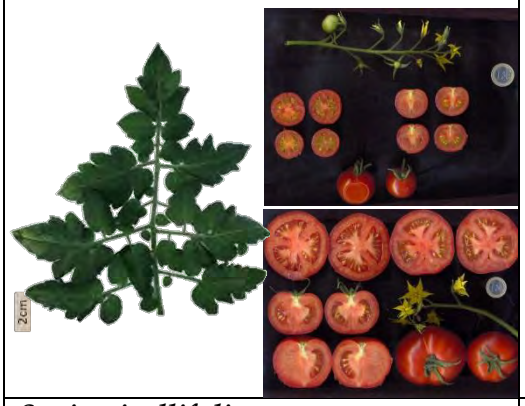
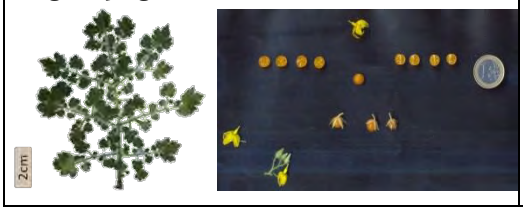
~omics approaches

Francis Crick started to elucidate this path for all species when he described the central dogma of molecular biology (Crick, 1970). Crick described how DNA is translated into RNA and which then is used to produce proteins. The research continued and the specific steps were elucidated. Nowadays, we have a differentiation of levels, methodologies and technologies applied into research to elucidate phenotypes. Each level receive the suffix ~omics, referring to the study of a level as a whole. There is an entire display of definitions for different ~omics approaches.

The list of ~omics topics in biology can be boundless today. However, for our purpose we will keep a simple explanation of how high-dimensional biology leads to phenotypes in general (Fig. 2). Genomics is related to studies at DNA level of an organism. Transcriptomics refers to the study of RNA molecules including messenger (mRNA), ribosomal (rRNA), transfer (tRNA) and non-coding RNA, the transcriptome within the cells. Proteomics focuses on proteins, the large biological molecules of organisms and their arrangement, expression and changes in cell's proteome. Then, metabolomics refers to the study of the metabolome; those chemical fingerprints derived from the organism processes (Bino *et al.*, 2004; Ravi *et al.*, 2014). All of these components are part of an events´ cascade leading to a certain phenotype (phenome) in each organism.

Ccombined ~omics approaches can offer a better understanding of the metabolism underlined by genes and having interactions with the environment to develop a phenotype (Osorio *et al.*, 2009).

Table 1. Main ecological, botanical, and reproductive features of the wild tomatoes from *Solanum* subsection Lycopersicon, group Lycopersicon. Adapted from Peralta *et al.*, 2005; Grandillo *et al.*, 2011 and Bergougnoux, 2014. SC: self-compatible. Leaves and fruits pictures from greenhouse grown plants.

Species	Distribution and habitats	Mating system; fruit colour	Distinguishing morphological features	Importance in breeding
S. lycopersicum 	Found world-wide in a variety of habitats (sea level to 4000 m). Plants with smaller fruits referred as *S. lycopersicum* var. ''cerasiforme''*. Usually found in moderate humid sites.	SC-autogamous, facultative allogamous; Red	Plants semi-erect to sprawling; fruits red, 1.5 up to 60cm with fused pistils.	Moisture-tolerance, resistance to wilt, root-rotting, and leaf-spotting fungi, size and shapes
S. pimpinellifolium 	Lowland Ecuador and coastal Peru (sea level – 500 m). Found in arid, sandy places, often near sources of water or on the edges of farm fields.	SC-autogamous, facultative allogamous; Red, orange	Plants semi-erect to sprawling, flower small-large; fruit red 0.5–1 cm	Colour and fruit quality; resistance to insect, nematode and disease
S. cheesmaniae 	Endemic to Galápagos Islands (sea level – 1500 m). Found in arid, rocky slopes, prefers shaded, cooler sites.	SC-autogamous; Yellow, orange	Plants semi-erect to sprawling, flowers very small, pale; fruit purple, greenish yellow, or orange, 0.5–1.5 cm	Salt tolerance; Lepidoptera and virus resistance
S. galapagense 	Endemic to Galápagos Islands (sea level – 650 m). Found in arid, rocky outcrops and slopes, sometimes near shoreline.	SC-autogamous; Yellow, orange	Plants erect; leaves highly subdivided; internodes short; flowers small, pale, fruit orange 0.5–1 cm	Salt tolerance; Lepidoptera and virus resistance

* *S. lycopersicum* "var. cerasiforme" is referred worldwide but is not validly published under the rules of botanical naming (Peralta *et al.*, 2008).

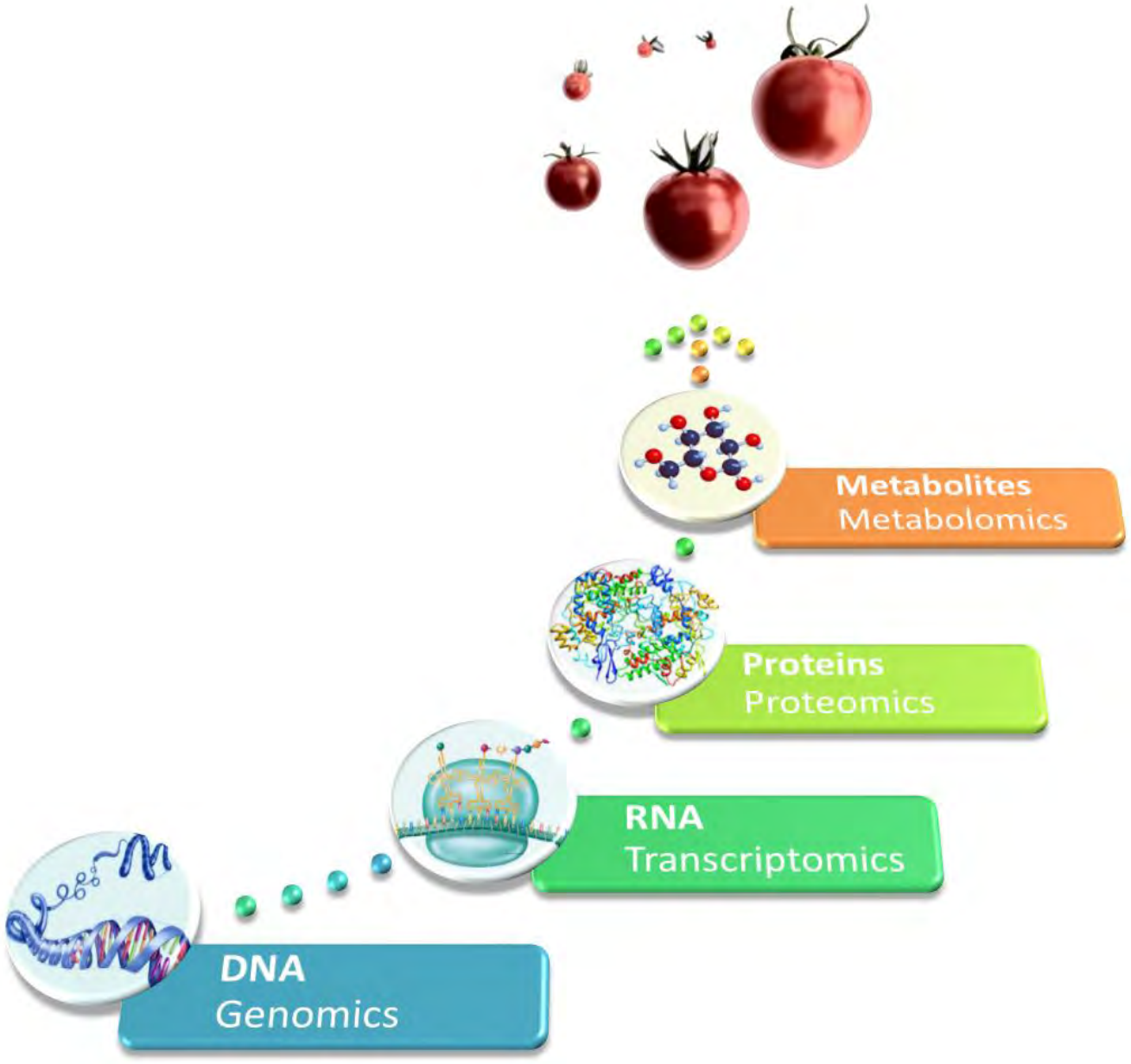


Figure 2. From genomics to metabolomics: Levels of research to understand phenotypes.

Molecular breeding in tomato

Tomato is a diploid species with 12 chromosome pairs. The predicted genome size for *S. lycopersicum* is approximately 900 megabase pairs (Mbp or Mb) and from those, currently approximately 760 Mbp are assembled into 91 scaffolds which is the basis of the published version of the tomato genome (Tomato Genome Consortium, 2012). Tomato became the model to study the genetics of solanaceous plants after being the first crop for which the use of marker assisted selection (MAS) in breeding was suggested (Rick and Fobes, 1974). In addition, tomato became the model for fleshy fruit development. Tomato is now studied worldwide also in collaborations such as the SOL Genomics Network (http://solgenomics.net/).

The capability of self-pollination in *S. lycopersicum* allows high levels of homozygosity in the plants. Therefore, the generation of a great diversity of genetically stable resources made tomato an ideal candidate to explore the genus. Researchers continue to work on the generation of genetic maps, advanced breeding populations, information about wild relatives, developed markers, annotated genes, microarrays and so on. As mentioned, all these resources are trying to elucidate, in one way or another, the relation between genotypes and phenotypes.

In the application of molecular breeding in tomato, genotyping has been a major bottle neck for a long time. Although several marker systems have been developed and applied, most of them fall short in the genomics era, either by lack of reproducibility or by being too low throughput (Agarwal *et al.*, 2008). This changed radically with the evolution of the genome (re)sequencing projects that resulted in large numbers of informative markers such as single nucleotide polymorphisms (SNP).

In our project we used a custom made Illumina Infinium Array to genotype tomato germplasm. A total of 6000 SNPs were selected between four commercial cultivars of *S. lycopersicum*, two round and two cherry tomatoes. From those, 5528 SNPs and their corresponding sequences passed the quality check according to the requirements of Illumina. Then, the protocol depicted in Figure 3 was executed for every sample in each corresponding SNP probe. With the SNP allele calls we performed a genomic analysis.

Furthermore, we genotyped by sequencing (GBS) a subset of a recombinant inbred line (RIL) population. Whole genomes were re-sequenced and the reads were aligned to the tomato reference genome. This information was used in the thesis to increase the robustness of a genetic map and the (fine) mapping analysis of QTLs.

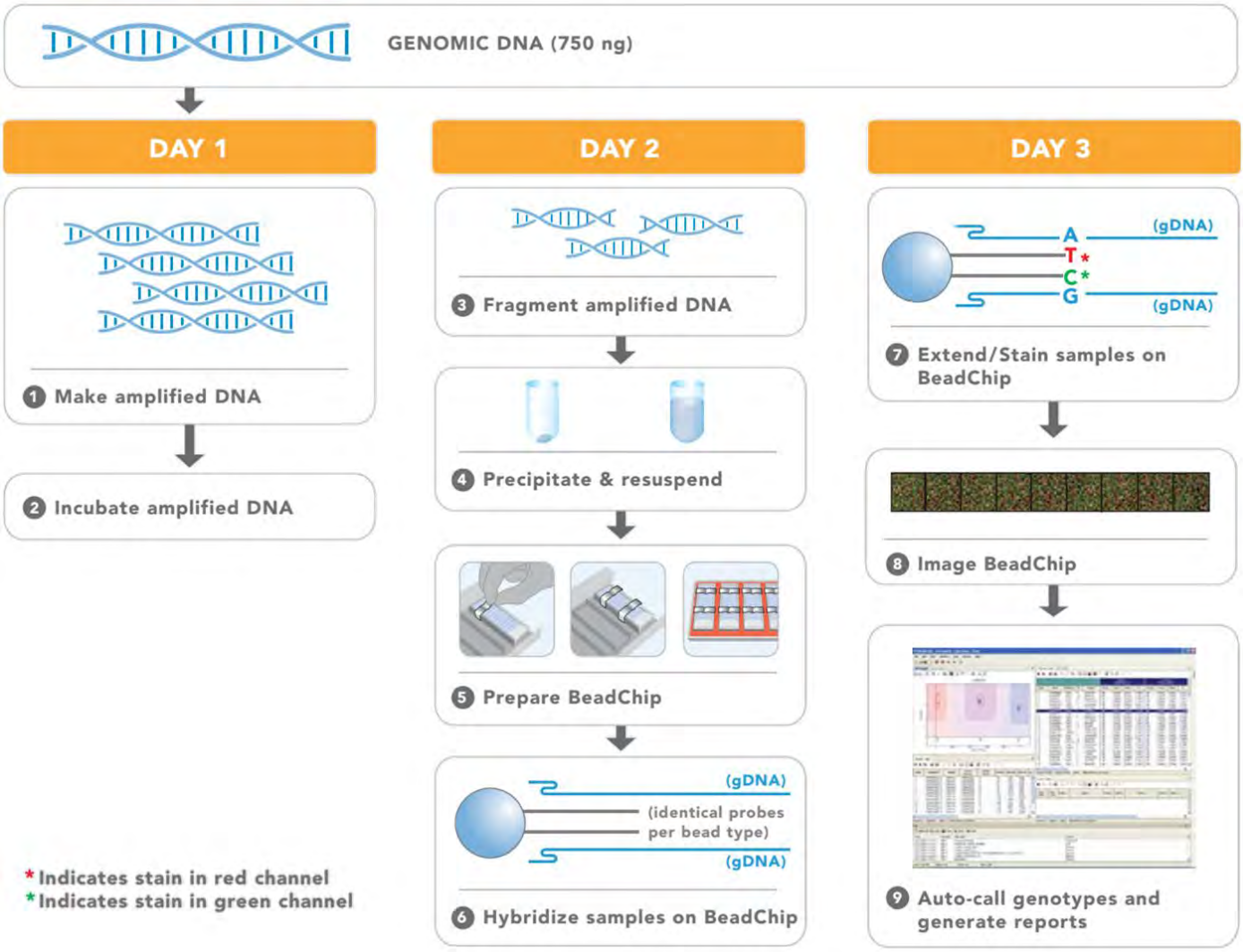


Figure 3. Infinium® Assay Workflow. Whole genome genotyping procedure used to evaluate SNPs in diverse tomato germplasm (http://res.illumina.com/documents/products/workflows/workflow_infinium.pdf).

Breeding for taste and nutritive value

Tomatoes are consumed fresh, processed, dried, and concentrated, in paste, juices and sauces. Breeding programs are adapted according to each market preference. If we talk about the organoleptic quality of a fresh fruit, flavour is one of the major attributes. In tomato, flavour is determined by the amount of sugars, acids, volatiles and texture-related compounds of the fruit. According to Tieman *et al.* (2012), the most important compound accounting for flavour in tomatoes is fructose. In addition, seven other compounds were correlated to flavour intensity: citric acid, 2-butylacetate, cis-3-hexen-1-ol, 3-methyl-1-butanol, 2-methylbutanal, 1-octen- 3-one, and trans-2,4-decadienal. Furthermore, several

volatiles are important for aroma such as geranial, 2-methylbutanal, and 3-methyl-1-butanol (Tieman *et al.*, 2012).

Tomato has a high nutritive value, micronutrients like carotenoids, vitamins, glycoalkaloids, minerals, amino acids, phenols and flavonoids are major compounds accounting for tomato nutritive value. These compounds present in the metabolome of tomatoes, can be targeted by different analytical approaches. These approaches can concentrate on a single class of metabolites like sugars or amino acids, or can be used for a broad untargeted profiling aimed at detecting as many as possible compounds in a given extract. In general, several analytical platforms are needed to cover the entire metabolome, since no analytical technique can detect all metabolites of a sample in one go. Therefore, users need to establish beforehand which classes of metabolites they aim to target in their analyses (Osorio *et al.*, 2009).

In our project, we used three different platforms to perform metabolic profiling of different samples: 1) Liquid chromatography (LC) coupled to mass spectrometry (MS). LC-QTOF-PDA-MS was used to extract, separate and detect semi-polar compounds such as flavonoids, alkaloids, and phenolic acids. We used an untargeted LC-MS metabolic profiling approach to get a broad overview of the variation in semi-polar metabolites, including possible health-related compounds such as flavonoids (De Vos *et al.*, 2007) and trying to reach beyond the known health benefits of carotenoids such as lycopene. 2) Gas chromatography (GC) coupled to electron impact time of flight (TOF)-MS (Lisec *et al.*, 2006). GC-TOF-MS can be used to detect primary metabolites, including organic and amino acids, sugar alcohols and sugars, after 'volatising' them through a chemical derivatization step. 3) Solid phase microextraction (SPME)-GC-MS. SPME-GC-MS allowed the detection of approximately 80% of the volatile compounds present in tomato according to Tikunov *et al.* 2005 and 2010. In this way, we tried to use the different metabolic profiling platforms as functional tools to explore *Lycopersicon* metabolomes. Together with further integration of genomic information we could compare germplasm and deduct characteristics underlying traits of interest.

These three analytical platforms can be used to obtain a broad overview of the metabolic composition of a given sample. A general procedure for the analysis of the samples apart from the use of each platform is exemplified in Figure 4. In all cases, quality control samples were included to evaluate the stability and accuracy of the measurements. Each platform required the application of different parameters for each process. Therefore, knowledge and expert advice was taken into account to choose the best options.

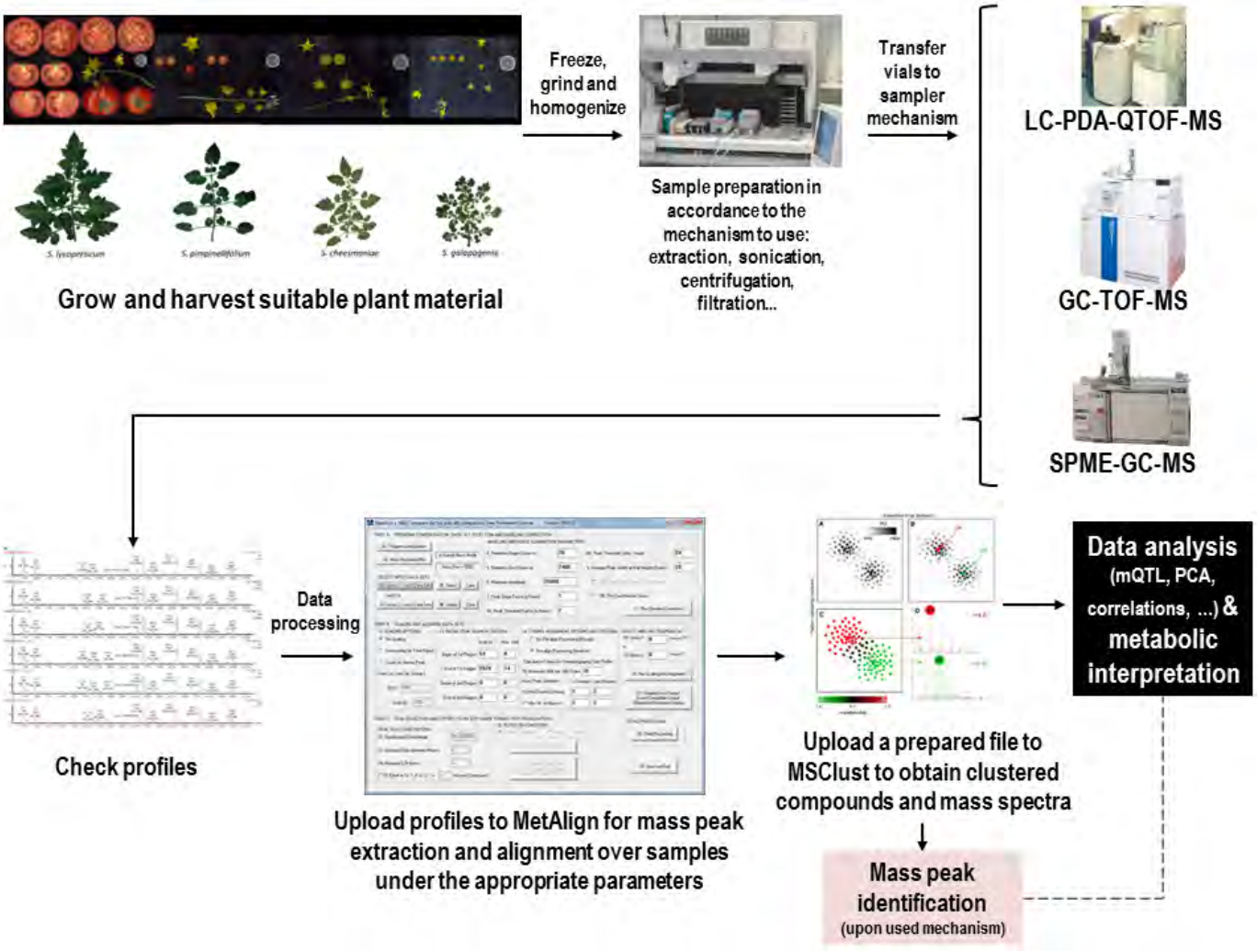


Figure 4. Schematic overview of the followed procedure to perform for untargeted metabolic profiling of plant materials (Adapted from de Vos, 2007).

Aim and outline of the thesis project

High-throughput genomic data make it possible to elucidate patterns and genetic characteristics underlying traits. In this project we first explored genomic information of different tomato accessions and explored the integration between the genomics and metabolic profiling platforms, particularly between *S. lycopersicum* and its wild relative *S. pimpinellifolium*.

In **Chapter 2** we described a creative way to use the genotyping data from a custom made Infinium bead array. We evaluated different tomato varieties to check the level of heterozygosity and introgressions. We discovered that cherry tomatoes were especially different from round/beef tomatoes on chromosomes 4, 5 and 12. We identified a set of 750

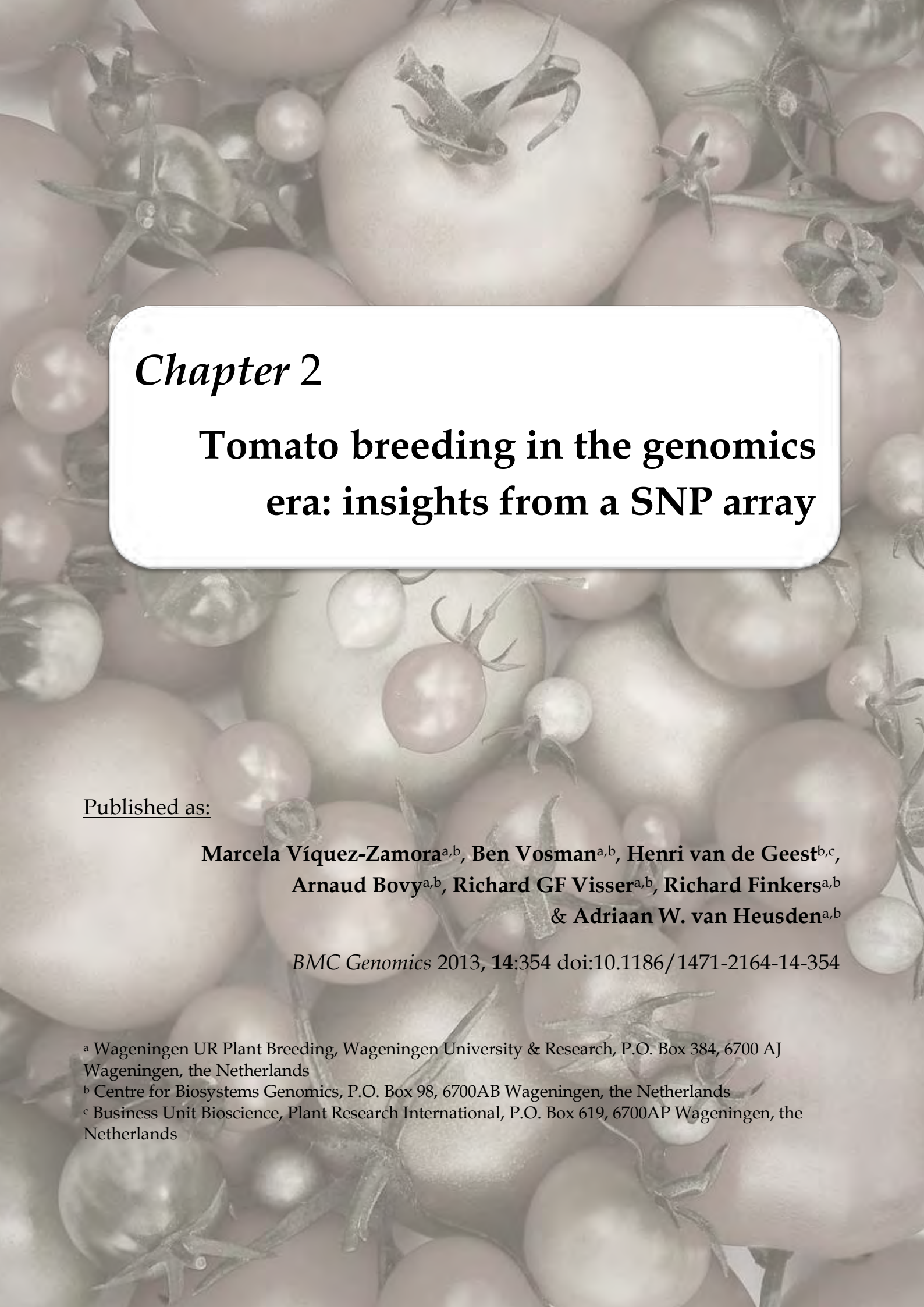
unique markers distinguishing *S. lycopersicum* 'Moneymaker' from all its distantly related wild relatives. Clustering and neighbour joining analysis among varieties and species showed expected grouping patterns, with *S. pimpinellifolium* as the most closely related to commercial tomatoes. However, we observed that the involved accessions are of influence in the clustering patterns of the *Lycopersicon* group.

For **Chapter 3** we used a RIL population between *Solanum lycopersicum* var. Moneymaker and *S. pimpinellifolium* G1.1554. The population was genotyped with the custom made Infinium bead array and, additionally, a subset of the lines was genotyped by sequencing (GBS). We were able to perform an *in silico* mapping improvement adding up the GBS data to the map developed with the array information and we confirmed genetic loci underlying resistance to Tomato Yellow Leaf Curl Virus (TYLCV). Furthermore we explored the population using untargeted metabolic profiling of leaves and checked for possible metabolites correlated with the resistance.

In **Chapter 4** we used the same RIL population but now targeted at fruits and we coupled the data with an untargeted metabolic profiling and genotyping of *S. pimpinellifolium*, *S. cheesmaniae* and *S. galapagense* accessions. This chapter will help breeders to profit from the analysed material to find interesting candidates for their breeding programmes.

In **Chapter 5** we analysed the effect of ripening on the metabolic composition of *S. pimpinellifolium* accessions. We followed the accumulation or decrease of several compounds and their relation to specific pathways and we targeted accessions that might support breeding for quality programmes.

Finally, **Chapter 6** is a general discussion aimed at conciliating the outcomes from the different chapters and integrating the results. We discuss the implications of what we explored and the overall prospects of breeding in relation to our findings.



Chapter 2

Tomato breeding in the genomics era: insights from a SNP array

Published as:

Marcela Víquez-Zamora[a,b], **Ben Vosman**[a,b], **Henri van de Geest**[b,c], **Arnaud Bovy**[a,b], **Richard GF Visser**[a,b], **Richard Finkers**[a,b] & **Adriaan W. van Heusden**[a,b]

BMC Genomics 2013, **14**:354 doi:10.1186/1471-2164-14-354

[a] Wageningen UR Plant Breeding, Wageningen University & Research, P.O. Box 384, 6700 AJ Wageningen, the Netherlands
[b] Centre for Biosystems Genomics, P.O. Box 98, 6700AB Wageningen, the Netherlands
[c] Business Unit Bioscience, Plant Research International, P.O. Box 619, 6700AP Wageningen, the Netherlands

Abstract

Background

The major bottle neck in genetic and linkage studies in tomato has been the lack of a sufficient number of molecular markers. This has radically changed with the application of next generation sequencing and high throughput genotyping. A set of 6000 SNPs was identified and 5528 of them were used to evaluate tomato germplasm at the level of species, varieties and segregating populations.

Results

From the 5528 SNPs, 1980 originated from 454-sequencing, 3495 from Illumina Solexa sequencing and 53 were known markers. Genotyping different tomato samples allowed the evaluation of the level of heterozygosity and introgressions among commercial varieties. Cherry tomatoes were especially different from round/beefs in chromosomes 4, 5 and 12. We were able to identify a set of 750 unique markers distinguishing *S. lycopersicum* 'Moneymaker' from all its distantly related wild relatives. Clustering and neighbour joining analysis among varieties and species showed expected grouping patterns, with *S. pimpinellifolium* as the most closely related to commercial tomatoes.

Conclusions

Our results show that a SNP search in only a few breeding lines already provides generally applicable markers in tomato and its wild relatives. It also shows that the Illumina bead array generated data are highly reproducible. Our SNPs can roughly be divided in two categories: SNPs of which both forms are present in the wild relatives and in domesticated tomatoes (originating from common ancestors) and SNPs unique for the domesticated tomato (originating from after the domestication event). The SNPs can be used for genotyping, identification of varieties, comparison of genetic and physical linkage maps and to confirm (phylogenetic) relations. In the SNPs used for the array there is hardly any overlap with the SolCAP array and it is strongly recommended to combine both SNP sets and to select a core collection of robust SNPs completely covering the entire tomato genome.

Keywords: Single Nucleotide Polymorphisms (SNP); Custom made infinium array; Tomato wild relatives

Background

Landraces and wild relatives constitute a vast genetic resource that can be tapped to introduce novel traits into tomato breeding programmes (Miller and Tanskley, 1990). During the last decades, the focus has mainly been on the introduction of disease resistance genes. But, within the breeding efforts, the lack of sufficient molecular markers in tomato has been a bottle neck in genetic and linkage studies. Although all known marker systems have been applied in tomato, most of them fall short in the genomics area mostly because they are too laborious and too low throughput (Agarwal *et al.*, 2008). These shortcomings are now being overcome by next generation sequencing projects and Single Nucleotide Polymorphisms (SNPs) identification (Tang *et al.*, 2006). The importance of SNPs as bi-allelic molecular markers is now widely recognized and their use is rapidly increasing (Labate and Baldo, 2005; Anithakumari *et al.*, 2010), since they have the advantage of being locus specific markers that can be scored co-dominantly in a flexible way. Technology has been developed for scoring single SNPs in thousands of different samples, all the way up to scoring millions of SNPs in a single sample (Shirasawa *et al.*, 2010).

Currently, the most widely used systems for high throughput SNP genotyping are the Illumina GoldenGate™, Infinium™ arrays and the KBioscience Competitive Allele-Specific PCR genotyping system (KASPar: www.kbioscience.co.uk) (Gunderson *et al.*, 2006; Steemers and Gunderson, 2007 Appleby *et al.*, 2009). The evolution of genotyping technologies has resulted in unprecedented possibilities for evaluating germplasm collections, characterizing populations, and finding markers linked to specific alleles of important genes. SNPs are also markers of choice for studying evolutionary processes (van Tienderen *et al.*, 2002). Characterization of a large set of tomato varieties with a large number of markers can show the impact of breeding on the molecular level and the extent to which these markers are useful for variety identification (Bredemeijer *et al.*, 1998 and 2002).

A whole genome tomato genotyping array (custom made) using the Illumina® Infinium Beadarray technology (Gunderson *et al.*, 2006 and 2009; Illumina®, 2009: www.illumina.com) was constructed to generate a multiplexing platform to analyse tomato germplasm. A set of 5528 SNPs was used to evaluate more than a thousand tomato samples. This enabled us to compare data at the level of species, varieties and segregating populations. Within the Solanaceae Coordinated Agricultural Project (SolCAP: http://solcap.msu.edu/), in 2012 Sim *et al.* also developed a genotyping array. However, they focused on different applications and, as we found out, with almost 100% different markers. We were interested in the question to what extent our SNP collection, which is based on a limited number of genotypes, can be

applied for variety identification, phylogenetic analysis, genetic mapping, evaluation of introgressions and germplasm identification.

Results

SNPs evaluation and distribution

A set of 5528 SNP oligos (92%) passed the quality check of Illumina. From those oligos, 1980 originated from 454-sequencing, 3495 from Illumina Solexa and 53 from other studies (Table 1).

As the SNPs were chosen before the tomato genome was publicly released (version 2.1) it was not completely clear how the markers would be distributed over the tomato chromosomes and what the marker density would be. Later all markers were assigned to their chromosomal position once the genome sequence (version SL2.30) was available and a good coverage and distribution of the markers over the physical map was observed (Figure 1). Some markers could not be placed on the genome and were placed on a pseudo molecule called chromosome 0. On chromosome 2, all markers are on the long arm because the short arm contains almost exclusively highly repetitive rDNA sequences (Tomato Genome Consortium, 2012). Overall, the data quality was very good: The variety Heinz was used as control on each microtiter plate (12) and of the 66120 data points scored for this cultivar only 145 were deviating (0.2%) and in most cases this was due to no calls (NC).

However, approximately 10% of the markers could not be reliably scored mainly because of wrong automatic clustering by the GenomeStudio software. Closely linked markers in segregating populations can be used to find the correct score and the reasons for the mistakes in the automatic clustering (Additional file 1). Six percent of the SNPs resulted in NCs. These markers were removed resulting in 4072 SNPs for further analysis. Forty eight percent of the monomorphic markers within *S. lycopersicum* still were useful because they were polymorphic within tomato wild relatives or between wild relatives and cultivated tomato.

Table 1. Validated SNPs and their distribution over the chromosomes. A full list of the SNPs can be found in Additional file3.

Chromosome	454-seq on cDNA from breeding lines	Illumina Solexa on gDNA from breeding lines	Illumina Solexa on gDNA from introgression free varieties	Markers from previous analysis	Total SNPs per chromosome
1	195	45	103	4	347
2	183	38	47	6	274
3	94	10	136	0	240
4	244	149	359	8	760
5	138	460	97	2	697
6	375	349	183	2	909
7	106	54	43	0	203
8	87	28	30	9	154
9	299	151	32	12	494
10	33	36	519	0	588
11	104	57	41	3	205
12	107	249	225	6	587
Unknown position	15	26	28	1	70
Total	1980	1652	1843	53	5528

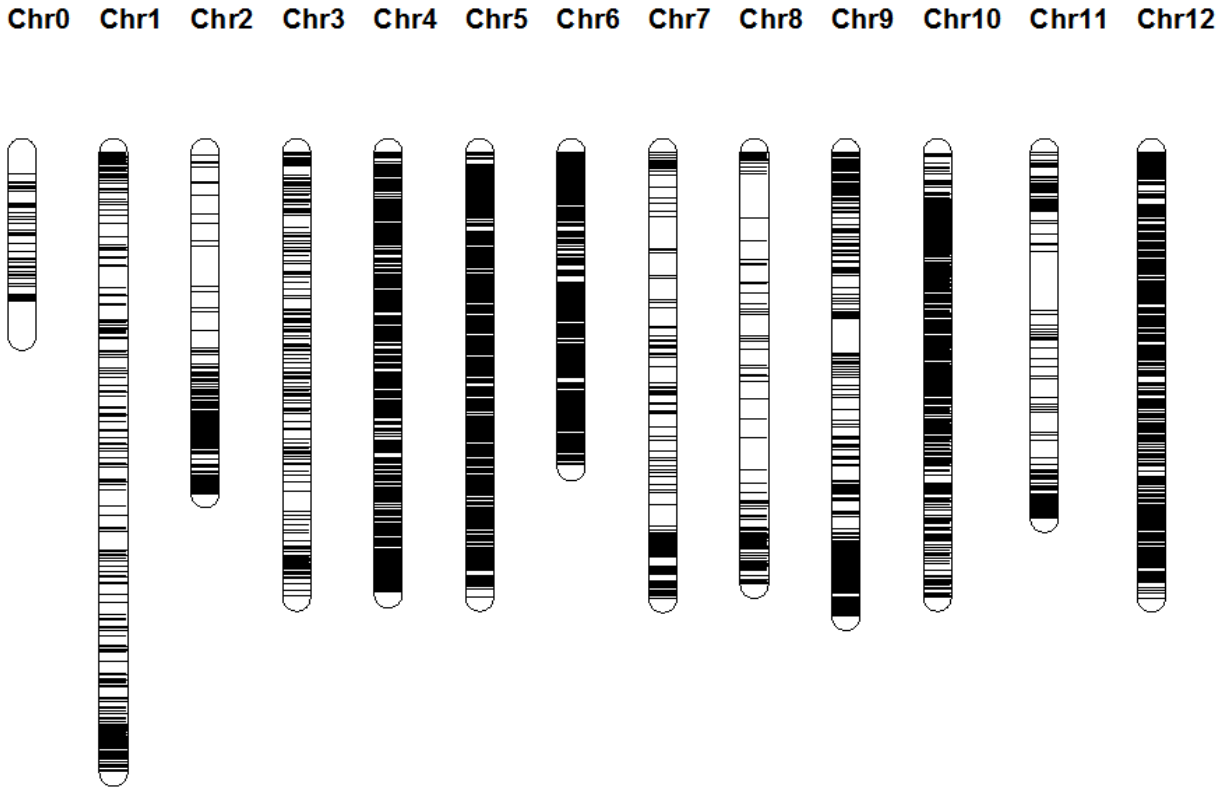


Figure 1. Distribution of the SNP markers along the genome. Physical positions according to the genome version in the SL2.30 version of the published tomato genome under the International Tomato Genome Sequencing Project (The Tomato Genome Consortium, 2012).

Constructing genetic maps

The SNPs were used to construct genetic maps in seven different mapping populations. For all of them the expected 12 linkage groups were found with most markers in the order as expected based on the tomato sequence (results not shown). Common ancestry of two parental lines resulted in regions without polymorphisms. Figure 2 shows an example of a linkage group created from the few SNPs showing recombination. On chromosome 8 of an F2 population between two cherry tomato breeding lines only 13 polymorphic markers were found. Although the genetic map still spanned 61.2 cM the physical map showed that only a small part of the chromosome is covered by the 13 markers, apparently this part of 5 Mbp has a high recombination frequency (Figure 2).

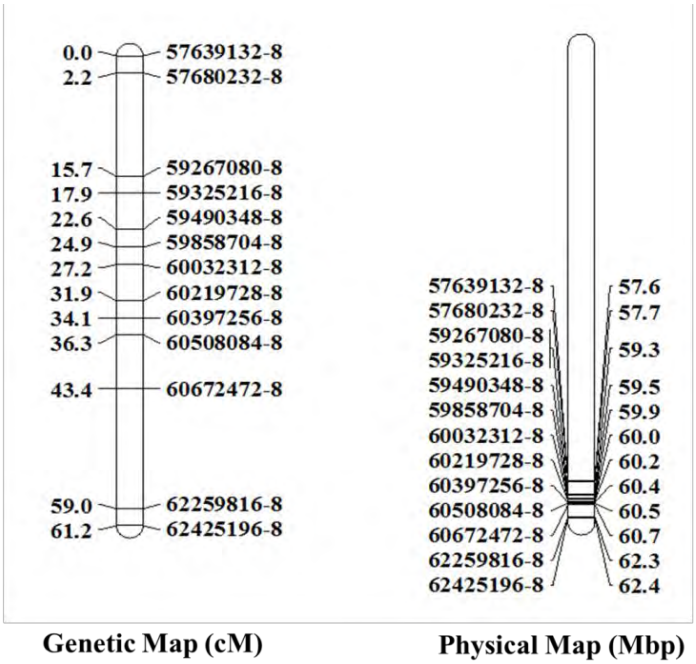


Figure 2. Comparison of genetic linkage map (centimorgans) with a physical map (megabase pairs on the right) of chromosome 8 for an F2 population between two cherry type tomatoes. Markers are indicated by the numbers ending on -8.

Variation among varieties

With the SNP markers we analysed 93 varieties plus some introgression free and other reference varieties (Additional file 2: Table S1). All varieties could be distinguished, although some were almost identical (Figure 3; Additional file 3 with the genotyping data for all materials used). Only the varieties Moneymaker and Moneyberg were completely identical. The percentage informative SNPs differed between the varieties.

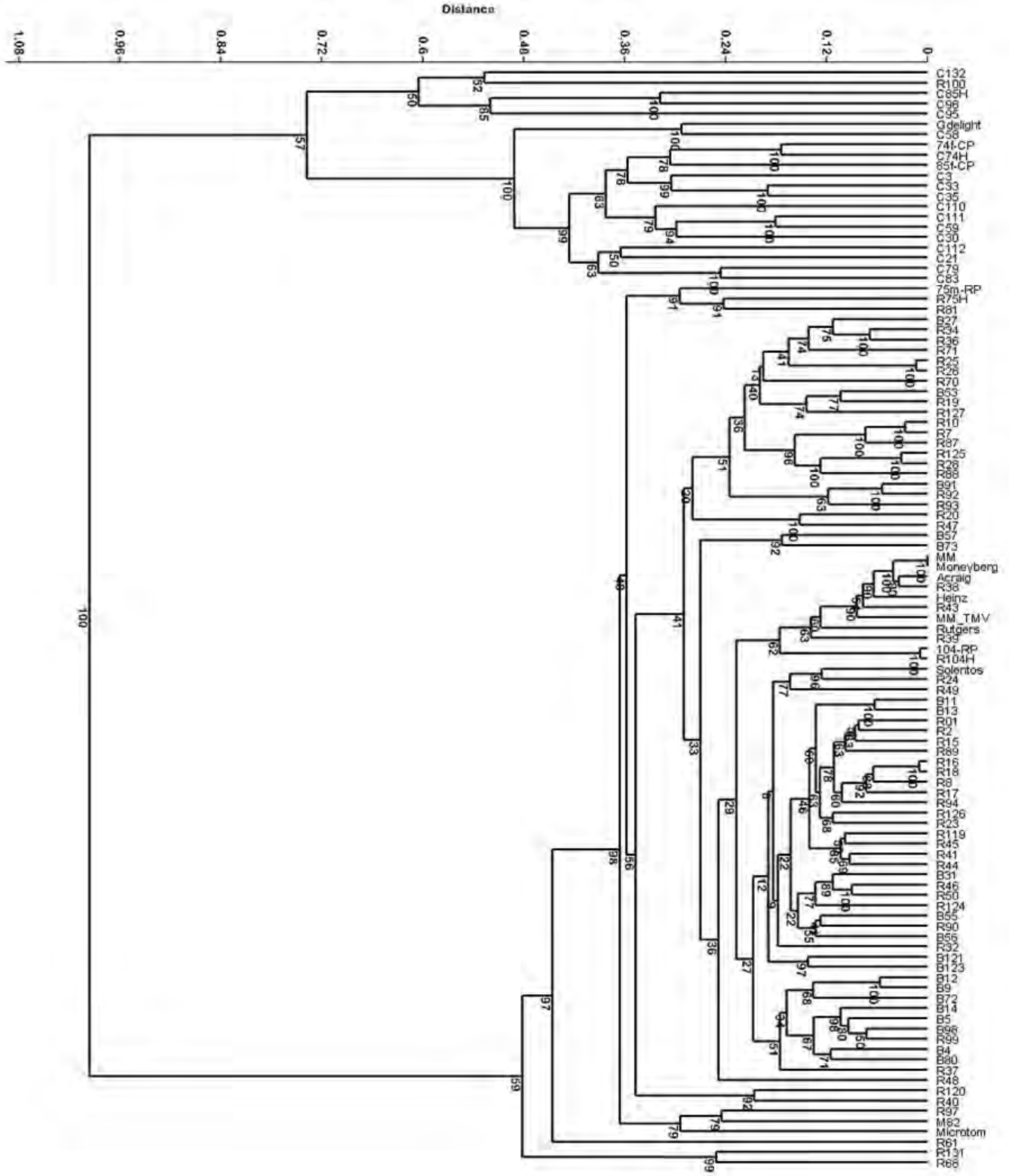


Figure 3. Cluster analysis of commercial hybrids and tomato lines. The Jukes-Cantor similarity measure with 1000 bootstraps was used.

When we compare Moneymaker and R38 we found about 5% polymorphic SNPs, between Moneymaker and R68 this was about 37%. The overall level of heterozygous markers within the varieties ranged from zero to almost 45% (Figure 4). The varieties included round, beef, and cherry types. In the dendrogram based on the SNP markers (Figure 3) the cherry tomatoes were clearly separated from the round/beef group, which were intermingled. Only the hybrid R100, classified as round, was in the group of cherry type tomatoes. This variety turned out to be a plum type of tomato and was misclassified as round. To see which markers contributed most to the separation of round/beef and cherry, we selected the markers distinguishing at least 90% of the cherries from round and beef tomatoes. This resulted in a selection of 955 SNPs that covered small areas of chromosomes 1 and 2 and the major central parts of chromosomes 4, 5 and 12 (Figure 5). However, also on chromosomes 3 and 10 small groups of markers specific to cherry tomatoes were found (results not shown).

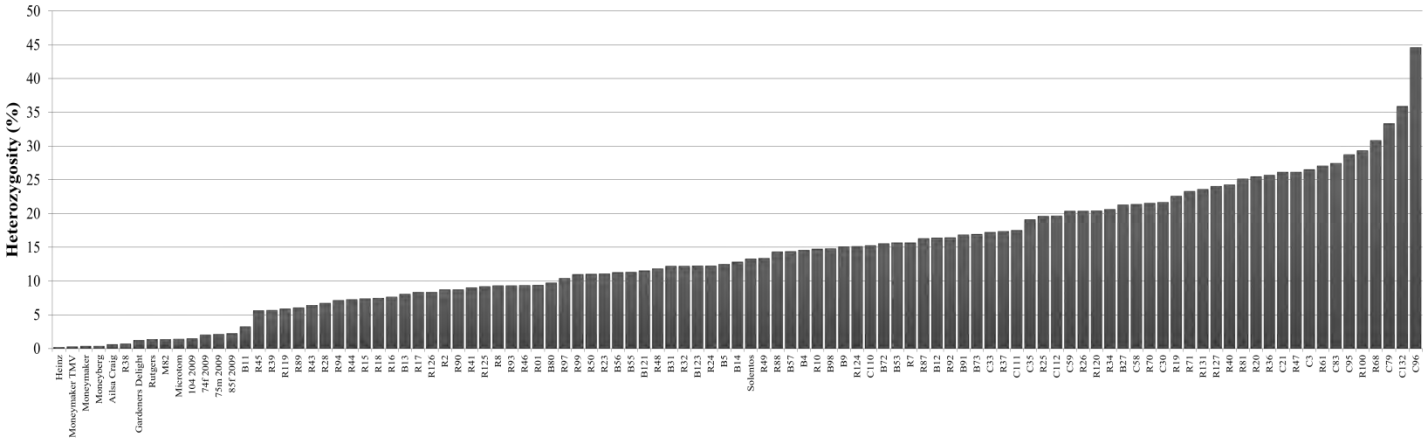


Figure 4. Percentage of heterozygous markers (among 4072) found in hybrids, introgression free varieties and known commercial lines.

Among the cherry varieties, Gardeners Delight did not have the cherry specific chromosome 5, and this variety has somewhat larger fruits than what we considered as cherry. In total four round tomatoes clearly had the cherry specific chromosome 5 (including R100), but after close inspection these were catalogued as deviating from round and more plum types.

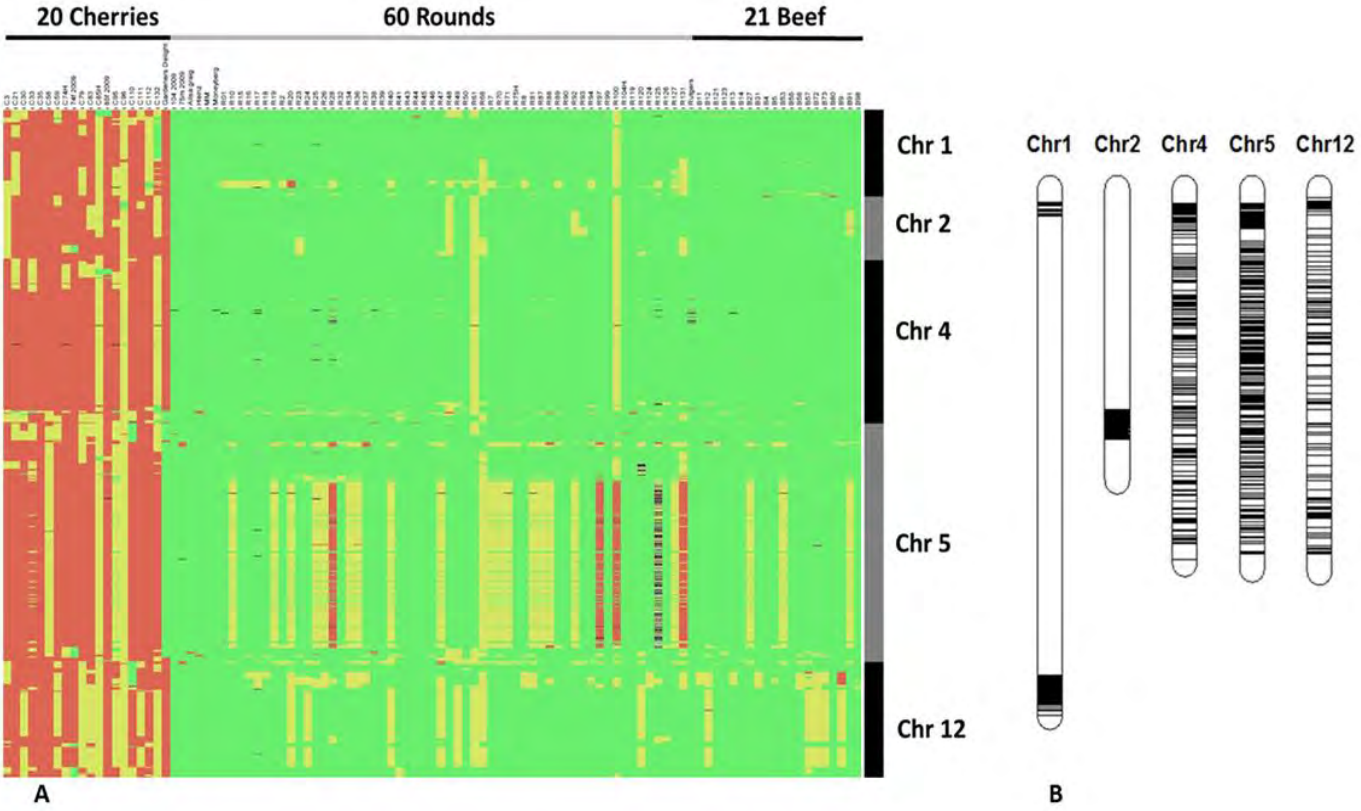


Figure 5. Graphical representation of 955 SNPs distinguishing round/beef from cherry tomatoes. Orange are cherry specific alleles, green round/beef alleles and yellow heterozygous calls (grey no-calls). B) Map-chart representation of the physical position of the 955 SNPs.

Identification of introgressions

Modern commercial varieties contain several introgressions from wild relatives. Most of these introgressions contain resistance genes (Labate and Robertson, 2012). We analysed a subset of varieties with known introgressions in detail and compared them with introgression-free varieties (Additional file 2: Table 1). Markers directly linked to known resistance-genes were selected and evaluated. Three markers were used on chromosome 6, two linked markers and one marker within the dominant Mi-1.2 gene (ITAG2.3 'Release: genomic annotations' at www.solgenomics.net), conferring root-knot nematode resistance. In 79 varieties the genotype was identical to the introgression-free genotypes and 28 varieties had an introgression in this region (Figure 6). In 6 varieties the introgression was homozygous and in the 22 others heterozygous. Tobacco Mosaic Virus resistance (tm2 gene) is located on chromosome 9 and two markers were selected for this introgression, one corresponding to the gene and one to in the flanking region to confirm the introgression pattern. Figure 6 (and Additional file 4) shows that the region containing the *tm2* gene was

present in 91 varieties taking into account the heterozygous introgressions. One of those varieties was R75, for which the introgression was not reported (Additional file 2 Table 3). There were 16 varieties that lacked the introgression. Differences in size of the introgressions were observed based on the polymorphism frequency among varieties. These differences were also compared with varieties annotated as introgression free and corroborated with known introgression (Additional file 2).

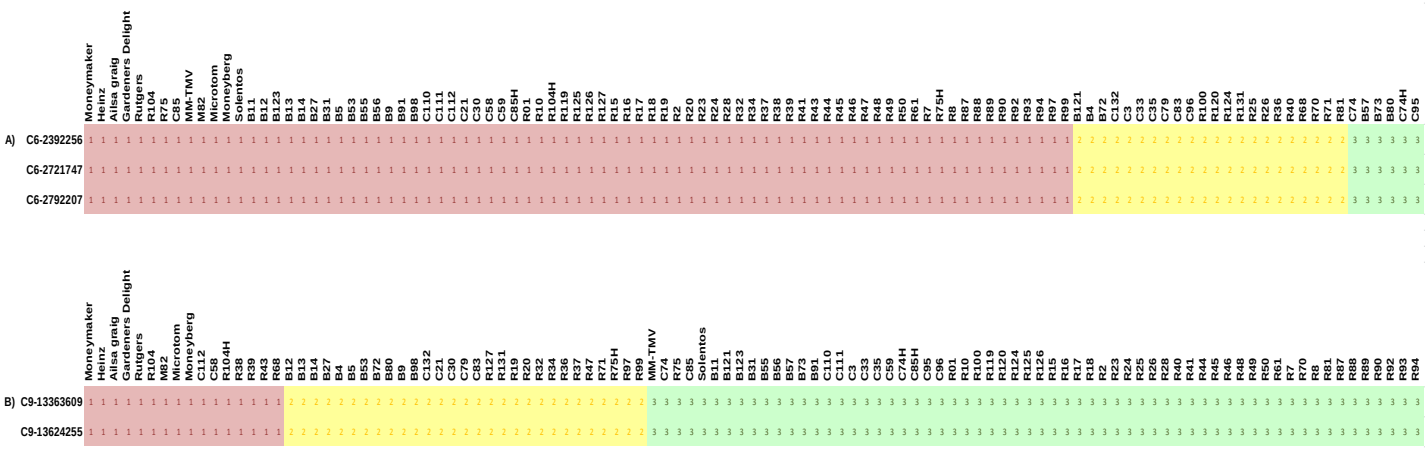


Figure 6. Heat map comparison of markers in the regions of known introgressions: **A)** Root-knot nematode resistance (Mi.1-2 gene) and **B)** Tobacco mosaic virus resistance (tm2 gene). Positions according to the ITAG2.30 'Release: genomic annotations' (www.solgenomics.net). Introgression-free varieties in light red (1), heterozygous introgressions in yellow (2) and homozygous introgression in green (3).

SNPs for interspecific crosses in tomato

Solanum lycopersicum cv. Moneymaker is the standard introgression free tomato used in our group to make mapping populations for breeding and genetic analysis. Therefore, we looked for markers that differentiate this *S. lycopersicum* cultivar from the majority of wild species accessions (Additional file 2: Table 2). A selection of 750 SNPs was polymorphic between Moneymaker and all screened accessions of the more distantly related wild relatives of tomato (*S. habrochaites* (2), *S. chmielewskii* (1), *S. neorickii* (2), *S. pennellii* (1), *S. arcanum* (2) and *S. chilense* (3)). Within this selection of markers, there were occasional (0.1%) non-polymorphisms with one or more of the 37 accessions of *S. cheesmaniae* and *S. galapagense* and a slight higher number of non-polymorphic cases (4%) within the 28 *S. pimpinellifolium* accessions (Figure 7).

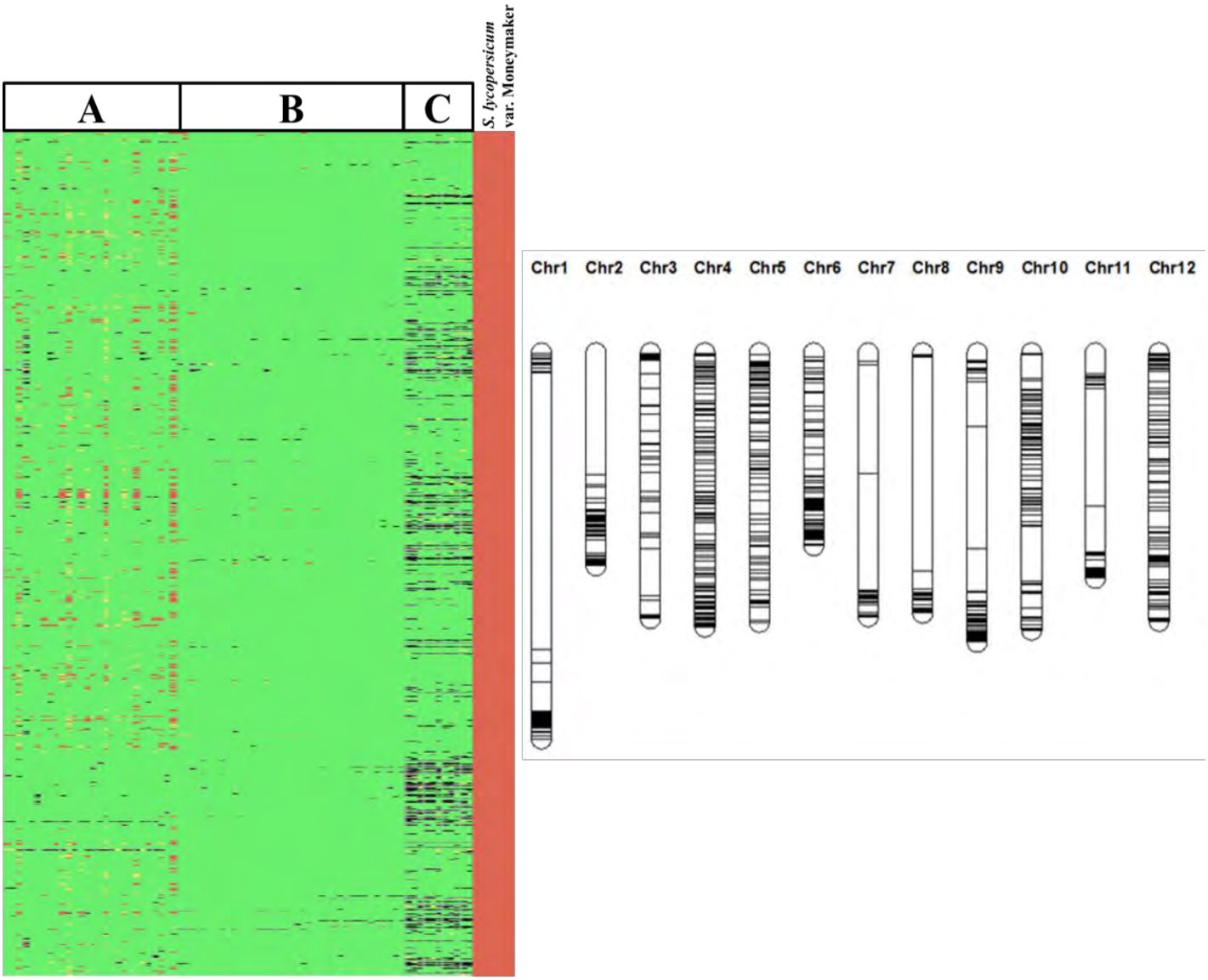


Figure 7. Heat map and distribution of 750 SNPs over the twelve chromosomes differentiating *S. lycopersicum* cv. Moneymaker and wild relatives. Genotype-calls: Red=[AA]; Yellow=[AB]; Green=[BB] and Black=no-call. A: 28 accessions from *S. pimpinellifolium*, B: 23 *S. cheesmaniae* and 14 *S. galapagense*. C: 2 *S. habrochaites*, 1 S.*chmielewskii*, 2 *S. neorickii*, 1 *S. pennellii*, 2 *S. arcanum* and 3 *S. chilense*.

SNPs among tomato species

Representative accessions of wild relatives of tomato were analysed with the SNP array to establish relationships in *Solanum* sect. *Lycopersicon*. The phenetic analysis was carried out using neighbour joining. The resulting tree is shown in Figure 8. These relations are in accordance with Rodriguez *et al.* (2009). A BioNJ tree (Gascuel, 1997) can be found as additional file 3. Figure 8 also shows that the number of NCs is becoming larger with increasing distance between the cultivated tomato and the wild relatives.

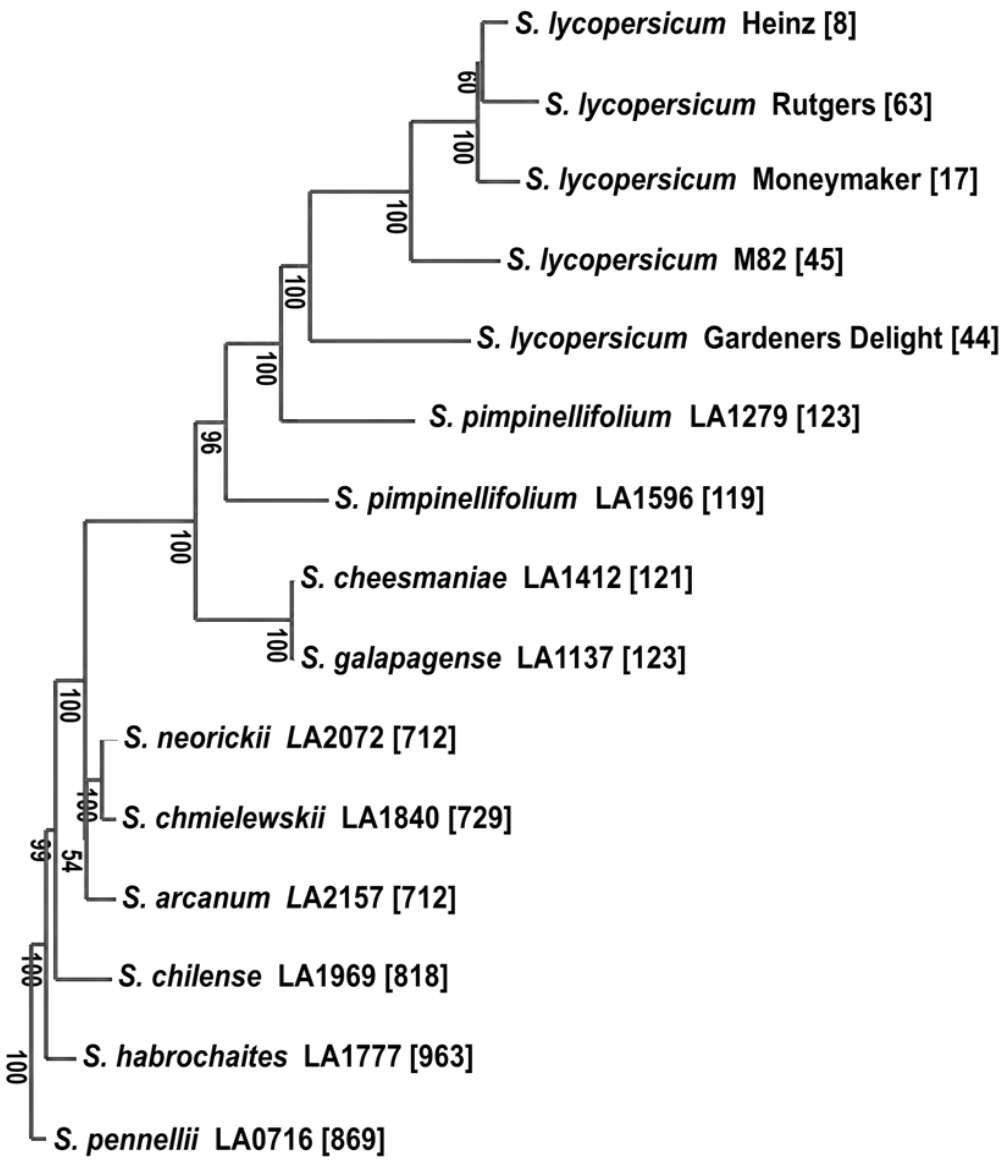


Figure 8. Neighbour joining analysis of representative samples of tomato species using 4072 SNPs. *Solanum pennellii* accession LA0716 was used as outgroup. Numbers at the nodes are bootstrap values for 1000 re-samplings. Numbers in brackets are the number of non-calls per genotype. Markers that were for more than 98% monomorphic, or had more than 25% heterozygous scores or more than 20% no-calls in the commercial hybrids or more than 50% in the wild relatives were removed from the dataset leaving a total of 4072 markers.

Accessions from *S. cheesmaniae* and *S. galapagense* clustered together (Figure 8). In our study only approximately 30 SNPs (from 5528) were found between the accessions of *S. cheesmaniae* and *S. galapagense* in spite of clear phenotypical differences in leaf structure and trichomes (Darwin *et al.*, 2003).

Discussion

Quality of SNP data

The high reproducibility of the results for the 12 Heinz samples shows the robustness of the data obtained with the Infinium array. This was also evident from the comparison between cv Moneymaker and cv Moneyberg where the only differences were a few NCs. Although the data were of high quality, individual SNP calls can be wrong. Wrong calls can be recognized in dense genetic linkage maps of a species from which the sequence is known. We observed errors in 10% of the SNPs, when using the standard settings of the Illumina Genome Studio software. Such errors can be corrected manually or 10% of the SNPs can be deleted (Steemers and Gunderson, 2007). Since the amount of data is vast, enough data remained after deleting 10% of the SNPs. Reasons for errors can be DNA quality, presence of outliers (Additional file 1) within the germplasm and, in few cases, double signalling due to duplications in the genome.

General applicability of the SNPs

Even though the SNPs were looked for in a limited number (4) of breeding lines of *S. lycopersicum* in combination with four introgression free varieties, they were polymorphic enough in the *Solanum* sect. *Lycopersicon* germplasm to discriminate varieties and species as well as to confirm phenetic relations. This implies that many of the SNPs originated from the time before domestication (Morin *et al.*, 2004; Street *et al.*, 2007; Peralta *et al.*, 2008).

The *S. lycopersicum* specific markers must have evolved after this species separated from the others. These markers will be polymorphic in any interspecific cross (Figure 7). A relatively cheap, SNP array with a limited number (as few as 20 per chromosome) of well distributed markers will be an excellent tool for a first fast characterization of any new interspecific mapping population involving *S. lycopersicum*. Based on our results such an array can be easily developed.

Furthermore, it is interesting to note that so many SNPs were found among the four breeding lines. This result was quite unexpected as *S. lycopersicum* is considered as a species with little genetic variation (Miller and Tanksley 1990).

The use of SNPs to improve the tomato genome sequence

One application of the SNP array was to compare genetic with physical positions when working with mapping populations. On the genetic linkage maps most of the markers were in the expected order; identical to the order in the assembled tomato sequence (Tomato Genome Consortium, 2012). This confirmed the accurateness of the assembled Tomato Genome.

Some unassigned markers could be mapped to specific chromosomal positions in one or more of the linkage maps that we produced (results not shown). Comparison of genetic linkage maps and the physical linkage map also pointed out a misassembly on the long arm of chromosome 12 (between 48.8 Mbp and 61.7 Mbp; Additional file 4) in version 2.4 of the tomato genome. Also the data published by Sim *et al.* (2012) suggest a disruption of marker order in the same region (see their Figure 3), but the conclusion that this might be due to a misassembly was not drawn. Markers should be used to genetically validate and further improve *de novo* genome assemblies.

Variety identification

Several DNA profiling techniques have been used for variety identification (Cooke, 1999). For tomato, one of the most extensive studies was done by Bredemeijer *et al.* in 2002 using simple sequence repeats (SSRs). They showed that 90% of the more than 500 varieties that were genotyped had a unique SSR profile using 20 markers (on average this is less than 2 markers per chromosome and one chromosome was even without markers). The SNP array covered between 150 and 900 markers per chromosome and all varieties could be distinguished, except the varieties Moneyberg and Moneymaker. That these two showed identical profiles means that they are highly related, if not identical. Both have been registered by the International Union for the Protection of New Varieties of Plants in the National Listing in Great Britain (UPOV: http://www.upov.int), so phenotypic differences must have been seen. Under the UPOV act of 1991 such varieties would likely be considered as essential derived varieties (UPOV 1991; Vosman *et al.*, 2004). The SNP markers developed

in our study will be very useful for establishing whether varieties are essentially derived from other varieties using the protocol developed for lettuce (Van Eeuwijk and Law, 2004).

The trend to exploit genes from tomato wild relatives for specific traits enlarges the variation in cultivated tomato and the differences among varieties (Labate and Robertson, 2012). Such introgressions can easily be detected using the SNP array as we have shown for the Mi1.2 and TMV gene. When gene-specific (or closely linked) SNP markers are used, genotyping may substitute phenotypic assays even in variety registration as was demonstrated by Arens *et al.* in 2010. The markers also allowed us to determine the level of heterozygous markers in present day varieties, which varied between zero and almost 45%. It is interesting to see that the highest numbers are found for some of the plum/cherry tomatoes. This is most likely because they are hybrids between round and cherry tomatoes and the 955 cherry specific SNPs will contribute to the large number of heterozygous markers (Figure 4 and 5). The high throughput SNP marker determination can be carried out at relatively low cost and is less laborious than other methods used. Therefore it is likely that SNP markers will be the markers of choice for variety identification and registration in future. However it may be anticipated that the SNP arrays will soon be replaced by complete sequencing.

Differences between round/beef and cherry tomatoes

Many of the polymorphisms located on chromosomes 4, 5 and 12 were between round/beef and cherry tomatoes. This suggests that regions on these chromosomes are essential to get the full cherry tomato phenotype and that there is selection for these regions in breeding programs for cherry tomatoes. The fact that whole chromosomes (4, 5 and 12) seem to be involved is possibly due to suppression of recombination in the large pericentromeric regions (Sherman and Stack, 1992; Stack *et al.*, 2009). This is not the case on chromosome 1 where the cherry region is a hotspot of recombination as shown in a RIL population of *S. lycopersicum* and *S. pimpinellifolium* (unpublished observations by the authors).

Cherry type tomatoes have more SNPs in common with *S. pimpinellifolium* accessions than the round/beef varieties indicating that cherry tomatoes are closer to this wild relative than round and beef commercial lines. The varieties chosen for SNP selection might have been the reason that so many cherry specific markers were found. The SolCAP array also revealed different patterns of genetic variation particularly for chromosomes 2, 4, 5, 6 and 11. For chromosome 4 and 5 this is probably also due to the cherry round differences we observed. In general, relatively little is known about genomic regions distinguishing cultivated tomato gene pools (Sim *et al.*, 2012b).

Some regions are known to contain genes/QTLs that are related to differences between cherry and round. For instance a QTL for fruit weight and soluble solids content, is found on chromosome 2, QTLs for yield, brix, fruit weight, fruit shape, colour and epidermal reticulation have been mapped on chromosome 4 (Monforte *et al.*, 2001). Chromosome 5 is known to harbour QTLs for fruit colour and QTLs for viscosity traits related to total red yield and pH in chromosome 12 are known (Tanksley *et al.*, 1996; Bernacchi *et al.*, 1998).

SNPs in Solanum ***sect.*** Lycopersicon

Our SNP based phenetic trees were comparable to the ones made by Bretó *et al.* in 1993 using isozymes, Palmer & Zamir in 1982 and Spooner *et al.* in 1993 with chloroplast DNA, McClean & Hanson in 1986 with mitochondrial DNA, Miller & Tanksley (1990) with genomic DNA, Marshall *et al.* in 2001 with internal transcribed spacer (ITS) region of nuclear ribosomal DNA sequences and, also Alvarez *et al.* in 2001 with microsatellite markers. Peralta *et al.* (2008) performed the most extensive taxonomic study of tomato and its wild relatives and our results confirm their findings.

In our analysis we found *S. pimpinellifolium* as the closest wild relative to *S. lycopersicum,* which is similar to observations made by Grandillo *et al.* (2011) and The Tomato Genome Consortium in 2012. The cherry tomato is considered either as a domesticated group or as an admixture of *S. pimpinellifolium* and *S. lycopersicum* (Ranc *et al.* 2008). *S. cheesmaniae* and *S. galapagense* are also very closely related to the domesticated tomato. Introgressions in the cultivated germplasm can affect the similarity weight in the relationships between *S. pimpinellifolium, S. galapagense* and *S. cheesmaniae* on one hand and *S. lycopersicum* hybrids on the other hand. For phylogenetic studies it is important to define the initial germplasm and its characteristics. In the case of *S. habrochaites* and *S. pennellii* the increased number of NCs decreased the resolution.

Prospects of SNP data in tomato

For our custom made array, the SNP selection was based on commercial breeding lines. Sim *et al.* (2012 and 2012b) developed a large SNP genotyping array using commercial varieties. To evaluate if the same SNPs were present, the precise SNP positions from both arrays were compared (allowing a window of ± 3 base pairs). Only 98 SNPs, less than 2% of our SNPs were found in the exact same position or within the allowed window. This means that there is still a large number of SNPs to be discovered in tomato. For further comparisons among

the two arrays we made the SNPs including the flanking sequences available at: http://www.plantbreeding.wur.nl/Publications/SNP/4072SNP-Sequences.xlsx.

Conclusions

Our results show that an SNP search in only a few breeding lines permitted the development of markers generally applicable in tomato and its wild relatives and furthermore that the Illumina bead array generated highly reproducible data. Our SNPs can be roughly divided in two categories: SNPs of which both forms are present in the wild relatives and in domesticated tomatoes and SNPs unique for the domesticated tomato. The SNPs can be used for genotyping, identification of varieties, comparison of genetic and physical linkage maps and to confirm phylogenetic relations. There is hardly any overlap with the SolCAP array and we suggest to combine both SNP sets and to select a core collection of robust SNPs completely covering the tomato genome for the development of future arrays.

Methods

Plant material

Tomato germplasm was obtained from the collection of Wageningen UR Plant Breeding, The Netherlands: the Tomato Genetics Resource Center (TGRC) at University of California, Davis; the Centre for Genetic Resources (CGN), The Netherlands; and from the breeding companies Monsanto, RijkZwaan, Takii, Vilmorin & Cie (VCo), ENZA and Syngenta. The evaluated material included hybrid varieties of the project within the Centre of Biosystems Genomics (CBSG: www.cbsg.nl). Based on QTL model predictions, four breeding lines were chosen to obtain a large diversity in taste related characteristics (van Berloo *et al.*, 2008). A half diallel was made with the four breeding lines resulting in six segregating populations. The parents were C74 (cherry, orange), C85 (cherry, red), R75 (round, yellow), and R104 (round, red). Further material included landraces, hybrids, commercial varieties, accessions of tomato wild relatives and mapping populations (Additional file 5 Table 1). The genotyping results of the varieties with the used SNPs can be found in Additional file 3.

DNA and RNA extraction

Genomic DNA from young leaflets was extracted following a CTAB based protocol (Steward and Via, 1993; Kabelka *et al.*, 2002) adjusted for high throughput isolation. Two young leaflets were ground with a Retsch 300 mm shaker (Retsch BV, Ochten, The Netherlands) using 1 ml micronic tubes (Micronic BV, Lelystad, The Netherlands). The DNA pellets were washed in 76% EtOH with 10mM NH_4Ac before re-suspending the DNA in TE buffer.

Total RNA was isolated using TRIzol reagent (Simms *et al.*, 1993) according to the manufacturer's instructions (Roche, Switzerland) and finally treated with DNaseI (Invitrogen).

SNPs identified through Roche/454-sequencing

Total RNA was isolated from the four chosen breeding lines (C74, C85, R75 and R104), and at Vertis Biotechnologie AG (Freising, Germany: http://www.vertis-biotech.com/) cDNA was made. The 454 Sequencing gave 1.3 $x10^6$ reads of a median length of 400 base pairs. The reads were aligned to the tomato genome (v2.10) and SNPs were called using QualitySNP^{ng} (Tang *et al.*, 2006) after being adapted for large numbers of reads (Nijveen *et al.*, 2013). After the Tomato v2.30 was available the SNP positions were renamed based on this version.

SNPs identified through Illumina/Solexa-sequencing

A potential risk with the four breeding lines was that primarily interspecific SNPs would be found due to introgressed regions originating from tomato wild relatives (Additional file 5 Table 3). To include additional intraspecific (*S. lycopersicum*) variation four introgression free varieties were also included in the Illumina/Solexa sequencing. To reduce the complexity, genomic DNA (gDNA) of the eight different samples (C74, C85, R75, R104 and the introgression free varieties, Ailsa Craig/round, Rutgers/beef and Gardeners Delight/cherry plus the reference line Heinz/round) was digested with restriction enzyme *Mbo*I (four cutter) and the 400-600 bp fraction was cut out of a 1.5 % agarose gel and purified. Theoretically, this should result in a coverage of at least 23x per fragment. After Illumina sequencing 15 x 10^6 fragments were blasted against the Heinz v2.10 contigs and compared. The Illumina reads of 72 basepairs were aligned with the software tool Bowtie (>95% similarity) (Langmead *et al.* 2009). After alignment SNPs were called with VarScan (variant detection in massively parallel sequencing data) (Koboldt *et al.* 2009). All SNPs with a minimal coverage of three in a genotype were listed in Excell. A SNP was called when it was present in at least six reads in one genotype and six reads in another genotype.

Allocation of SNPs

Putative SNPs and their flanking regions were blasted against the then available contig sequences of tomato (Tomato WGS contigs v2.10) in order to choose SNPs as dispersed over the genome, when possible at least one SNP per contig. Later the availability of the tomato genome sequence (Tomato WGS chromosomes v2.3) allowed us to assign the SNPs to their physical location. A total of 6000 SNPs with two times 50 bp flanking sequences of Heinz were used for designing the oligo's for the Illumina beadarray (Illumina, 2009). After the oligo's were synthesized, ~8% of them did not comply to the quality standards set by Illumina and were discarded leaving 5528 SNP markers per array.

Illumina® Infinium Bead Array analysis

Solanum sp. DNA samples with a concentration of 50 ng/µl were sent to ServiceXS, Leiden, The Netherlands, where 4µl was processed according to the Infinium HD Ultra Assay protocol (Illumina®, 2009) and used for hybridization onto the BeadChip (ServiceXS, 2010).

Genotyping data processing

All the SNPs were named after their position on the SL2.30 version of the tomato genome sequence published online by the International Tomato Genome Sequencing Project (http://solgenomics.net/). This version contains approximately 85% of the tomato genome sequence. The lacking sequences are mostly highly repetitive or heterochromatic regions (The Tomato Genome Consortium, 2012).

The Genotyping Module 1.9.4 of the Illumina's software GenomeStudio® V2011.1 software package was used to analyse the genotyping results under default settings. The software assigned allele calls ('GeneCall') according to the intensity signals obtained, resulting in a [AA], [BB], [AB] or a non-call for each SNP. Advanced assembling within each correspondent analysis was performed and manual inspection and adjustment were performed in order to optimize call rates in the case of questionable SNPs. In particular those cases, and based on the knowledge on segregation patterns within the material, clustering errors were identified and amended (Teo *et al.*, 2007).

Before further analysis, markers that were more than 98% monomorphic, were removed, as well as markers with more than 25% heterozygosity in accessions or breeding lines. Finally, also markers with a large number of NCs were removed. Two thresholds for the percentage NCs were used: more than 20% NCs among the commercial hybrids and/or more than 50% among wild relatives.

When specific populations were evaluated, synchronization of parental lines together with the corresponding offspring was performed. This means that, for each analysis alleles were sorted according to the parent lines and replaced by a specific allele designation (A or B) for each parent.

Data Analysis

For cluster analysis the genotype calls were converted into numerical values: [AA]=1, [AB]=2, [BB]=3. Cluster analyses were done using the Jukes-Cantor similarity measure with 1000 bootstraps. Neighbour joining analysis using the Manhattan similarity measure with an out-group rooting and 1000 bootstraps was performed using the statistical package PAST version 2.12 (Hammer *et al.*, 2001). The BioNJ analysis was carried out using SplitsTree version 4.6 with 1000 bootstraps.

Data visualization heat maps were made in GeneMaths XT 2.12 (Applied Maths). Linkage maps were constructed using JoinMap® version 4.1 (Kyazma©; Van Ooijen, 2011). The default calculation parameters were adjusted to cope with the large number of markers. In the similarity thresholds the option 'show individual pairs with a similarity larger than' was decreased from 0.95 to 0.7. Recombination frequency was used as a grouping parameter and the linkage parameters were set to take all LOD values from 0 to 100. The 'Show strong linkages with a rec. freq. larger/smaller than' were set to 0.5/0. The number of maximum linkages to show per locus was set to 0. As algorithm we used the ML (Maximum Likelihood) mapping option, and within the map building, the spatial sampling thresholds were set one to 0.1 the first and the rest to 0. The 'Number of map optimization rounds per sample' was fixed to 1. Thereafter, linkage groups were compared with chromosomal distribution in the physical maps using MapChart 2.2 (Voorrips, 2002).

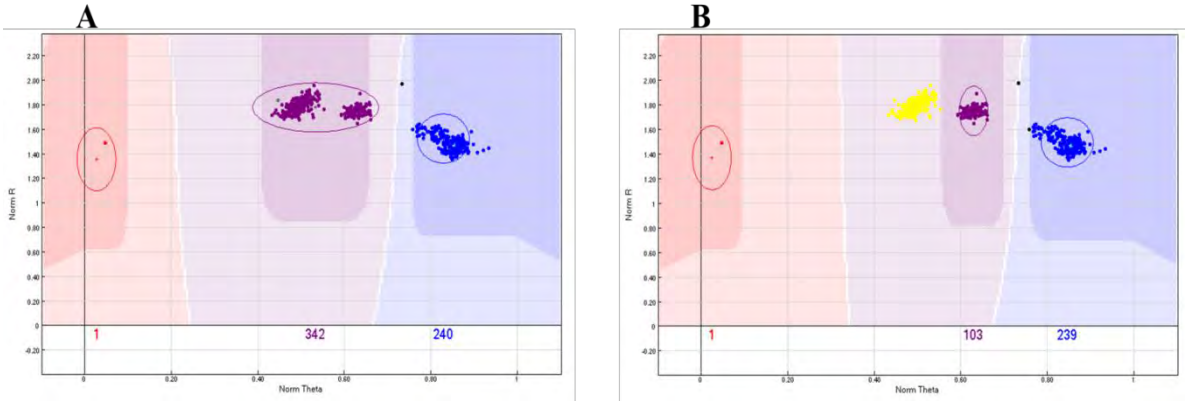
Acknowledgements

This project was carried out within the research programme of the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

We would like to acknowledge Fien Meijer-Dekens for collecting and maintaining all the tomato samples at WUR-Plant Breeding. The department of Bioinformatics of Wageningen UR for their support in the development of the SNP markers and in allocating them into the different versions of the Tomato Genome Sequence.

Additional Files

Additional file 1. Example of genotyping graphs in GenomeStudio®. SNP marker within one population in which two different groups were clustered automatically by the program in one group (the heterozygous group) due to an outlier sample (NTC). The right grouping is in figure 1B, this was confirmed by flanking markers in a segregating population. The red circle exemplifies an outlier sample.



Additional file 2: Table S1. Varieties from *S. lycopersicum* used for comparisons (also used by van Berloo, 2008).

Introgression free varieties	*Reference varieties*	*Breeding lines*	*Varieties Cherry*	*Varieties Round*		*Varieties Beef*
Ailsa Craig	Heinz	C74	C110	R01	R39	B11
Gardeners Delight	Moneymaker-TMV	C85	C111	R10	R40	B12
Rutgers	Moneyberg	R75	C112	R100	R41	B121
Moneymaker	Microtom	R104	C132	R104H	R43	B123
	M82		C21	R119	R44	B13
	Solentos		C3	R120	R45	B14
			C30	R124	R46	B27
			C33	R125	R47	B31
			C35	R126	R48	B4
			C58	R127	R49	B5
			C59	R131	R50	B53
			C74H	R15	R61	B55
			C79	R16	R68	B56
			C83	R17	R7	B57
			C85H	R18	R70	B72
			C95	R19	R71	B73
			C96	R2	R75H	B80
				R20	R8	B9
				R23	R81	B91
				R24	R87	B98
				R25	R88	
				R26	R89	
				R28	R90	
				R32	R92	
				R34	R93	
				R36	R94	
				R37	R97	
				R38	R99	

Additional file 2: Table S2. Accessions used in analysis

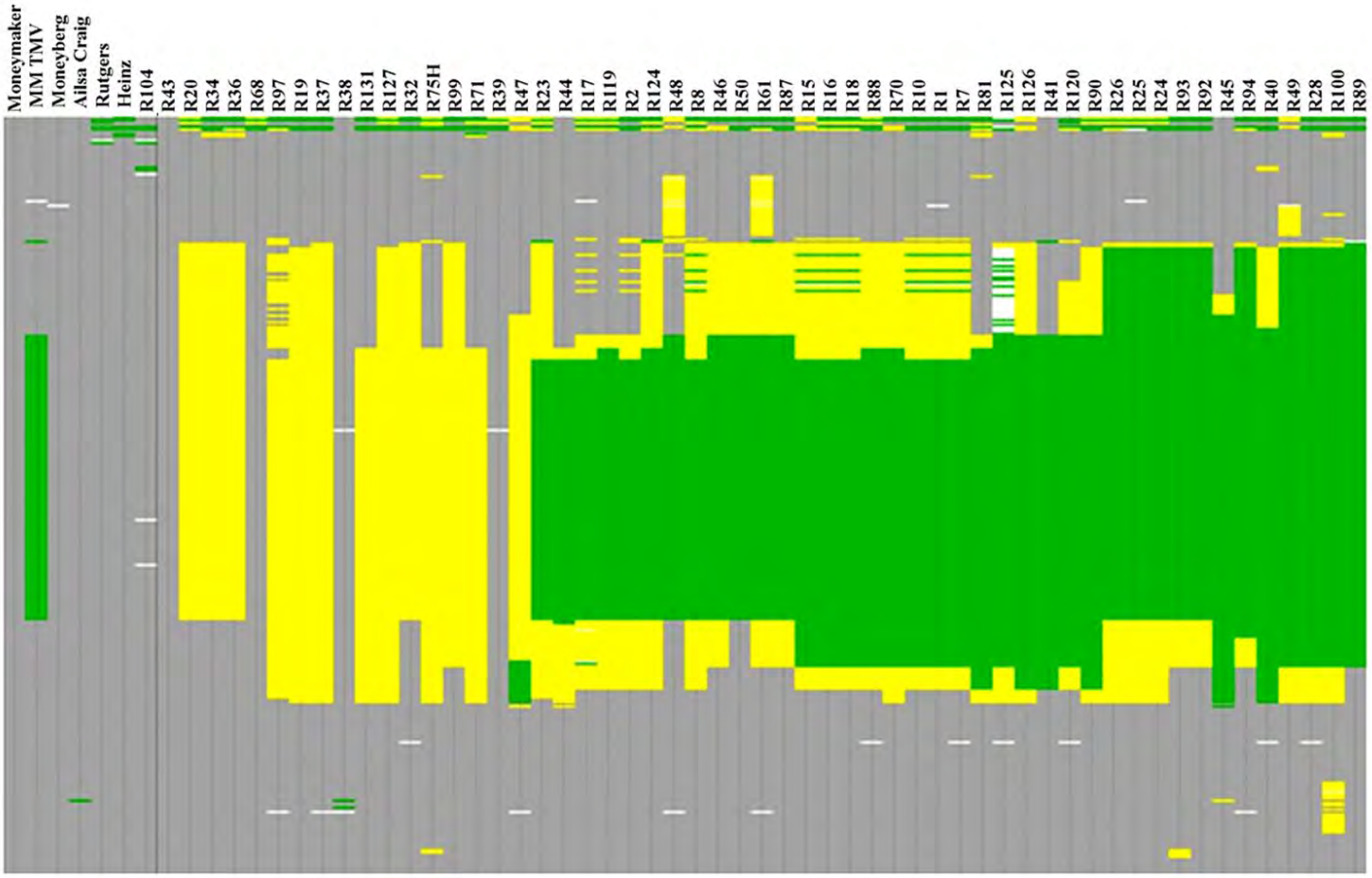
S. pimpinellifolium accessions	*S. cheesmaniae* accessions	*S. galapagense* accessions	Further distant accessions
G1.1554	LA0421	LA0480A	*S. habrochaites* LA1777
G1.1589	LA0422	LA0438	*S. habrochaites* Lyc4
LA1246	LA0428	LA0483	*S. chmielewskii* LA1840
LA1280	LA0437	LA0528	*S. neorickii* LA2072
LA1345	LA0521	LA0530	*S. neorickii* LA735
LA1349	LA0522	LA0532	*S. pennellii* LA716
LA1355	LA0524	LA0748	*S. arcanum* LA2172
LA1374	LA0529	LA0929	*S. arcanum* LA2157
LA1472	LA0746	LA1137	*S. chilense* LA1556
LA1478	LA0927	LA1401	*S. chilense* LA1558
LA1547	LA0932	LA1408	*S. chilense* LA1969
LA1577	LA1035	LA1452	
LA1580	LA1039	LA1508	
LA1584	LA1040	LA1627	
LA1596	LA1041		
LA1599	LA1042		
LA1601	LA1043		
LA1611	LA1139		
LA1645	LA1404		
LA1660	LA1409		
LA1670	LA1412		
LA1719	LA1447		
LA1924	LA1450		
LA1936			
LA1993			
LA2097			
LA2533			
LA2854			

Additional file 2: Table S3. Introgressions known to be present in the initial breeding lines.

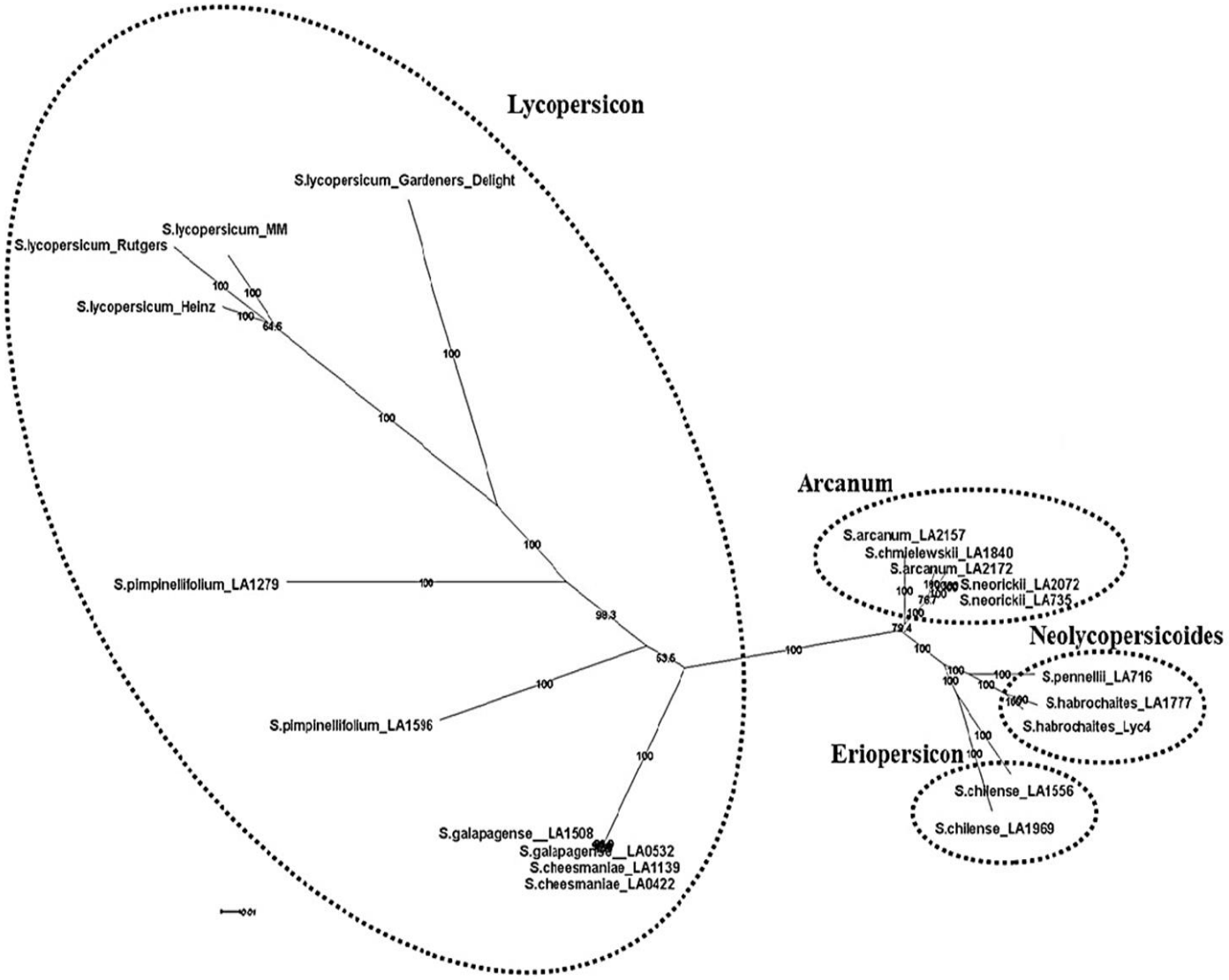
Introgressions	**Chr**	**C74**	**C85**	**R75**	**R104**	**Origin**
Tomato Ripening-Inhibitor (Rin)	5					*S. lycopersicum*
Cladosporium 5	5					*L. peruvianum*
TYLCV	6					*S. chilense / L. hirsutum*
Nematode	6					*L. peruvianum*
ToMV	9					*L. peruvianum*
ToMV (race 0,1,2)	9					*L. peruvianum*
Verticillium albo atrum	9					*S. lycopersicum var. cerasiforme*
Leaf mold (A,B,C,D,E)	9					*L. peruvianum*
Fusarium race 0 (ex1)	11					*S. pimpinellifolium / S. pennellii*
Fusarium crown and rootrot	?					*L. peruvianum*

Additional file 3. All SNP scores in the varieties from *S. lycopersicum* used for comparisons (also used by van Berloo, 2008). Available at:
http://www.biomedcentral.com/content/supplementary/1471-2164-14-354-S3.xlsxq

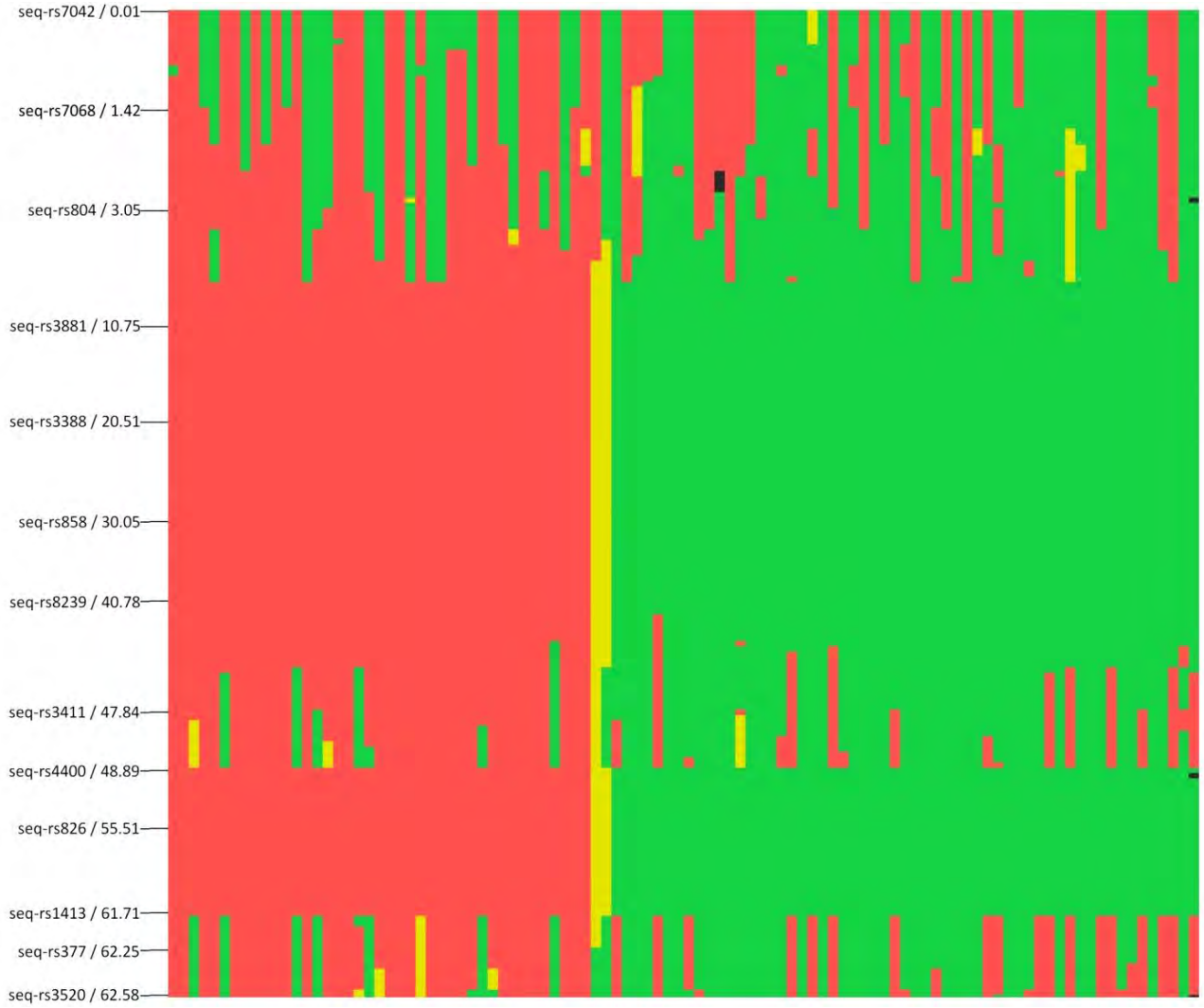
Additional file 4. Heat map representation of polymorphisms found in the TMV region of chromosome 9. *Solanum lycopersicum* allele - gray background), yellow heterozygous and homozygous wild relative allele – green background.



Additional file 5. BioNJ tree with 1000 bootstrap analysis showing an implicit relation of the available species according the different tomato groups.



Additional file 6. Heat map of the genotype call of 188 markers distributed along chromosome 12 of 100 RILs (horizontal) from a cross between *S. lycopersicum* cv Moneymaker (red) and *S. pimpinellifolium* G1.1554 (green). Heterozygous calls (yellow) and NCs (black) are also included. Certain loci marked for reference as: sequence name / position (Mbp). The positions were blasted towards the published tomato genome version 2.4 (The Tomato Genome Consortium, 2012).



Chapter 3

Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in *Solanum pimpinellifolium*

Published as:

Marcela Víquez-Zamora[a,b,*], Myluska Caro[a,b,*], Richard Finkers[a,b], Yury Tikunov[a,b], Arnaud Bovy[a,b], Richard GF Visser[a,b], Yuling Bai[a,b] and Sjaak van Heusden[a,b]

BMC Genomics 2014, **15**:1152 doi:10.1186/1471-2164-15-1152

[a] Wageningen UR Plant Breeding, Wageningen University & Research, P.O. Box 384, 6700 AJ, Wageningen, the Netherlands
[b] Centre for Biosystems Genomics, P.O. Box 98, 6700AB, Wageningen, the Netherlands
*Equal contributors

Abstract

Background

A RIL population between *Solanum lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 was genotyped with a custom made SNP array. Additionally, a subset of the lines was genotyped by sequencing (GBS).

Results

A total of 1974 polymorphic SNPs were selected to develop a linkage map of 715 unique genetic loci. We generated plots for visualizing the recombination patterns of the population relating physical and genetic positions along the genome.

This linkage map was used to identify two QTLs for TYLCV resistance which contained favourable alleles derived from *S. pimpinellifolium*. Further GBS was used to saturate regions of interest, and the mapping resolution of the two QTLs was improved. The analysis showed highest significance on chromosome 11 close to the region of 51.3 Mb (*qTy-p11*) and another on chromosome 3 near 46.5 Mb (*qTy-p3*). Furthermore, we explored the population using untargeted metabolic profiling, and the most significant differences between susceptible and resistant plants were mainly associated with sucrose and flavonoid glycosides.

Conclusions

The SNP information obtained from an array allowed a first QTL screening of our RIL population. With additional SNP data of a RILs subset, obtained through GBS, we were able to perform an *in silico* mapping improvement to further confirm regions associated with our trait of interest. With the combination of different ~omics platforms we provide valuable insight into the genetics of *S. pimpinellifolium*-derived TYLCV resistance.

Keywords: SNPs, *S. pimpinellifolium*, *in silico*, TYLCV, flavonoids, hexose, genotype by sequencing (GBS).

Background

Solanum pimpinellifolium is a source for introgression breeding in tomato *(S. lycopersicum)*. This species is one of the closest wild relatives of *S. lycopersicum*, and it is present in the pedigree lineage of some commercial cultivars such as the sequenced 'Heinz 1706' (The Tomato Genome Consortium, 2012). Linkage maps from crosses between *S. lycopersicum* and *S. pimpinellifolium* were generated by various researchers (Grandillo and Tanksley 1996, Chen and Foolad 1999, Lippman and Tanksley 2001, Doganlar *et al.,* 2002, Sharma *et al.,* 2008, Ashrafi *et al.,* 2009, Sim *et al.,* 2012). Their work represents a small piece of the successful use of genome-wide linkage analyses to map underlying genetic factors of traits between the two species.

Recombinant inbred lines (RILs) derived from interspecific crosses consist of individuals with parental mosaics and are an efficient resource for mapping quantitative trait loci (QTL) (Broman, 2005). Genotyping with molecular markers allows the visualization of recombination patterns which is crucial for the elucidation of loci associated with segregating traits (Paran *et al.*, 1995; Mézard, 2006). This has become more efficient due to the availability of vast numbers of markers such as single nucleotide polymorphisms (SNPs). In tomato, the availability of high throughput SNP arrays allows massive parallel whole-genome screening of genotypes (Sim *et al.*, 2012, Viquez-Zamora *et al.*, 2013).

Nowadays, next generation sequencing technologies are offering new ways to increase genotyping throughput by several orders of magnitude (Huang *et al.*, 2009). Even more, it is possible to combine different genotyping platforms to increase the power of the analyses. Furthermore, due to published complete tomato genomes (The Tomato Genome Consortium, 2012), next generation re-sequencing approaches can be applied in related germplasm (Causse *et al.*, 2013). Studies on evolution and domestication, as well as the genetic basis underlying important traits can benefit from these genomic tools (Aflitos *et al.*, 2014).

TYLCV is the causal agent of an aggressive tomato disease that can result in production losses up to one hundred percent, and its rapid spread worldwide is threatening the production of tomatoes. Development of TYLCV resistant tomato cultivars is an important strategy to avoid the damaged cause by TYLCV. However, no TYLCV resistance has been identified in the cultivated tomato germplasm, except for the resistance allele of *ty-5* which is possibly originated from a mutation in the cultivated tomato (Anbinder *et al.*, 2009). Breeding for resistance to TYLCV has been focused on the introgression of tolerance or resistance genes from tomato wild relatives such as *S. pimpinellifolium, S. chilense*, *S. habrochaites* and *S. peruvianum* (Picó *et al.*, 2001, Verlaan *et al.*, 2013). Several *S. pimpinellifolium* accessions are

known to confer resistance to the virus (Banerjee *et al.,* 1987, Kasrawi *et al.,* 1988, Chagué *et al.*, 1997, Picó *et al.*, 2000, Pilowski and Cohen, 2000, Pérez de Castro *et al.,* 2007), but attempts to map the causal factor in this species were not very successful. Thus, *S. pimpinellifolium*-derived TYLCV resistance is currently not well-exploited in tomato breeding programs (Ji *et al.*, 2007). In our study we genotyped a RIL population between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 with a custom made SNP array (Viquez-Zamora *et al.*, 2013), and a subset of 60 lines was also genotyped by sequencing using Illumina HiSeq 2000 (150 Tomato Genome ReSequencing project; www.tomatogenome.net). Furthermore, we explored the population with an untargeted metabolic profiling and compared resistant vs. susceptible lines in order to get more insights on compounds that might play a role in the resistance. Our study shows how we can combine different ~omics approaches to identify genetic loci underlying resistance to Tomato Yellow Leaf Curl Virus (TYLCV) in *S. pimpinellifolium* using a RIL population.

Methods

Recombinant Inbred Lines (RILs)

From a cross between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 (CGN reference CGN 15528) a set of 100 RILs was generated through single seed descent (SSD) until the sixth generation (Voorrips *et al.*, 2000). These RILs, which have been used for many different experiments e.g. Khan *et al.* 2012, were used in this study.

DNA extraction

Genomic DNA from young leaflets was extracted using a CTAB based protocol (Stewart and Via, 1993; Kabelka *et al.*, 2002) adjusted for high throughput isolation. Two young leaflets were ground with a Retsch 300 mm shaker (Retsch BV, Ochten, The Netherlands) using 1 ml micronic tubes (Micronic BV, Lelystad, The Netherlands). DNA pellets were washed in 76% EtOH with 10mM NH_4Ac before re-suspending the DNA in TE buffer.

Genome wide genotyping

Genome wide genotyping was done as described by Víquez-Zamora *et al.* (2013). In short, DNA samples were sent to ServiceXS (http://www.servicexs.com/), Leiden, the Netherlands. A custom made Infinium HD Ultra Assay protocol (2009) was used for

hybridization onto a BeadChip. The Genotyping Module 1.9.4 of Illumina's GenomeStudio® V2011.1 software package was used to analyse the genotyping results under default settings. All samples corresponding to the RIL population and the parents were selected for a separate analysis in which manual inspection and adjustment were performed in order to discard questionable SNPs for the population and to optimize call rates. All polymorphic SNPs for the RIL population were named after their position on the SL2.40 version (http://solgenomics.net/) of the tomato genome sequence published online (The Tomato Genome Consortium, 2012).

Genotype by sequencing (GBS)

A subset of 60 lines was selected for resequencing (lines with extreme values for TYLCV resistance were included). Whole genomic DNA was isolated from each line (see above). Shallow sequencing of 500 bp inserts was carried out using Illumina HiSeq 2000 (100 bp paired end reads) at an average coverage of 3x. Bases with Q < 20 were trimmed before read mapping with BWA (Li *et al.*, 2009; Li and Durbin, 2010) against the SL2.40 genome sequence of *S. lycopersicum* cv. Heinz with a maximum insert size of 750 bp (50% deviation), reporting at most 30 hits and removing PCR duplicates. SAMTOOLS (Li *et al.*, 2009) was used for variant calling without skipping InDels and a minimum gap distance of 5bp. In addition, GATK (McKenna *et al.*, 2010), was used to call variants for all 60 genotypes in one single analysis.

The JBrowse by Skinner *et al.* (2009) was used for the embedding and visualization of the SNP variants. The available gene models (ITAG 2.3) were obtained from the Sol Genomics Network (http://solgenomics.net/). Subsequently, a script was generated in order to combine the information of SNPs within the RILs. Access to the JBrowse with the information of the sequences can be obtained through: http://www.tomatogenome.net/ril_variants. Furthermore, the program Marker2sequence (Chibon *et al.*, 2012) was used to look for genes between specific genome coordinates based on their annotation.

TYLCV screening

Virus inoculation

Agrobacterium-mediated inoculation was performed to infect plants with TYLCV. Plantlets at the 3-4 leaf stage (approximately 21 days after sowing) were inoculated with *A. tumefaciens* LBA4404 bearing a tandem repeat of an infectious TYLCV-IL (Israel isolate) clone. Bacterial growth was performed as previously described by Verlaan *et al.* (2011) and bacteria were

injected into true leaves using syringes without needle. Plants were grown under greenhouse conditions at 23 °C, 60% humidity and 16-h/8-h day/night cycle.

Disease *test*

Disease symptoms were recorded 20, 25, 35 and 45 days post inoculation. Plants were scored for symptom severity according to the scale described by Friedmann *et al.* (1998). A first screening of the RILs was conducted using one plant per line. Thereafter, a second screening followed for the RILs classified as resistant to confirm the phenotype where four plants per resistant line were assessed. TYLCV disease symptoms rating was: 0 = no visible symptoms, inoculated plants show same growth and development as non-inoculated plants; 1 = very slight yellowing and minor curling of leaflet margins on apical leaf; 2 = some yellowing and minor curling of leaflet ends; 3 = a wide range of leaf yellowing, curling and cupping, with some reduction in size, yet plants continue to develop; 4 = very severe plant stunting and yellowing, and pronounced cupping and curling; plants cease to grow (Additional file 1: Figure S1).

Metabolic profiling

The RIL population was grown in triplicate under the same greenhouse conditions. Seven weeks after sowing, fully developed leaves were detached and main veins were removed. Samples were frozen in liquid nitrogen and thereafter ground into fine powder.

Untargeted metabolic profiling of leaves was performed with three platforms: 1) Liquid chromatography (LC), using a C18-reversed phase column, coupled to a Quadrupole-time-of-flight (TOF) mass spectrometer (MS) and a photodiode array detector (PDA) to detect semi-polar compounds such as flavonoids, alkaloids, phenylpropanoids, saponins, phenolic acids and polyamines according to De Vos *et al.* (2007). 2) Gas chromatography (GC) coupled to electron impact time of flight (TOF)-MS for detection of primary metabolites according to Lisec *et al.* (2006). 3) Solid phase microextraction (SPME)-GC-MS for the analysis of volatiles according to Tikunov *et al.* (2005 and 2010).

Metabolomics data processing

Metabolites were quantified and identified according to Tikunov *et al.* (2010). Each dataset was processed using MetAlign (www.metalign.nl) for baseline correction, noise estimation, and ion-wise mass spectral alignment of the corresponding chromatograms. MSClust software was used to extract compounds mass spectra and for data reduction (Tikunov *et al.*, 2012).

The putative identification of metabolites was based upon their spectra, retention time, molecular weight and fragmentation patterns. For LC-MS data, compound characteristics were analysed and compared using the Dictionary of Natural Products (http://dnp.chemnetbase.com) and in-house tomato metabolite databases. GC-MS data were annotated using the NIST Mass Spectral Search Program v2.0 (http://chemdata.nist.gov/mass-spc/ms-search/) by matching mass spectra extracted to the NIST mass spectra collection and the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/) for mass spectra matching followed by retention index comparison.

Linkage Analysis

Linkage maps were constructed using JoinMap® 4.1 (Kyazma©; Van Ooijen, 2011) with the specifications by Víquez-Zamora *et al.* (2013) using the Haldane's mapping function. Genetic linkage groups were compared to the physical maps based on the tomato genome version SL2.40 using MapChart 2.2 (Voorrips, 2002). The software GenStat 16th edition was used to perform mapping of QTLs for TYLCV resistance and the MapQTL software was used to map metabolite QTLs (mQTLs; van Ooijen, 2009). The genotypic and phenotypic information is available at: http://www.plantbreeding.wur.nl/Publications/SNP/RILs_genotype-TYLCVphenotype.xlsx. Identified QTLs for TYLCV resistance were named according to their chromosomal position as in Kadirvel *et al.* (2013); *qTy-p3* and *qTy-p11* (*p* as from *S. pimpinellifolium*) for QTLs on chromosomes 3 and 11, respectively. The Marker2sequence application was used to mine regions for candidate genes (Chibon *et al.*, 2012).

Furthermore, the information of the sequences was embedded into JBrowse 1.11.1 (Skinner *et al.*, 2009) to visualize the detected structural variants. The SL2.40 tomato genome assembly and ITAG 2.31 tomato genome annotation was loaded together with the BAM and VCF files of the 60 genotypes.

Results

Linkage map and genome-wide visualizations

A custom made SNP Array was assembled from polymorphisms mainly found between two cherry and two round tomatoes (Víquez-Zamora *et al.*, 2013). This array was used to genotype a RIL population between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554. A total of 1974 polymorphic SNPs were identified between the parents. These SNPs were used to develop a linkage map based on their segregation patterns among the 100 RILs. The resulting map included 715 loci with an average distance of 1.85 cM between loci (Fig. 1). The greatest gap was approximately 40 cM on chromosome 1 and covered the region between 76 and 83 Mb.

In order to visualize the recombination patterns along each chromosome, the physical positions of the SNP markers were determined using the published tomato genome (The Tomato Genome Consortium, 2012). For each SNP and its flanking sequence, a BLAST was performed to the genome sequence version SL2.40. Except for markers on chromosome 12, colinear orders were observed between the genetic and physical maps, as shown in scatter plots per chromosome between the linkage (cM) and physical map (Mb) (Fig. 2). These scatter plots further allowed the visualization of cold- and hot-spots of recombination. When a large physical distance corresponds to only a small difference in cM, we can assume cold-spots of recombination. These cold-spots were always the heterochromatin pericentromeric regions and could be as long as 50 to 80 Mb. In contrast, hot-spots of recombination could be present if there is a large cM difference corresponding to small physical distance between markers.

The mosaic pattern of each RIL was calculated and composition of lines varied between 20% and 80% of alleles coming from each parent. In addition, we calculated the SNP allele frequency within the RIL population per marker location along each chromosome. The frequency distribution was mostly 50-50% as expected. However, we found skewness in the distribution of two regions. A preference for *S. pimpinellifolium* alleles was seen near the centromere of chromosome 2, and a preference for *S. lycopersicum* alleles on chromosome 9 (Additional file 2: Figure S2).

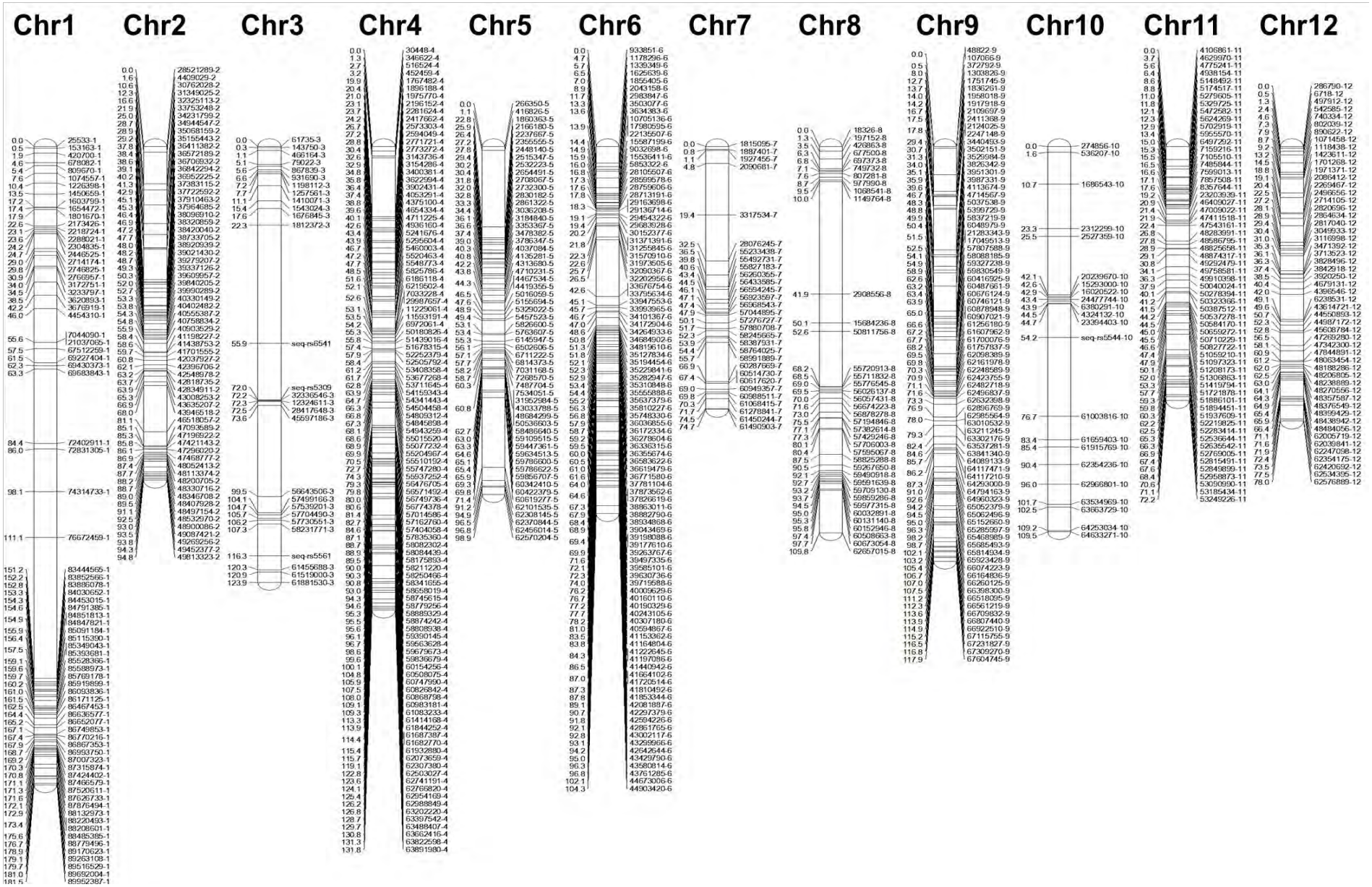


Figure 1. Linkage map of a RIL population originating from a cross between *Solanum lycopersicum* cv. Moneymaker and *Solanum pimpinellifolium* G1.1554. The map shows 715 SNPs representing single recombination positions. Markers are named according to their physical positions.

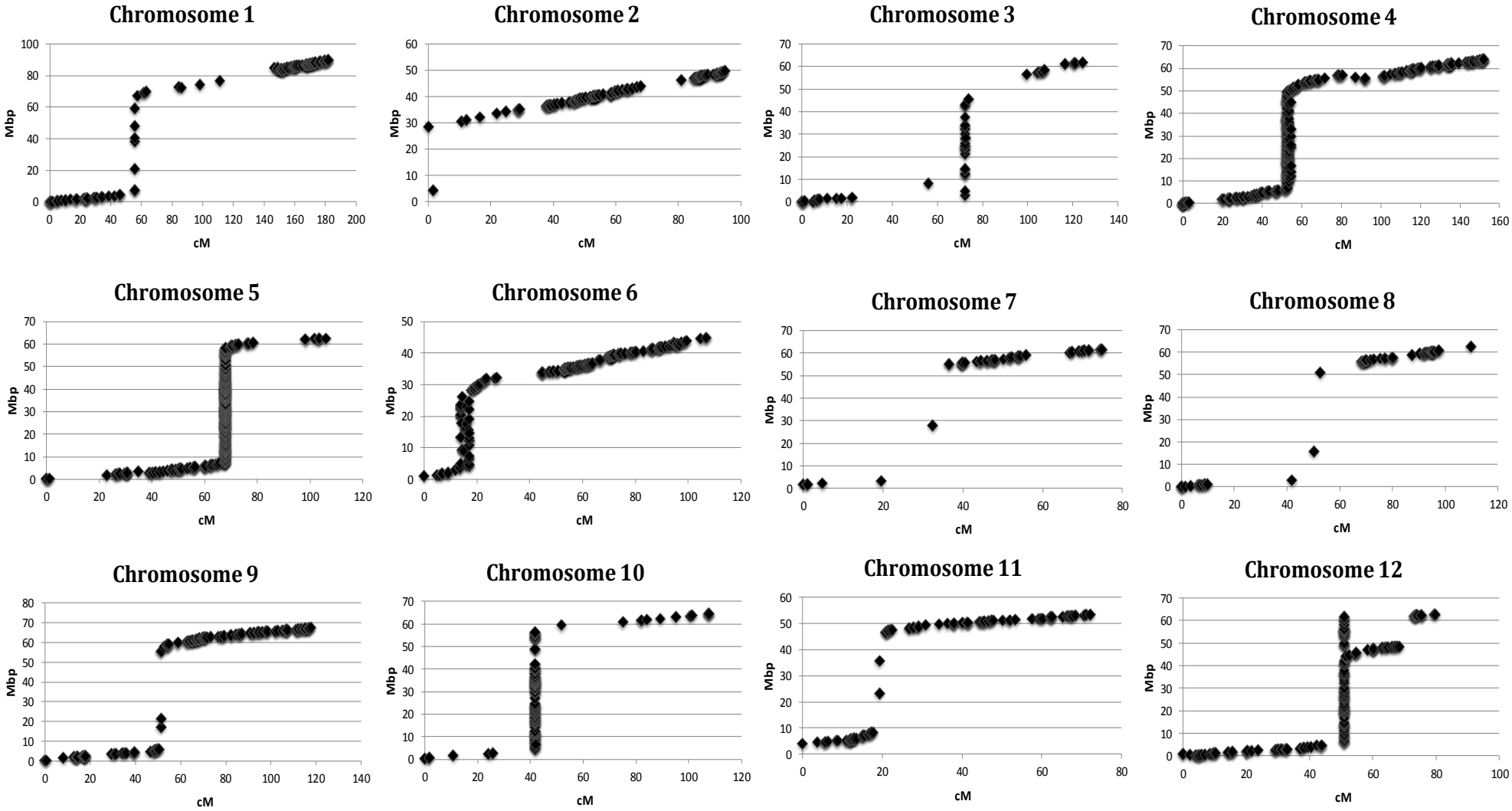


Figure 2. Scatter plots combining linkage maps (genetical positions in cM) and physical positions (Mb) from the RIL population created from a cross between *Solanum lycopersicum* cv. Moneymaker and *Solanum pimpinellifolium* G1.1554.

QTLs and in silico ***mapping***

The genotypic file and the linkage map obtained above were then used to map multiple traits. One of the traits screened using our RIL population was TYLCV resistance. Eighty-one RILs were infected with TYLCV. Typical virus symptoms appeared from 30 days after inoculation (dpi); plants were scored according to their symptom development up to 45 dpi and classified as Resistant (R) or Susceptible (S). The susceptible parent 'Moneymaker', as expected, displayed severe TYLCV symptoms such as plant stunting and reduced leaf size with upwards curling and yellowing. The resistant parent, *S. pimpinellifolium* G1.1554, remained without symptoms until the end of the experiment. Five out of 81 tested RILs showed no symptoms after virus inoculation (disease score = 0), and four RILs showing very mild symptoms (disease score ≤ 1) were considered resistant. The remaining 72 RILs were classified as susceptible, showing clear TYLCV symptoms including the characteristic leaf curling and yellowing with disease scores ranging from 2 to 4 (Additional file 1: Figure S1).

In order to identify the genomic regions involved in the resistance, single trait QTL analysis was performed. Two putative QTLs associated with the resistance were identified, one on chromosome 3, hereafter referred to as *qTy-p3*, and one on chromosome 11, hereafter referred to as *qTy-p11* (Fig. 3A). For *qTy-p3,* 20 markers showed significant association with a LOD value ranging from 3.68 to 3.81, locating the QTL between 4.74 and 45.59 Mb of chromosome 3; the most significant marker for *qTy-p3* was L_45597186-3. For *qTy-p11*, 6 significantly associated markers were identified with a LOD value from 3.79 to 4.04, in a region between 50.82 and 51.20 Mb of chromosome 11. The most significant marker for this QTL was L_51208173-11 (Fig. 3; Additional file 3: Figure S3).

Sixty lines from the RIL population were re-sequenced, and the resulting genome sequences were aligned to the published tomato genome, version SL2.40 (The Tomato Genome Consortium, 2013). The fully resistant lines were included among the 60 sequenced RILs. JBrowse (Skinner *et al.*, 2009) was used to visualize SNP variants within the RILs and allowed us to retrieve the corresponding SNP information of all aligned reads in regions of interest.

We selected 43 additional SNPs to saturate chromosome 3 resulting in approximately one marker per 0.6 Mb. For chromosome 11, we included two markers in the region of 7.5-8.3 Mb and 27 in the region between 49-53Mbp. As result, the chromosomes 3 and 11 linkage groups were improved, as was the *in silico* mapping for the subset of 60 lines.

The outcome of the QTL analysis with the enriched genotypic data and improved genetic map is depicted in Figure 3B. Using this extended dataset, the analysis confirmed the QTLs *qTy-p3* and *qTy-p11*. The calculated threshold was very similar to the previous calculated threshold (3.64). For *qTy-p3* the LOD values ranged from 3.7 to 4.5, comprising a region with

53 significantly linked markers. The most significantly linked marker position for *qTy-p3* was then refined from 45597186 bp in the first QTL mapping to 46454095 bp and 46520535bp (both LOD of 4.46) in the improved version. For *qTy-p11* the LOD values for the 26 significantly linked markers (in the improved map) ranged from 3.86 to 4.86, and the most significant marker position was refined from 51208173 bp to 51347236 bp and 51373277 bp (both LOD of 4.86). Together, both QTLs explained almost 28% of the phenotypic effect (13.46 for *qTy-p3* and 14.18 for *qTy-p11*).

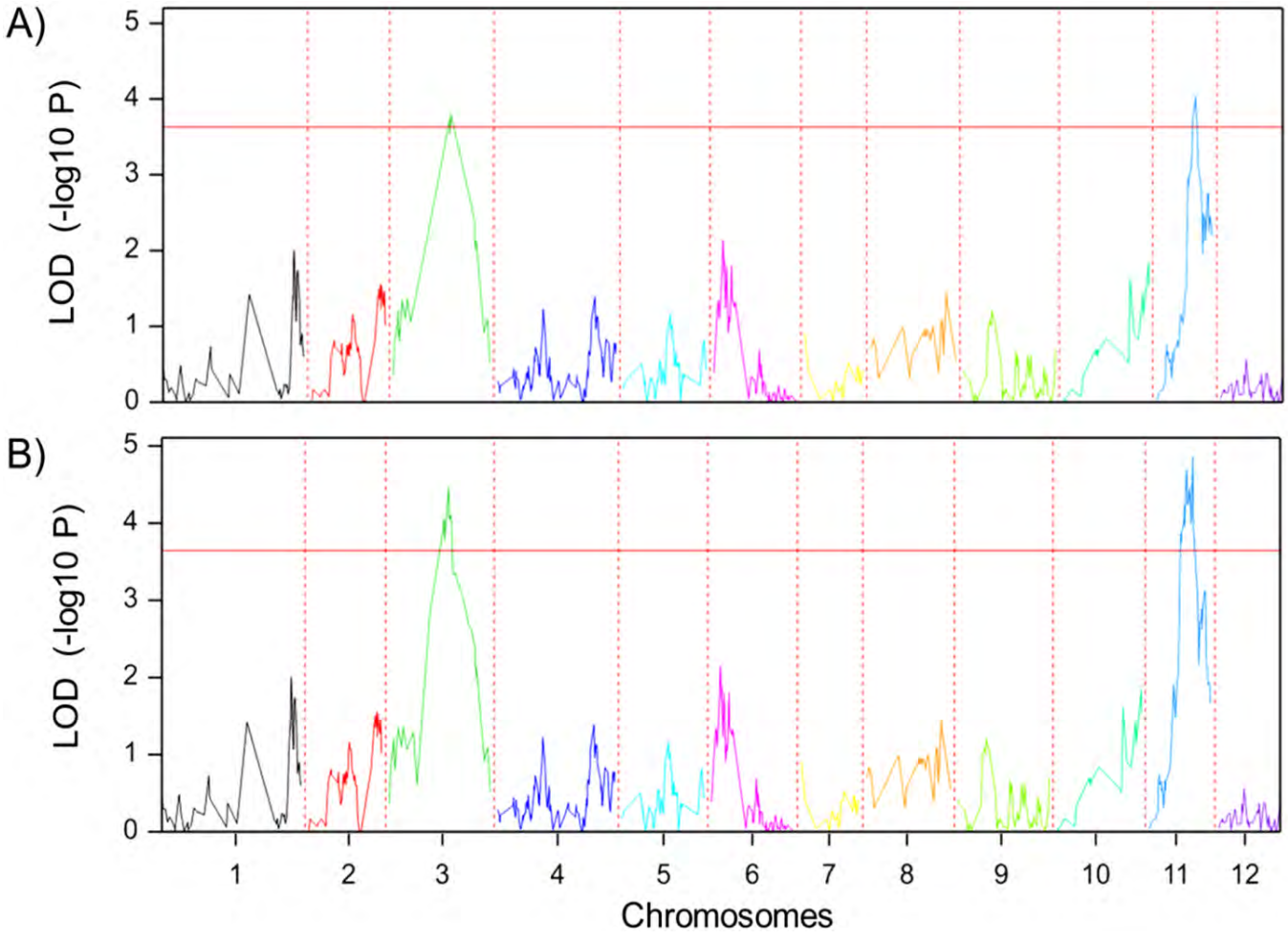


Figure 3. QTL mapping of *qTy-p3* and *qTy-p11* (chromosome 3 and chromosome 11) conferring resistance to TYLCV from *S. pimpinellifolium* G1.1554. Y-axis represents values according to the interval mapping, horizontal red line delimits threshold of 3.6. A) QTL mapping in GenStat only with the SNPs obtained from the SNP array. B) QTL mapping after the inclusion of more SNP information obtained from sequences in chromosomes 3 and 11.

A QTL analysis using cofactors (MQM) was performed. When the most significant markers of chromosome 3 were used as cofactors, the LOD values of *qTy-p11* decreased but were still significant. However when the most significant markers of chromosome 11 were used as cofactors, the values of *qTy-p3* decreased to non-significant levels. Therefore, the greater impact of *qTy-p11*for the resistance was confirmed. Although all resistant RILs were homozygous for the *S. pimpinellifolium* allele at both QTLs, 14 RILs had disease scores of 2-4 (susceptible). Thus both QTLs with the favourable alleles are necessary for resistance, but their presence did not necessarily result in resistant plants.

Identification of candidate genes

In order to identify candidate genes for TYLCV resistance, we re-explored the QTL regions using the physical positions of the SNP markers flanking the QTLs. For chromosome 11, we targeted the region between 50.2 and 51.4 Mb. For *qTy-p11*, a total of 124 predicted genes were identified using Marker2sequence (Chibon *et al.*, 2012) based on the tomato genome sequence (Sol Genomics Network, SGN). Four putative disease-resistance proteins were predicted in the *qTy-p11* region, three of them clustering in the region from position 51347236 to 51373277. Furthermore, approximately 74.9 kb of *qTy-p11* overlaps with the region reported to contain the *Ty-2* resistance allele from S. *habrochaites* accession B6013 (Yang *et al.*, 2014).

The *qTy-p3* QTL region is physically large, from 2.48 to 47.44 Mb (45 Mb), including the centromeric region. This QTL region harbours more than six hundred annotated genes. In the vicinity of position 46454095 bp (the marker with the highest LOD score) there are genes related to sugars (e.g. high-affinity sugar transporters) and flavonoids (e.g. flavanone 3-hydroxylase-like protein).

RIL population metabolic profiling

Using the RIL population (not TYLCV infected), we performed untargeted metabolic profiling on leaf material. Primary metabolites were evaluated using GC-TOF-MS. Few differences were observed between parents and individuals of the population showing a similarity in the primary metabolism. However, the LC-TOF-MS and the SPME-GC-MS platforms uncovered more differences and revealed several QTLs for secondary metabolites and volatiles. More than 200 QTLs were found with putatively identified compounds; an mQTL for sucrose was mapped near *qTy-p11,* and several mQTLs for flavonoid glycosides were present near the region of *qTy-p3* (Supplemental Table 1).

Furthermore, since there were TYLCV-susceptible and resistant lines with both QTLs having the homozygous *S. pimpinellifolium* alleles, we performed a t-test with all metabolic data in

order to find metabolites that were significantly different between the two groups of RILs. Five compounds showed significant differences (p-value lower than 0.05) and had higher accumulations in the resistant plants. Three of them were putatively identified as glycosylated forms of kaempferol (LCS146), laricitrin (LCS149) and quercetin (LCS151) having a 4.3, 3.8 and 2.8-fold change, respectively. The other two compounds were acetoxytomatine (C724) and sucrose (C121) with 1.6 and 1.5-fold difference, respectively.

Discussion

High-throughput genetic mapping

The custom made SNP array was designed to distinguish different *S. lycopersicum* cultivars, nevertheless a vast amount of polymorphisms were detected between *S. pimpinellifolium* and *S. lycopersicum* cv. Moneymaker making it possible to construct a high density genetic linkage map. In general, positions on the genetic linkage map were consistent with the physical positions on the tomato genome showing the accuracy and robustness of the map and the quality of the tomato sequence.

High and low recombination rates were consistent with the known distribution of euchromatic and heterochromatic regions, as shown by Sim *et al.* (2012). Chromosomes 1, 3, 4, 5 and 10 had large regions without recombination including the centromeres. Centromeric patterns were also observed for chromosomes 6, 7, 8, 9 and 11, but there were some possible distortions that could profit from more markers in the region. Still, the distortions of chromosome 6 might be influenced by the distinct heterochromatin distribution that follows an alternating pattern (Iovene *et al.*, 2008). Chromosome 12 also showed a non-recombining centromeric pattern, but this is a clear representation of the likely scaffold misalignment reported previously (Víquez-Zamora *et al.*, 2013). Strong clustering of markers on the genetic map but with a clear physical distance between these markers shows a suppression of recombination in these areas (Fig. 2).

The allele frequencies showed a preference for the *S. pimpinellifolium* alleles near the centromere on chromosome 2. This part of the chromosome is linked to rDNA genes. Therefore, there could be a preference for *S. pimpinellifolium* rDNA. A preference was also found for the 'Moneymaker' alleles on chromosome 9 which might be related to deleterious effects of carrying the *S. pimpinellifolium* alleles in this region or to structural DNA differences. Species in the same genus can have DNA configuration differences generating structural changes in the rearrangement of chromosomes after a cross (Mézard, 2006). Differences in local recombination frequencies could be related to the pairing of homologous

chromosomes, DNA sequence similarity or divergence, including the presence or absence of genes involved in the recombination process, chromatin conformation or to differences in timing during meiosis (Tam *et al.*, 2011).

Actual research is enriched by the combination of different software packages. The combination of JBrowse (Skinner *et al.*, 2009), loaded with gene models from the Sol Genomics Network (http://solgenomics.net/), with previous information of possible genes of interest obtained from Marker2sequence (Chibon *et al.*, 2012) allowed an efficient targeted *in silico* mapping.

TYLCV resistance mapping and ~omics platforms combination

The sequenced subset of 60 lines created suitable tools for mapping regions of interest. We enriched regions on chromosome 3 and chromosome 11 that were associated with TYLCV resistance, and the *in silico* approach proved to be successful in increasing the power of QTL detection. After the addition of more SNPs coming from the known sequences, we confirmed that *qTy-p3* and *qTy-p11* were not artefacts but had real effects. This allowed us to target the location of the QTL region for *qTy-p11* and it showed the most significant region for *qTy-p3* (Figure 3), even though a large region of chromosome 3, including the centromere, looks to have an essential impact on the expression of the resistance.

The effect of both QTLs together explained only 28% of the phenotypic effect on the resistance of our RIL population, suggesting additional genetic factors playing a role on the resistance which might have been undetected in our analysis. The accuracy of QTL localization using RILs depends on population size, where a genome-wide coverage of the parents should be present in the mapping population (Keurentjes *et al.*, 2007). The fact that both *qTy-p3* and *qTy-p11* were needed for resistance but their presence does not necessarily lead to resistant plants also suggests the possible interaction of extra factors. TYLCV resistance derived from a number of *S. pimpinellifolium* accessions (e.g. LA121, LA373, UPV16991) has been previously suggested to be quantitatively inherited and to show variable gene penetrance (Pérez de Castro *et al.*, 2007). Further genotyping, targeting the regions of low marker coverage, is being assessed in order to detect the presence of one or more additional QTLs, or potential modifier genes. These interactions might be associated with the secondary metabolism of the plants.

A number of TYLCV resistance loci have been reported from different wild *Solanum* species, including *S. chilense*, *S. habrochaites* and *S. peruvianum* (Ji *et al.*, 2007). Recently, the *Ty-1* gene from *S. chilense* LA1969 has been cloned and is a representative for a novel class of resistance genes, an RNA-dependent RNA polymerase of the RDR_{γ} class (Verlaan *et al.*, 2013, Butterbach *et al.*, 2014). TYLCV resistance in *S. chilense* accessions LA1932 and LA2779, *S.*

habrochaites accession B6013 and TY172, a tomato line derived from different accessions of *S. peruvianum* have been mapped to chromosomes 3 and 10 (*Ty-4*, Ji *et al.*, 2009 and *Ty-6*, Hutton 2013), chromosome 11 (*Ty-2*, Yang *et al.*, 2014) and chromosome 4 (*ty-5*, Anbinder *et al.*, 2009), respectively.

Several accessions from *S. pimpinellifolium* have been screened and identified to confer resistance to TYLCV (Banerjee & Kalloo 1987, Kasrawi *et al.,* 1988, Picó *et al.,* 2000, Pilowski & Cohen 2000, Ji *et al.,* 2007, Pérez de Castro *et al.,* 2007). However, the genetics of the trait are complex and only one report on mapping resistance originating from *S. pimpinellifolium* (accession 'Hirsute INRA') has been reported using RAPD markers (Chagué *et al.*, 1997). This resistance was mapped to chromosome 6, close to the *Ty-1* gene. The QTLs identified in the present study represent newly mapped loci conferring resistance derived from *S. pimpinellifolium* G1.1554 and provide a starting point for assessing putative candidate genes in the identified regions. A cluster of disease resistance-like proteins is present near *qTy-p11* (based on the cultivated tomato genome sequence). Furthermore, this region on chromosome 11 overlaps with 75 kb of the upper part of the mapped region of *Ty-2*, a TYLCV resistance allele derived from *S. habrochaites* accession B6013 (Yang *et al.,* 2014). Although *Ty-2* has not yet been cloned, annotated genes in this common region (e.g. elongation factor 1-alpha) might provide further insights for assessing candidate genes for TYLCV resistance derived from these wild tomato species, and/or additional genes involved in the resistance pathway. Plant defense mechanisms are the result of complex gene networks which trigger or mediate the signaling pathways leading to resistance. Besides the reported *Ty*-loci, genes playing a role in these networks have been identified from their differential expression in resistant vs. susceptible genotypes and induced by TYLCV infection, e.g. *Permease I*-like protein and the hexose transporter *LeHT1* (Eybishtz *et al.*, 2009, Eybishtz *et al.*, 2010). Silencing these genes through Virus-induced gene silencing (VIGS) in a resistant genotype led to the collapse of the resistance, demonstrating the role and importance of these genes in the defense network of the plant.

In general, the presence of compounds such as amino acids and organic acids was very similar between the two species. Differences are more pronounced in the secondary metabolism. Our metabolic data show that the compounds present at higher amounts in the resistant plants are mainly flavonoid glycosides (Table 1). Flavonoids are phenolic compounds known to be involved in resistance to diverse stress conditions, including plant viruses (Bol *et al.*, 1990). For instance quercetin, one of the metabolites detected at higher levels in the resistant lines is a flavonoid known to inhibit HSP70 (Heat-shock protein 70) transcription in animal and plant cells. In *N. benthamiana,* Tomato yellow leaf curl Sardinia virus (TYLCSV) had a delayed infection speed after silencing a member of the HSP70 family,

showing that high levels of this protein are required for infection of the virus (Czosnek *et al.*, 2013). Inhibition of HSP70 expression by quercetin resulted in decreased amounts of nuclear TYLCV coat protein in tomato, demonstrating the potential involvement of this flavonoid in the virus resistance pathway (Gorovits *et al.*, 2013). Furthermore, an additional QTL analysis suggests that glycosides of the flavonoid kaempferol co-localise with the TYLCV resistance QTL on chromosome 3 and that sucrose could be related to the QTL on chromosome 11 (Supplemental Table 1). Kaempferol is known for its antibacterial properties. Besides, we observed the presence of this compound and other flavonoids attached to hexoses in the resistant RILs; transporters of hexoses have been reported to play crucial roles in disease resistance (Eybishtz *et al.*, 2010, Sade *et al.*, 2013). Some of these compounds likely linked to the resistance also showed an mQTL on chromosome 1 besides the one on chromosome 3, and the mQTL of sucrose also showed significance on chromosome 7. These regions will be further targeted in a fine mapping effort following up this research.

It should be noted that the different concentrations of the compounds observed in resistant vs. susceptible lines were measured prior to TYLCV infection. Sade *et al.* (2015) showed that the expression of genes controlling the synthesis of these phenolic compounds is associated with TYLCV resistance. Genes in the flavonoid biosynthesis pathway of a resistant line derived from *S. habrochaites* increased their expression after TYLCV infection leading to the accumulation of flavonoids and contributing to the resistance.

Conclusion

A RIL population obtained from a cross between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 was successfully genotyped with a custom made SNP array. Furthermore, the re-sequencing of a subset of the RILs allowed the possibility of *in silico* mapping of TYLCV resistance. Two QTLs were related to the resistance, one showing the highest significance on chromosome 11 close to the region of 51.3 Mb and the other close to 46.5Mbp on chromosome 3. However, there might be extra loci or genetic factors playing a role that could be unravelled if the population size is increased or when advanced populations are further explored. The resistance towards TYLCV suggests an interaction between flavonoids and hexoses favouring the trait.

We concluded that investments in sequencing can redeem the value of screenings of germplasm due to the fact that both SNPs and sequences can be targeted at the same time. Therefore, screenings can start with a defined number of retrieved SNPs per chromosome,

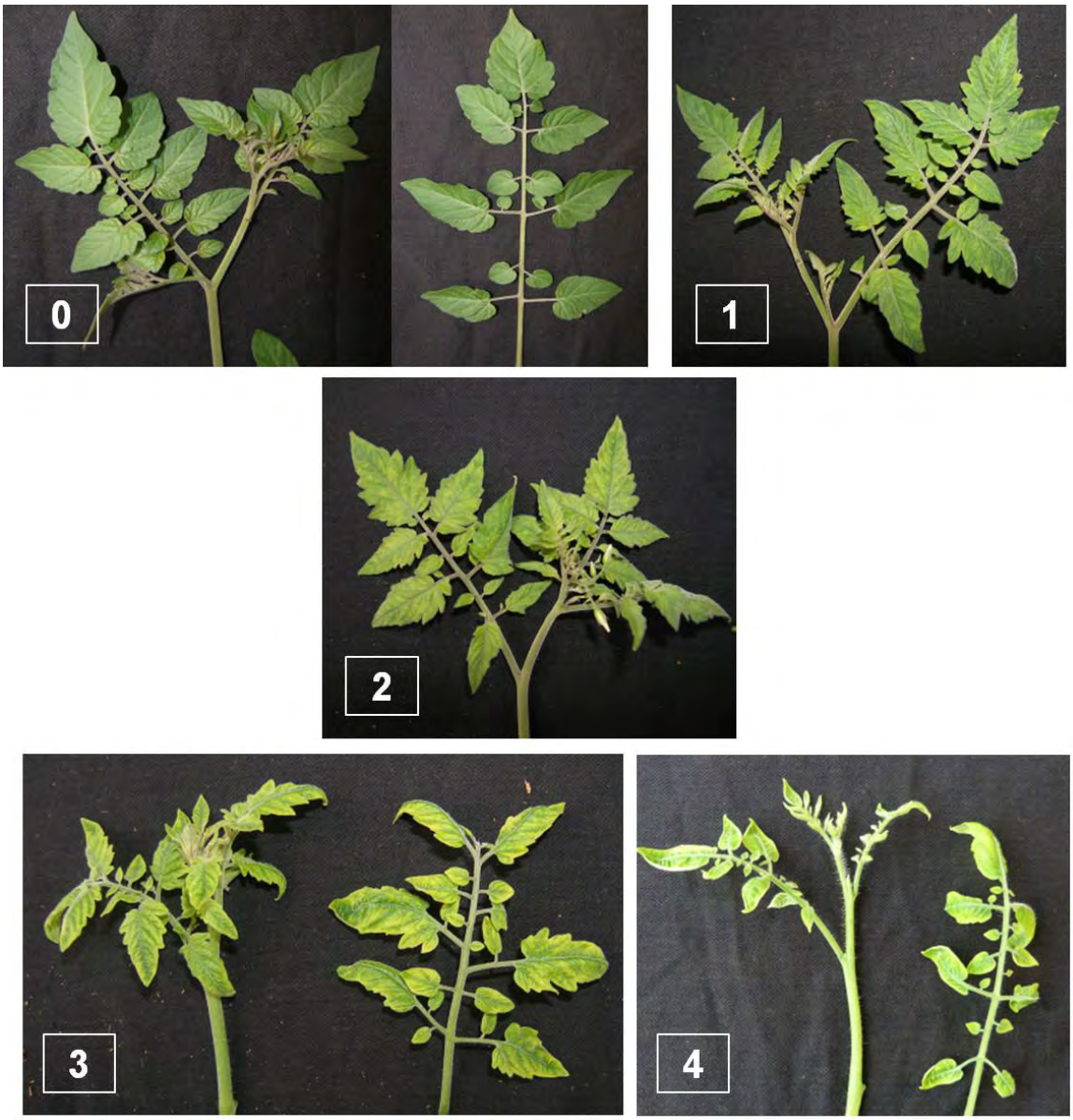
and thereafter, regions of interest can be further targeted. However, data storage, software acquisition and qualified human resources for data analysis and interpretation of combined ~omics platforms are going to make the difference to get robust analyses.

Acknowledgements

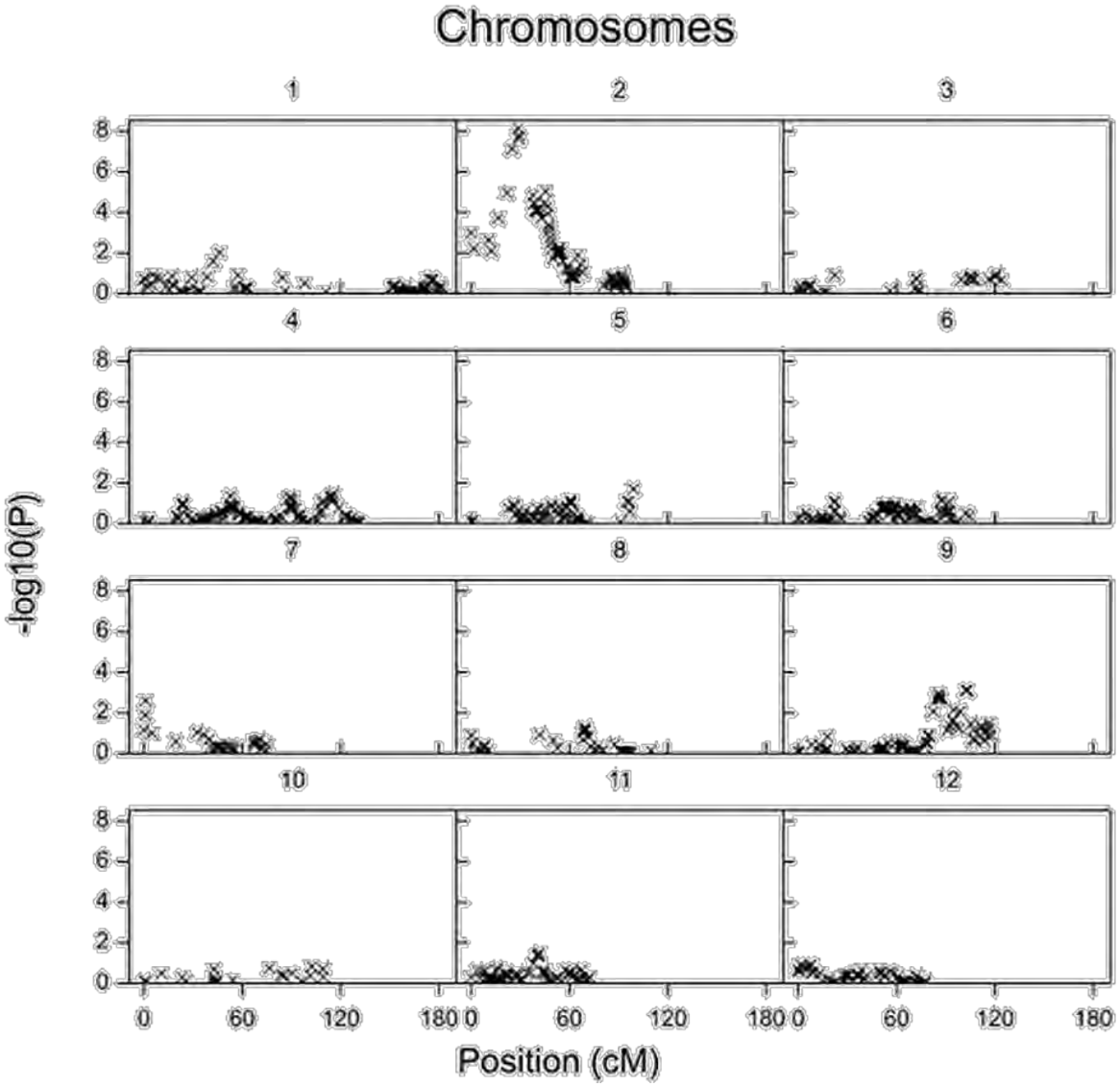
This project was carried out within the research programme of the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. And the metabolic profiling of the leaves was carried out within a project developed in collaboration with the Netherlands Metabolomics Centre (NMC). Furthermore, we received support from the COST action 'Quality Fruit' (COST FA1106).

We would like to thank Fien Meijer-Dekens for collecting and maintaining all the tomato samples at Wageningen UR-Plant Breeding, Ben Vosman for his support with the NMC project and Maarten Verlaan for his support and guidance in the elaboration of the TYLCV resistance experiments. We also would like to thank various people within the department of Bioinformatics of Wageningen UR for their support in the development of the SNP markers and in allocating them into the different versions of the Tomato Genome Sequence. We would like to make a special acknowledgement to the Business Unit Bioscience at Wageningen UR, especially to Ric de Vos, Henriette van Eekelen, Harry Jonker, Roland Mumm and Robert Hall for all their help and support in running the samples. Additionally, MVZ acknowledges the 'Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria' in Costa Rica, for their support through part of the doctoral project.

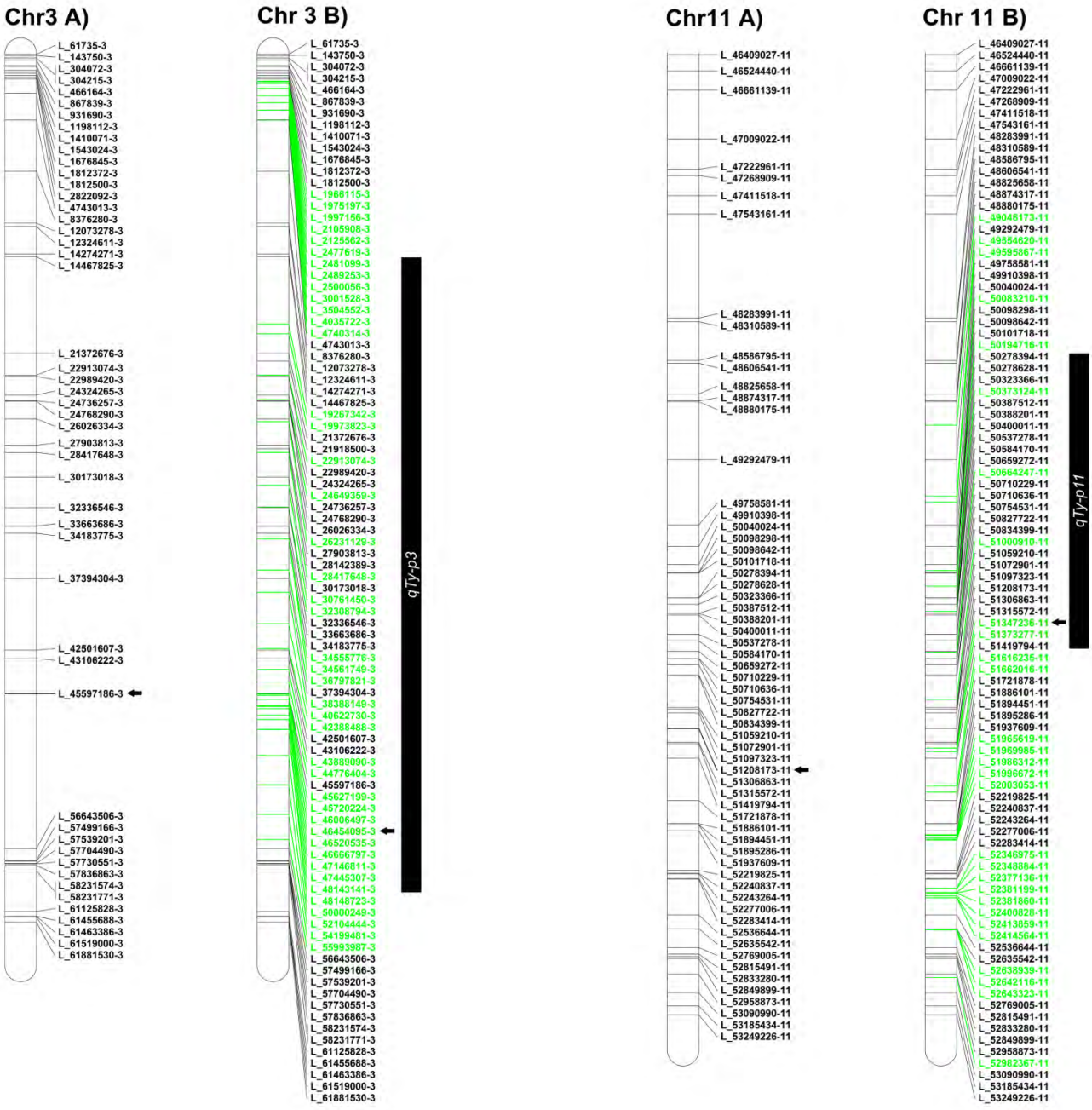
Additional Files



Supplemental Figure 1. Disease scores based on TYLCV symptom development. Plants were scored according to symptom severity: 0, no visible symptoms; 1, very slight yellowing and minor curling of leaflet margins; 2, yellowing and minor curling of leaflet ends; 3, leaf yellowing, curling and cupping; 4, severe leaf yellowing, curling and cupping, plant stunting (Friedmann *et al*, 1998).



Supplemental Figure 2. Probabilities of marker frequencies calculated in GenStat. A skewness in the direction of the chromosome region from *S. pimpinellifolium* G1.1554 is observed for chromosome 2. A skewness in the direction of the chromosome region from *S. lycopersicum* cv. Moneymaker is observed for chromosome 9.



Supplemental Figure 3. Physical map of chromosome 3 and chromosome 11 between 49-53 Mb. A) Chromosomes with initial SNPs. B) Chromosomes with incorporated SNPs in green. Black arrows indicate the most significant marker related to TYLCV resistance on each case. Black frames indicate the length covering significant markers for *qTy-p3* and *qTy-p11*.

Supplemental Table 1. QTLs found in non-infected leaves among the population between *S. lycopersicum* var. Moneymaker and *S. pimpinellifolium* G1.1554.

Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
Delphinidin deoxyhexose-feruloyl-hexose	5	947.2435	Flavonoid	LC-QTOF	LCS131	Chr1	7.44	1074557-1	**3.01**	13.30
Cinnamicacid	5	147.0427	Phenolic Acid	LC-QTOF	LCS2	Chr1	26.69	2446525-1	**3.05**	13.50
Caryophyllene oxyde	3	161	Sesquiterpene	GC-SPME	SPME12691	Chr1	36.48	3233797-1	**3.24**	14.3
Humulene	2	146	Sesquiterpene	GC-SPME	SPME11646	Chr1	37.48	3620893-1	**4.32**	18.6
Caryophyllene	2	124	Sesquiterpene	GC-SPME	SPME11232	Chr1	38.48	3620893-1	**4.57**	19.5
O-Cymene	2	91	Terpenoid/Alkylbenzene	GC-SPME	SPME3329	Chr1	38.51	3620893-1	**3.25**	14.3
Myrcene	2	119	Terpenoid	GC-SPME	SPME3543	Chr1	40.51	3676919-1	**3.61**	15.7
Glycoalkaloid	5	1344.6145	Alkaloid	LC-QTOF	C816	Chr1	51.96	7044090-1	**3.59**	15.70
Quercetin3-O-rutinoside-7-O-glucoside	2	771.1974	Flavonoid	LC-QTOF	C365	Chr1	55.58	7044090-1	**3.38**	14.80
Kaempferol-hexose-hexose-hexose	5	771.1974	Flavonoid	LC-QTOF	LCS52	Chr1	55.58	7044090-1	**3.49**	15.30
N296	4	693.3505	n.a.	LC-QTOF	C435	Chr1	107.12	76672459-1	**4.15**	17.90
N740	4	493.2294	n.a.	LC-QTOF	LCS171	Chr1	134.15	83444565-1	**9.05**	34.90
Geraniol	4	51	Terpenoid	GC-SPME	SPME9044	Chr1	139.15	83444565-1	**5.47**	22.9
3-Methyl-2-butenal	2	56	Leucine/Isoleucine derivative	GC-SPME	SPME774	Chr1	151.15	83444565-1	**4.42**	18.9
(E)-4-Oxo-2-hexenal	3	57	Lipid derivative alcohol	GC-SPME	SPME2724	Chr1	152.75	83886078-1	**6.7**	27.2
(Z)-2-Hexenol	2	100	Lipid derivative alcohol	GC-SPME	SPME1864	Chr1	152.75	83886078-1	**18.36**	58.2
1-Penten-3-one	1	51	Lipid derivative alcohol	GC-SPME	SPME536	Chr1	152.75	83886078-1	**7.27**	29.2
2,2,6-Trimethylcyclohexanone	2	69	Cyclic molecule	GC-SPME	SPME4755	Chr1	152.75	83886078-1	**6.69**	27.2
T-2-hexenal	2	84	Lipid derivative alcohol	GC-SPME	SPME1506	Chr1	152.75	83886078-1	**23.68**	67.5
(4Z)-Heptenal	2	83	Lipid derivative alcohol	GC-SPME	SPME2271	Chr1	154.28	84453015-1	**13.11**	46.3
(E)Hex-3-enol	1	70	Lipid derivative alcohol	GC-SPME	SPME1557	Chr1	154.28	84453015-1	**14.49**	49.7
2-ethylthiophene	2	52	Heterocyclic compound	GC-SPME	SPME2001	Chr1	154.28	84453015-1	**14.58**	50
E-2-pentenal	1	85	Lipid derivative alcohol	GC-SPME	SPME826	Chr1	154.28	84453015-1	**17.24**	55.9
Pentanal	1	50	Lipid derivative alcohol	GC-SPME	SPME602	Chr1	154.28	84453015-1	**4.07**	17.6
(Z)-2-pentenol	2	63	Lipid derivative alcohol	GC-SPME	SPME944	Chr1	154.34	84453015-1	**7.86**	31.2
Heptanal	1	71	Lipid derivative alcohol	GC-SPME	SPME2290	Chr1	154.34	84453015-1	**7.91**	31.3
Hexa-2,4-dienal	1	61	Lipid derivative alcohol	GC-SPME	SPME2379	Chr1	154.34	84453015-1	**11.93**	43.2
Penten-3-ol	2	37	Lipid derivative alcohol	GC-SPME	SPME486	Chr1	154.34	84453015-1	**9.41**	36
3-penten-2-one	2	69	Lipid derivative alcohol	GC-SPME	SPME744	Chr1	154.60	84791385-1	**3.53**	15.4
Hexanoic acid, 2-oxo-, methyl ester	4	97	Carboxilic fatty acid	GC-SPME	SPME4900	Chr1	154.60	84791385-1	**3.87**	16.8
β-Acoradien-15-ol	4	63	Sesquiterpene	GC-SPME	SPME10498	Chr1	154.60	84791385-1	**6.56**	26.8
2-ethylfuran	2	49	Lipid derivative alcohol	GC-SPME	SPME621	Chr1	154.86	84851813-1	**12.59**	45
3-Hexenoic acid, (E)	3	99	Fatty acid	GC-SPME	SPME3127	Chr1	154.86	84851813-1	**13.22**	46.6
4-methylpentanol	2	41	Lipid derivative alcohol	GC-SPME	SPME1419	Chr1	154.86	84851813-1	**6.29**	25.8
Hexanal	1	61	Lipid derivative alcohol	GC-SPME	SPME1223	Chr1	154.86	84851813-1	**4.16**	17.9
Pentanol	1	53	Lipid derivative alcohol	GC-SPME	SPME912	Chr1	154.86	84851813-1	**5.03**	21.3
Phenylethanal	1	90	Phenolic	GC-SPME	SPME4932	Chr1	154.86	84851813-1	**5.63**	23.5

Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
(Z)-2-pentenol	4	67	Lipid derivative alcohol	GC-SPME	SPME853	Chr1	155.86	85091184-1	**7.96**	31.5
2-Heptanol	2	70	Alcohol	GC-SPME	SPME2235	Chr1	156.44	85115390-1	**3.52**	15.4
Cis-3-nonen-1-ol	4	45	Lipid derivative alcohol	GC-SPME	SPME5848	Chr1	157.51	85349043-1	**5.68**	23.6
Ethyl Acetate	2	62	Ester	GC-SPME	SPME281	Chr1	157.51	85349043-1	**4.82**	20.4
Phenol	1	37	Phenolic	GC-SPME	SPME3091	Chr1	157.51	85349043-1	**5.32**	22.3
Linalyl oxide	1	72	Terpenoid	GC-SPME	SPME5527	Chr1	159.12	85528366-1	**5.05**	21.3
P-mentha-1,5-dien-8-ol	4	68	Terpenoid	GC-SPME	SPME7758	Chr1	159.12	85528366-1	**4.35**	18.7
Trans linalool furanoxide	4	111	Terpenoid	GC-SPME	SPME5785	Chr1	159.12	85528366-1	**4.15**	17.9
Linalool	2	86	Terpenoid	GC-SPME	SPME5953	Chr1	160.17	85919899-1	**15.39**	51.9
α-terpinol	2	62	Terpenoid	GC-SPME	SPME7947	Chr1	160.17	85919899-1	**8.96**	34.7
1-p-Menthen-9-al	2	84	Terpenoid	GC-SPME	SPME8586	Chr1	160.96	86093836-1	**5.48**	22.9
Quercetin3-O-glucoside	1	463.0887	Flavonoid	LC-QTOF	C554	Chr1	161.49	86171125-1	**4.04**	17.40
Kaempferol-hexose-deoxyhexose,-hexose-coumaroyl	5	901.2403	Flavonoid	LC-QTOF	LCS146	Chr1	167.14	86749853-1	**4.92**	20.80
Laricitrin-deoxyhexose-coumaroyl	5	785.1927	Flavonoid	LC-QTOF	LCS149	Chr1	168.71	86993750-1	**3.64**	15.90
Quercetin-hexose-deoxyhexose,-hexose, -C10H8O3 (176)	2	947.2434	Flavonoid	LC-QTOF	C625	Chr1	169.23	87007323-1	**5.13**	21.60
Quercetin-hexose-deoxyhexose,-hexose,-coumaroyl	2	917.2349	Flavonoid	LC-QTOF	C643	Chr1	169.23	87007323-1	**6.93**	28.00
Delphinidin-deoxyhexose-coumaroyl-hexose	5	917.2350	Flavonoid	LC-QTOF	LCS136	Chr1	169.23	87007323-1	**4.86**	20.60
Camphene	4	92	Terpenoid	GC-SPME	SPME8873	Chr1	171.60	87626733-1	**7.95**	31.4
Isocitricacid	1	191.0191	organic acid	LC-QTOF	C240	Chr1	178.87	89170623-1	**3.47**	15.20
N458	4	623.1622	n.a.	LC-QTOF	C596	Chr2	20.60	33753248-2	**3.04**	13.40
Glucose	3	157	Sugar	GC-TOF	GCTOF6232	Chr2	21.92	33753248-2	**3.05**	13.4
Glucopyranose	4	204	Sugar	GC-TOF	GCTOF9109	Chr2	21.92	33753248-2	**3.65**	15.7
(E)-Geranylacetone	4	109	Acyclic carotenoids	GC-SPME	SPME11158	Chr2	30.20	35155443-2	**3.5**	15.3
Eugenol	1	117	Phenylpropanoid	GC-SPME	SPME10297	Chr2	45.34	37964685-2	**27.22**	72.5
Eugenol-hexose-pentose	1	457.1724	Phenylpropanoid glycosilated volatile	LC-QTOF	LCS132	Chr2	46.34	38096910-2	**12.91**	45.80
Pentadecanal	2	124	Lipid derivative	GC-SPME	SPME13089	Chr2	48.23	39021430-2	**3.06**	13.5
Tridecanal	4	79	Lipid derivative	GC-SPME	SPME13102	Chr2	48.23	39021430-2	**3.03**	13.4
Quercetin-dihexose-deoxyhexose	5	771.1979	Flavonoid	LC-QTOF	LCS79	Chr2	77.96	46518057-2	**3.68**	16.00
Kaempferol-hexose-deoxyhexose,-hexose-coumaroyl	2	901.2407	Flavonoid	LC-QTOF	C728	Chr2	89.51	48407928-2	**3.47**	15.20
Methylbutenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME254	Chr2	91.12	48497154-2	**3.56**	15.5
Acetoxy-tomatine+FA	1	1136.5490	Alkaloid	LC-QTOF	C724	Chr2	94.81	49813323-2	**3.63**	15.80
4-Oxoisophorone	2	152	Cyclic ketone	GC-SPME	SPME6959	Chr3	77.05	46454095-3	**4.79**	20.3
Methylbutenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME254	Chr3	77.05	46454095-3	**3.16**	13.9
N238	4	431.1921	n.a.	LC-QTOF	C416	Chr3	80.31	47146811-3	**4.15**	17.90
1-Nonanol	4	98	Lipid derivative alcohol	GC-SPME	SPME7260	Chr3	93.81	54199481-3	**3.06**	13.5
Geraniol	4	51	Terpenoid	GC-SPME	SPME9044	Chr3	97.33	54199481-3	**3.51**	15.3
Isopentanol	1	54	Leucine/Isoleucine derivative	GC-SPME	SPME702	Chr3	101.33	55993987-3	**5.09**	21.5
Laricitrin-hexose,hexose	3	665.1724	Flavonoid	LC-QTOF	LCS88	Chr3	111.25	57499166-3	**3.16**	13.90
Quercetin-deoxyhexose-feruloyl	5	785.1929	Flavonoid	LC-QTOF	LCS151	Chr3	111.25	57499166-3	**4.07**	17.60
N338	4	793.1805	n.a.	LC-QTOF	LCS102	Chr3	113.37	57730551-3	**4.78**	20.30
Quercetin-3-O-glucoside	1	463.0887	Flavonoid	LC-QTOF	C554	Chr3	114.43	58231574-3	**3.02**	13.40
Kaempferol-3-O-rutinoside	1	593.1501	Flavonoid	LC-QTOF	C585	Chr3	114.43	58231574-3	**3.61**	15.80

Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
Quercetin-hexose-deoxyhexose,-hexose, -C10H8O3 (176)	2	947.2434	Flavonoid	LC-QTOF	C625	Chr3	114.43	58231574-3	**3.02**	13.40
Quercetin-hexose-deoxyhexose, -C12H12O5(236)	2	845.2148	Flavonoid	LC-QTOF	C773	Chr3	114.43	58231574-3	**3.51**	15.40
N458	4	623.1622	n.a.	LC-QTOF	C596	Chr3	114.43	58231574-3	**3.79**	16.50
Quercetin-hexose,-hexose (3,7-O)	5	625.1405	Flavonoid	LC-QTOF	C362	Chr3	114.43	58231574-3	**4.06**	17.50
N429	5	773.1933	n.a.	LC-QTOF	C466	Chr3	114.43	58231574-3	**3.60**	15.70
Kaempferol-3-O-glucoside	1	447.0937	Flavonoid	LC-QTOF	C601	Chr3	115.43	58231574-3	**4.86**	20.60
Quercetin-hexose-deoxyhexose,-pentose	2	741.1871	Flavonoid	LC-QTOF	C473	Chr3	115.43	58231574-3	**6.83**	27.70
Kaempferol-hexose	5	447.0937	Flavonoid	LC-QTOF	LCS125	Chr3	115.43	58231574-3	**4.67**	19.90
Kaempferol-hexose-deoxyhexose,-pentose	2	725.1921	Flavonoid	LC-QTOF	LCS101	Chr3	116.43	58231574-3	**6.64**	27.10
Laricitrin-deoxyhexose-coumaroyl	5	785.1927	Flavonoid	LC-QTOF	LCS149	Chr3	116.43	58231574-3	**4.89**	20.70
Kaempferol3-O-rutinoside-7-O-glucoside	1	755.2031	Flavonoid	LC-QTOF	LCS71	Chr3	117.43	58231574-3	**6.01**	24.80
Kaempferol3-O-rutinoside	5	593.1516	Flavonoid	LC-QTOF	C406	Chr3	117.43	58231574-3	**6.00**	24.80
Kaempferol -hexose-deoxyhexose,-hexose-coumaroyl	5	901.2403	Flavonoid	LC-QTOF	LCS146	Chr3	117.43	58231574-3	**8.92**	34.50
Heptahydroxyflavone,-trimethylether	3	375.0712	Flavonoid	LC-QTOF	LCS53	Chr4	92.78	58658019-4	**3.36**	14.80
P-mentha-1,5-dien-8-ol	4	94	Terpenoid	GC-SPME	SPME7486	Chr4	99.56	59836679-4	**3.23**	14.2
1,3,8-p-Menthatriene	4	78	Terpenoid	GC-SPME	SPME5859	Chr4	99.62	59836679-4	**3.94**	17
Quercetin-hexose-deoxyhexose,-pentose	2	741.1871	Flavonoid	LC-QTOF	C473	Chr4	125.10	62954169-4	**3.14**	13.80
Methyl salicylate	1	104	Phenylpropanoid	GC-SPME	SPME8127	Chr4	125.44	62954169-4	**3.75**	16.3
Hexanoic acid, 2-oxo-, methyl ester	4	97	Carboxilic fatty acid	GC-SPME	SPME4900	Chr5	38.98	3786347-5	**3.27**	14.4
Threitol	5	103	Sugar alcohol	GC-TOF	GCTOF2469	Chr5	38.98	3786347-5	**3.18**	13.9
N-Acetylglutamic acid	1	174	Amino acid	GC-TOF	GCTOF2777	Chr5	38.98	3786347-5	**3.39**	14.7
Pentanal	1	50	Lipid derivative alcohol	GC-SPME	SPME602	Chr5	90.42	62101535-5	**3.34**	14.7
L-Glutamic acid	1	246	Amino acid	GC-TOF	GCTOF3032	Chr5	91.18	62101535-5	**3.26**	14.2
Methylheptenone	4	77	Organic compound	GC-SPME	SPME3411	Chr5	96.75	62456014-5	**3.86**	16.8
Benzophenone	2	181	Phenolic ketone	GC-SPME	SPME12882	Chr6	16.83	28105507-6	**3.08**	13.6
Quercetin-hexose-deoxyhexose,-hexose,-coumaroyl	2	917.2349	Flavonoid	LC-QTOF	C643	Chr6	46.97	34172904-6	**3.07**	13.60
Isopentanol	1	54	Leucine/Isoleucine derivative	GC-SPME	SPME702	Chr6	52.86	35282947-6	**3.41**	14.9
Pentanol	1	53	Lipid derivative alcohol	GC-SPME	SPME912	Chr6	52.86	35282947-6	**3.39**	14.9
4-methylpentanol	2	41	Lipid derivative alcohol	GC-SPME	SPME1419	Chr6	69.43	39198088-6	**3.94**	17
Laricitrin-hexose,hexose	3	665.1724	Flavonoid	LC-QTOF	LCS88	Chr6	89.11	42081887-6	**3.59**	15.70
(E)Hex-3-enol	1	70	Lipid derivative alcohol	GC-SPME	SPME1557	Chr7	21.43	3317534-7	**4.4**	18.9
Sucrose	1	341.1074	Sugar	LC-QTOF	C121	Chr7	70.35	61068415-7	**5.82**	24.20
Fenchene	4	52	Terpenoid	GC-SPME	SPME6919	Chr8	2.34	197152-8	**3.89**	16.9
1,3,8-p-Menthatriene	4	78	Terpenoid	GC-SPME	SPME5859	Chr8	3.34	426863-8	**8.93**	34.5
(E)-Ocimene	2	66	Terpenoid	GC-SPME	SPME4844	Chr8	3.53	426863-8	**15.37**	51.8
2-Carene	2	66	Terpenoid	GC-SPME	SPME3962	Chr8	3.53	426863-8	**22.67**	65.9
Camphene	2	107	Terpenoid	GC-SPME	SPME2895	Chr8	3.53	426863-8	**6.58**	26.8
Limonene	1	65	Terpenoid	GC-SPME	SPME4484	Chr8	3.53	426863-8	**11.01**	40.7
P-Cymen-9-ol	2	132	Terpenoid	GC-SPME	SPME7806	Chr8	3.53	426863-8	**6.7**	27.2
P-Cymol	2	66	Terpenoid	GC-SPME	SPME4397	Chr8	3.53	426863-8	**19.42**	60.2
Pinene	2	74	Terpenoid	GC-SPME	SPME2621	Chr8	3.53	426863-8	**5.15**	21.7

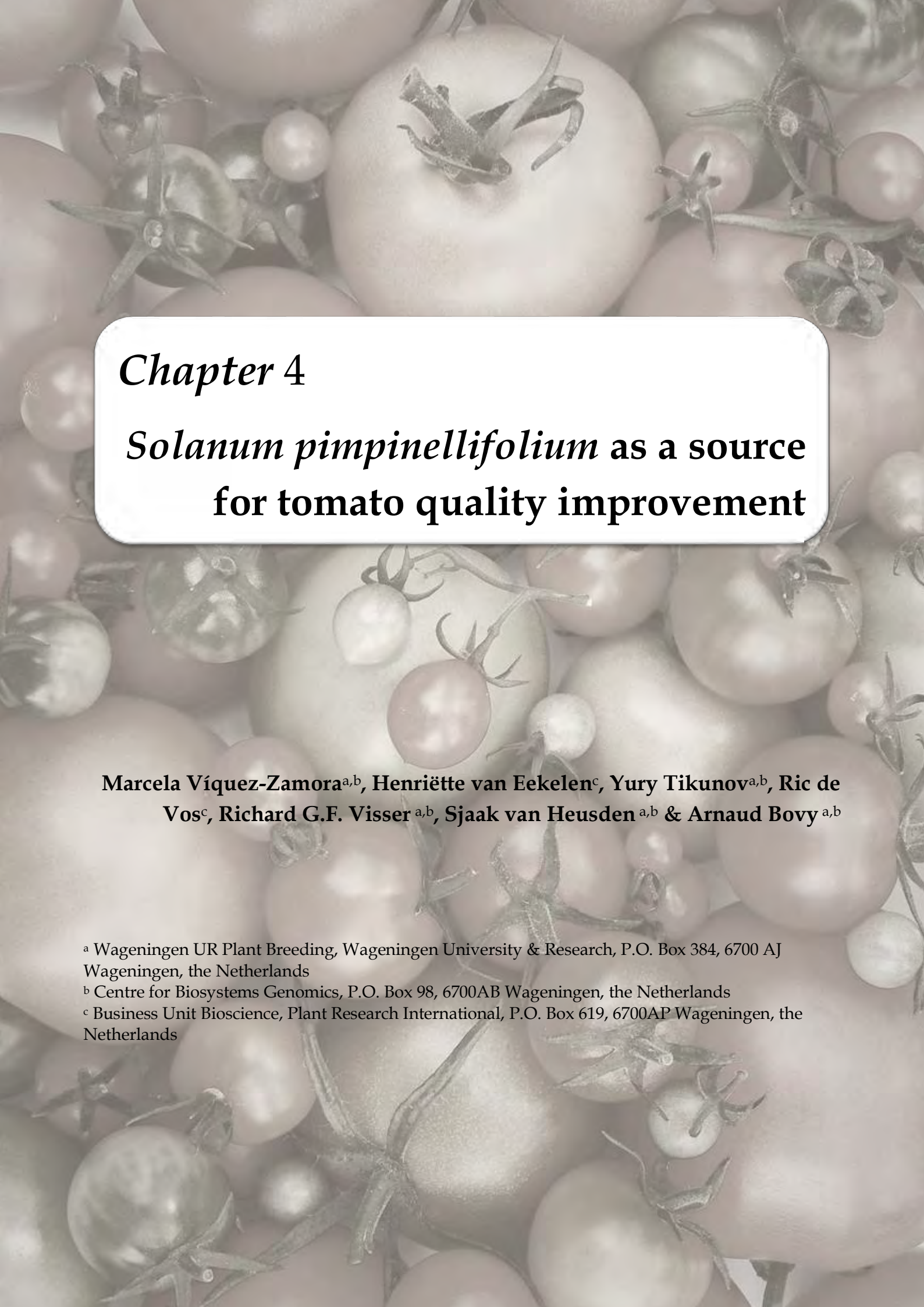
Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
Pinene	4	136	Terpenoid	GC-SPME	SPME6601	Chr8	3.53	426863-8	**13.24**	46.7
P-mentha-1,5-dien-8-ol	4	94	Terpenoid	GC-SPME	SPME7486	Chr8	3.53	426863-8	**9.72**	37
Verbenone	4	108	Terpenoid	GC-SPME	SPME9327	Chr8	3.53	426863-8	**6.78**	27.5
α-Phellandrene	2	107	Terpenoid	GC-SPME	SPME4012	Chr8	3.53	426863-8	**25.74**	70.5
α-Terpinene	2	119	Terpenoid	GC-SPME	SPME4204	Chr8	3.53	426863-8	**20.65**	62.5
β-Phellandrene	2	123	Terpenoid	GC-SPME	SPME4643	Chr8	3.53	426863-8	**27.11**	72.4
O-Cymene	2	50	Terpenoid/Alkylbenzene	GC-SPME	SPME3280	Chr8	4.53	426863-8	**5.98**	24.7
O-Cymene	2	91	Terpenoid/Alkylbenzene	GC-SPME	SPME3329	Chr8	4.53	426863-8	**3.77**	16.4
P-mentha-1,5-dien-8-ol	4	51	Terpenoid	GC-SPME	SPME7353	Chr8	4.53	426863-8	**6.88**	27.9
L-Glutamine	5	145.0609	Amino acid	LC-QTOF	C88	Chr8	70.02	56057431-8	**3.69**	16.10
LycoperosideHorHydroxytomatineIV +FA	5	1094.5402	Alkaloid	LC-QTOF	C567	Chr8	72.64	56878278-8	**3.23**	14.20
Citricacid	1	191.0200	organic acid	LC-QTOF	C291	Chr8	75.46	57194846-8	**3.16**	13.90
3-Caffeoylquinicacid (Chlorogenicacid)	5	353.0876	phenolic acid	LC-QTOF	C395	Chr8	75.46	57194846-8	**3.04**	13.40
LycoperosideHorHydroxytomatineI	5	1048.5354	Alkaloid	LC-QTOF	C765	Chr8	83.37	57595067-8	**3.58**	15.60
Dehydrotomatine (S)I	5	1076.5258	Alkaloid	LC-QTOF	C652	Chr8	88.46	58825288-8	**3.01**	13.30
Benzylalcohol-hexose-pentose	2	401.1456	Phenolic glycosilated volatile	LC-QTOF	C380	Chr8	94.51	59977315-8	**3.22**	14.20
Protocatechuicacid	5	153.0204	phenolic acid	LC-QTOF	C311	Chr8	94.51	59977315-8	**3.65**	15.90
α-tomatin	1	1078.5415	Alkaloid	LC-QTOF	C734	Chr8	97.69	60673054-8	**3.32**	14.60
Tomatidinedihexosedipentose +FA	1	1048.5322	Alkaloid	LC-QTOF	C749	Chr8	97.69	60673054-8	**3.28**	14.40
Glycoalkaloid	5	1344.6145	Alkaloid	LC-QTOF	C816	Chr8	97.69	60673054-8	**3.32**	14.60
β-Damascenone	2	190	Cyclic carotenoids	GC-SPME	SPME10714	Chr9	6.52	1303826-9	**4.2**	18.1
(Z)-2-Hexenol	2	100	Lipid derivative alcohol	GC-SPME	SPME1864	Chr9	8.02	1303826-9	**4.34**	18.6
2-ethylthiophene	2	52	Heterocyclic compound	GC-SPME	SPME2001	Chr9	8.02	1303826-9	**3.08**	13.6
E-2-pentenal	1	85	Lipid derivative alcohol	GC-SPME	SPME826	Chr9	8.02	1303826-9	**3.69**	16.1
L-Aspartic acid	1	100	Amino acid	GC-TOF	GCTOF2612	Chr9	38.88	4113674-9	**4.21**	17.9
Phenylethanol	1	37	Aromatic alcohol	GC-SPME	SPME6326	Chr9	52.47	57807588-9	**6.15**	25.3
Quercetin-hexose,-hexose (3,7-O)	5	625.1405	Flavonoid	LC-QTOF	C362	Chr9	63.94	60746121-9	**4.29**	18.40
Kaempferol-hexose-deoxyhexose,-hexose-coumaroyl	2	901.2407	Flavonoid	LC-QTOF	C728	Chr9	66.64	61256180-9	**3.16**	13.90
N152	4	443.1924	n.a.	LC-QTOF	LCS41	Chr9	67.16	61607962-9	**6.97**	28.20
(E)-Geranylacetone	4	109	Acyclic carotenoids	GC-SPME	SPME11158	Chr9	69.21	62098389-9	**3.05**	13.5
Laricitrin-deoxyhexose-coumaroyl	5	785.1927	Flavonoid	LC-QTOF	LCS149	Chr9	70.33	62248589-9	**3.16**	13.90
Kaempferol-3-O-rutinoside-7-O-glucoside	1	755.2031	Flavonoid	LC-QTOF	LCS71	Chr9	70.85	62423755-9	**9.64**	36.70
Quercetin3-O-rutinoside-7-O-glucoside	2	771.1974	Flavonoid	LC-QTOF	C365	Chr9	70.85	62423755-9	**12.99**	46.00
Quercetin-dihexose-deoxyhexose-pentose	3	903.2413	Flavonoid	LC-QTOF	LCS54	Chr9	70.85	62423755-9	**12.84**	45.60
Quercetin-dihexose-deoxyhexose-pentose	5	903.2408	Flavonoid	LC-QTOF	C353	Chr9	70.85	62423755-9	**9.99**	37.80
Kaempferol3-O-rutinoside	5	593.1516	Flavonoid	LC-QTOF	C406	Chr9	70.85	62423755-9	**10.15**	38.20
Kaempferol-hexose-hexose-hexose	5	771.1974	Flavonoid	LC-QTOF	LCS52	Chr9	70.85	62423755-9	**12.04**	43.50
Isorhamnetin-hexose-hexose (3-O)	3	639.1605	Flavonoid	LC-QTOF	LCS111	Chr9	76.87	62896769-9	**3.06**	13.50
Isorhamnetin-hexose,-hexose (3,7-O)	5	639.1574	Flavonoid	LC-QTOF	LCS107	Chr9	76.87	62896769-9	**3.53**	15.40
Hexanol-pentose-hexose	3	395.1934	Lipid glycosilated volatile	LC-QTOF	C572	Chr9	104.21	65923428-9	**3.86**	16.80
Glycoalkaloid	5	1344.6145	Alkaloid	LC-QTOF	C816	Chr9	104.21	65923428-9	**3.16**	13.90

Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
Hexanoic acid, 2-oxo-, methyl ester	4	97	Carboxilic fatty acid	GC-SPME	SPME4900	Chr9	105.40	66074223-9	**3.45**	15.1
β-Acoradien-15-ol	4	63	Sesquiterpenes	GC-SPME	SPME10498	Chr9	105.40	66074223-9	**3.13**	13.8
N427	5	773.1922	n.a.	LC-QTOF	C480	Chr9	106.40	66164836-9	**4.63**	19.80
5-Caffeoylquinicacid	1	353.0875	phenolic acid	LC-QTOF	C337	Chr10	1.00	536207-10	**3.03**	13.40
4-Caffeoylquinicacid	5	353.0883	Acid	LC-QTOF	C361	Chr10	1.00	536207-10	**8.44**	33.00
Citricacid	1	191.0200	organic acid	LC-QTOF	C291	Chr10	3.62	536207-10	**3.72**	16.20
N143	5	402.9155	n.a.	LC-QTOF	C29	Chr10	18.65	2312299-10	**3.09**	13.70
2-methylbutanol	1	45	Leucine/Isoleucine derivative	GC-SPME	SPME726	Chr10	23.33	2312299-10	**3.71**	16.2
Butanol	2	55	Alcohol	GC-SPME	SPME429	Chr10	23.33	2312299-10	**3.47**	15.2
Methylbutenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME254	Chr10	23.33	2312299-10	**3.66**	16
Caffeicacid	5	179.0362	organic acid	LC-QTOF	C295	Chr10	25.33	2527359-10	**6.42**	26.30
Kaempferol-hexose-deoxyhexose,-hexose-coumaroyl	2	901.2407	Flavonoid	LC-QTOF	C728	Chr10	25.52	2527359-10	**4.22**	18.20
N71	4	337.0771	n.a.	LC-QTOF	LCS20	Chr10	42.90	16020522-10	**3.46**	15.20
β-Damascenone	2	190	Cyclic carotenoids	GC-SPME	SPME10714	Chr10	44.48	4324132-10	**8.36**	32.8
3-methyl-2-butenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME1019	Chr10	47.74	23394403-10	**3.52**	15.4
Limonene	1	65	Terpenoid	GC-SPME	SPME4484	Chr10	49.74	seq-rs5544-10	**3.84**	16.7
Coumaroylquinicacid	1	337.0940	phenolic acid	LC-QTOF	C414	Chr10	62.21	59477572-10	**4.51**	19.30
Coumaroylquinicacid	5	337.0936	phenolic acid	LC-QTOF	C462	Chr10	63.21	59477572-10	**4.56**	19.40
4-Oxoisophorone	2	152	Cyclic ketone	GC-SPME	SPME6959	Chr10	95.36	62966801-10	**3.18**	14
2-ethylhexanol	2	81	Lipid derivative alcohol	GC-SPME	SPME4243	Chr11	0.00	4106861-11	**9.81**	37.2
2-methylbutanol	1	45	Leucine/Isoleucine derivative	GC-SPME	SPME726	Chr11	0.00	4106861-11	**10.82**	40.2
3-methyl-2-butenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME1019	Chr11	0.00	4106861-11	**3.2**	14.1
Butanol	2	55	Alcohol	GC-SPME	SPME429	Chr11	0.00	4106861-11	**11.66**	42.5
Geranial	2	137	Terpenoid	GC-SPME	SPME9389	Chr11	0.00	4106861-11	**3.3**	14.5
Heptanol	2	68	Lipid derivative alcohol	GC-SPME	SPME2968	Chr11	0.00	4106861-11	**4.78**	20.3
Hexanol	1	70	Lipid derivative alcohol	GC-SPME	SPME1892	Chr11	0.00	4106861-11	**6.12**	25.2
Isopentanol	1	54	Leucine/Isoleucine derivative	GC-SPME	SPME702	Chr11	0.00	4106861-11	**7.54**	30.1
Methylbutenol	2	68	Leucine/Isoleucine derivative	GC-SPME	SPME254	Chr11	0.00	4106861-11	**3.48**	15.2
Pentanol	1	53	Lipid derivative alcohol	GC-SPME	SPME912	Chr11	0.00	4106861-11	**7.3**	29.3
Benzylalcohol-hexose-pentose	2	401.1456	Phenolic glycosilated volatile	LC-QTOF	C380	Chr11	0.00	4106861-11	**5.60**	23.30
Xylose	1	103	Sugar	GC-TOF	GCTOF3183	Chr11	3.71	4629970-11	**3.16**	13.8
2-Heptanol	2	70	Alcohol	GC-SPME	SPME2235	Chr11	8.86	5174517-11	**4.23**	18.2
Caryophyllene	2	124	Sesquiterpenes	GC-SPME	SPME11232	Chr11	11.07	5279605-11	**3.58**	15.6
Caryophyllene oxyde	3	161	Sesquiterpenes	GC-SPME	SPME12691	Chr11	11.07	5279605-11	**3.78**	16.4
Humulene	2	146	Sesquiterpenes	GC-SPME	SPME11646	Chr11	11.07	5279605-11	**3.11**	13.7
Benzaldehyde	1	63	Phenolic/Aromatic aldehyde	GC-SPME	SPME2995	Chr11	11.86	5329725-11	**4.02**	17.4
β-Ionone	2	145	Cyclic carotenoids	GC-SPME	SPME11902	Chr11	18.82	23203939-11	**3.13**	13.8
O-Feruloylquinicacid	3	367.1035	phenolic acid	LC-QTOF	C449	Chr11	21.02	47009022-11	**3.90**	16.90
3-O-Feruloylquinicacid	5	367.1040	Acid	LC-QTOF	C491	Chr11	21.02	47009022-11	**3.42**	15.00
Methyl salicylate	1	104	Phenylpropanoid	GC-SPME	SPME8127	Chr11	46.45	50710636-11	**3.74**	16.3
N50	4	609.1888	n.a.	LC-QTOF	LCS17	Chr11	56.66	51347236-11	**3.75**	16.30
Sucrose	1	341.1074	Sugar	LC-QTOF	C121	Chr11	73.08	52635542-11	**4.26**	18.30
Protocatechuicacid	5	153.0204	phenolic acid	LC-QTOF	C311	Chr11	73.08	52635542-11	**3.68**	16.00
1-Nonanol	4	98	Lipid derivative alcohol	GC-SPME	SPME7260	Chr12	48.99	6238531-12	**3.31**	14.5

Putative Compound Identification	Id. Level*	Specific negative ion, m/z**	Compound class	Platform -MS	Compound ID	Chr	Position (cM)	Marker (closest)	LOD	% Expl.
Octanol	2	71	Lipid-derived	GC-SPME	SPME5262	Chr12	49.12	6238531-12	**4.2**	18.1
Quercetin3-O-rutinoside	1	609.1450	Flavonoid	LC-QTOF	C512	Chr12	49.12	6238531-12	**4.04**	17.50
Kaempferol3-O-rutinoside	1	593.1501	Flavonoid	LC-QTOF	C585	Chr12	51.70	44987172-12	**3.42**	15.00
Kaempferol-3-O-glucoside	1	447.0937	Flavonoid	LC-QTOF	C601	Chr12	51.70	44987172-12	**4.28**	18.40
N429	5	773.1933	n.a.	LC-QTOF	C466	Chr12	51.70	44987172-12	**3.37**	14.80
Kaempferol-hexose	5	447.0937	Flavonoid	LC-QTOF	LCS125	Chr12	51.70	44987172-12	**4.29**	18.40
Petunidin-deoxyHexose-coumaroyl-hexose +H2O	5	949.2614	Flavonoid	LC-QTOF	LCS89	Chr12	52.32	44987172-12	**3.21**	14.10
Guaiacol	1	37	Phenylpropanoid	GC-SPME	SPME5805	Chr12	74.45	62420692-12	**5.65**	23.5

*Annotation level: 1=Identified compounds. 2=Putatively annotated compounds (e.g. without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries). 3=Putatively characterized compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class). 4. Unknown compounds—although unidentified or unclassified these metabolites can still be differentiated and quantified based upon spectral data. 5. Unknown compounds—similar mass to a putatively characterized compound.

**Compounds analysed using LC-QTOF platform represented by measured accurate masses of corresponding negatively charged parent molecule ions or their formic acid adducts (denoted by +FA). Volatile compounds and primary metabolites measured by SPME-GC and GC-TOF, respectively, represented by selected nominal negative mass ion fragments picked automatically by MSClust software.



Chapter 4

Solanum pimpinellifolium as a source for tomato quality improvement

Marcela Víquez-Zamora[a,b], Henriëtte van Eekelen[c], Yury Tikunov[a,b], Ric de Vos[c], Richard G.F. Visser [a,b], Sjaak van Heusden [a,b] & Arnaud Bovy [a,b]

[a] Wageningen UR Plant Breeding, Wageningen University & Research, P.O. Box 384, 6700 AJ Wageningen, the Netherlands
[b] Centre for Biosystems Genomics, P.O. Box 98, 6700AB Wageningen, the Netherlands
[c] Business Unit Bioscience, Plant Research International, P.O. Box 619, 6700AP Wageningen, the Netherlands

Abstract

Tomato breeding has always been focused on yield, resistance and appearance. Until recent years, almost no attention was given to flavour or nutritional value. The wild species *S. pimpinellifolium* represents a source to expand the possibilities to improve several quality traits of tomatoes. We applied high throughput genotyping technologies and three different platforms for metabolic profiling to a recombinant inbred line population developed from a cross between *S. pimpinellifolium* and *S. lycopersicum* var. Moneymaker. We found clear QTL metabolite hotspots on chromosomes 1 and 10. According to our results, to increase the antioxidant properties of tomato, it would be more beneficial if the region between 71-87 Mb on chromosome 1 originates from Moneymaker while regions of chromosome 6 (35-44 Mb), the bottom of chromosome 10 (~44.3 Mb) and chromosome 12 (~48 Mb) would be of *S. pimpinellifolium* origin. The above-mentioned region on chromosome 6 also affected the concentration of malic acid in the fruits. Sugars can be increased by combining the wild alleles on chromosome 2 (~41.7Mb) for sucrose and chromosome 10 (~1.7 Mb) with the Moneymaker allele on the hotspot region of chromosome 1 for fructose and chromosome 4 (~55Mb) for glucose. Off flavour regions that should be avoided in crosses with *S. pimpinellifolium* are the ones on top of chromosome 1 and on chromosome 9 around 65Mb where we found regions associated with the presence of the compounds putrescine and dimethyl disulfide. An aromatic boost to the fruits can be given by introgressing the *S. pimpinellifolium* of chromosome 8 which results in an increase of phenolic VOCs. In general, our results give an insight in the physical positions of metabolite related QTLs that could be used by breeders to improve tomato quality.

Keywords: Metabolomics, flavonoids, phenylpropanoids, QTLs

Introduction

Solanum lycopersicum is one of the top commodities in the world. The global production of tomatoes has gone up to 164 million tonnes per year (FAO, 2014). The available genetic resources in certain places and time, the history of breeding activities and different environmental conditions made the plants survive or the growers select and breed for specific traits (Nakazato *et al.*, 2008 and 2010; Moyle, 2010; Gross and Olsen, 2010). Selection for the fresh market was generally based on traits like yield, size, resistance and appearance (Causse *et al.,* 2013). Major attention was given to increase in production and almost none to flavour or nutritional value (Bennett, 2012). Domestication and breeding goals have narrowed the variation within the cultivated germplasm. Thus, plant breeders are trying to find ways to enlarge genetic variability in order to expand breeding opportunities.

Solanum pimpinellifolium represents a source for tomato quality improvement (Lin *et al.,* 2014). In different accessions of this wild relative of tomato favourable traits have been identified. Examples are texture, sweetness (Voorrips *et al.,* 2000), fruit shape (Rodríguez *et al.,* 2011), colour, seed quality (Khan *et al.,* 2012), flowering behaviour, curliness in leaves (Lippman and Tanksley, 2001), shelf life (Pereira da Costa *et al.,* 2013), resistance to pests (Alba *et al.,* 2009; Salinas *et al.,* 2013) and resistance to diseases (Sharma *et al.*, 2008; Víquez-Zamora *et al.,* 2014).

Knowledge about associations of quality traits and specific genomic regions is the starting point to efficiently introgress alleles of interest into commercial cultivars (Sim *et al.,* 2012b). Currently high throughput genotyping and metabolomics platforms are available to characterize populations. Metabolic profiling of wild species (Schauer *et al.,* 2005), recombinant inbred lines (RIL) populations between cherry and large-fruited tomatoes (Saliba-Colombani *et al.,* 2001; Lippman *et al.,* 2007), admixtures with wild germplasm (Ranc *et al.,* 2012) as well as introgression lines (ILs) with fragments of *S. pennellii* wild chromosomes within the background of a commercial tomato (Schauer *et al.,* 2006; Bermudez *et al.,* 2008; Perez-Fons *et al.,* 2014) have been subject of extensive metabolic profiling and phenotyping. These studies demonstrated the benefits of metabolic profiling approaches to reveal loci linked to nutritional or organoleptic quality. Consumers worldwide are making demands about the quality of the products. Meanwhile, researchers are trying to elucidate the biochemical pathways underlying quality-related metabolites and isolate the key genes controlling them (Causse *et al.,* 2001; Fernie *et al.,* 2006; Bovy *et al.,* 2007; Kamenetzky *et al.*, 2010, de Vos *et al.,* 2011; Klee and Tieman., 2013). This knowledge is required to improve the metabolic composition of tomato and meet the demands of better informed consumers worldwide.

As the basic taste qualities are considered to be sweet, salty, sour and bitter. Sugars and organic acids are essential contributors to the taste of tomatoes. Other experts point to umami, associated with the amino acid glutamate, as a fifth basic taste quality (Mouritsen and Styrbæk, 2014). In addition, through interactions between taste and retronasal olfaction different signals are generated in the brain (Small *et al.,* 2004). By comparing large numbers of samples, researchers try to link flavour preferences with sensory attributes and metabolite content (Bartoshuk *et al.,* 2013). In our research, we applied genomics and metabolomics to expand breeders' possibilities to develop better quality tomatoes. Three different platforms were used to screen a RIL population for presence/absence and concentration differences of these taste related metabolites. In this paper we describe the variation in metabolite content and concentration in fruits and we determined the regions harbouring the genes underlying these differences. In this way we identified those regions in *S. pimpinellifolium* that might be of interest for quality improvement of tomato.

Material and Methods

Plant material

The RIL population (Voorrips *et al.*, 2000), derived from a cross between *S. pimpinellifolium* G1.1554 and the elite cultivar *S.* lycopersicum cv Moneymaker, was sown in greenhouses from Unifarm, Wageningen University and Research Centre, the Netherlands. Ninety four RIL lines were grown in triplicate and six lines in duplicate. From each plant, five to ten ripened fruits were collected, frozen in liquid nitrogen, ground to fine powder, pooled, weighed and stored at -80 °C.

DNA extraction

Genomic DNA was extracted from young leaflets with a CTAB based protocol (Steward and Via, 1993; Kabelka *et al.,* 2002). Young plant material was frozen and ground with a Retsch 300 mm shaker (Retsch BV, Ochten, The Netherlands) using 1 ml micronic tubes (Micronic BV, Lelystad, The Netherlands). The DNA pellets were washed with 76% EtOH, 10 mM NH_4Ac before re-suspending in TE buffer.

Metabolic profiling

LC-QTOF

Liquid Chromatography (LC) in combination with Mass Spectrometry (MS) was used to detect semi-polar compounds such as flavonoids, alkaloids, phenylpropanoids, saponins, phenolic acids, polyamines and products thereof (De Vos *et al.,* 2007). To each sample of 100

mg frozen ripened-fruits powder, 300 µl of 75% aqueous methanol with 0.133% formic acid were added. The samples were vortexed and after a treatment in an ultrasonic bath for 10 minutes centrifuged (~15000 rpm) for 10 minutes. The supernatants were transferred to glass vials via 0.2 µm polytetrafluoroethylene (PTFE) filters. Quality control samples were prepared by pooling fruit material from the parents of the population and analyzed after every 15 samples to monitor the performance of the system. Extracts were analysed via reversed phase liquid chromatography coupled to a photodiode array detector, using C_{18}-reversed phase in a Quadrupole-time-of-flight Ultima V4.00.00 mass spectrometer equipped with an electrospray ionization (ESI) source, and separate lock mass spray inlet system in negative mode for deprotonated molecular masses. LC profiles were handled for data processing and mass peak extraction with MetAlign (see section data processing).

GC-TOF

Gas Chromatography (GC) coupled with electron impact Time of Flight (TOF) can detect a large variety of non-volatile primary metabolites. The detection of a large variety of non-volatile primary metabolites was according to Lisec *et al.* (2006). This technique targets particularly polar compounds such as amino acids, sugars and organic acids. From tomato frozen powder, 100 mg was mixed with 1 ml of ribitol in MeOH. Samples were sonicated for 30 min and then centrifuged. A half millilitre of the supernatant was transferred to a separate vial and 450 µl of water and 250 µl of chloroform were added. After vortexing, samples were incubated for 5 min at room temperature, and then centrifuged at maximum speed for 10 min. The supernatant is enough to put 50 µl into two separate 2 ml glass vials (one as backup sample). The vials with the derivatized extracts were dried overnight in a speedvac. Finally vials were capped under argon with magnetic crimp caps and used for GC-TOF analysis.

SPME-GC

Solid-Phase Microextraction (SPME)-GC was used for the analysis of volatiles (Tikunov *et al.*, 2005). In brief, 750 mg of frozen powder of fruits were weighed and transferred to a 5 ml screw-cap glass vial, closed and incubated for 10 min at 30°C. Afterwards, 750 µl of a EDTA-NaOH aqueous solution was added to get a final EDTA concentration of 50 mM. Immediately, 1100 mg of solid $CaCl_2$ made the final concentration 5 M. Vials were closed and the samples sonicated for 5 minutes. From the solution, a pulp aliquot of 1 ml was transferred into a 10 ml crimp cap vial (Waters) and capped for the SPME-GC analysis. A combi PAL autosampler (CTC Analytics) was used for sampling. Under continuous shaking and heating (50 °C), one vial at the time released its headspace to a 65-mm polydimethylsiloxane-divenylbenzene SPME fiber (Supelco) for 20 min. After desorption of the fibre into the injection port, volatiles were transferred to the GC and separated in an HP-5 column with

helium as carrier gas. Quality control samples from pooled material were included in duplicate in every batch of 12 samples.

Soluble solids content

Soluble solids content was measured with thawed tomato powder using a digital ATAGO refractometer. The results were used to calculate brix degrees, one degree BRIX represents approximately 1 gram of soluble solids (predominantly sugars) in 100 grams of solution.

Data processing

Mass peaks were handled according to Tikunov *et al.* (2005). Each dataset was processed using MetAlign (www.metalign.nl) for baseline correction, noise estimation, and ion-wise mass spectral alignment. The noise was decreased depending on the compound detection threshold of each machine. Prepared data were analysed with MSClust software to extract compounds mass spectra and to reduce the data (Tikunov *et al.*, 2012). Then, raw amplitudes of metabolites were processed in GenStat 17th edition (www.vsni.co.uk/software/genstat/) to extract best linear unbiased estimators (BLUEs) and to correct for population structure.

The putative identification of metabolites was based upon their spectra, molecular weight and fragmentation patterns. For LC data, compound characteristics were analysed and compared using the Dictionary of Natural Products (http://dnp.chemnetbase.com) and in-house tomato metabolite databases. For the GC data, the NIST Mass Spectral Search Program v2.0 was used for identification.

Genomic profiling, genotype by sequencing and genetic analysis

Genomic profiling of the parental lines and the RIL population was done by a custom made Infinium array as described by Víquez-Zamora *et al.* (2013). Additionally, a subset of 59 lines was selected for resequencing (Víquez-Zamora *et al.,* 2014). The information of the sequences was embedded into JBrowse 1.11.1 (Skinner *et al.,* 2009) to visualize the structural variants. The SL2.40 tomato genome assembly and ITAG 2.31 tomato genome annotation was loaded together with the BAM and VCF files of the 59 genotypes.

Linkage maps were made using JoinMap® version 4.1 (Kyazma©; Van Ooijen, 2011, Víquez-Zamora *et al.,* 2013). Physical maps with mQTLs distribution were depicted with MapChart 2.2 (Voorrips, 2002). QTL mapping and permutation tests were performed using MapQTL® 6 (Van Ooijen, 2009). The program MQ2 (Chibon *et al.,* 2013) was used to extract all significant QTLs from MapQTL®6 in one go. The application Marker2sequence (Chibon *et al.,* 2012) was used to mine regions for candidate genes (Finkers, 2009: http://www.plantbreeding.wur.nl/BreeDB/).

Results and discussion

The aim of this QTL study was to unravel the genetic regions underlying metabolic differences between cultivated tomato and its wild relative, *S. pimpinellifolium*. To achieve this, we analysed fruits from almost hundred lines with three different platforms for untargeted metabolic profiling of primary metabolites (GC-TOF), secondary metabolites (LC-QTOF) and volatile organic compounds (VOCs; SPME-GC). Using the MetAlign and MSClust algorithms, mass ions were grouped into mass clusters, representing reconstructed metabolites. From each mass cluster, the most abundant ion was selected as the representative of each putative metabolite and subsequently used for further analysis. The first QTL screening was done with a total of more than one thousand putative metabolites. Two metabolite QTL hotspots were identified (on chromosome 1 and 10), where more than one hundred metabolite QTLs (mQTLs) clustered (Fig. 1). The genome sequences of 59 RILs (http://www.tomatogenome.net/ril_variants) were used to saturate these hotspot regions to make a more precise *in silico* mapping as previously reported (Víquez-Zamora *et al.*, 2014; Additional Fig. 1). We included 79 and 69 SNPs in the regions between 69.7-83.4 Mb of chromosome 1 and between 0.05-5.0 Mb on chromosome 10. Additionally, another 27 markers were added in order to improve the marker coverage and to saturate the rest of chromosome 10 (Fig. 2). A large proportion of the mQTLs in these two hotspots consisted of different (forms of) phenylpropanoids and flavonoids, suggesting that these regions either encode pathway regulators, structural genes acting at early steps of these metabolic pathways or modifying genes, such as glycosyl-, acyl-, or methyltransferases, influencing the decoration of these compounds.

Other mapping studies with physical and chemical traits in tomato such as the study by Saliba-Colombani *et al.*, 2001 did not report mQTLs on chromosomes 1 or 10 even though one of the parents of their population was considered wild: *S. lycopersicum* var. *cerasiforme*. The mapping populations of three independent crosses of a cherry tomato and a large fruit cultivar published by Zanor *et al.* (2009) showed a mQTL on the long arm of chromosome 1, but they did not refer to phenylpropanoids. Lin *et al.* (2014) evaluated tomato evolution characteristics of 360 accessions including 53 *S. pimpinellifolium*. They found large differences related to the origin of chromosome 1 and the influence of *SlMYB12* in getting the pink colour of tomatoes. From their figure it looks like there were more QTLs in this region. However, they did not discuss further variation in their manuscript rather than the pink colour and its relation to the accumulation of naringenin chalcone. Schauer *et al.* (2005) reported differences between the accession LA3475 of *S. lycopersicum* and the accession LA1589 of *S. pimpinellifolium* related to the accumulation of shikimic acid in leaves and chlorogenic acids in pericarp of fruits. The presence of mQTL hotspots on chromosomes 1 and 10 related to the

phenylpropanoid pathway might be due to intrinsic genetic differences between *S. lycopersicum* and *S. pimpinellifolium*.

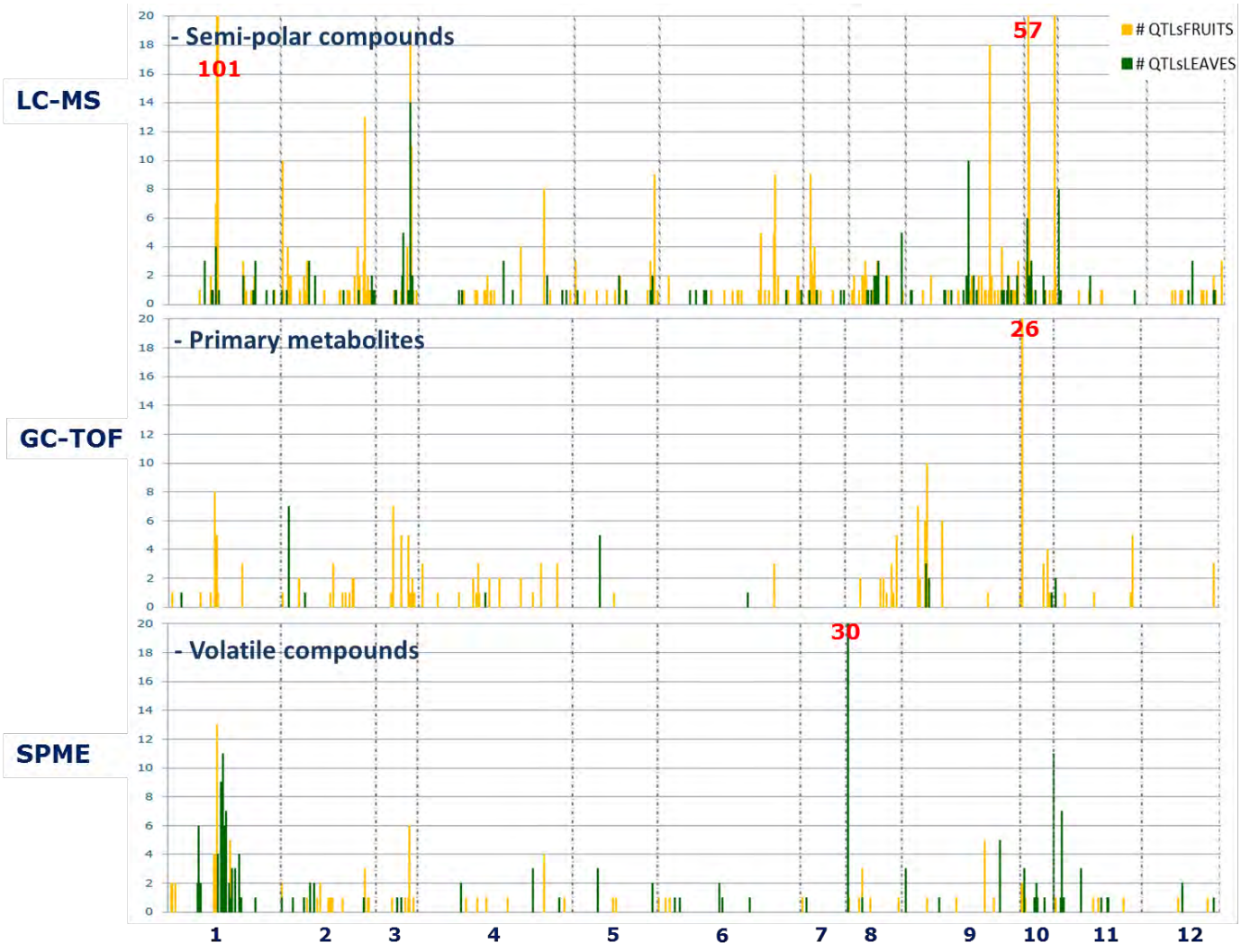


Figure 1. Comparison of the number of mQTLs found in a first screening of all metabolites in fruits (in yellow) and leaves (in green; Víquez-Zamora *et al.,* 2014). The numbers on the X-axis are the tomato chromosomes and the width is their relative length. The values on the Y-axis represent the number of QTLs. When there were more as 20 the actual numbers are given in red.

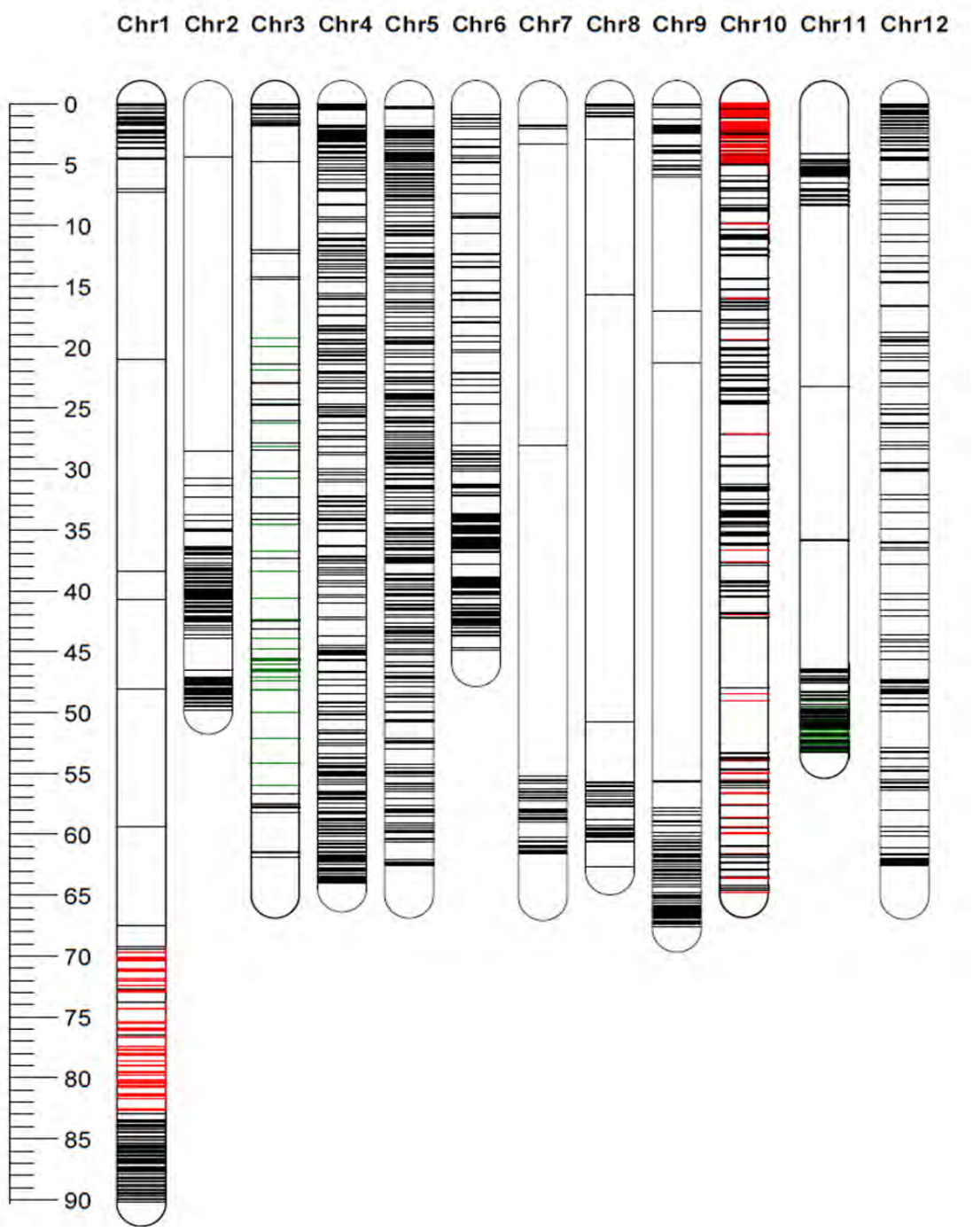


Figure 2. Physical distribution of SNPs covering the tomato genome (version SL2.4). Black markers originate from a custom made SNP array (VíquezZamora *et al.,* 2013) and coloured markers from Genotyping By Sequencing (GBS) in a subset of RILs (green scored for the study by VíquezZamora *et al.,* 2014 and red for this study).

For the detailed mapping analysis, we restricted our analyses to known compounds with a putative identification. Eighty eight primary metabolites found with GC-TOF were annotated and for 24 of these compounds significant QTLs were detected. The low number of QTLs for primary metabolites was expected because of the high level of similarity between *S. lycopersicum* Moneymaker and *S. pimpinellifolium* G1.1554 (Víquez-Zamora *et al.,* 2014) rather than greater differences with other species like *S. pennellii* (Toubiana *et al.*, 2012). Furthermore, over 100 secondary metabolites could be annotated and 74 mQTLs were found of which many were related to the phenylpropanoid pathway and the alkaloid composition of the fruits. Finally QTLs were identified for 38 of the measured 101 volatiles (Table 1). Thanks to the availability of a reference genome sequence (Tomato Genome Consortium, 2012) and the use of SNP markers with known physical positions, our results allowed us to link genetic data to physical genomic positions and to define regions of importance to breeders (possible metabolites of interest are described in Fig. 3).

Sugars

Sugars are essential elements for the taste of fresh market tomatoes. But high levels of sugars are also an advantage for processing tomatoes (Bennett, 2011). In general, for humans, fructose is perceived sweeter than sucrose which at a time is perceived sweeter than glucose (Baldwin, 2002). Schauer *et al.* (2005) found that fructose and glucose concentrations were higher in some accessions of *S. pimpinellifolium* than in *S. lycopersicum* genotypes. In our studies a QTL for fructose content was found in the hot-spot region between 69.7 and 83.4 Mb on chromosome 1 and for another isomer of fructose on the top of chromosome 10. The mQTL for fructose on chromosome 1 co-localised with many other QTLs like the ones for quinic acid and caffeoylquinic acid. The precise location of the three QTLs on the bottom of chromosome 1 could be determined more precisely through the *in silico* improvement of the map with additional markers obtained by GBS. The mQTL hotspot on the top of chromosome 10 (1.7 - 1.9 Mb) was associated with three sugar related compounds (fructose, threitol and lactose). The *S. pimpinellifolium* allele of the mQTL hotspot on the top of chromosome 10 leads to 4 times more fructose. This QTL will have a large effect on sugar perception and will result in sweeter tomatoes.

We also found a QTL for higher sucrose (GC-TOF) and also one for Brix on chromosome 2. The variation in Brix gave an average increase of 1.4 °Brix with the alleles of *S. pimpinellifolium*. The peak QTL marker for Brix was near 40.6 Mb and this is close to the one for sucrose at 41.7 Mb. This QTL region was previously reported by Bernacchi *et al.* (1998), Saliba-Colombani *et al.* (2001), Zanor *et al.* (2009) and Ranc *et al.* (2012). In that chromosome 2 region, QTLs named ssc2.2 and suc2.2 were described. With our results we can pinpoint

the genetic position to the physical position 39-42 Mb on chromosome 2. Within that region, a good candidate to explore further would be the gene for sucrose synthase (Solyc02g081300) located in the region between position 39893591 and 39897878 bp. Chromosome 2 harbours also genes involved in fruit weight (the gene *fw2.2*) and locule number (*lcn2.1*) (Muños *et al.,* 2011; Ranc *et al.,* 2012; Lin *et al.,* 2014). Therefore, attention should be paid to avoid linkage drag with the introgression of the QTL improving sugar content.

For glucose, a QTL was found on chromosome 4 (52-57 cM) in a region with QTLs related to myo-inositol, allantoin and N-acetylglutamic acid. Others (Causse *et al.,* 2004; Baxter *et al.,* 2005) reported this Brix-related QTL on chromosome 4 as well. Possible candidate genes in this QTL are the high-affinity glucose transporter (Solyc04g074070) and 6-phosphofructokinase 2 (Solyc04g072580). The LOD2 interval from 54.8-55.1 Mb however does not contain the fructokinase. Sugar levels can be increased by combining the wild allele on chromosome 2 (~41.7 Mb) for sucrose and the Moneymaker allele on chromosome 4 (~55 Mb) for glucose.

Organic acids

Apart from sugars, organic acids are essential elements of tomato flavour. The most abundant acids in tomato are citric and malic acid. The citric to malic acid ratio generally is >2 and it has been reported that citric acid enhances the sweetness effect of glucose more than that of fructose (Petro-Turza, 1986). Baldwin *et al.* (2008) studied the effect of adding acids to tomato puree. They found that adding acids decreased sweet and bitter taste, green and floral aromas, while enhancing sour, tropical, ripe tomato and citrus taste. In our study, no mQTLs for citric acid concentration were found. However, one for citramalic acid was found on chromosome 12 in the region around 62 Mb. In addition, we found a QTL for malic acid on chromosome 6 (41.1 - 42 Mb), explaining 44 % of the variation. This QTL was also found in the progeny of a *S. lycopersicum* cherry versus round cross (Bovy *et al.,* unpublished results). The QTL peak was near two aluminium-activated malate transporter-like proteins (Solyc06g072910 and Solyc06g072920). These genes might influence malic acid accumulation by regulating its transport to the vacuole. The introgression of this region of *S. pimpinellifolium* in tomato is likely to have an impact on taste.

Amino acids

Receptors in our tongue respond to sugars, acids and free amino acids, mainly to glutamate. Glutamate is the compound responsible for the so-called fifth flavour: umami (Jinap and Hajeb, 2010; Mouritsen and Styrbæk, 2014). In our research we did not find QTLs for

glutamate itself but we found QTLs for two precursors of glutamate. The first, L-glutamine (Chr. 3, 58.2 Mb), reacts in water with the enzyme glutaminase and is converted to glutamate and ammonia. The second, N-acetylglutamic acid (Chr. 4, 56.5 Mb), gives through the action of the enzyme N-acetyl-glutamate synthase glutamate and acetate. Both compounds had greater relative abundances in the presence of the Moneymaker alleles. These QTLs could be the same as those previously found by Schauer *et al.,* 2006. In addition, we found a QTL with relative greater abundance in the presence of the *S. pimpinellifolium* alleles for the amino acid proline, which is derived from glutamate. This QTL region on chromosome 2 includes a proline synthetase gene (Solyc02g080940). Even though Schauer *et al.* (2005) did not specify QTL regions, they found in another accession of *S. pimpinellifolium* increased concentrations of glutamine, proline, phosphate, 3-phosphoglyceric acid (3PG), chlorogenic acid and shikimic acid. The latter is the precursor of chorismic acid, which is the precursor of aromatic amino acids (phenylalanine, tyrosine, and tryptophan) and several secondary metabolites.

Using data of two platforms, LC-QTOF and GC-TOF, we detected a QTL for the aromatic amino acid phenylalanine on chromosome 10 (around 64.4 Mb). Phenylalanine is the direct precursor of all phenylpropanoids. Its deamination by the enzyme phenylalanine ammonia lyase (PAL) is the first step in the pathway leading to phenylpropanoids and flavonoids. There is a *PAL* gene (Solyc10g086180) located in the region from 64422862 to 64426099 bp on chromosome 10. Causse *et al.* (2004) suggested before that variation in phenylalanine content might be due to differences in the expression and/or activity of this *PAL* gene. Other candidate genes in this region are a glucosyltransferase (Solyc10g086240) and several MYB transcription factors (Solyc10g086250, Solyc10g086260, Solyc10g086270, Solyc10g086290).

Health related compounds

The phenylpropanoid pathway leads to the production of antioxidants, which are not only beneficial for the plant itself, since these compounds play a role in defence mechanisms against pest and diseases, but are also important in relation to human health (Dixon *et al.*, 2002, Sade *et al.,* 2015). The most important genomic region affecting the phenylpropanoid pathway (80-81.4 Mb) could be traced back within the mQTL hotspot on chromosome 1. This region affects the relative abundance of different forms of chlorogenic acid, a well-known antioxidant. The Moneymaker allele led to increased levels of 3-caffeoylquinic acid and the more complex di- and tri-caffeoylquinic acids, whereas the *S. pimpinellifolium* allele promoted the production of 4- and 5-caffeoylquinic acid (both with a LOD2 interval from 80.67-81.30 Mb). A second mQTL for 5-chlorogenic acid was found on chromosome 4. Chlorogenic acid is obtained from the esterification of caffeic acid and quinic acid and the different forms of chlorogenic acid are caused by differences in the esterification position on the quinic acid

moiety and the number of caffeic acid moieties esterified on the quinic acid molecule. This esterification reaction is catalysed by the enzyme hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferases (HQT), which belongs to the acyltransferase family. Three HQT candidate genes (Solyc01g105550, Solyc01g105580 and Solyc01g105590) are located at 85 Mb on chromosome 1. This is outside the LOD2 interval of the mQTL region, so we have no supportive evidence for a direct role of these genes in determining this QTL effect.

Xu *et al.* (2012) compared chlorogenic acid and some of its isomers for antioxidant profiles and DNA-protective activities. They observed differences in the antioxidant capacity of different chlorogenic acid isomers and speculated that these differences could be due to steric hindrance in some forms. Xu *et al.* (2012) also observed that dicaffeoylquinic acids enhanced antioxidant activities due to the presence of more hydroxyl groups. Therefore, the accumulation of more complex forms of chlorogenic acid, such as dicaffeoylquinic acids and tricaffeoylquinic acids promotes the antioxidant properties of tomatoes. Within the chromosome 1 mQTL region we observed several so-called cinnamoyl related proteins (Solyc01g087640, Solyc01g102730, Solyc01g107050, Solyc01g107070, Solyc01g107080 and Solyc01g107590). In addition to the above-mentioned HQT genes, these cinnamoyl related proteins could represent a promising target to further explore in this region.

In addition to caffeoylquinic acids, QTLs for other phenylpropanoids, such as glycosides of the caffeic acid and coumaric acid, as well as coumaroylquinic acid, were associated with QTLs on chromosomes 1 (~86Mb), 2 (~48 Mb), 3 (~34.5 Mb) and 10 (~64.4 Mb). The QTL at 86 Mb on chromosome 1 co-localises with the above-mentioned HQT candidate genes at 85Mb. For all these QTLs, the Moneymaker allele was associated with increased levels of these compounds.

Chromosome 1 also harbours important flavonoid mQTLs influencing the accumulation of several kaempferol glycosides. These QTLs co-localise with the MYB12 transcription factor (Solyc01g079620 between 71255600 and 71258882) and the *S. lycopersicum* allele leads to increased kaempferol glycoside levels. This SIMYB12 effect on flavonoid accumulation was also reported before by Adato *et al.* 2009, Ballester *et al.* (2010) and Lin *et al.* (2014). The *MYB12* gene is located outside the QTL intervals for phenylpropanoids and therefore we consider it unlikely that this gene plays a major role in determining the observed variation in phenylpropanoids in the chromosome 1 QTL hotspot region.

When the alleles of *S. lycopersicum* are replaced by the ones of *S. pimpinellifolium* in the core of the QTL hotspot on chromosome 1 (from 75-79 Mb), the result is an accumulation of alkaloids. This is also the case for the other QTL hotspot on the top of chromosome 10. The

bitter taste or poisonous effects of alkaloids could be a reason for selection against those regions during breeding.

In contrast to the MYB12 region on chromosome 1, most quercetin, naringenin and kaempferol (exceptions on chromosome 1) aglycones and glycosides increased in the presence of *S. pimpinellifolium* alleles. All significant mQTLs found on chromosome 6 (35-43.6 Mb) resulted in an increase in flavonoid content in the presence of S. *pimpinellifolium* alleles. Within that region on chromosome 6 many different genes are located such as flavonol synthase/flavonone 3-hydroxylase (Solyc06g073080), several cytochrome P450 genes (Solyc06g060000, Solyc06g060020, Solyc06g060040, Solyc06g060190, Solyc06g065420, Solyc06g065430, Solyc06g066230, Solyc06g066240, Solyc06g067930, Solyc06g074180, Solyc06g076160) and several MYB transcription factors (Solyc06g071230, Solyc06g071690, Solyc06g073640, Solyc06g074910, Solyc06g074920, Solyc06g075660, Solyc06g075670). Furthermore, the three QTLs on chromosome 12 (2 Mb, 48 Mb and 62.5 Mb) showed an enhanced accumulation of flavonoids if the regions originated from *S. pimpinellifolium*. There, a cytochrome P450 (Solyc12g056810) and a chalcone synthase (Solyc12g098090) are near the QTL region. Those could be promising candidate genes to further explore.

According to our results, in order to promote antioxidant activity of tomatoes, it is more favorable to combine the region between 71-87 Mb on chromosome 1 with the *S. lycopersicum* alleles for complex forms of chlorogenic acid and the regions between 35-44 Mb on chromosome 6, the bottom of chromosome 10 and chromosome 12 around 48 Mb from *S. pimpinellifolium*.

Flavour volatiles

Taste perception in the brain is based on the interaction of sugars, organic acids and amino acids with receptors in the mouth, but also on signals perceived through retronasal olfaction of volatile compounds (Small *et al.,* 2004). The amino acid phenylalanine is the precursor of several aromatic volatiles, including the important flavour volatile 2-phenylethanol, which is perceived as a pleasant rosy-sweet smell. We found a QTL on chromosome 8 (~50.8Mb) for 2-phenylethanol and 1-nitro-2-phenylethane. Also the latter compound can lead to a similar sweet-floral aroma. Tieman *et al.* (2006) described the function of three decarboxylases (LeAADC1A, LeAADC1B, and LeAADC2) located on chromosome 8 and demonstrated that down-regulation of these genes led to a reduced production of phenylethanol in transgenic plants (Tieman *et al.*, 2006). It was hypothesized that (one of) these genes could be the gene underlying the *Malodorous* QTL present on *S. pennellii* chromosome 8 (IL8-2-1) (Tadmor *et al.*, 2002), leading to extremely high levels of the phenolic volatiles 2-phenylethanol and

phenylacetaldehyde. The position of the *S. pennellii* introgression assigned responsible for *Malodorous* on top of chromosome 8 fits with the QTL for phenylacetaldehyde (at approximately 3Mb) of our study with *S. pimpinellifolium* where nearby there are 2 decarboxylases (Solyc08g006740- the mentioned LeAADC2- and Solyc08g006750). But on the other hand, our QTLs for the pleasant aromas from 2-phenylethanol and 1-nitro-2-phenylethane were localised at the end of chromosome 8 (the region around 50.8 Mb). This last region perhaps coincides with the region described as 8D by Tieman *et al*. (2006b) for *S. pennellii*. Near our QTL-maximum, at 1.1 Mb upstream, there are other decarboxylases (Solyc08g066220, Solyc08g066240, Solyc08g066250) and about 4.2 Mb upstream the QTL maximum there is the region of the mentioned LeAADC1A (synonym of AADC1B: Solyc08g068680, described by Tieman *et al*., 2006) and more decarboxylases in the region (Solyc08g068600, Solyc08g068610, Solyc08g068620, Solyc08g068630, Solyc08g068640, Solyc08g068670). Perhaps the genes involved are the same ones as the ones found for *S. pennellii*, but even in this case, the further exploration of decarboxylases from even more *S. pimpinellifolium* accessions may help to increase the aroma of tomato fruits.

The bottom of chromosome 1 presented a hotspot for volatile QTLs. All volatiles derived from branched chain amino acids, terpenes and lipids had higher concentrations in the presence of a Moneymaker background. Compounds like linalool are known to promote good floral aromas, but compounds like trans, trans-2,4-heptadienal are known for its rancid flavour. Linkage drag of co-localised QTLs in QTL-rich regions can give rise to a combination of desired and undesired metabolites. This makes it very difficult, laborious and time consuming to breed for new varieties with increased levels of desired metabolites, while selecting against the presence of off-flavours.

Other volatile QTLs that might be influencing flavour were found on chromosome 2 (39-49 Mb), including volatiles such as beta-cyclocitral, 2-heptenal, 3-hexen-1-ol and 1-hexanol. These volatiles were more abundant due to *S. pimpinellifolium* alleles. Furthermore, on chromosome 11 (49-50 Mb), QTLs were found for the volatiles 3-methylbutanal, 1-nitro-pentane and 3-methylbutanoic acid of which higher levels were due to Moneymaker alleles. Isovaleric acid (3-Methylbutanoic acid) is known to cause a cheese or dirty socks smell as pure compound. Nevertheless this compound was found to contribute to the perception of sweetness, therefore it can be considered as positive trait (Bartoshuk and Klee, 2013). A QTL for the lipid derived 3-methyl-1-pentanol, causing a whiskey-like taste and characteristic of tabasco pepper, was found on chromosome 12 near 62 Mb.

As for off-flavours, we found the strong smoky flavour QTLs related to the production of the phenylpropanoid-derived volatiles guaiacol, methyl salicylate and eugenol around 64.7 Mb, as decribed by Tikunov *et al.,* 2013. The *S. lycopersicum* allele was responsible for the

production of cleavable diglycoside precursors (hexose-pentose) of these three volatiles, whereas the *S. pimpinellifolium* allele led to the production of non-cleavable, triglycoside precusors (dihexose-pentose), due to the action of an active *NSGT1* gene that adds the third sugar group. Other QTLs influencing the levels of these phenylpropanoid volatiles were found on chromosome 2: a QTL, located near 30.7 Mb led to increase levels of methylsalicylate when the *S. pimpinellifolium* allele was present, while a QTL at 37.7-39.3 Mb increase the levels of eugenol and eugenol glycosides when the *S. lycopersicum* allele was present.

Upstream of the smoky flavour QTL on chromosome 9 we detected a QTL for the carotenoid-derived volatile beta-damascenone. Although encoded by the same *S. lycopersicum* allele as smoky volatiles Tieman *et al.* (2012) showed that β-damascenone is not associated with tomato flavour intensity. Therefore, major efforts to separate the β-damascenone QTL from the smoky off-flavour region on chromosome 9 should not be a priority.

Downstream of the smoky flavour QTL on chromosome 9, we detected a QTL for the off-flavour dimethyl disulphide (at 65.4 Mb) caused by the presence of the *S. pimpinellifolium* allele. This QTL co-localises with a QTL for the off-flavour putrescine. We detected another *S. pimpinellifolium* putrescine QTL at the top of chromosome 1 (around 0.7 Mb) as well. Avoidance of the off-flavour of putrescine could also have been a selection pressure applied during breeding.

Conclusion

Genetical metabolomics, combining genetics with metabolic profiling, can guide breeding for quality strategies. According to our results, to increase the antioxidant properties of tomato, it could be more beneficial if the region between 71-87 Mb on chromosome 1 is Moneymaker while the regions on chromosome 6 (35-44 Mb; flavonoids and phenylpropanoids) the bottom of chromosome 10 (phenylpropanoids) and chromosome 12 (around 48 Mb; flavonoids) should be from *S. pimpinellifolium*. The above-mentioned region on chromosome 6 could also affect the concentration of malic acid in the fruits. Therefore, the ratios between other acids should be observed, as well as their effects on taste. Sugars can be increased by combining the wild alleles on chromosome 2 (~41.7 Mb) for sucrose and chromosome 10 (~1.7 Mb) with the Moneymaker allele on the hotspot region of chromosome 1 for fructose and chromosome 4 (~55 Mb) for glucose. Off flavour regions that should be avoided in crosses with *S. pimpinellifolium* are the ones on top of chromosome 1 and on chromosome 9 around 65 Mb

where we found the compounds putrescine and dimethyl disulfide. An aromatic boost to the fruits will be given by introgressing parts of the *S. pimpinellifolium* chromosome 8 this will increase phenolic VOCs, such as phenylethanol and phenylacetaldehyde. Possible specific candidate gene expression patterns, additivity and epistatic interactions (as observed by Causse *et al.*, 2007; Muir and Moyle, 2009) should be further evaluated. In general, our results give an insight in the physical positions of metabolite related QTLs that could be used by breeders that would like to use *S. pimpinellifolium* to improve tomato quality.

Acknowledgements

This project was carried out within the research programme of the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. The metabolic profiling in this project was funded by the Netherlands Metabolomics Centre (NMC). Furthermore, the work benefited from the networking activities within the European COST Action FA1106. We would like to make a special acknowledgement to the Business Unit Bioscience at Wageningen UR, besides our co-authors Ric de Vos and Henriëtte van Eekelen, we thank Harry Jonker, Roland Mumm and Robert Hall for all their help and support. The sequencing of the RILs was performed within the 150 Tomato Genome ReSequencing project (http://www.tomatogenome.net/). Finally, we would like to also give special thanks to Dr. Joao Paulo Caldas from Biometris at Wageningen UR for all the support and knowledge in developing a script to extract the BLUEs from the metabolomic data.

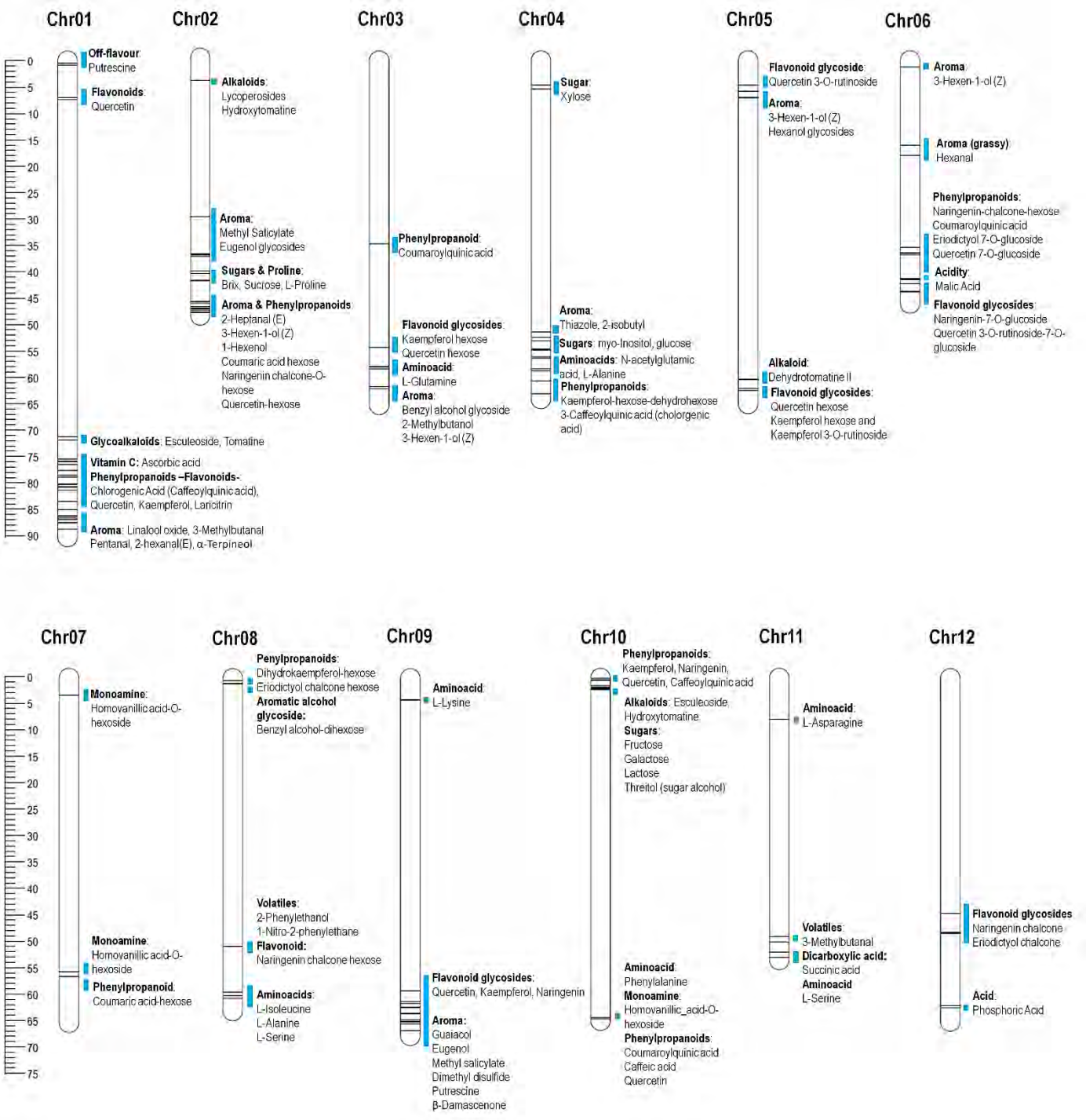


Figure 3. Graphical distribution of a number of mQTLs along the tomato genome in a population between *S. lycopersicum* cv Moneymaker and *S. pimpinellifolium* G1.1554.

Table 1. Metabolite QTLs. The averages of the individuals with a QTL region coming from the tomato (Moneymaker) or *S. pimpinellifolium* parent are coloured according to concentration (green: lower, red: higher). Average values are log10 transformed where a change of 0.301 means a 2x difference in relative abundance of the compounds.

Putative Compound Identification	Ann. Level*	Compound class	Specific negative ion, m/z**	Retention Time (min)**	Ref. Name	Chr	Position (cM)	Marker (closest) Position Mb-Chr	LOD	muA Money-maker	muB Pimp G1.1554	% Explained variace
Putrescine	I	Polyamine	114	15.2030	GCTOF6069	1	3.38	678082-1	8.66	3.27	3.57	34.9
Octanal	II	Lipid	56	20.4879	SPME3494	1	3.38	420700-1	3.41	4.19	4.3	15.5
2-Decenal, (Z)-	II	Lipid	42	29.6403	SPME7200	1	5.16	678082-1	5.15	3.96	4.11	22.5
2-Penten-1-ol, (E)-	III	Lipid	65	11.2882	SPME946	1	5.16	678082-1	5.09	4.68	4.84	22.3
Kaempferol 3-O-rutinoside-7-O-glucoside	I	Flavonoid: Flavonol	846.2003	17.5705	LC1561	1	77.86	71250373-1	4.06	2.3	1.67	18.0
Kaempferol 3-O-rutinoside	I	Flavonoid: Flavonol	683.1442	27.0634	LC3346	1	80.92	71257624-1	5.64	2.6	1.94	24.1
3-Methylbutanal	II	Branched chain amino acid	85	7.3937	SPME367	1	110.84	75536195-1	4.16	4.71	4.3	18.6
Pentane, 1-nitro-	II	Lipid	63	16.6168	SPME2256	1	110.84	75536195-1	6.22	5.31	4.69	26.5
Esculeoside A	I	Alkaloid	657.7862	27.0634	LC3339	1	110.84	75536195-1	5.11	2.32	2.75	22.1
Glycoalkaloid	III	Alkaloid	1269.5953	25.187	LC2811	1	111.84	75536195-1	3.5	2.75	3.06	15.8
Glycoalkaloid	III	Alkaloid	1331.59	23.3999	LC2471	1	111.84	75536195-1	4.37	3.22	3.44	19.3
Hydroxytomatine I + FA	II	Alkaloid	1284.5905	26.6309	LC3254	1	111.84	75536195-1	5.96	2.35	2.91	25.3
Glycoalkaloid	III	Alkaloid	1210.5605	35.3125	LC4704	1	111.84	75536195-1	6.23	2.17	2.78	26.3
Alpha-tomatin	II	Alkaloid	1096.5371	32.0458	LC4157	1	112.93	76030524-1	4.67	1.04	1.45	20.4
Alpha-Tomatin + FA	II	Alkaloid	1147.5287	33.3092	LC4396	1	130.33	78617409-1	4.11	0.85	1.19	18.2
Naringenin-hexose, - C3H7O2NS	II	Flavonoid: Flavonone	984.3438	15.7477	LC1332	1	121.85	77713469-1	4.99	1.3	1.79	21.7
Fructose	III	Sugar	177	16.6671	GCTOF9613	1	112.84	76030524-1	3.43	6.08	6.03	15.6
3-Methylbutanoic acid	II	Branched chain amino acid	45	13.448	SPME1349	1	112.84	76030524-1	6.99	4.79	4.22	29.3
3-Methyl-1-butanol	II	Branched chain amino acid	47	9.9546	SPME644	1	116.77	77712624-1	5.68	5.2	4.67	24.5
3-Methylbutanenitrile	III	Branched chain amino acid	37	9.8425	SPME614	1	119.54	77714028-1	4.1	4.66	4.32	18.4
Homovanillic acid-O-hexoside	II	Phenylpropanoid	344.1076	13.2207	LC1022	1	144.49	80197071-1	3.99	1.87	1.31	17.7
Linalool	II	Terpene	82	24.1111	SPME4976	1	146.49	80347083-1	8.2	4.54	4.18	33.4
3,4 or 3,5-Dicaffeoylquinic acid	II	Phenylpropanoid	354.0917	28.2554	LC3537	1	152.11	80347083-1	14.99	2.26	0.92	52.0
D(-)-Quinic acid	I	Phenylpropanoid	221	16.4179	GCTOF8561	1	159.73	80768288-1	7.14	3.85	2.57	29.8
3-Caffeoylquinic acid	II	Phenylpropanoid	707.1808	14.9166	LC1210	1	159.73	80768288-1	19.96	2.77	1.1	62.4
Tricaffeoylquinic acid II	II	Phenylpropanoid	677.1515	41.1974	LC5547	1	159.73	80768288-1	4.96	1.34	0.82	21.6
Laricitrin-Hexose, Hexose	III	Flavonoid: Flavonol	666.1787	22.8227	LC2320	1	159.73	80768288-1	6.17	1.41	0.81	26.1
Tricaffeoylquinic acid II	II	Phenylpropanoid	677.1508	40.4378	LC5468	1	159.73	80768288-1	6.72	1.63	0.9	28.1
Furostane-tetrol, Hexose, Hexose, Hexose	III	Saponin	353.0874	34.4636	LC4573	1	159.73	80768288-1	9.28	1.8	0.88	36.5
4,5-Dicaffeoylquinic acid	I	Phenylpropanoid	173.0457	30.3856	LC3922	1	159.73	80768288-1	14	2.39	0.96	49.6
3,4,5-Tricaffeoylquinic acid	II	Phenylpropanoid	691.1284	39.1029	LC5318	1	159.73	80768288-1	14.91	2.58	1.02	51.8
Caffeic acid-hexose IV	II	Phenylpropanoid	281.0672	11.7768	LC915	1	159.73	80768288-1	15.48	1.96	0.83	53.2
5-Caffeoylquinic acid	II	Phenylpropanoid	708.1845	13.8872	LC1109	1	159.73	80768288-1	15.6	3.49	4.61	53.4
4-Caffeoylquinic acid	II	Phenylpropanoid	353.0878	17.2452	LC1489	1	159.73	80768288-1	9.11	2.86	3.43	36.0
2,4-Heptadienal, (E,E)-	II	Lipid	109	20.2991	SPME3447	1	177.55	85115390-1	4.4	3.97	3.9	19.6
2-Hexenal, (E)-	II	Lipid	66	14.3154	SPME1554	1	177.55	85115390-1	4.16	5.23	5.11	18.6
alpha-Terpineol	II	Terpene	66	27.7048	SPME6395	1	177.55	85115390-1	8.52	4.65	4.26	34.4
Linalool oxide B	II	Terpene	107	23.3794	SPME4696	1	177.55	85115390-1	10.58	4.81	4.33	40.8
p-Menth-1-en-9-al	II	Terpene	68	28.6667	SPME6791	1	177.55	85115390-1	9.21	4.01	3.67	36.6
Caffeic acid-hexose V	II	Phenylpropanoid	179.0364	12.949	LC998	1	184.65	86636577-1	3.92	2.38	2.02	17.5
Dehydrotomatine (S) I + FA	II	Alkaloid	1183.5585	27.696	LC3467	1	182.58	86171125-1	4.05	1.2	1.62	18.0

Putative Compound Identification	Ann. Level*	Compound class	Specific negative ion, m/z**	Retention Time (min)**	Ref. Name	Chr	Position (cM)	Marker (closest) Position Mb-Chr	LOD	muA Money-maker	muB Pimp G1.1554	% Explained variace
Dehydroesculeoside A I + FA	II	Alkaloid	656.2778	25.3675	LC2890	1	192.44	87520611-1	3.37	2.66	2.92	15.2
Lycoperoside H or Hydroxytomatine II + FA	III	Alkaloid	1097.5463	25.9804	LC3097	2	1	4409029-2	5.23	1.84	1.47	22.6
Methyl salicylate	II	Phenylpropanoid	108	27.8995	SPME6482	2	7.62	30762028-2	5.77	4.55	5.33	24.8
Acetic acid	II	Acid	43	5.2516	SPME124	2	39.15	36842294-2	3.25	3.75	3.91	14.9
Eugenol malonyl dihexose-pentose	II	Glycosylated VOC	706.2286	27.5334	LC3439	2	45.08	37910463-2	10.53	1.75	0.97	40.3
Eugenol dihexose pentose	II	Glycosylated VOC	620.2272	24.8617	LC2736	2	45.34	37964685-2	3.51	1.35	0.96	15.8
Eugenol	III	Phenylpropanoid	150	32.9212	SPME7920	2	46.34	38096910-2	6.1	4.34	4.08	26.1
Eugenol-hexose-pentose	II	Glycosylated VOC	293.0887	29.7886	LC3842	2	46.4	38096910-2	4.87	1.27	0.84	21.2
beta-Cyclocitral	II	Carotenoid	81	28.8201	SPME6854	2	48.23	39021430-2	3.28	4.24	4.37	15.0
L-Proline	I	Amino acid	127	9788256	GCTOF2556	2	56.9	40903529-2	3.3	3.33	3.69	15.1
Sucrose	I	Sugar	213	23.2077	GCTOF 16178	2	59.64	41701555-2	5.01	5.07	5.26	22.0
2-Heptenal, (E)-	II	Lipid	111	18.7412	SPME2771	2	66.29	43635207-2	4.15	5.34	5.44	18.6
Succinic acid	I	Organic acid	247	10.0057	GCTOF2716	2	86.12	47296020-2	3.12	2.74	2.95	14.3
3-Hexen-1-ol, (Z)-	II	Lipid	72	14.5102	SPME1619	2	86.38	47421143-2	4.87	4.59	4.8	21.4
Quercetin	I	Flavonoid: Flavonol	977.4944	37.2801	LC5063	2	87.68	48113374-2	3.89	1.24	1.58	17.4
Beta-tomatine	II	Alkaloid	900.4957	33.2002	LC4366	2	87.68	48113374-2	5.77	1.08	1.68	24.6
Benzyl alcohol-hexose-pentose + FA	II	Glycosylated VOC	775.3373	16.1801	LC1376	2	91.13	48497154-2	3.41	2.15	2.5	15.4
Coumaric acid-hexose I	II	Phenylpropanoid	651.1918	10.0968	LC866	2	92.99	48900086-2	3.36	2.84	2.07	15.2
1-Hexanol	II	Lipid	84	15.1297	SPME1868	2	93.77	49269256-2	4.66	4.81	5.1	20.6
Coumaroylquinic acid	II	Phenylpropanoid	337.0936	17.9136	LC1615	3	58.99	34561749-3	3.58	2.82	1.94	16.1
3-Heptanone	II	Lipid	114	15.9205	SPME2058	3	101.2	52104444-3	3.65	4.46	4.41	16.5
Kaempferol-hexose-deoxyhexose, -hex.-C10H8O3	II	Flavonoid: Flavonol	932.2512	31.6848	LC4094	3	105.2	54199481-3	4.14	1.41	2.05	18.4
Methylthioacetaldehyde	III	Aldehyde	47	11.0168	SPME882	3	126.34	56643506-3	3.49	3.82	3.6	15.9
Hexanol - pentose - Hexose I	III	Glycosylated VOC	428.2016	24.465	LC2643	3	132.21	57704490-3	5.98	1.93	2.23	25.4
Camphor	II	Terpene	97	26.3889	SPME5910	3	132.58	57704490-3	4.74	4.91	4.88	20.9
Ethanol, 2-(methylthio)-	II	Phenolic	48	14.1561	SPME1497	3	132.58	57704490-3	3.42	3.86	3.67	15.6
Kaempferol 3-O-rutinoside-7-O-glucoside	I	Flavonoid: Flavonol	756.2051	18.2746	LC1641	3	132.58	57704490-3	6.34	1.71	2.22	26.7
L-Glutamine	III	Amino acid	189	15.0847	GCTOF5968	3	137.63	58231771-3	3.09	3.95	3.61	14.2
Quercetin-hexose-deoxyhexose, -pentose	II	Flavonoid: Flavonol	741.187	22.0273	LC2192	3	137.63	58231771-3	8.51	4.01	3.36	34.1
Quercetin-hexose-deoxyhexose, -hexose-C9H10O2	II	Flavonoid: Flavonol	922.2689	34.5905	LC4589	3	138.63	58231771-3	9.62	2	1.16	37.6
Benzophenone	II	Phenolic	181	41.0585	SPME8866	3	139.63	58231771-3	3.45	4.33	4.28	15.7
Benzyl alcohol-hexose-pentose + FA	II	Glycosylated VOC	775.3373	16.1801	LC1376	3	139.63	58231771-3	5.3	2.1	2.56	22.9
3-Hexen-1-ol, (Z)-	II	Lipid	72	14.5102	SPME1619	3	139.63	58231771-3	3.93	4.59	4.8	17.7
Quercetin	I	Flavonoid: Flavonol	977.4944	37.2801	LC5063	3	142.63	61455688-3	3.46	1.24	1.59	15.6
Naringenin-hexose, -hexose II	III	Flavonoid: Flavonone	694.1218	20.3136	LC1950	3	143.63	61455688-3	3.43	1.32	1.86	15.5
2-Methyl-1-butanol	II	Branched chain amino acid	54	10.1022	SPME681	3	151.05	61881530-3	3.71	5.17	4.84	16.8
myo-Inositol	I	Sugar	139	18.7220	GCTOF 13742	4	60.1	52505792-4	4.43	4.9	4.64	19.7
Glucose	III	Sugar	105	16.7779	GCTOF 10367	4	68.77	54943259-4	8.71	3.77	2.45	35.0
Allantoin	III	Organic acid	246	16.7679	GCTOF 10286	4	69.55	55077232-4	7.2	4.39	2.84	30.0
N-Acetylglutamic acid	I	Amino acid	157	12.8523	GCTOF4344	4	79.92	56476705-4	4.75	3.92	3.75	21.0
1-Pentanol, 3-methyl-	III	Lipid	53	14.2387	SPME1519	4	83.36	57162760-4	7.26	4.47	3.98	30.2

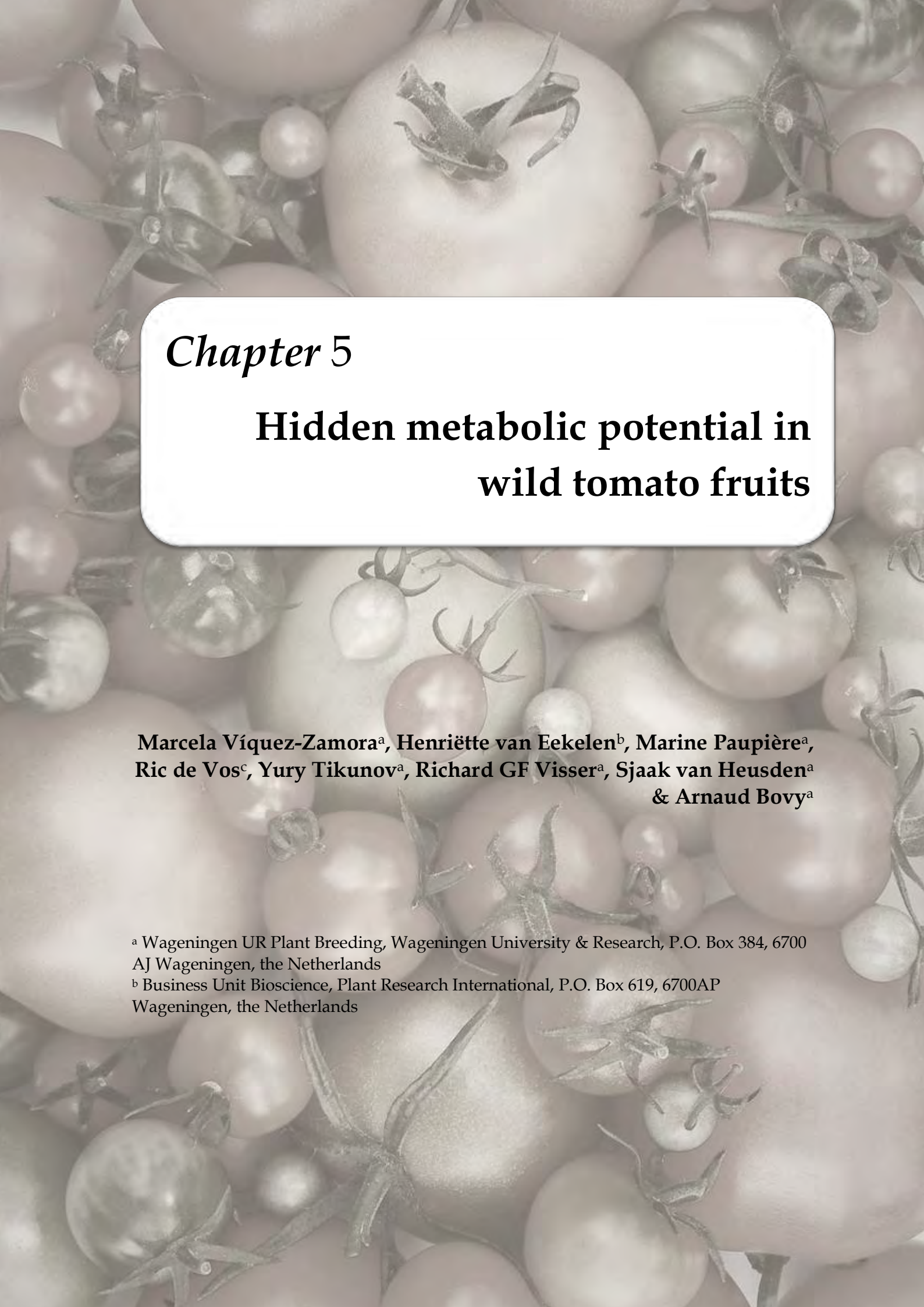
Putative Compound Identification	Ann. Level*	Compound class	Specific negative ion, m/z**	Retention Time (min)**	Ref. Name	Chr	Position (cM)	Marker (closest) Position Mb-Chr	LOD	muA Money-maker	muB Pimp G1.1554	% Explained variace
L-Alanine	I	Amino acid	119	6.9658	GCTOF1102	4	94.97	58745615-4	6	3.51	3.82	25.7
3-Caffeoylquinic acid	II	Phenylpropanoid	708.1845	13.8872	LC1109	4	109.72	60983181-4	3.58	3.67	4.21	16.1
3-trans-Caffeoylquinic acid	I	Phenylpropanoid	267	26.6075	GCTOF 17957	4	110.98	61083233-4	3.43	2.72	3.36	15.6
Linalool	II	Terpene	82	24.1111	SPME4976	4	113.98	61414168-4	4	4.51	4.31	18.0
p-Menth-1-en-9-al	II	Terpene	68	28.6667	SPME6791	4	113.98	61414168-4	3.24	3.98	3.75	14.8
Dehydrotomatine II + FA	II	Alkaloid	1106.5372	30.9271	LC5010	4	130.31	63488407-4	3.38	0.93	1.41	15.3
Quercetin 3-O-rutinoside	I	Flavonoid: Flavonol	610.1487	24.0683	LC2554	5	43.75	4710231-5	3.51	4.35	4.63	15.8
3-Hexen-1-ol, (Z)-	II	Lipid	72	14.5102	SPME1619	5	53.08	5826600-5	4.75	4.81	4.6	21.0
Hexanol - pentose - Hexose I	III	Glycosylated VOC	428.2016	24.465	LC2643	5	58.13	7031168-5	7.24	2.26	1.93	29.9
Dehydrotomatine II + FA	II	Alkaloid	11.0654	30.9271	LC4010	5	69.24	60342410-5	4.69	1.39	0.98	20.5
Kaempferol-hexose-deoxyhexose, -pentose	II	Flavonoid: Flavonol	794.1761	24.3202	LC2621	5	83.38	62101535-5	3.61	1.98	2.42	16.2
Quercetin-dihexose-deoxyhexose-pentose	II	Flavonoid: Flavonol	393.0733	14.0677	LC1142	5	95.84	62370844-5	3.79	1.63	1.94	16.9
Kaempferol 3-O-rutinoside	I	Flavonoid: Flavonol	683.1442	27.0634	LC3346	5	96.46	62370844-5	5.66	1.97	2.57	24.2
4-Caffeoylquinic acid	II	Phenylpropanoid	353.0878	17.2452	LC1489	5	96.46	62370844-5	3.64	3.29	2.94	16.3
Esculeoside A	I	Alkaloid	657.7862	27.0634	LC3339	5	96.46	62370844-5	4.82	2.76	2.33	21.0
3-Hexen-1-ol, (Z)-	II	Lipid	72	14.5102	SPME1619	6	4.67	1178296-6	5.16	4.61	4.84	22.5
Hexanoic acid	II	Lipid	67	18.871	SPME2808	6	10.96	2983847-6	3.56	5.42	5.31	16.2
Hexanal	II	Lipid	85	12.4979	SPME1198	6	13.91	17980595-6	4.66	7.21	7.11	20.6
Kaempferol 3-O-rutinoside	I	Flavonoid: Flavonol	683.1442	27.0634	LC3346	6	53.97	35310848-6	5.85	2.02	2.62	24.9
Kaempferol 3-O-rutinoside-7-O-glucoside	I	Flavonoid: Flavonol	846.2003	17.5705	LC1561	6	60.57	36355674-6	4.87	1.71	2.35	21.2
Quercetin 3-O-rutinoside	I	Flavonoid: Flavonol	610.1487	24.0683	LC2554	6	62.14	36771580-6	4.09	4.36	4.67	18.2
Naringenin-hexose, -C3H7O2NS	II	Flavonoid: Flavonone	271.0625	16.2515	LC1387	6	72.93	39630736-6	4.06	1.22	1.86	18.1
Naringenin chalcone-hexose	II	Flavonoid: Flavonone	867.2342	33.3271	LC4411	6	84.35	41164804-6	3.23	2.3	2.87	14.6
Eriodictyol chalcone-hexose	II	Flavonoid: Flavonone	449.1096	28.8147	LC3624	6	84.35	41164804-6	3.83	1.8	2.42	17.1
O-acetylprunin	II	Flavonoid: Flavonone	519.1193	33.7971	LC4470	6	84.35	41164804-6	4.03	1.92	2.77	17.9
Naringenin-hexose III	III	Flavonoid: Flavonone	433.1143	23.7609	LC2524	6	84.35	41164804-6	4.28	1.89	2.47	18.9
Eriodictyol 7-O-glucoside	II	Flavonoid: Flavonone	288.0603	24.4471	LC2635	6	84.35	41164804-6	4.5	1.47	2.06	19.8
Dihydrokaempferol-hexose or Eriodictyol chalcone-hexose	II	Flavonoid: Flavonone	449.1099	28.2911	LC3545	6	84.35	41164804-6	5.21	1.29	1.87	22.5
Phloretin-C-diglycoside	II	Flavonoid: Flavonol	478.1429	25.3675	LC2887	6	84.87	41222645-6	3.79	2.16	2.48	16.9
Naringenin-C3H7NO2S, -C9H12O8	III	Flavonoid: Flavonone	641.1375	22.0273	LC2190	6	84.87	41222645-6	4.09	1.05	1.61	18.2
Naringenin-hexose, -C3H7O2NS	II	Flavonoid: Flavonone	271.0619	17.029	LC1463	6	84.87	41222645-6	5.75	1.45	2.13	24.5
Coumaroylquinic acid	II	Phenylpropanoid	1341.1058	18.3996	LC1663	6	84.87	41222645-6	6.27	1.91	2.43	26.5
Malic acid	I	Organic acid	232	12.2498	GCTOF3585	6	86.87	41440942-6	11.68	5.18	4.84	43.9
Naringenin-7-O-glucoside	II	Flavonoid: Flavonone	433.1146	29.14	LC3715	6	91.32	42297379-6	3.65	2.77	3.08	16.4
Naringenin-O-dihexose III	II	Flavonoid: Flavonone	658.163	25.0243	LC2770	6	96.88	43580814-6	4.64	1.46	2.15	20.3
Homovanillic acid-O-hexoside	II	Phenylpropanoid	344.1076	13.2207	LC1022	7	22.47	3317534-7	3.95	1.87	1.42	17.6
Homovanillic acid-O-hexoside	II	Phenylpropanoid	687.2127	12.4987	LC964	7	39.88	55492731-7	4.25	1.68	1.23	18.8
Dihydrokaempferol-hexose or Eriodictyol chalcone-hexose III	III	Flavonoid: Flavonol	450.1126	30.0246	LC3878	8	4.53	426863-8	3.83	1.31	1.87	17.1

Putative Compound Identification	Ann. Level*	Compound class	Specific negative ion, m/z**	Retention Time (min)**	Ref. Name	Chr	Position (cM)	Marker (closest) Position Mb-Chr	LOD	muA Money-maker	muB Pimp G1.1554	% Explained variace
2-Methyl-1-butyl acetate	II	Branched chain amino acid	42	15.5723	SPME1982	8	5.53	677500-8	4.95	3.72	3.49	21.7
Benzyl alcohol-dihexose + FA	II	Glycosylated VOC	467.132	13.2922	LC1036	8	9.99	1149764-8	5.3	2.26	1.8	22.9
Ethanol, 2-(methylthio)-	II	Phenolic	48	14.1561	SPME1497	8	26.99	2908556-8	3.96	3.62	3.9	17.8
Phenylacetaldehyde	II	Aldehyde	75	22.3172	SPME4236	8	26.99	2908556-8	3.15	4.6	5.1	14.4
2-Phenylethanol	II	Phenolic	45	24.8841	SPME5278	8	52.1	50811756-8	4.34	4.93	5.38	19.3
1-Nitro-2-phenylethane	II	Lipid	118	31.2512	SPME7566	8	55.58	50811756-8	3.09	4.14	4.44	14.2
Naringenin chalcone-hexose,-deoxyhexose I	II	Flavonoid: Flavonone	580.1754	29.5367	LC3783	8	52.58	50811756-8	5.45	1.6	1.04	23.4
2-Methylfuran	II	Furanoid	39	6.2902	SPME257	8	70.83	56674223-8	9.18	3.77	3.48	36.5
L-Isoleucine	I	Amino acid	160	9.6907	GCTOF2334	8	91.34	59490918-8	3.4	3.03	3.29	15.5
L-Serine	I	Amino acid	66	10.5782	GCTOF2882	8	96.88	60673054-8	3.33	3.66	3.96	15.2
L-Lysine	I	Amino acid	102	17.1279	GCTOF 12036	9	30.72	3502151-9	4.17	3.9	3.99	18.7
3-Methylbutanenitrile	III	Branched chain amino acid	37	9.8425	SPME614	9	39.56	4113674-9	3.37	4.33	4.64	15.4
Dihydrokaempferol-hexose or Eriodictyol chalcone-hexose III	III	Flavonoid: Flavonol	450.1126	30.0246	LC3878	9	54.93	59327238-9	3.21	1.87	1.37	14.5
Naringenin-hexose, -C3H7O2NS	II	Flavonoid: Flavonone	271.0619	17.029	LC1463	9	55.93	59327238-9	3.2	2.08	1.55	14.5
beta-Damascenone, (Z)-	III	Carotenoid	65	33.883	SPME8054	9	63.16	60487661-9	4.72	4.42	4.24	20.8
Quercetin 3-O-rutinoside-7-O-glucoside	I	Flavonoid: Flavonol	862.1947	15.2955	LC1245	9	71.62	62496837-9	5.47	2.31	1.75	23.5
Eugenol malonyl dihexose-pentose	II	Flavonoid: Flavonol	706.2286	27.5334	LC3439	9	82.33	63537281-9	4.8	0.97	1.5	21.0
Methyl salicylate malonyl dihexose-pentose	II	Glycosylated VOC	693.1875	18.7249	LC1723	9	89.25	64794163-9	5.46	2.02	2.6	23.5
Guaicol malonyl dihexose-pentose	II	Glycosylated VOC	667.2048	17.4078	LC1513	9	90.25	64794163-9	23.52	1.15	2.76	68.4
Guaiacol dihexose-pentose	II	Glycosylated VOC	625.1973	13.671	LC1066	9	90.96	64794163-9	20.74	0.92	2.15	63.8
Methyl salicylate	II	Phenylpropanoid	108	27.8995	SPME6482	9	90.25	64794163-9	9.88	5.39	4.49	38.7
Phenol, 2-methoxy- 'o-Guaiacol'	II	Phenylpropanoid	96	23.9872	SPME4894	9	90.96	64794163-9	16.67	5.85	4.69	56.2
Benzyl alcohol-dihexose + FA	II	Glycosylated VOC	467.132	13.2922	LC1036	9	91.96	64960323-9	3.97	2.2	1.78	17.7
Guaiacol-hexose-pentose	II	Glycosylated VOC	417.141	16.3784	LC1408	9	91.96	64960323-9	5.27	1.52	0.91	22.8
Putrescine	I	Polyamine	114	15.2030	GCTOF6069	9	94.19	65052379-9	5.03	3.32	3.56	22.0
Dimethyl disulfide	II	Amino acid	95	10.6332	SPME774	9	98.21	65468989-9	4.75	4.17	4.43	20.9
Kaempferol-hexose-deoxyhexose, -hexose,-C9H10O2	II	Flavonoid: Flavonol	906.2756	36.5224	LC4909	9	113.88	66807440-9	4.73	1.36	1.95	20.7
Glycoalkaloid	III	Alkaloid	1210.5605	35.3125	LC4704	10	0	134598-10	3.58	2.24	2.71	16.1
4-Caffeoylquinic acid	II	Phenylpropanoid	353.0878	17.2452	LC1489	10	2.99	307305-10	3.32	2.96	3.28	15.0
Glycoalkaloid	III	Alkaloid	1269.5953	25.187	LC2811	10	17.37	1548569-10	3.35	2.76	3.08	15.1
Glycoalkaloid	III	Alkaloid	1331.59	23.3999	LC2471	10	17.37	1548569-10	3.41	3.24	3.45	15.4
Esculeoside A	I	Alkaloid	657.7862	27.0634	LC3339	10	18.22	1662341-10	3.47	2.36	2.76	15.6
Dehydroesculeoside A I + FA	II	Alkaloid	656.2778	25.3675	LC2890	10	18.22	1662341-10	3.57	2.67	2.96	16.0
Hydroxytomatine I + FA	II	Alkaloid	1284.5905	26.6309	LC3254	10	18.22	1662341-10	4.42	2.4	2.92	19.5
Fructose	III	Sugarl'--	141	22.5019	GCTOF 15486	10	22.05	1768385-10	3.89	2.17	2.82	17.5
Threitol	I	Sugar alcohol	115	22.2285	GCTOF 15311	10	23.05	1835448-10	4	2.34	3.1	18.0
Kaempferol-3-O-glucoside	I	Flavonoid: Flavonol	1285.5933	28.1285	LC3516	10	23.93	1835448-10	4.22	1.66	2.11	18.7
Naringenin-hexose IV	III	Flavonoid: Flavonone	435.1231	30.4035	LC3942	10	24.42	1836624-10	3.48	1.4	1.93	15.7
Lactose	II	Sugar	132	22.5644	GCTOF 15629	10	25.42	1887782-10	3.33	2.51	3.25	15.2

Putative Compound Identification	Ann. Level*	Compound class	Specific negative ion, m/z**	Retention Time (min)**	Ref. Name	Chr	Position (cM)	Marker (closest) Position Mb-Chr	LOD	muA Money-maker	muB Pimp G1.1554	% Explained variace
Adenosine-5-monophosphate	II	Nucleoside	243	26.3142	GCTOF 17838	10	25.42	1887782-10	6.26	1.87	2.44	26.7
Naringenin chalcone-hexose	II	Flavonoid: Flavonone	867.2342	33.3271	LC4411	10	27.24	2002067-10	3.25	2.29	2.91	14.7
O-acetylprunin	II	Flavonoid: Flavonone	519.1193	33.7971	LC4470	10	27.24	2002067-10	4.3	1.9	2.86	19.0
Quercetin 3-O-glucoside	I	Flavonoid: Flavonol	461.2217	25.0243	LC2762	10	28.13	2082269-10	5.54	1.57	2.02	23.8
Naringenin-7-O-glucoside	II	Flavonoid: Flavonone	433.1146	29.14	LC3715	10	32.03	2180741-10	3.83	2.77	3.1	17.1
Coumaroylquinic acid	II	Phenylpropanoid	337.0936	17.9136	LC1615	10	116	64442975-10	5.96	2.83	2.07	25.3
Quercetin-dihexose-deoxyhexose-pentose	II	Flavonoid: Flavonol	393.0733	14.0677	LC1142	10	118.37	64442975-10	3.51	1.94	1.65	15.8
Homovanillic acid-O-hexoside	II	Phenylpropanoid	687.2127	12.4987	LC964	10	118.37	64442975-10	3.71	1.69	1.26	16.6
Caffeic acid-hexose V	II	Phenylpropanoid	179.0364	12.949	LC998	10	118.37	64442975-10	5.94	2.43	2	25.2
Homovanillic acid-O-hexoside	II	Phenylpropanoid	344.1076	13.2207	LC1022	10	118.37	64442975-10	6.98	1.91	1.38	29.0
Caffeic acid-hexose II	II	Phenylpropanoid	387.0943	9.827	LC842	10	118.37	64442975-10	10.68	2.4	1.58	40.7
Coumaric acid-hexose I	II	Phenylpropanoid	651.1918	10.0968	LC866	10	118.37	64442975-10	17.12	3.16	1.72	56.8
L-Phenylalanine	I	Amino acid	118	13.9947	GCTOF5324	10	118.37	64442975-10	3.59	3.17	3.71	16.3
Phenylalanine	II	Amino acid	147.0448	4.7375	LC736	10	118.37	64442975-10	7.35	1.11	1.64	30.2
Benzyl alcohol-dihexose + FA	II	Glycosylated VOC	467.132	13.2922	LC1036	10	118.63	64633271-10	3.36	2.24	1.87	15.2
L-Asparagine	I	Amino acid	220	14.4480	GCTOF5545	11	12.35	5624269-11	3.22	4.18	4.41	14.8
3-Methylbutanal	II	Branched chain amino acid	85	7.3937	SPME367	11	31.95	49046173-11	3.39	4.69	4.33	15.5
Pentane, 1-nitro-	II	Lipid	63	16.6168	SPME2256	11	31.95	49046173-11	3.26	5.23	4.77	14.9
3-Methylbutanoic acid	II	Branched chain amino acid	45	13.448	SPME1349	11	40.25	50083210-11	4.35	4.78	4.32	19.4
Acetic acid	II	Acid	43	5.2516	SPME124	11	70.68	52219825-11	3.92	3.95	3.8	17.6
Succinic acid	I	Organic acid	247	10.0057	GCTOF2716	11	78.11	52982367-11	4.53	2.99	2.75	20.1
L-Serine	I	Amino acid	66	10.5782	GCTOF2882	11	78.85	52958873-11	3.56	3.97	3.66	16.2
Quercetin-dihexose-deoxyhexose-pentose	II	Flavonoid: Flavonol	393.0733	14.0677	LC1142	12	28.89	2864634-12	3.2	1.64	1.91	14.5
Acetic acid	II	Acid	43	5.2516	SPME124	12	40.4	4679131-12	4.95	3.77	3.94	21.7
Naringenin-hexose IV	III	Flavonoid: Flavonone	435.1231	30.4035	LC3942	12	61.2	48063454-12	3.69	1.39	1.88	16.5
Dihydrokaempferol-hexose or Eriodictyol chalcone-hexose III	III	Flavonoid: Flavonol	450.1126	30.0246	LC3878	12	61.99	48188286-12	3.21	1.34	1.84	14.5
O-acetylprunin	II	Flavonoid: Flavonone	519.1193	33.7971	LC4470	12	61.99	48188286-12	3.05	1.96	2.72	13.9
Naringenin chalcone-O-hexose	II	Flavonoid: Flavonone	434.1178	32.8571	LC4323	12	64.36	48357587-12	3.45	2.32	2.75	15.6
Citramalic acid	III	Organic acid	115	12.0398	GCTOF3460	12	71.1	62005719-12	3.7	2.18	2.67	16.7
1-Pentanol, 3-methyl-	III	Lipid	53	14.2387	SPME1519	12	71.1	62005719-12	4.79	3.99	4.4	21.1
L-Isoleucine	I	Amino acid	158	9.3874	GCTOF2234	12	76.47	62534395-12	3.32	4.89	4.99	15.1
Kaempferol 3-O-rutinoside-7-O-glucoside	I	Flavonoid: Flavonol	756.2051	18.2746	LC1641	12	77.47	62534395-12	7.22	1.75	2.29	29.8

*Annotation level: I = Identified compounds (e.g. with chemical reference standards and for GC with: RT less than 1min difference and match factors (MF) over 800). II= Putatively annotated compounds (e.g. without chemical reference standards but still with RT and MF over our thresholds). III= Putatively characterized compounds or compound classes (e.g. based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class. Compounds with RT difference larger than 1min, but with good MF or MF between 800-600).

**Compounds analysed using LC-QTOF platform represented by measured accurate masses of corresponding negatively charged parent molecule ions or their formic acid adducts (denoted by +FA). Volatile compounds and primary metabolites measured by SPME-GC and GC-TOF, respectively, represented by selected nominal negative mass ion fragments picked automatically by MSClust software.



Chapter 5

Hidden metabolic potential in wild tomato fruits

Marcela Víquez-Zamora[a], Henriëtte van Eekelen[b], Marine Paupière[a], Ric de Vos[c], Yury Tikunov[a], Richard GF Visser[a], Sjaak van Heusden[a] & Arnaud Bovy[a]

[a] Wageningen UR Plant Breeding, Wageningen University & Research, P.O. Box 384, 6700 AJ Wageningen, the Netherlands
[b] Business Unit Bioscience, Plant Research International, P.O. Box 619, 6700AP Wageningen, the Netherlands

Abstract

Diversity in cultivated tomato has been assumed to be reduced because of domestication and humans breeding efforts. Reclaiming this 'lost' diversity can be done via exploring wild *Solanum* species. A non-targeted approach allowed the identification of interesting traits throughout the whole process of fruit ripening between two *S. lycopersicum*, two *S. cheesmaniae* and eighteen *S. pimpinellifolium* genotypes. In all untargeted metabolic-analyses the different ripening stages were the factors explaining the greatest percentage of variance at the first component. However, a significant amount of variation was explained by the genetic diversity of the accessions. The ripening-dependent changes in the metabolome of *S. cheesmaniae* accessions lagged behind those of *S. lycopersicum* and *S. pimpinellifolium*; ripe fruits of *S. cheesmaniae* had a metabolite composition comparable to that of turning staged *S. pimpinellifolium* and *S. lycopersicum* fruits. Certain accessions looked promising as source for quality traits. For example, *S. pimpinellifolium* accessions such as 4_25_ LA1349, 4_13_LA1924, 2_25_PI126925, 2_9_G1.1589 and PimpG1.1554 are good sources to improve flavonoid content and thus diversity in tomato. Our research offers better insight in the metabolic variation in certain closely related wild accessions of tomato.

Keywords: *S. pimpinellifolium*, *S. cheesmaniae*, flavonoids, metabolic profiling, tomato aroma.

Introduction

Cultivated tomato is estimated to comprise less than five percent of the available genetic variation within the tomato clade (Miller and Tanksley, 1990). Wild relatives are a key to expand and combine morphological and physiological traits of the tomatoes we consume today (Bai and Lindhout, 2007).

Solanum pimpinellifolium is a tomato wild relative that can be found in lowlands from Ecuador to central Chile. The species is self-compatible but outcrossing can also happen even with other species from the Lycopersicon group. *S. pimpinellifolium* is the only wild species having red fruits. Several *S. pimpinellifolium* accessions have been used as a source of resistance genes and different fruit attributes for introgression into commercial tomato cultivars (Peralta *et al.*, 2005; Grandillo *et al.*, 2011 and Bergougnoux, 2014).

Tomato has high nutritional value. Carotenoids, vitamins, glycoalkaloids, minerals, amino acids, phenols and flavonoids are major compounds accounting for that. Tomato has also

become one of the most important components in many cuisines worldwide and therefore the flavour of tomato is considered a very important characteristic, which has become a highly investigated subject over the last decade. Tomato flavour is perceived by receptors in our nose and mouth and is mainly influenced by the amount of sugars, acids, volatiles and texture-related compounds of the fruits. The main sugars in tomato fruit are fructose and glucose. Additionally, acids like citrate, malate, glutamate and ascorbate play a major role in taste perception of tomatoes (Vogel *et al.*, 2010).

There is great metabolic complexity underlying tomato taste and nutritional value. The accumulation or degradation of metabolites in fruit is mediated through several parallel pathways, whose activities are regulated during fruit development and ripening (Klee and Tieman, 2013; Wahyuni, 2014). Compounds in the metabolome of tomatoes can be targeted by different analytical procedures. Some procedures concentrate on a single class of metabolites like sugars or amino acids, and others are suitable for a broad untargeted profiling (Osorio *et al.*, 2009). In this research we used a multi-instrumental metabolomics approach to analyse metabolic perturbations taking place in ripening fruits of a diverse set of 18 *S. pimpinellifolium* accessions and we compared their fruit metabolite content to that of two cultivated *S. lycopersicum* accessions and two accessions of another tomato wild relative, *S. cheesmaniae*. Our research provides insight in the variation between accessions of *S. pimpinellifolium* that can be used to further improve the quality of the domesticated tomato.

Material and methods

Plant material

Eighteen accessions of *S. pimpinellifolium* were grown in three-fold in a greenhouse at Unifarm, Wageningen University & Research centre, The Netherlands. The *S. pimpinellifolium* accessions G1.1554 (Pimp), G1.1704 (1_31), G1.1589 (2_9), G1.1593 (2_17), G1.1594 (2_19), PI126933 (2_7), PI126925 (2_25) were obtained from the Centre for Genetic Resources in The Netherlands and LA0114 (3_3), LA0121 (3_5), LA1596 (3_31), LA1279 (3_13), LA1374 (3_17), LA1472 (3_19), LA1645 (4_7), LA1924 (4_13), LA2839 (4_19), LA2854 (4_21), LA1349 (4_25) from the Tomato Genetics Resource Center at University of California, Davis. In addition, two cultivars of *S. lycopersicum* (Moneymaker and cherry tomato breeding line developed at Wageningen UR Plant Breeding: F6_Pop20_10) and two accessions from *S. cheesmaniae* (5_17_LA0421 and 6_17_LA1139) were included as controls. Two internal sequential numbers next to the names of the accessions were included as references for better visualization of results.

All genotypes were randomized in three blocks and surrounded by border plants. During two months, fruits from the third truss onwards were harvested per genotype per block in different stages (Additional Fig. 1). All fruits were cut in pieces and frozen in liquid nitrogen and later ground to powder.

Metabolic profiling

Methanol extraction

To 1 g frozen powder of each fruit sample 3 ml methanol containing 0.133 percent formic acid and 12.5 µg/ml ribitol was added as a standard. Samples were shortly vortexed, sonicated for 30 minutes, and centrifuged at maximum speed (~3,000 g) for 10 minutes. The supernatant (aqueous-methanol extract) of each sample was divided into two parts. One part was used for further extraction of polar compounds to be analysed by GC-TOF, the other was directly filtered through a Captiva 96-wells 0.45 µm polypropylene filter plate (Agilent Technologies Netherlands B.V., Amstelveen) for analysis of semi-polar compounds using LC-PDA-QTOF.

GC-TOF

The detection of polar metabolites was performed using GC-TOF. This technique mainly detects polar primary compounds such as amino acids, sugars and organic acids (Lisec *et al.* 2006). For the extraction of these polar compounds, 500 µL of the methanol extract was transferred to a new vial to which 450 µL of water and 250 µL of chloroform were added. The samples were vortexed, left at room temperature for 5 minutes and then centrifuged at maximum speed for 10 min. Twenty five µl of the polar phase was transferred into a 2 ml glass vial with a 100 µL glass insert, dried in a speedvac overnight and capped under argon using magnetic crimp caps. Samples were derivatized online and analysed by GC-TOF as described before (Carreno-Quintero *et al.*, 2012) using a detector voltage of 1700 V. Leco ChromaTOF software 2.0 was used for pre-processing of the raw data.

LC-QTOF MS

For analysis of semi polar metabolites, the filtered aqueous-methanol extract were diluted 6 fold in 75% methanol with 0.1% formic acid. These diluted extracts were analysed on a LC-PDA-QTOF MS system, using C_{18}-reversed phase chromatography (Luna C18, 2.0x150 mm, 3 µm particles; Phenomenex) coupled to both a photodiode array (PDA) detector and a Quadrupole time-of-flight (QTOF) MS instrument (QTOF Ultima, V4.00.00; Waters)

equipped with an electrospray ionization (ESI) source, and a separate lock mass spray, in negative mode (De Vos *et al.*, 2007).

SPME-GC

Volatile metabolites were analysed according to the process described by Tikunov *et al.* (2005). In brief, 1 g frozen powder of each sample of ripened-fruits were weighed into a 5 ml screw-cap glass vial, closed and incubated for 10 min at 30°C. Enzymatic reactions were stopped by adding 1 ml of a EDTA-NaOH aqueous solution and 2.2 g of solid $CaCl_2$ giving a final concentration of 50 mM EDTA and 5 M $CaCl_2$. Vials were immediately closed and then sonicated for 5 minutes. A 1 ml aliquot of the pulp was transferred into a 10 ml crimp cap vial (Waters) and capped for SPME-GC analysis. A Combi PAL autosampler (CTC Analytics) was used for headspace sampling on a 65 μm polydimethylsiloxane-divinylbenzene SPME fiber (Supelco), under continuous shaking and heating at 50 °C for 20 min. Volatiles trapped were separated onto an HP-5 column (30 m 0.25 mm) with helium as carrier gas. Quality control samples from pooled material were added in duplicate for every batch of 12 samples.

Data processing

MS data files of all three platforms were processed in an untargeted manner using MetAlign software (www.metalign.nl; Lommen, 2012) for baseline correction, noise estimation and ion-wise mass spectral alignment (Tikunov *et al.*, 2005; De Vos *et al.* 2007). Next, MSClust software was used to group individual mass signals belonging to the same metabolite and for the reconstruction of mass spectra (Tikunov *et al.*, 2012). The putative identification of metabolites was based upon their mass spectral data (molecular weight and fragmentation patterns). For LC data, compound characteristics were compared using the LC based metabolite database for tomato fruit (MoTo; Moco *et al.*, 2006), the Dictionary of Natural Products (http://dnp.chemnetbase.com), the KNApSAcK database (http://kanaya.naist.jp/) and in-house LC metabolite databases. For both GC-TOF and SPME-GC data, the mass spectra and retention indices obtained were matched with those of authentic standards and available spectral databases (NIST, Golm metabolome DB), using NIST Mass Spectral Search Program v2.0. For the last two platforms a strict Match Factor (MF) threshold of 800 was used to annotate putative compounds. For LC the annotation was restricted to a mass difference of only 5ppm plus a restrained range difference in retention time according to the databases. Metabolic data was normalized by log transformation and a batch effect correction was performed for the SPME samples. The average concentrations of the fruits at ripe stage are shown in the Additional Table 1.

Genetic and metabolic analysis

All accessions were genotyped with a custom made SNP Array (Viquez-Zamora *et al.*, 2013). Using a script in GenStat 16th Edition (VSN International: https://www.vsni.co.uk/software/genstat/) Single Nucleotide Polymorphisms (SNPs) were selected approximately every 0.5Mb through the genome, in total 461 SNPs were chosen to calculate genetic similarities. A neighbour joining analysis using the Manhattan similarity measure with a final branch and 1000 bootstraps was performed using the statistical package PAST version 2.12 (Hammer *et al.*, 2001). Principal components analysis (PCA), hierarchical cluster analysis (HCA) and heat maps with log transformed and normalized data were performed in GeneMaths XT created by Applied Maths NV (http://www.applied-maths.com). ANOVAs were calculated and no block effects were found within genotypes. Averages, standard deviations and t-test were calculated when we wanted to define (non-)significant differences of compounds between genotypes, generally at ripe stage.

Results and Discussion

Genetic similarities

A neighbour joining analysis based on 461 SNPs distributed evenly through the genome was used to assess the genetic similarity between the genotypes (Fig. 1). The *S. pimpinellifolium* accessions 4_19_LA2839 and 1_31_G1.1703 were the closest to *S. lycopersicum* cultivar Moneymaker. The *S. cheesmaniae* accessions were also genetically close to Moneymaker and even closer to Moneymaker than the *S. lycopersicum* cherry breeding line F6_20_010. According to the revised key for the *Lycopersicon* and related *Solanum* species (http://tgrc.ucdavis.edu/key.aspx: Additional Fig. 2) tomato plants with serrate leaves of which the ripe fruits are red with a diameter of more than 1.5 cm and with seeds of 1.5 mm or longer, should be classified as *S. lycopersicum* or *S. lycopersicum* var. *cerasiforme*. Accession 4_19_LA2839 had fruits with a diameter larger than 1.5 cm is likely to be a misclassification and should be considered *S. lycopersicum*. This probable misclassification was already noticed in 2009 when it was mentioned that it could be a *S. lycopersicum* var. *cerasiforme* (TGCR: http:// tgrc.ucdavis.edu/key.aspx). Other red fruited accessions with serrated leaves like 1_31_G1.1703, 4_21_LA2854, 2_17_G1.1593 and 3_13_LA1279 had fruits with diameters smaller than 1.5 cm and were therefor rightfully considered *S. pimpinellifolium*. The cherry line F6_20_010 was genetically more similar to *S. pimpinellifolium* accessions than to the round

Moneymaker. It is likely that the line F6_20_10 has gained *S. pimpinellifolium* alleles through breeding. In that case, we could expect that this line based on a cherry-cherry cross is genetically closer to other *S. pimpinellifolium* accessions than to Moneymaker.

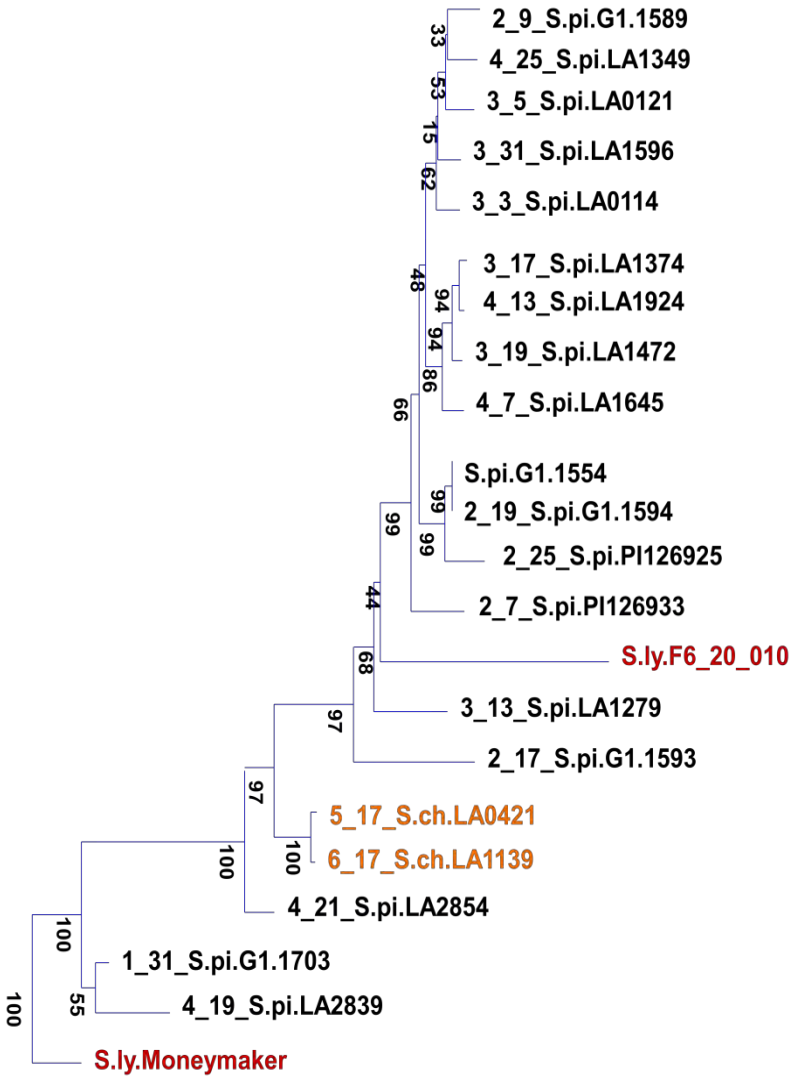


Figure 1. Neighbour joining (NJ) analysis based on 461 markers along the genome using 1000 bootstraps under the Manhattan similarity measure with *S. lycopersicum* Moneymaker as an outgroup. *S. lycopersicum* genotypes are red and *S. cheesmaniae* accessions in orange.

The cluster of *S. pimpinellifolium* and *S. cheesmaniae* accessions which was genetically positioned between the two cultivated *S. lycopersicum* suggests a strong admixture between these three species. The accessions flanked by 2_7_PI126933 and 2_9_G1.15589 seem to represent a clade of true *S. pimpinellifolium* accessions beyond the cherry type *S. lycopersicum*. Genetically, this cluster of *S. pimpinellifolium* accessions was further apart from Moneymaker than the *S. cheesmaniae* accessions.

Metabolic profiling during ripening

Different analytical platforms were used to study three different classes of metabolites: GC-TOF for primary metabolites, such as sugars, organic and amino acids, headspace SPME-GC for various types of volatile organic compounds and LC-qTOF for non-volatile secondary metabolites mostly represented by alkaloids, flavonoids and other phenolic compounds.

Principal Components Analyses (PCA) were performed with all compounds detected (including compounds of unknown identity) to visualize the differences between samples analysed using each individual analytical platform along ripening (Fig. 2). For all platforms used, the first principal component reflected the metabolic changes taking place in fruits of all genotypes during ripening. Although the ripening process was the major source of variation in the dataset, this represented only about 1/5 of the total variation, indicating that a significant amount of the variation might be due to the genetic diversity of the accessions studied. The ripening-dependent changes in the metabolome of *S. cheesmaniae* accessions lagged behind those of *S. lycopersicum* and *S. pimpinellifolium* and ripe fruits of *S. cheesmaniae* had a metabolite composition comparable to that of turning staged *S. pimpinellifolium* and *S. lycopersicum* fruits, despite the presence of fully ripe seeds (data not shown). This suggests that several aspects of the ripening program in *S. cheesmaninae* are delayed or arrested. This is not restricted to colour formation, but is reflected in a broad metabolic phenotype. Such differences in the timing and ripening programs might reflect different adaptation strategies of the species.

Primary metabolites

The GC-TOF platform allowed the detection of 124 primary metabolites from which 36 could be putatively annotated. A Hierarchical Cluster Analysis (HCA) of the genotypes analysed at mature green and ripe fruit stage was performed based on annotated compounds (Fig. 3). The two major clusters of the HCA corresponded to the stages where green and red fruits were separated. In general, the differences between stages were larger than the differences between genotypes in the same stage.

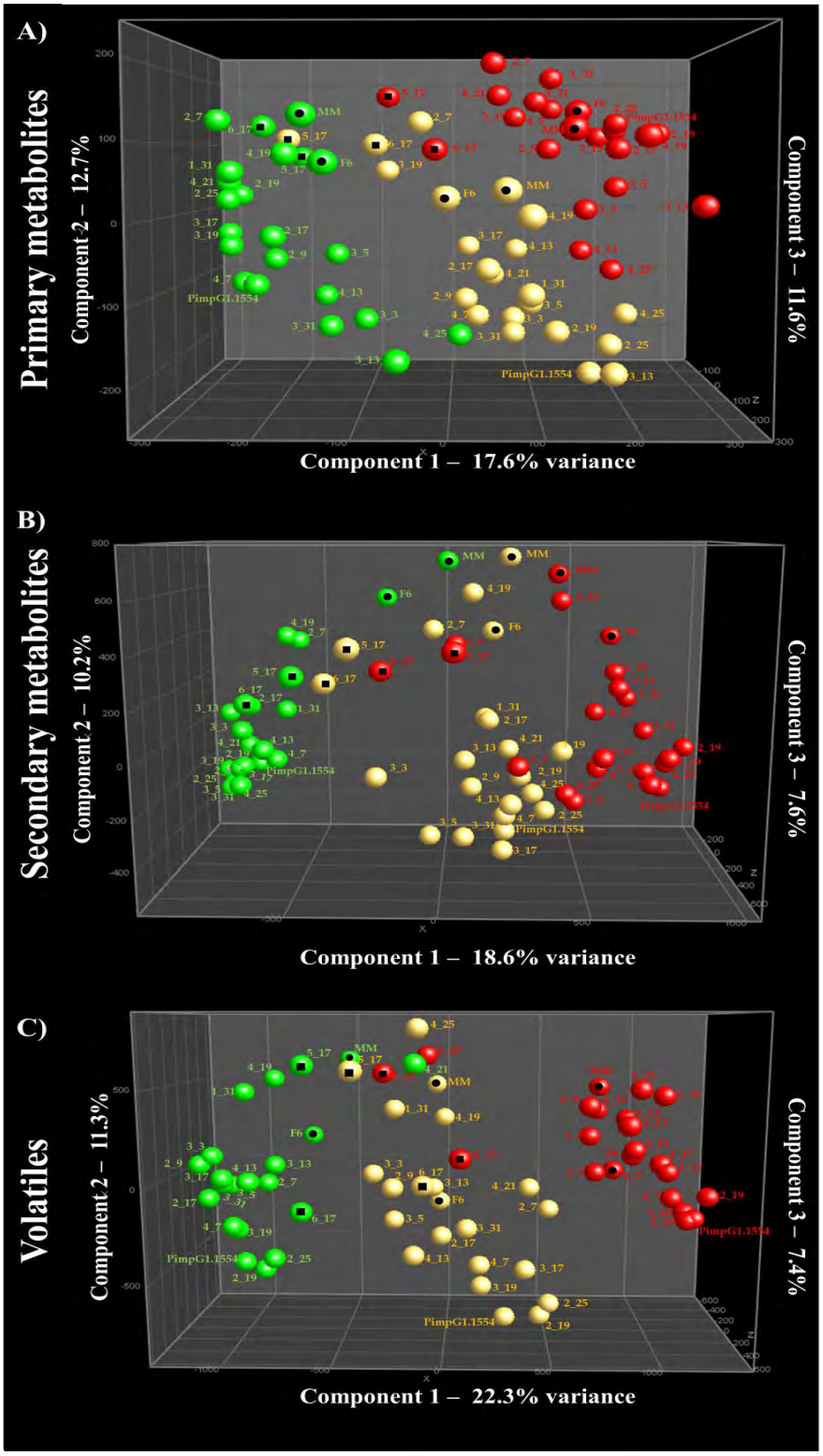


Figure 2. PCAs of all metabolites at the different stages (log transformed and normalized values). A) primary metabolites from GC-TOF. B) secondary (semipolar) metabolites from LC. C) volatiles from SPME-GC. All PCAs show the percentage of the variance within the first three dimensions. *S. lycopersicum* genotypes have a distinct circle symbol (●), *S. cheesmaniae* accessions a squared one (■) and *S. pimpinellifolium* accessions have no inner symbol. Spheres represent the different stages: green (green), turning (yellow) and ripe (red).

Only one accession did not follow this pattern. In *S. cheesmaniae* 6_17_LA1139 the differences between stages were smaller than the differences with the other genotypes. In this accession the red and green stages clustered together, suggesting that the metabolic changes ongoing during ripening in this accession are less pronounced than in the other accessions tested. This is in line with the observations made in the PCA analysis.

Both *S. lycopersicum* genotypes were in the same cluster at green stage together with 4_19_LA2839 and 1_31_G1.1703. However, at ripe stage, the cherry line F6_20_10 was in a distant cluster together with more *S. pimpinellifolium* accessions (Fig. 3). This suggests that the cherry line followed a different ripening pattern than Moneymaker.

Sugars

Sugars directly contribute to quality of tomato, but also play a role in energy storage and as building blocks for other important fruit components. Absolute concentrations and ratios between sucrose, glucose and fructose in tomato fruit are the major determinants of fruit flavour. The levels of sugars like fructose, sucrose, xylose and myo-inositol varied between different genotypes (Additional Fig. 3). The accessions 1_31_G1.1703, 2_25_PI126933, 3_31_LA1596 and 4_25_LA1349 showed no significant differences for the concentrations of fructose, glucose and sucrose when compared among them or with the cherry line F6_20_10, which was known for its superior accumulation of sugars. Therefore, these accessions could be used to improve tomato taste. Sucrose was the sugar with the largest variation between genotypes. This variation might be related to the efficiency of the plants to obtain and convert the monosaccharides glucose and fructose into sucrose. UDP-glucose and fructose 6-phosphate are needed to produce sucrose, however, there were no correlations between the level of UDP-glucose and sucrose. Also, differences in the activity of enzymes like sucrose-6-phosphate synthase can also result in differences in the disaccharide production (Galtier *et al.*, 1993). In addition, there is a chance that the effects of different invertases influence the accumulation of sugars. In this respect, Husain *et al.* (2001) already reported that one invertase alone might not be sufficient to increase the total content of soluble solids of fruits.

In *S. cheesmaniae* 5_17_LA0421 the greatest change in the accumulation of sugars from green to ripe stage was detected. In this *S. cheesmaniae* accession the accumulation of xylose was the highest. Xylose can be an abundant monosaccharide in hemicelluloses and is one of the compounds of cell walls together with lignin. It might be that this *S. cheesmaniae* accession had less hemicellulose reduction during ripening resulting in a high accumulation of free xylose (Lunn *et al.*, 2013).

Organic acids

The two most important organic acids determining the acidity of tomato fruits are citric acid and malic acid. In general, the level of citric acid was not much affected by the ripening stages, while levels of malic acid decreased upon ripening in most accessions (Additional figure 4). It has been shown that there is a preference in taste for higher levels of citric acid relative to malic acid (Carli *et al.*, 2011). Therefore, if the ratio between malic and citric acid changes to more citric acid the tomato flavour improves. Ripe Moneymaker fruits contained the highest level of malic acid. Several accessions showed an increased level of citrate and and increased citrate/malate ratio compared to Moneymaker (e.g. 1_31_G1.1703, 2_17_G1.1593 and 3_5_LA0121). These accessions could be promising donors for genetic improvement of the organic acid levels and balance and could be targeted to study differences in enzymatic activity within the Tricarboxylic Acid Cycle (TCA or Citric Acid Cycle).

Tomatoes are also a source of ascorbic acid, better known as vitamin C. Different relative abundances were measured. The vitamin C content in genotypes F6_20_10, 2_25_PI126933, 3_13_LA1279, 2_9_G1.1589 were at least 2-fold higher compared to Moneymaker. Further exploration should be done taking into account the oxidation of ascorbic acid to dehydroascorbate during the analysis. Despite this there is a possibility that genotypes like 2_25_PI126933, 3_13_LA1279, 2_9_G1.1589 can be used to enhance vitamin C content in tomato fruits.

Amino acids

Sweet, salty, bitter and sour are four of our basic taste experiences. Recently a fifth one was added: umami, the pleasant savoury taste. Glutamate is the amino acid that makes it possible that we can detect the umami flavour with our tongues (Mouritsen and Styrbæk, 2014). All accessions accumulated glutamate (L-Glutamic acid) during ripening. A higher accumulation of glutamate improves the umami taste. Some accessions accumulated 4-fold more than Moneymaker and 2-fold more than the F6_20_10 line and these accessions (4_7_LA1645 and 3_17_LA1374) should be explored for further use in breeding programs. Other amino acids serve as sprecursors for secondary metabolite pathways. For instance, the essential precursor of the phenylpropanoid pathway is phenylalanine. Starting from this amino acid, the plants produce phenylpropanoids. Different enzymes and regulators in the phenylpropanoid pathways can enhance the production of flavonoids, phenylpropanoids, esters and/or monolignols (Additional Fig. 5).

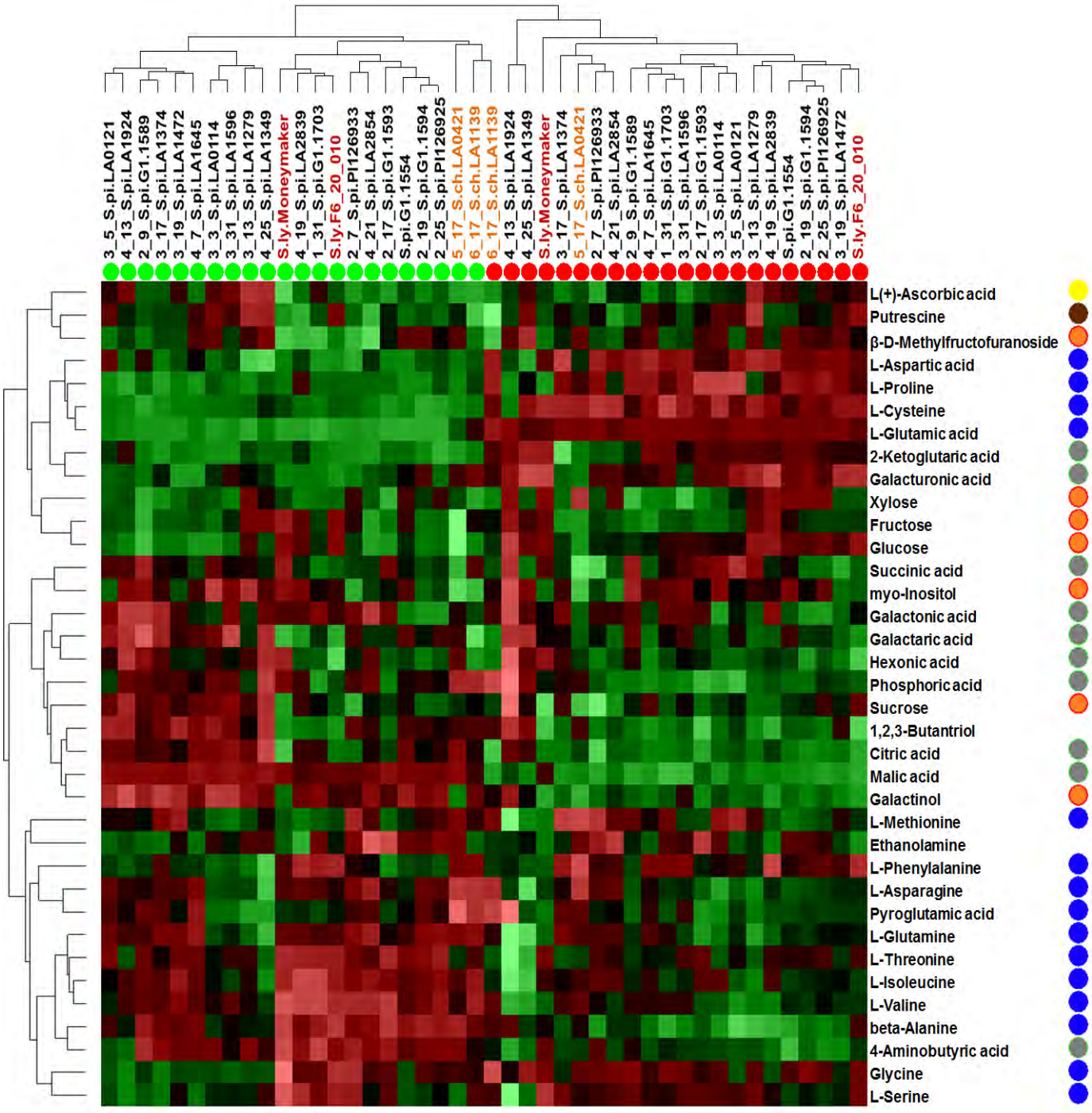


Figure 3. HCA with entries compared under Pearson correlation and UPGMA for clustered entries among annotated primary metabolites. In the X-axis are genotypes in two stages: green (green dot next to the genotype name) and ripe (red dot next to the genotype name). *S. lycopersicum* is in red fonts, *S. cheesmaniae* in orange and in black *S. pimpinellifolium*. The Y-axis represents sugars (orange), amino acids (blue), acids/organic acids (grey) except for ascorbic acid in yellow and putrescine (brown).

Moneymaker accumulated low amounts of free phenylalanine during its ripening stage while the F6_20_10 line accumulated approximately 6-fold more (Additional Fig. 6). This suggests a difference in the production or usage of phenylalanine. The accessions 4_25_ LA1349 and 4_13_LA1924 had the lowest accumulation of free phenylalanine in combination with relatively high levels of flavonoids. Low levels of free phenylalanine might be caused by a higher efficiency in the conversion of the precursor into phenylpropanoids. These differences might reflect variation in the expression and/or coding sequences of one or more phenylalanine ammonia lyase (*PAL)* genes. The degradation of amino acids can result into the presence of off-flavour compounds as putrescine. As can be deduced from its name, putrescine is responsible for a putrefying smell. The *S. lycopersicum* line F6_20_10 and 3_13_LA1279 had the highest levels of putrescine. The *S. cheesmaniae* accession 6_17_LA1139 had 7-fold less putrescine than Moneymaker and 45-fold less than the line F6_20_10. Knowledge about the mechanism in this accession leading to the very low levels of putrescine could lead to a better taste (less bad smell) in new tomato cultivars.

Secondary metabolites

With LC-TOF 379 compounds were detected of which 74 could be putatively identified. An HCA analysis was performed with the identified secondary semi-polar metabolites in green and ripe stages (Fig. 4). The two biggest clusters were based on the green and red stages. For both *S. cheesmaniae* accessions, the differences with the other genotypes were larger than the differences between the ripening stages and they clustered together towards the green stage of other accessions. The secondary semi-polar metabolic profile of both *S. lycopersicum* genotypes made them cluster together at green stage. But, once ripe and as happened with primary metabolites, the secondary semi-polar metabolites of Moneymaker clustered close to the metabolites of *S. pimpinellifolium* 4_19_LA2839, whereas line F6_20_10 was in a distant cluster with the accession 2_17_G1.1593.

Flavonoids and other phenylpropanoids

Compared to Moneymaker, the relative abundance of most phenylpropanoids and flavonoids was higher in the *S. pimpinellifolium* accessions. However, the level of more complex glycosylated forms of flavonoids (e.g. quercetin-hexose-deoxyhexose-hexose attached to the compound –C9H6O2–) varied among the different genotypes. Accession 4_25_ LA1349 showed the largest accumulation of flavonoids. Accessions 4_13_LA1924, 2_25_PI126925 and PimpG1.1554 had high levels of naringenin chalcone and naringenin-glycosides. Accession 2_9_G1.1589 was the one with the highest level of quercetin-3-O-glucoside and quercetin-3-O-rutinoside.

According to the accumulation of the different compounds, we found that Moneymaker accumulated glycoside forms of coumaric acid and caffeic acid. This is in agreement with the observation that chlorogenic acid (3-O-caffeoylquinic acid) and crypto-chlorogenic acid (4-O-caffeoylquinic acid) were present in more complex isomer forms like 4,5-dicaffeoylquinic acid and tricaffeoylquinic acid. For Moneymaker, the higher levels of caffeic acid might result in more complex and stable isomers (Additional Fig. 6). Due the high concentrations of chlorogenic and caffeic acid, Moneymaker has a high antioxidant value (Sato *et al.*, 2011) and might even have anti-obesity properties (Cho *et al.*, 2010). This trend towards high antioxidant levels was also observed in the *S. pimpinellifolium* accession 4_19_LA2839 that was found to be genetically close to Moneymaker. On the other side, these two genotypes had significantly lower accumulation of flavonoids, ferulic acid and sinapic acid. This supports the hypothesis that *S. lycopersicum* like genotypes have a different phenylpropanoid biosynthesis than the other S. *pimpinellifolium* accessions. An accession that behaved similar to *S. lycopersicum* was *S. cheesmaniae* 6_17_LA1139, but *S. cheesmaniae* 5_17_LA0421 behaved quite different. Other S. *pimpinellifolium* accessions and F6_20_10 had intermediate patterns.

The only orange fruited *S. pimpinellifolium,* 4_25_LA1349, had a large accumulation of most metabolites within the phenylpropanoid pathway, especially flavonoids. Also, this was one of the accessions accumulating the least amount of phenylalanine, the precursor of the phenylpropanoid pathway (Additional Fig. 6). On the other hand accessions such as 5_17_LA0421 accumulated the greatest amount of the precursor. This might suggest that 4_25_LA1349 is able to rapidly convert available phenylalanine into different types of phenylpropanoids. Perhaps, as mentioned before, differences might be due to structural or expression variation in phenylalanine ammonia lyase (PAL) genes.

If there is an interest in higher levels flavonoids, the *S. pimpinellifolium* accessions 2_9_G1.1589, 4_25_LA1349, 4_13_LA1924 and 2_25_PI126925 could be interesting sources for a QTL mapping strategy. This could be combined with a candidate gene approach, targeting variation in flavonoid pathway genes, since transgenic approaches have shown that increasing expression of flavonoid pathway genes, such as those encoding chalcone synthase, chalcone isomerase and flavanone 3-dioxygenase or flavonoid transcription factors leads to higher flavonoid content in tomato fruit (Schijlen *et al.*, 2004).

Figure 4. HCA with entries compared under Pearson correlation and UPGMA for clustered entries among annotated of putatively identified secondary semi-polar metabolites. On the X-axis are genotypes in two stages: green (green dots) and ripe (red dots). *S. lycopersicum* in red fonts, *S. cheesmaniae* in orange and in black *S. pimpinellifolium*. The Y-axis represents phenylpropanoids (purple), flavonoids (pink), alkaloids (green), phenolics compounds (yellow), glycosylated volatiles (light blue) and uridinedphosphate-glucose (red).

Alkaloids

Alkaloid concentrations are generally higher in wild accessions (Schwahn *et al.*, 2014). Differences in the levels of green fruit-specific alkaloids, so called tomatine-type alkaloids, could be one of the reasons why in the PCA the green fruits of the *S. lycopersicum* genotypes cluster more towards the turning stage of the wild accessions (Fig. 2). Ripe-fruit type alkaloids, such as lycoperosides and esculeosides, accumulate upon ripening and result from the conversion of tomatine-like alkaloids into lycoperosides and esculeosides. Both *S. lycopersicum* genotypes contained the lowest amounts of alkaloids, while in the *S. cheesmaniae* accessions and 4_25_LA1349 high amounts of all tomatine related compounds were found even after ripening, resulting in fruits with high levels of tomatine and tomatidine.

Glycoalkaloids can play a role in defence mechanisms against fungi, insects, viruses, etc. These compounds are also related with cardioprotective, anticarcinogenic and other health benefits for humans (Friedman, 2002 & 2013). Accession 4_25_LA1349 can be a source for further exploration of both flavonoid and glycoalkaloid related issues.

Volatiles

With the SPME-GC platform we retrieved 429 volatile compounds from which 125 were putatively annotated. An HCA analysis was performed with the identified volatiles in green and ripe stages (Fig. 5). Like with the primary and secondary metabolites the differences in volatile concentration between the *S. cheesmaniae* accessions and the other genotypes were larger than the difference between red and green stages. The *S. cheesmaniae* accessions clustered together with accession 4_25_LA1349, but unfortunately there were no data of the green stage for this accession. In this cluster relatively high amounts of volatiles derived from branched chain amino acids were detected and low amounts of lipid derived volatiles. The fruits of these three accessions did not develop a red colour at ripe stage and were clearly different regarding the volatiles derived from the carotenoid biosynthesis pathway.

Volatiles play an important role in taste perception (Goff and Klee, 2006). Bartoshuk *et al.* (2013a) showed that volatiles can influence flavour and even sweetness perception independently from the real sugar concentration of tomatoes (Bartoshuk and Klee, 2013). Retronasal olfaction happens when, while chewing, volatiles bind to receptors sending signals to the brain (Baldwin *et al.*, 2000). In our research we characterized volatiles derived from lipids, phenylpropanoids, branched chain amino acids, phenols, terpenes, esters and carotenoids. The method of extraction (Tikunov *et al.*, 2005; Tikunov *et al.*, 2010) slightly

stimulated the disruption of the cells in order to estimate a better comparison between the samples and the expectations in taste. In the case of lipids, when cells are disrupted there is an oxidation of compounds through lipoxygenase, hydroperoxide lyase and hydroperoxy cleavage enzymes. Both *S. cheesmaniae* accessions and the *S. pimpinellifolium* 4_25_LA1349 might have a deficiency in the activity of those enzymes or they could have less availability of linoleic acid, the precursor compound for this pathway. Alternatively, the low levels of lipid-derived volatiles in these orange accessions might reflect an arrest of several ripening-dependent developmental programs in these accessions, including the induction of lipid-derived volatile production.

Bartoshuk *et al.* (2013b) concluded in their study that there were volatiles contributing to sweetness perception independently of sugar concentration. From those volatiles, we have identified in our material (Additional Fig. 7): neral, 1-nitro-2-phenylethane, 3-methyl-1-butanol and isovaleric acid. Plus an extra two compounds in the same pathways that might be correlated (2-carene and 6-methyl-5-hepten-2-one). However, these two together and isovaleric acid were found not to have congruence (experienced together) of the odorant and tastant. For all these volatiles, the *S. pimpinellifolium* accessions 4_21_LA2854 and 3_31_LA1596 can provide suitable input for breeding.

Finally, among our set of genotypes we confirmed that some wild accessions have the smoky taste due to guaiacol, eugenol and methyl salicylate which is characteristic of certain modern cultivars (Tikunov *et al.*, 2013). This indicates that the hypothesis that this trait, which is present in modern cultivars, is the result of a spontaneous mutation duringbreeding, can be rejected.

Conclusion

Cultivated tomato 'lost' diversity through domestication and breeding. Many of the lost alleles we could try to (re-)introduce. Our research offers better insight in the large metabolic variation in wild accessions. A non-targeted approach allowed the identification of interesting traits during fruit ripening. An increase in the perception of taste, availability of health-related compounds or targeted resistance-related metabolites could be achieved through the application of obtained knowledge into a more targeted approach.

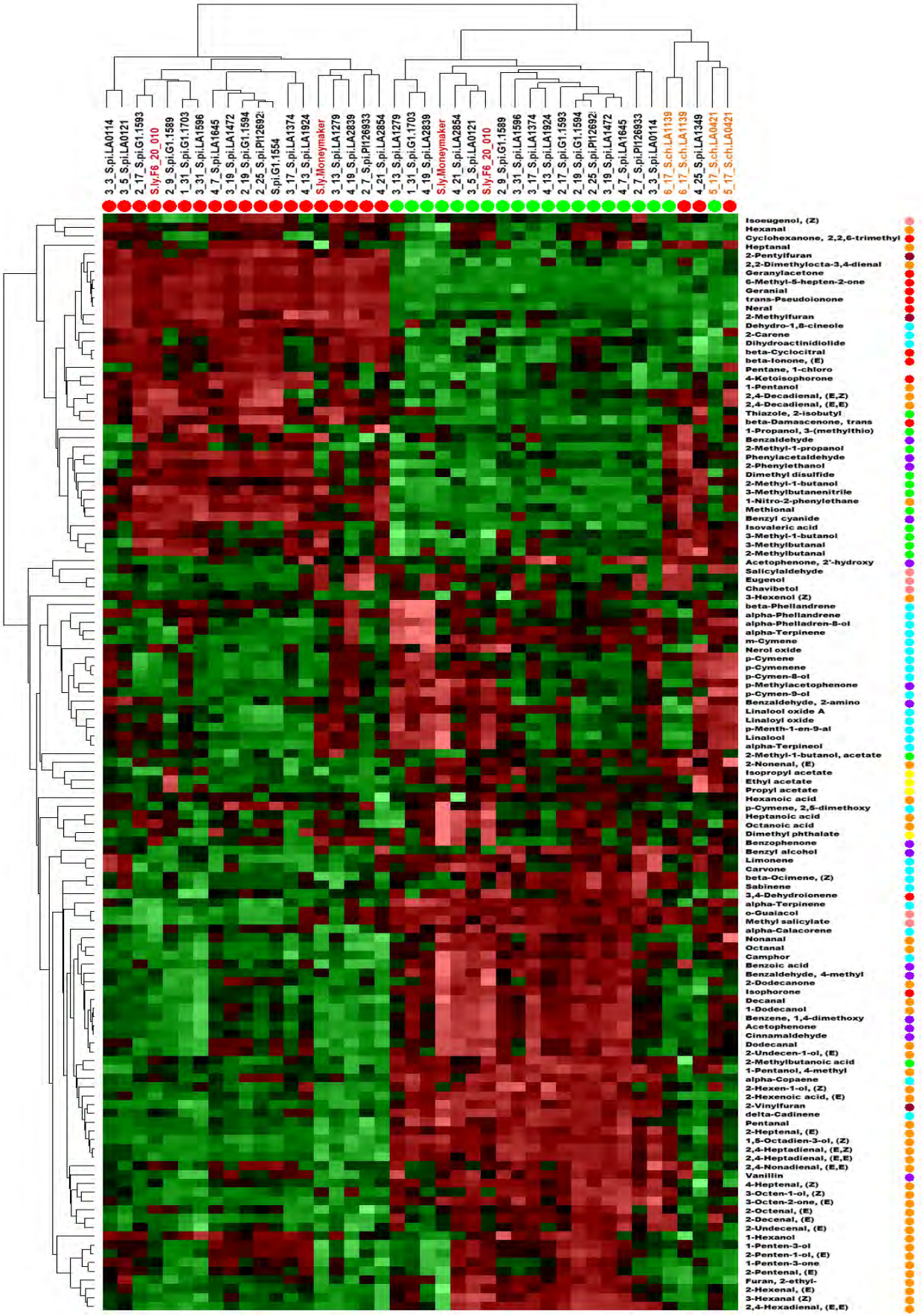


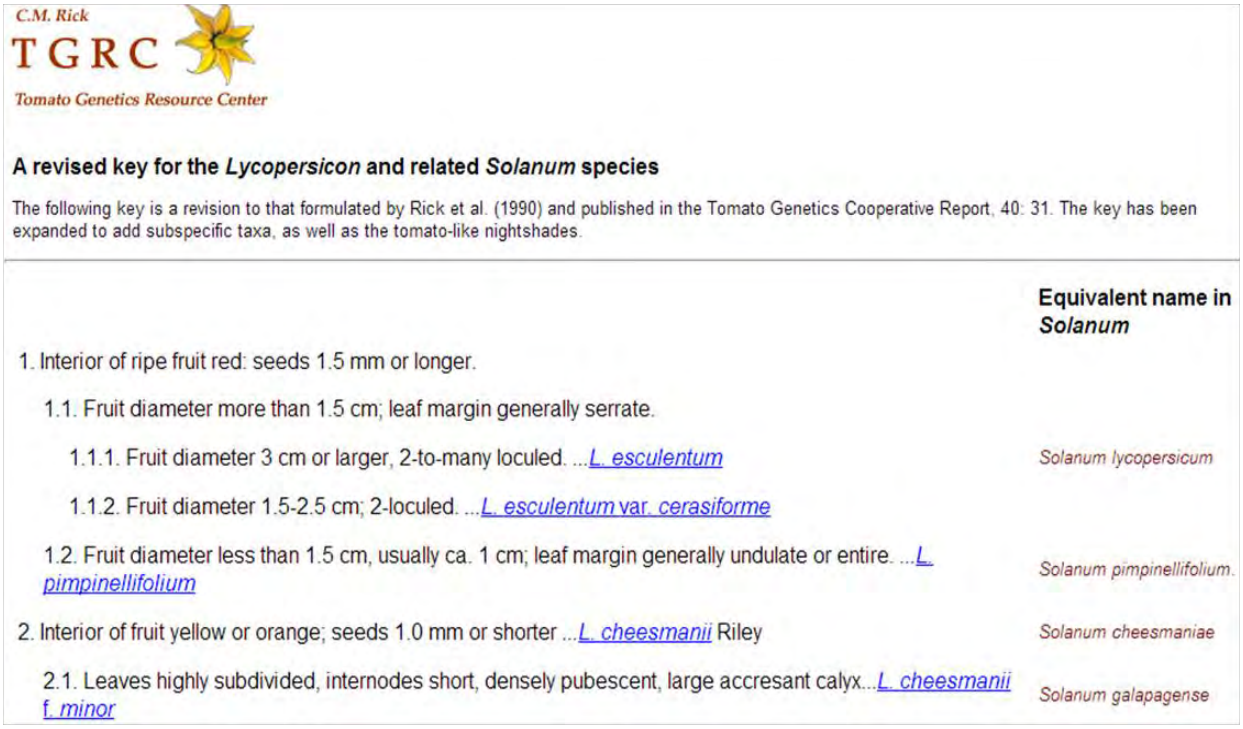
Figure 5. HCA with entries compared under Pearson correlation and UPGMA for clustered entries among secondary semi-polar metabolites. In the X-axis are genotypes in two stages: green (green dot next to the genotype name) and ripe (red dot next to the genotype name). *S. lycopersicum* is in red fonts, *S. cheesmaniae* in orange and in black *S. pimpinellifolium*. The Y-axis represents lipid (orange), phenylpropanoid (pink), branched chain amino acid (green), phenolic (purple), terpene and terpenoid (light blue), ester (yellow), furanoid (brown) and carotenoid (red) derived volatiles.

Acknowledgements

This research was carried out with support of the Netherlands Metabolomics Centre (NMC). The analysed material was developed as part of the research programme of the Dutch plant genomics initiative Centre for BioSystems Genomics (CBSG). Besides our co-authors Ric de Vos and Henriëtte van Eekelen, we would like to express our gratitude to the Buisiness unit Bioscience at Wageningen UR. Special thanks to Carmen Díez-Simón, Roland Mumm and Robert Hall for all their support within this project.

Additional Figures

Additional Figure 1. Pictures as examples of 8 genotypes in different ripening stages. One euro coins (diameter 23.25 mm) and a ruler with centimetres were placed next to the fruits for size estimation.



C.M. Rick

TGRC

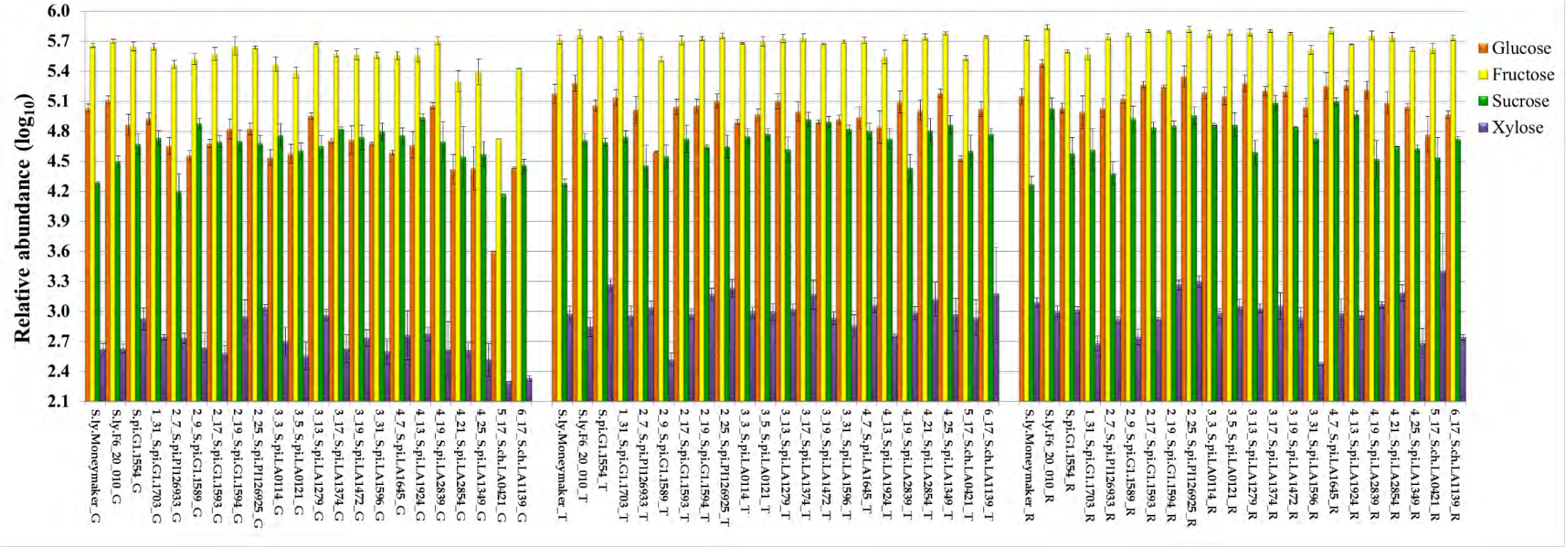
Tomato Genetics Resource Center

A revised key for the *Lycopersicon* and related *Solanum* species

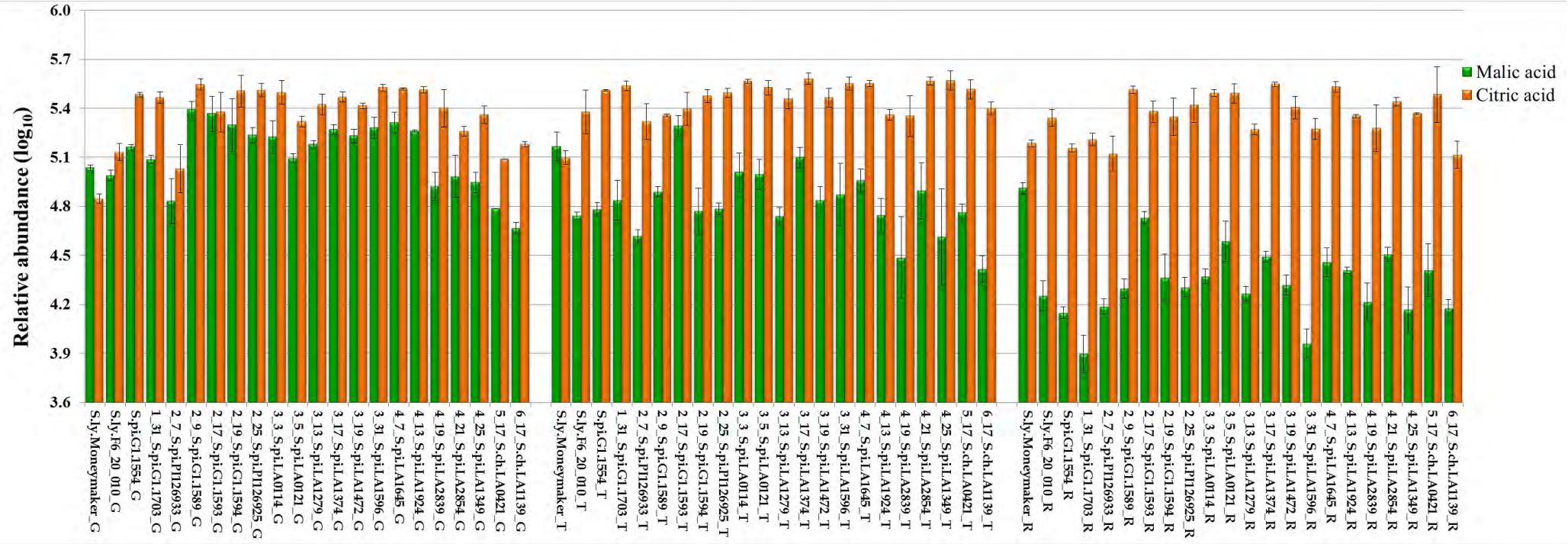
The following key is a revision to that formulated by Rick et al. (1990) and published in the Tomato Genetics Cooperative Report, 40: 31. The key has been expanded to add subspecific taxa, as well as the tomato-like nightshades.

	Equivalent name in *Solanum*
1. Interior of ripe fruit red: seeds 1.5 mm or longer.	
1.1. Fruit diameter more than 1.5 cm; leaf margin generally serrate.	
1.1.1. Fruit diameter 3 cm or larger, 2-to-many loculed. ...*L. esculentum*	*Solanum lycopersicum*
1.1.2. Fruit diameter 1.5-2.5 cm; 2-loculed. ...*L. esculentum* var. *cerasiforme*	
1.2. Fruit diameter less than 1.5 cm, usually ca. 1 cm; leaf margin generally undulate or entire. ...*L. pimpinellifolium*	*Solanum pimpinellifolium*
2. Interior of fruit yellow or orange; seeds 1.0 mm or shorter ...*L. cheesmanii* Riley	*Solanum cheesmaniae*
2.1. Leaves highly subdivided, internodes short, densely pubescent, large accresant calyx...*L. cheesmanii* f. *minor*	*Solanum galapagense*

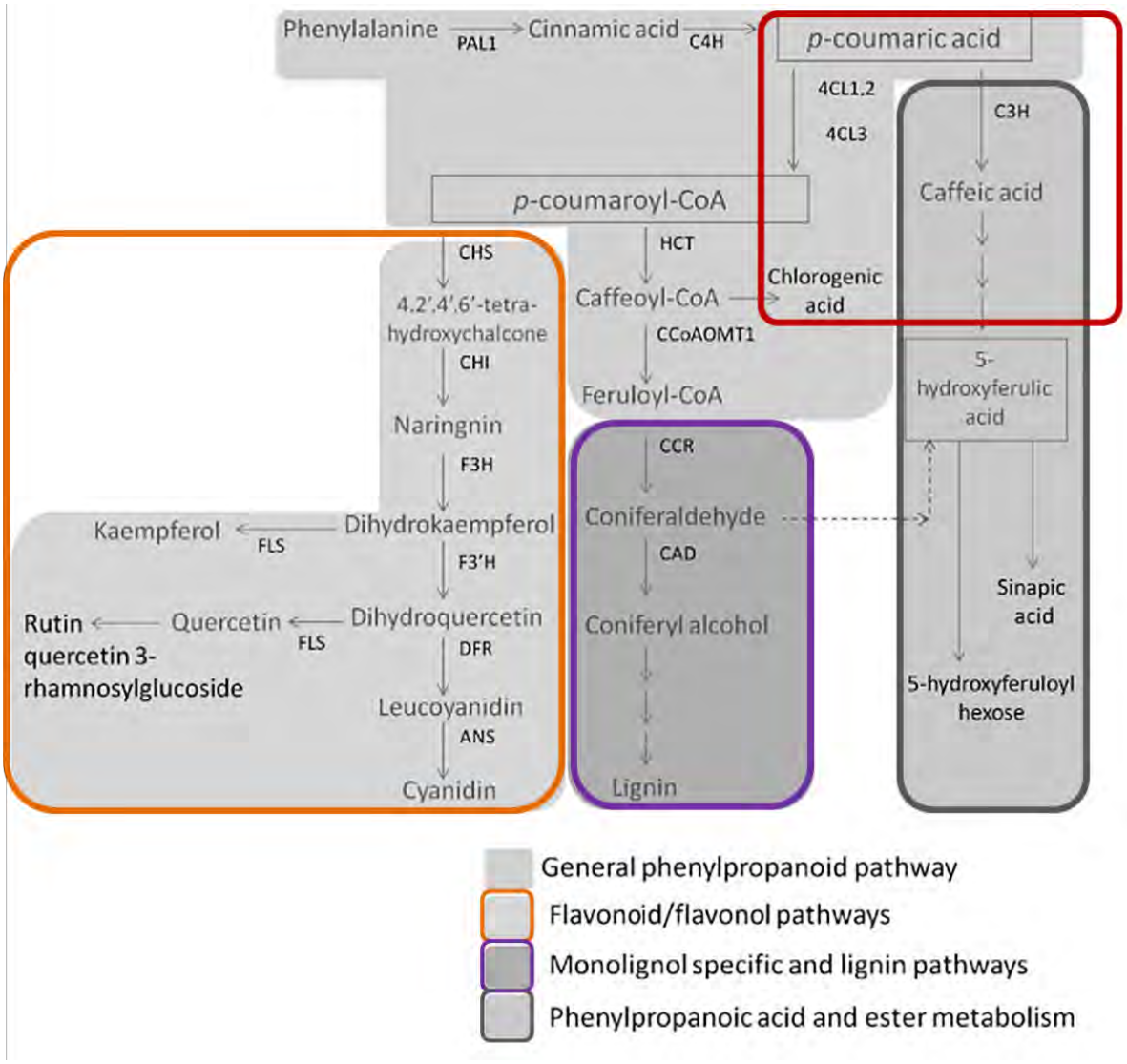
Additional Figure 2. Depiction of sections 1 and 2 of the revised key for the *Lycopersicon* and related *Solanum* species available at the Tomato Genetics Resource Center (Rick *et al.*, 1990. TGRC: http://tgrc.ucdavis.edu/key.aspx).



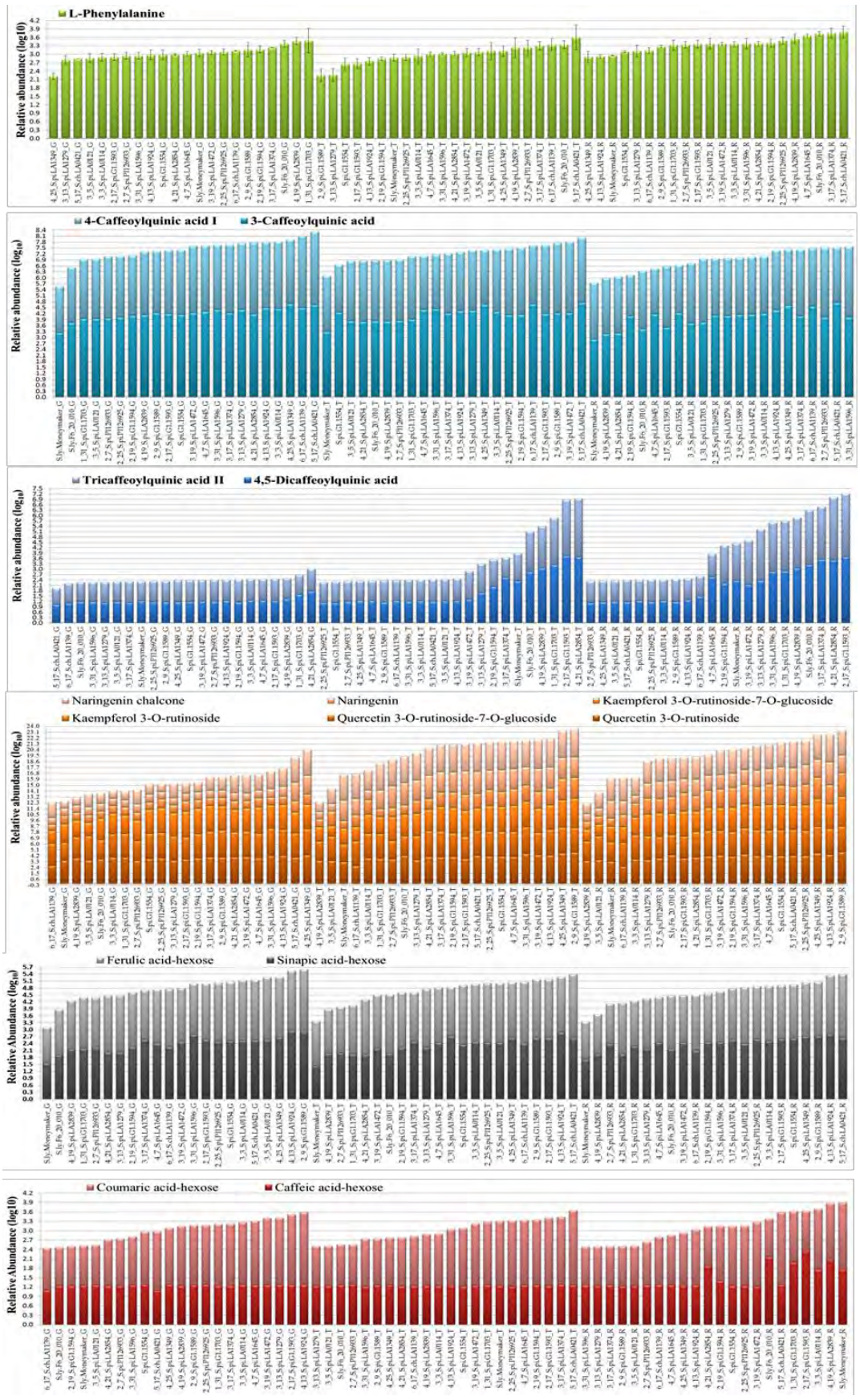
Additional Figure 3. Bar chart of the relative abundance of Glucose, Fructose, Sucrose and Xylose per genotype at different stages: green (G), turning (T), ripe (R). A difference of 0.30103 in log transformed data corresponds approximately to 2x fold-change in relative abundance.



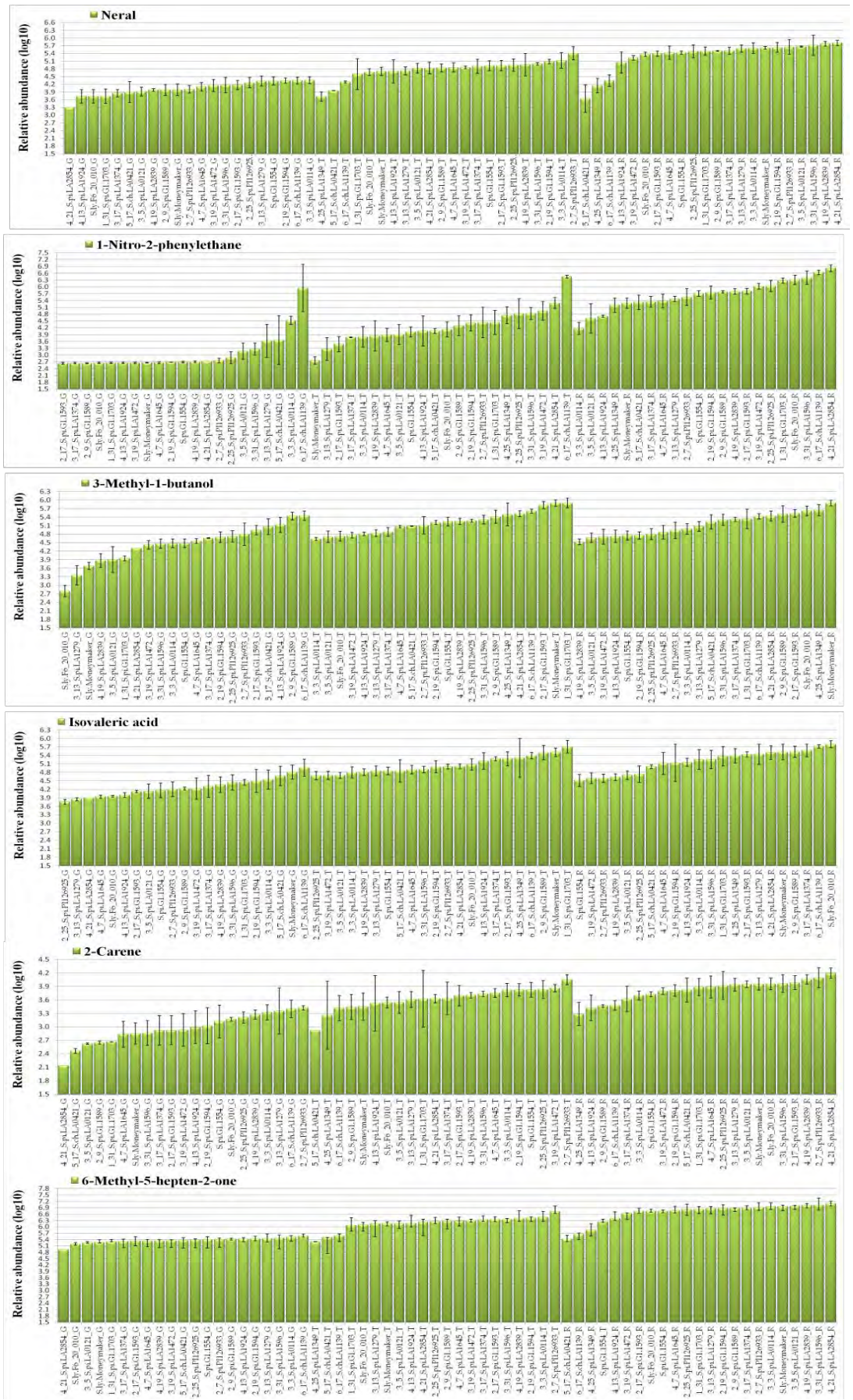
Additional Figure 4. Bar chart of the relative abundance of Malic and Citric Acid per genotype at different stages: green (G), turning (T), ripe (R).



Additional Figure 5. Representation of the metabolic phenylpropanoid pathway in tomato (Adapted from Sade *et al.*, 2015). Inside the red frame are compounds for which *S. lycopersicum* Moneymaker present higher accumulation.



Additional Figure 6. Relative abundance of metabolites within the phenolics metabolism of tomato.



Additional Figure 7. Relative abundance of volatiles associated with the perception of sweetness of tomato according to Bartoshuk *et al.* (2013b).

Additional Table 1. Average relative abundances of identified compounds along three metabolic profiling platforms for fruits of 22 *Solanum* genotypes at ripe stage. Averages of relative abundances are in log 10 transformed values differing in colour intensities going from low (green) to high (red).

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
GCMS3371	Amino acid	**L-Glutamic acid**	4.7	5.0	4.9	5.0	5.2	5.1	5.1	5.1	5.2	5.0	5.3	5.1	5.0	4.9	5.1	4.9	4.8	4.9	4.6	5.3	5.1	5.2
GCMS3820	Amino acid	**L-Phenylalanine**	3.0	3.7	3.1	3.3	3.4	3.4	3.5	3.3	3.3	3.1	3.8	3.4	3.4	3.4	3.4	2.9	3.5	3.4	2.9	3.7	3.8	3.1
GCMS1178	Amino acid	**L-Isoleucine**	3.1	3.2	2.9	3.0	3.5	3.2	3.3	3.4	3.3	2.8	3.4	3.2	3.5	3.1	3.1	2.2	3.1	3.4	2.4	3.5	3.3	3.1
GCMS1240	Amino acid	**L-Proline**	2.6	3.6	3.7	3.5	4.4	3.9	4.0	3.8	4.0	3.0	4.0	3.8	4.3	3.8	4.3	2.5	3.3	3.3	3.1	4.1	3.5	3.9
GCMS1310	Amino acid	**Glycine**	2.9	3.2	2.8	2.9	3.1	3.1	3.1	3.0	3.2	2.9	3.0	2.9	3.1	2.9	2.8	2.6	2.6	3.1	3.0	3.1	3.1	3.3
GCMS1431	Amino acid	**L-Serine**	3.6	3.6	3.0	3.2	3.6	3.4	3.3	3.4	3.6	3.4	3.7	3.3	3.5	3.4	3.5	1.9	3.2	3.5	2.9	3.6	3.4	3.4
GCMS1545	Amino acid	**L-Threonine**	3.1	3.2	2.9	2.9	3.3	3.1	3.2	3.1	3.1	2.8	3.1	3.0	3.0	3.0	3.1	1.5	2.9	3.1	2.4	3.2	3.0	2.8
GCMS1766	Amino acid	**beta-Alanine**	2.4	2.7	1.8	1.8	2.5	2.3	2.2	2.1	2.6	1.7	2.5	2.3	2.2	2.0	1.8	2.4	1.7	2.5	2.2	2.4	2.6	2.6
GCMS2385	Amino acid	**L-Aspartic acid**	4.2	4.5	4.2	4.4	4.5	4.5	4.6	4.6	4.6	4.5	4.9	4.5	4.4	4.4	4.6	3.9	4.4	4.5	3.7	4.8	4.2	4.5
GCMS2458	Amino acid	**L-Methionine**	2.4	2.8	2.4	2.8	2.8	2.7	2.8	3.1	3.0	2.7	3.3	2.8	3.1	2.6	3.1	1.9	2.5	2.9	2.3	3.0	3.2	2.9
GCMS2530	Amino acid	**Pyroglutamic acid**	3.9	4.4	4.0	4.1	4.2	4.3	4.3	4.3	4.3	4.0	4.7	4.3	4.1	4.3	4.3	4.7	4.0	4.4	4.0	4.6	4.5	4.8
GCMS2957	Amino acid	**L-Cysteine**	2.3	2.5	2.0	2.4	2.4	2.3	2.3	2.5	2.3	2.3	2.6	2.4	2.6	2.3	2.3	1.5	2.2	2.5	2.2	2.5	2.4	2.2
GCMS4024	Amino acid	**L-Asparagine**	3.4	3.7	3.3	3.4	3.6	3.6	3.6	3.8	3.6	3.2	3.9	3.6	3.3	3.5	3.5	3.2	3.4	3.8	2.7	3.9	4.2	4.0
GCMS5101	Amino acid	**L-Glutamine**	4.1	4.5	4.2	4.4	4.2	4.4	4.5	4.6	4.5	4.0	4.9	4.5	4.3	4.3	4.5	3.1	4.1	4.6	3.6	4.8	4.6	4.5
GCMS720	Amino acid	**L-Valine**	2.8	3.2	2.9	3.1	3.1	3.1	3.2	3.2	3.3	2.8	3.4	3.1	3.1	3.1	2.8	2.3	2.6	3.1	2.5	3.4	3.4	3.3
GCMS7228	Ascorbic acid	**L(+)-Ascorbic acid**	3.5	3.8	3.5	3.5	3.8	3.7	3.8	3.5	3.8	3.8	3.7	3.7	3.7	3.4	3.7	3.3	3.6	3.7	3.5	3.7	3.6	3.5
GCMS952	Acid	**Phosphoric acid**	4.4	4.6	4.3	4.4	4.6	4.6	4.7	4.6	4.6	4.4	4.8	4.6	4.6	4.4	4.5	4.6	4.4	4.5	4.5	4.7	4.7	4.7
GCMS14084	Alcohol	**1,2,3-Butantriol**	2.5	2.7	2.7	2.7	2.7	2.9	2.9	2.8	3.0	2.7	3.1	3.1	3.0	2.8	3.0	3.1	2.5	3.0	3.0	3.0	3.1	3.1
GCMS1934	Organic acid	**Malic acid**	4.9	4.3	4.2	3.9	4.7	4.4	4.3	4.2	4.3	4.3	4.5	4.3	4.4	4.0	4.6	4.4	4.2	4.5	4.2	4.5	4.4	4.2
GCMS6494	Organic acid	**Citric acid**	5.2	5.3	5.2	5.2	5.4	5.4	5.4	5.1	5.5	5.3	5.6	5.4	5.5	5.3	5.5	5.4	5.3	5.4	5.4	5.5	5.5	5.1
GCMS10075	Organic acid	**Galacturonic acid methoxyamine**	3.4	3.5	2.7	2.7	2.9	3.0	2.9	2.7	2.7	3.2	2.4	3.4	2.9	2.2	3.2	2.8	3.4	2.8	3.3	2.8	2.0	2.6
GCMS10982	Organic acid	**Galactonic acid**	2.7	3.0	2.6	2.9	3.1	2.6	2.6	2.9	3.1	2.7	3.3	2.8	2.6	2.9	2.9	3.1	3.0	3.0	2.8	3.0	2.7	2.9
GCMS11053	Organic acid	**Hexonic acid**	2.8	2.6	2.6	2.6	2.9	2.7	2.8	2.7	2.9	2.6	3.0	2.7	2.8	2.7	2.8	2.9	2.6	2.8	2.7	2.8	2.8	2.6
GCMS11302	Organic acid	**Galactaric acid**	2.4	2.4	2.4	2.4	2.6	2.5	2.6	2.5	2.8	2.3	2.8	2.5	2.5	2.5	2.6	2.6	2.3	2.5	2.6	2.6	2.5	2.3
GCMS1344	Organic acid	**Succinic acid**	2.9	3.0	2.8	2.8	3.2	2.9	2.9	2.6	3.2	3.0	3.1	2.8	3.2	2.9	3.2	2.9	2.7	2.9	2.6	3.1	2.6	2.8
GCMS2679	Organic acid	**4-Aminobutyric acid**	4.7	4.6	3.8	4.4	4.6	4.4	4.3	4.8	4.8	4.4	4.7	4.3	4.6	4.4	4.5	3.8	4.2	4.5	4.2	4.7	4.5	4.5
GCMS3023	Organic acid	**2-Ketoglutaric acid methoxyamine**	2.5	2.4	2.3	2.4	2.6	2.5	2.4	2.1	2.3	2.3	1.6	2.3	2.5	2.2	2.5	2.2	2.2	2.1	2.5	2.7	2.0	2.3
GCMS4536	Polyamine	**Putrescine**	3.0	3.9	3.0	3.3	3.5	3.4	3.4	3.4	3.2	3.9	3.6	3.6	3.6	3.1	3.6	2.9	3.0	2.9	3.4	3.7	3.3	2.2

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
GCMS861	Primary amine + alcohol	**Ethanolamine**	2.6	2.8	3.1	3.1	3.8	3.3	3.3	3.5	3.2	2.9	3.2	3.0	3.4	3.1	3.0	2.7	2.6	3.7	2.9	3.1	3.4	3.0
GCMS11473	Sugar	**myo-Inositol**	4.1	4.3	4.3	4.3	4.5	4.4	4.5	4.3	4.7	4.3	4.6	4.3	4.6	4.4	4.4	4.4	4.2	4.2	4.3	4.6	4.1	4.3
GCMS14401	Sugar	**Sucrose**	4.3	5.0	4.6	4.6	4.8	4.9	5.0	4.4	4.9	4.6	5.1	4.8	4.9	4.7	4.9	5.0	4.5	4.7	4.6	5.1	4.5	4.7
GCMS3912	Sugar	**Xylose methoxyamine**	3.1	3.0	3.0	2.7	2.9	3.3	3.3	2.9	2.7	3.0	3.1	2.9	3.0	2.5	3.1	3.0	3.1	3.2	2.7	3.0	3.4	2.7
GCMS4852	Sugar	**beta-D-Methylfructofuranoside**	2.5	3.3	3.3	2.7	3.4	3.4	3.4	2.9	3.3	3.4	3.3	3.5	3.5	2.8	2.9	3.1	2.8	2.8	3.8	3.6	3.5	2.0
GCMS7692	Sugar	**Fructose methoxyamine**	5.7	5.8	5.6	5.6	5.8	5.8	5.8	5.7	5.8	5.8	5.8	5.8	5.8	5.6	5.8	5.7	5.8	5.7	5.6	5.8	5.6	5.7
GCMS9597	Sugar	**Glucose methoxyamine**	5.2	5.5	5.0	5.0	5.3	5.2	5.4	5.0	5.1	5.3	5.2	5.2	5.2	5.0	5.2	5.3	5.2	5.1	5.0	5.3	4.8	5.0
GCMS15918	Sugar alcohol	**Galactinol**	1.8	2.1	2.1	2.2	2.2	2.2	2.4	2.2	2.4	2.2	2.5	2.2	2.5	2.5	2.3	2.8	1.8	2.5	2.3	2.4	1.9	2.5
LCMS111	Alkaloid	**Glycoalkaloid +FA**	1.3	1.2	1.3	1.3	1.3	1.2	1.2	1.3	1.3	1.3	1.2	1.3	1.2	1.3	1.3	1.2	1.3	1.2	1.3	1.3	1.2	1.3
LCMS2499	Alkaloid	**Esculeoside isomer**	1.3	2.6	3.5	3.3	3.1	3.2	3.5	1.2	3.1	2.9	3.3	3.0	3.3	3.3	2.7	3.5	2.4	3.4	3.6	3.1	3.6	3.3
LCMS2570	Alkaloid	**Esculeoside isomer + FA**	2.2	3.5	4.2	4.0	4.0	4.1	4.0	1.1	4.0	3.8	4.1	3.9	4.1	4.1	3.8	4.0	3.5	4.2	4.3	4.0	4.2	4.0
LCMS2683	Alkaloid	**Glycoalkaloid**	2.7	3.7	4.2	4.0	4.1	4.1	4.1	1.3	4.1	3.8	4.2	4.1	4.2	4.1	4.1	4.2	3.6	4.1	4.5	4.1	4.8	4.7
LCMS3177	Alkaloid	**Hydroxytomatine + FA**	2.6	3.0	2.8	3.1	3.2	2.5	2.8	3.6	3.1	2.9	3.1	2.8	3.5	3.2	3.0	3.3	3.1	3.3	3.6	3.0	3.8	3.6
LCMS3271	Alkaloid	**Esculeoside**	4.2	4.8	4.9	4.9	4.9	4.9	4.9	2.6	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.9	4.8	4.9	4.9	4.9	4.9	4.9
LCMS3466	Alkaloid	**Leptinidine; Hexose , Hexose, Hexose + FA**	1.2	1.2	1.1	2.1	1.2	1.2	1.1	1.7	1.1	1.2	1.4	1.2	1.0	1.9	1.2	1.7	1.2	1.2	1.2	1.2	2.8	1.4
LCMS3804	Alkaloid	**Dehydrotomatine + FA**	1.1	1.6	1.4	1.2	1.1	1.2	1.2	2.1	2.4	1.2	1.2	1.2	2.3	1.4	1.7	1.3	1.3	1.9	3.2	1.2	3.9	3.7
LCMS3888	Alkaloid	**Tomatidine-tetrahexose + FA**	1.2	1.2	1.1	1.1	1.2	1.2	1.2	1.3	1.3	1.2	1.2	1.2	1.3	1.2	1.2	1.3	1.3	1.9	2.9	1.3	3.6	3.7
LCMS3913	Alkaloid	**beta-tomatine**	1.2	2.2	1.3	3.0	2.7	1.5	1.2	2.8	2.8	1.3	2.5	1.2	1.7	3.1	2.0	2.0	1.2	2.2	2.7	2.0	1.2	2.5
LCMS4036	Alkaloid	**alpha-Tomatin**	2.3	2.8	3.0	2.6	2.9	3.0	2.8	3.3	3.7	2.4	2.8	2.8	3.4	2.6	3.0	3.0	2.4	2.9	4.4	2.5	4.8	4.8
LCMS4144	Alkaloid	**Acetoxy-tomatine III + FA**	1.1	1.2	1.1	1.2	1.8	1.2	1.3	3.3	2.5	1.6	1.4	1.3	2.1	1.2	1.2	1.3	1.3	1.2	3.2	1.2	4.0	3.6
LCMS4178	Alkaloid	**Tomatidine-dihexose-dipentose + FA**	1.1	1.1	1.1	1.2	1.3	1.2	1.3	1.2	1.2	1.0	1.2	1.2	1.2	1.2	1.2	1.3	1.2	1.2	1.2	1.3	2.1	2.4
LCMS103	Flavonoid	**Quercetin-dihexose-deoxyhexose-pentose**	1.2	1.3	1.3	1.7	1.3	1.2	1.9	1.3	1.3	1.3	1.4	1.2	1.8	1.3	1.2	2.0	1.4	1.3	1.2	1.7	1.9	1.2
LCMS104	Flavonoid	**Quercetin-dihexose (3,7-O)**	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.2	2.2	1.3	1.3	1.3	1.5	1.2	1.2	1.2	1.2	1.3	2.1	1.3	1.3	1.3
LCMS105	Flavonoid	**Myricetin - dihexose-Deoxy**	1.3	1.3	1.3	1.2	1.3	1.2	1.3	1.3	1.3	1.3	1.3	1.2	1.3	1.3	1.3	1.3	1.2	1.2	1.3	1.3	1.2	1.3

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
LCMS107	Flavonoid	Naringenin hexose-pentose	1.3	1.3	1.9	1.3	1.3	1.7	2.0	1.3	1.2	1.2	1.3	1.3	1.3	1.3	1.3	1.7	1.3	1.4	1.9	1.4	1.9	1.3
LCMS110	Flavonoid	Kaempferol-hexose-deoxyhexose, -pentose, -C9H6O2_	1.3	1.2	2.1	1.3	1.9	2.1	2.1	1.3	1.2	1.2	1.2	1.3	1.2	1.2	1.3	1.3	1.2	1.2	1.3	1.3	1.3	1.3
LCMS112	Flavonoid	Methyl ether of hydroxylated naringenin chalcone	1.3	1.4	1.4	1.3	1.2	1.4	1.4	1.3	1.2	1.3	1.3	1.3	1.3	1.3	1.3	2.2	1.3	1.3	1.2	1.3	1.6	1.2
LCMS113	Flavonoid	Quercetin 3-O-glucoside	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.3	2.1	1.1	1.1	1.1	1.1	1.0	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
LCMS1963	Flavonoid	Quercetin 3-O-rutinoside-7-O-glucoside	3.0	3.1	4.0	3.9	3.3	3.7	3.9	3.3	3.9	3.6	4.0	3.8	3.7	4.0	3.6	3.9	3.0	3.9	4.0	3.8	3.9	3.7
LCMS1980	Flavonoid	Benzyl alcohol-hexose-pentose + FA	1.2	1.3	2.1	1.8	2.1	1.9	2.0	1.6	1.6	2.0	1.6	1.3	1.9	1.9	1.4	1.9	1.4	1.8	2.3	1.7	3.3	2.5
LCMS1999	Flavonoid	Naringenin-C-diglycoside	2.1	2.3	2.7	2.5	2.6	2.7	2.8	2.6	2.7	2.6	2.7	2.8	1.2	2.6	1.2	3.0	1.1	2.7	2.8	2.6	2.9	2.5
LCMS2036	Flavonoid	Dihydrokaempferol-hexose or Eriodictyol chalcone-hexose	1.2	1.3	2.0	1.8	1.9	1.8	2.2	1.2	3.2	1.4	1.8	1.8	1.2	2.0	1.2	2.2	1.2	1.1	3.3	2.2	2.3	1.2
LCMS2112	Flavonoid	Kaempferol 3,7-di-O-glucoside	1.1	1.2	1.4	1.4	1.1	1.1	1.6	1.1	3.0	1.2	1.4	1.1	1.2	1.6	1.1	1.5	1.1	1.3	2.6	1.8	1.1	1.2
LCMS2146	Flavonoid	Kaempferol 3-O-rutinoside-7-O-glucoside	1.2	2.6	2.9	2.6	2.8	3.0	3.1	2.2	3.7	2.5	2.9	2.9	2.2	3.0	1.3	3.0	1.2	2.3	3.4	3.1	2.8	1.9
LCMS2358	Flavonoid	Naringenin-dihexose II	1.2	2.1	3.4	1.8	1.7	3.4	3.6	1.2	3.1	1.3	2.6	2.7	1.2	2.1	1.2	3.3	1.2	1.2	3.7	3.3	1.9	1.1
LCMS2425	Flavonoid	Quercetin 3-O-sophoroside or Myricetin Deoxy - Hexose	1.2	1.2	1.1	1.8	1.5	1.5	1.2	1.2	1.8	1.3	1.8	1.6	2.8	1.9	2.2	1.7	1.0	2.4	2.5	1.7	1.9	1.5
LCMS2443	Flavonoid	Eriodictyol 7-O-glucoside	1.1	1.2	1.5	1.5	1.9	1.2	1.6	1.2	3.1	1.2	1.2	1.3	1.2	1.7	1.2	1.9	1.2	1.2	3.0	2.0	2.0	1.2
LCMS2469	Flavonoid	Dihydrokaempferol-O-diglucoside	1.1	1.3	1.2	1.2	1.3	1.2	1.7	1.2	1.1	1.2	1.5	1.2	1.7	1.3	1.2	1.8	1.2	1.8	2.3	1.2	2.3	1.2
LCMS2596	Flavonoid	Quercetin-hexose-deoxyhexose, -pentose	3.0	3.2	3.3	3.5	2.9	3.2	3.3	2.1	3.4	2.8	3.6	3.7	3.2	3.6	2.8	3.9	3.6	1.2	3.6	3.4	3.9	1.9
LCMS2753	Flavonoid	Isorhamnetin 3-gentiobioside	1.2	1.3	1.9	1.7	1.8	2.0	2.4	1.6	2.1	2.0	2.4	1.8	2.0	1.9	2.3	2.6	1.2	1.2	2.6	2.3	1.8	1.8
LCMS2856	Flavonoid	Quercetin 3-O-rutinoside	3.4	4.3	4.2	4.2	3.9	4.1	4.2	3.9	4.8	3.6	4.2	4.3	3.7	4.4	3.6	4.4	3.4	4.0	4.6	4.3	4.3	2.6
LCMS2917	Flavonoid	Kaempferol-hexose-deoxyhexose, -pentose	2.6	3.1	3.6	3.2	3.5	3.5	3.5	2.9	3.6	3.3	3.8	3.5	3.2	3.3	2.4	3.6	2.7	1.1	3.6	3.7	3.5	2.2

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
LCMS3075	Flavonoid	Naringenin-dihexose I	1.1	1.5	1.6	1.9	2.5	1.6	1.1	1.9	1.4	2.0	2.4	2.1	1.8	1.8	1.8	1.5	1.5	1.9	2.2	1.9	3.7	2.4
LCMS3096	Flavonoid	Naringenin-7-O-glucoside	1.2	2.0	3.4	1.6	1.5	3.3	3.4	1.3	3.3	1.3	2.8	2.7	1.2	2.0	1.1	3.3	1.2	1.2	3.7	2.8	1.7	1.2
LCMS3135	Flavonoid	Phloretin-C-diglycoside	2.8	3.8	3.3	3.0	3.1	3.3	3.4	2.9	3.2	3.1	3.3	3.5	1.1	3.1	1.3	3.6	1.3	3.2	3.2	3.4	3.6	2.6
LCMS3483	Flavonoid	Kaempferol 3-O-rutinoside	2.4	3.3	3.9	3.5	3.8	3.8	3.9	3.6	4.5	3.3	3.7	3.5	2.7	3.7	1.8	3.7	1.8	3.2	4.2	3.9	3.7	2.4
LCMS3566	Flavonoid	Naringenin-hexose, -pentose	1.2	1.4	2.5	1.4	1.3	2.2	2.5	1.3	1.5	1.4	2.5	2.5	1.2	1.6	1.2	2.6	1.2	1.1	2.3	2.1	1.9	1.1
LCMS3597	Flavonoid	Kaempferol-3-O-glucoside	1.2	1.2	1.7	1.5	1.2	1.2	1.5	2.2	2.8	1.2	1.2	1.1	1.2	1.5	1.2	1.5	1.1	1.2	2.0	1.9	1.4	1.2
LCMS3659	Flavonoid	Naringenin-hexose	2.3	3.0	3.6	3.0	3.1	3.6	3.7	2.9	3.4	3.1	3.3	3.2	2.8	3.1	2.9	3.7	2.5	2.8	3.9	3.4	2.9	1.3
LCMS3772	Flavonoid	Quercetin-hexose-deoxyhexose, -hexose, -C9H6O2 (146) II	1.3	2.4	2.0	1.8	1.2	1.6	1.9	1.2	1.2	1.2	1.3	1.2	2.3	1.9	2.1	1.2	2.0	1.2	1.2	1.2	1.2	1.3
LCMS3876	Flavonoid	Kaempferol-hexose-deoxyhexose, -hexose, -C10H8O3 (176) I	1.2	1.2	2.3	1.1	2.5	2.2	2.2	2.2	1.2	2.5	1.2	1.2	1.6	1.2	1.4	1.2	1.2	1.2	1.2	1.2	1.2	1.2
LCMS4164	Flavonoid	Kaempferol-hexose-deoxyhexose, -hexose-C9H6O2 (146) II	1.3	1.9	2.2	1.1	1.5	2.0	2.2	1.9	1.1	1.6	1.2	1.2	2.3	1.2	1.3	1.2	1.2	1.3	1.2	1.2	1.2	1.2
LCMS4219	Flavonoid	Naringenin chalcone-hexose	1.2	2.4	3.2	1.9	2.6	3.1	3.1	2.6	2.4	2.3	2.9	2.6	1.2	2.0	1.2	3.3	1.2	2.3	3.3	3.0	3.0	1.6
LCMS4253	Flavonoid	O-acetylprunin	1.2	2.5	3.7	2.4	2.1	3.6	3.7	2.8	3.4	2.0	3.1	3.1	1.1	2.8	1.2	3.5	1.1	2.0	3.7	2.4	2.5	2.0
LCMS5115	Flavonoid	Naringenin	2.7	2.5	2.8	2.3	2.2	2.2	3.0	2.7	2.7	2.3	2.5	2.2	1.2	2.3	1.2	3.5	1.2	2.5	2.7	2.3	2.9	2.1
LCMS5136	Flavonoid	Naringenin chalcone	3.5	3.4	3.8	3.2	3.3	3.7	3.7	3.6	3.9	3.4	3.9	3.8	2.8	3.4	2.4	4.2	1.7	3.5	3.8	3.7	4.1	3.5
LCMS1573	Nucleoside diphosphate	Uridinediphosphate-glucose	3.9	4.2	4.0	4.1	4.0	4.0	4.0	4.0	4.0	4.0	4.1	3.9	4.0	4.2	4.1	3.9	4.0	4.1	3.9	4.0	3.9	3.8
LCMS1753	Phenolic	Benzyl alcohol-dihexose + FA	2.8	1.8	2.5	2.3	2.5	2.2	2.4	2.6	2.3	1.9	2.3	2.2	2.2	2.5	1.7	2.6	2.8	2.7	2.5	2.3	2.5	3.1
LCMS1983	Phenolic	Jasmonic acid - glucoside	2.2	2.5	3.2	3.1	2.7	3.0	3.0	2.6	3.1	3.0	3.1	3.0	2.9	3.2	2.9	3.0	2.4	3.1	3.3	2.9	3.0	2.7
LCMS100	Phenylpropanoid	Pantothenic acid-hexose	1.6	1.8	2.3	2.3	2.3	2.2	2.3	2.2	2.4	2.2	2.5	2.3	2.4	2.4	2.2	2.5	1.9	2.4	2.5	2.4	2.3	2.1
LCMS101	Phenylpropanoid	Caffeic acid-hexose	1.8	2.2	1.3	2.0	2.4	1.4	1.3	1.3	1.2	1.3	1.3	1.2	1.8	1.3	1.3	1.3	2.1	1.9	1.3	1.3	1.3	1.3
LCMS102	Phenylpropanoid	Coumaric acid-hexose	2.1	1.2	1.9	1.6	1.3	1.8	1.9	1.4	1.3	1.3	1.2	2.1	2.0	1.2	1.3	1.8	1.8	1.3	1.7	1.6	2.3	1.5
LCMS106	Phenylpropanoid	Eugenol malonyl dihexose-pentose	1.3	1.2	1.3	1.7	1.5	1.3	1.3	1.3	2.2	2.1	1.3	1.9	1.8	1.3	1.2	1.3	1.3	1.7	1.5	1.3	1.3	1.3
LCMS109	Phenylpropanoid	Feruloyl tyramine	1.3	1.3	1.3	1.2	1.3	1.3	1.2	1.7	1.3	1.3	1.3	1.3	1.3	1.2	1.3	1.3	1.3	1.3	1.3	1.2	1.2	1.3

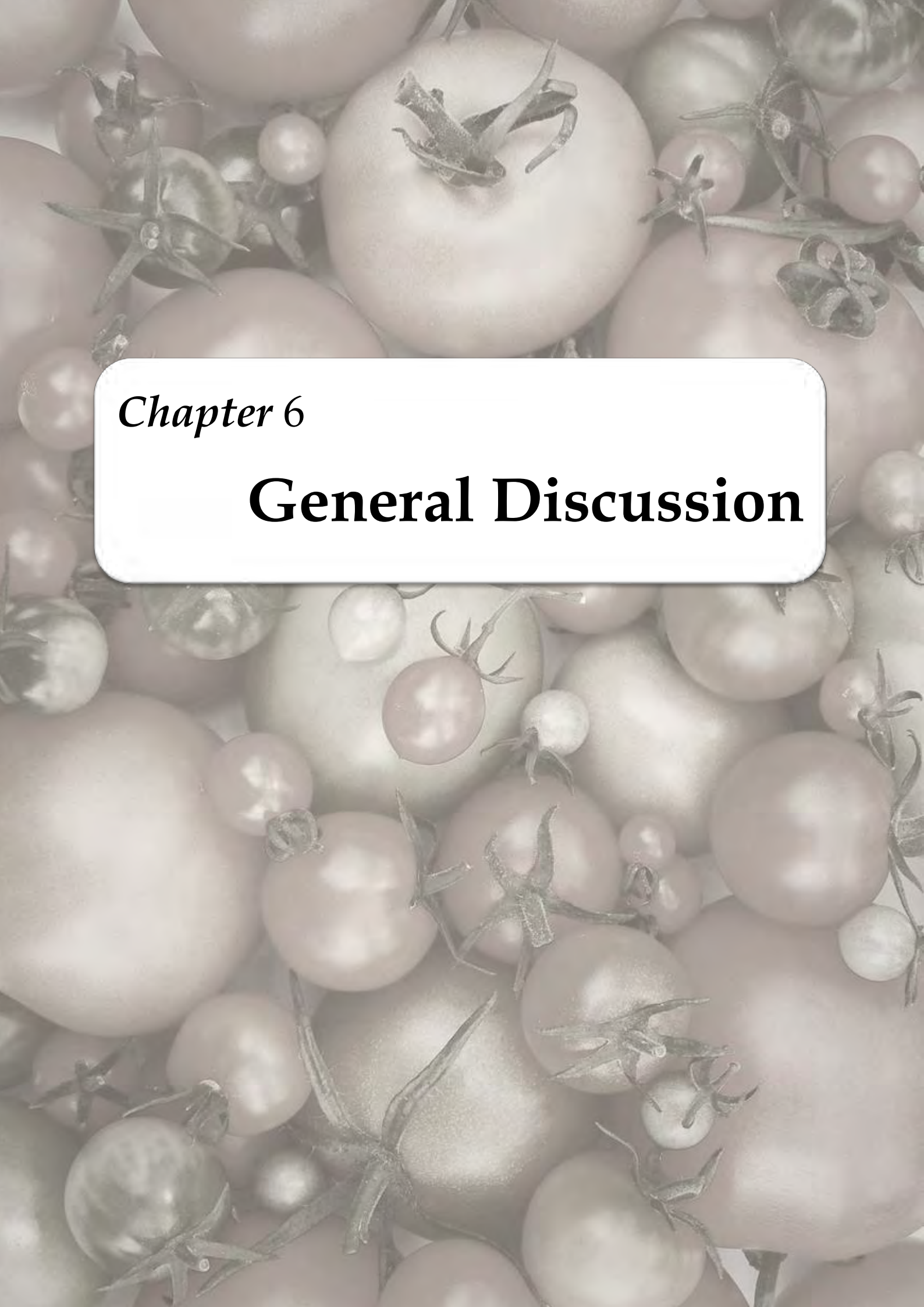
ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
LCMS1761	Phenylpropanoid	Homovanillic acid-O-hexoside	1.9	1.6	1.8	2.2	2.0	2.0	1.7	1.7	2.9	2.0	1.9	2.0	2.7	2.3	2.3	2.0	1.9	1.7	2.8	1.9	2.2	2.2
LCMS1771	Phenylpropanoid	Guaiacol dihexose-pentose	1.2	2.0	2.5	2.7	2.3	2.3	2.5	1.2	2.0	2.7	2.4	2.3	2.0	2.8	2.3	1.2	1.1	2.0	1.4	2.4	2.7	1.1
LCMS1858	Phenylpropanoid	3-Caffeoylquinic acid	2.9	3.4	4.2	3.8	3.5	4.1	4.1	4.0	4.1	4.1	4.1	4.2	4.2	4.0	3.7	4.4	3.2	3.2	4.6	4.2	4.8	4.6
LCMS1897	Phenylpropanoid	Sinapic acid-hexose	1.7	2.2	2.6	2.3	2.6	2.5	2.6	2.4	2.7	2.2	2.6	2.4	2.5	2.5	2.4	2.8	2.0	2.0	2.7	2.4	2.7	2.1
LCMS1978	Phenylpropanoid	Feruloylquinic acid-O-hexoside	1.7	1.3	1.2	1.4	1.6	1.1	1.2	1.8	2.6	1.1	1.9	1.5	2.0	1.3	2.5	1.7	1.3	1.5	2.3	1.5	1.9	1.2
LCMS2011	Phenylpropanoid	Guaiacol-hexose-pentose	1.1	1.2	1.2	1.2	1.1	1.1	1.2	2.0	1.2	1.0	1.1	1.2	1.2	1.2	1.2	1.9	1.1	1.8	1.1	1.2	1.2	1.5
LCMS2020	Phenylpropanoid	Ferulic acid-hexose	1.6	2.3	2.3	2.0	2.3	2.1	2.3	1.7	2.3	2.2	2.2	2.0	2.4	2.2	2.4	2.5	1.7	2.2	2.3	2.0	2.8	2.4
LCMS2075	Phenylpropanoid	Guaicol malonyl dihexose-pentose	1.2	2.7	2.9	3.1	2.9	2.7	2.9	1.2	2.7	2.8	2.7	2.8	2.5	3.2	2.5	1.2	1.1	2.7	1.9	1.7	3.0	1.2
LCMS2124	Phenylpropanoid	4-Caffeoylquinic acid I	2.1	2.4	3.1	3.0	2.8	2.9	2.9	2.9	2.8	2.9	2.9	2.9	3.2	2.9	2.9	3.0	2.3	2.8	3.5	2.9	3.6	3.4
LCMS2196	Phenylpropanoid	Coumaroylquinic acid	2.3	2.6	2.0	2.3	3.0	1.7	1.9	1.2	1.2	2.7	1.9	1.9	1.2	2.5	1.2	2.2	1.8	1.3	1.8	1.6	2.2	1.6
LCMS2222	Phenylpropanoid	Methyl salicylate malonyl dihexose-pentose	1.2	2.6	1.8	2.3	2.5	2.1	1.8	1.2	2.2	1.8	2.4	2.5	1.1	2.1	1.2	1.2	1.3	2.1	1.8	1.2	2.9	1.3
LCMS2264	Phenylpropanoid	Methyl salicylate hexose-pentose + FA	2.4	1.3	1.2	1.2	1.1	1.1	1.2	2.0	1.1	1.2	1.2	1.1	1.2	1.2	1.2	2.1	2.1	2.0	1.2	1.2	1.3	2.1
LCMS2386	Phenylpropanoid	-O-Feruloylquinic acid	2.1	2.4	2.6	2.9	2.7	2.6	2.5	2.6	2.9	2.6	3.1	2.6	3.2	2.9	3.0	3.0	2.1	3.1	3.0	2.9	2.6	2.8
LCMS2514	Phenylpropanoid	Eugenol dihexose pentose	1.2	2.8	2.7	3.0	3.0	2.7	2.5	1.1	3.2	2.5	3.4	3.2	2.2	3.1	2.4	1.2	1.1	3.0	2.6	1.7	2.8	1.2
LCMS3717	Phenylpropanoid	Eugenol-hexose-pentose + FA	2.0	1.3	1.2	1.2	1.6	1.2	1.1	2.8	1.1	1.2	1.1	1.2	1.2	1.2	1.1	1.2	2.5	1.2	1.1	1.2	1.1	1.2
LCMS3779	Phenylpropanoid	4,5-Dicaffeoylquinic acid	2.4	3.2	1.2	2.9	3.7	2.2	1.2	1.1	1.2	2.4	3.6	2.1	1.2	2.9	1.2	1.3	3.0	3.5	1.2	2.6	1.2	1.5
LCMS4948	Phenylpropanoid	Tricaffeoylquinic acid II	2.1	3.1	1.2	2.8	3.6	2.1	1.2	1.2	1.3	2.9	3.0	2.5	1.2	2.8	1.2	1.2	2.9	3.5	1.2	1.3	1.2	1.1
LCMS4210	Saponin	Furostane-tetrol-trihexose	1.5	2.2	1.3	2.8	1.2	1.2	1.3	1.1	3.3	1.8	2.3	1.2	2.0	2.7	1.7	2.2	1.2	2.2	2.5	1.8	2.6	2.0
SPME11754	Branched chain amino acid	Thiazole, 2-isobutyl	6.2	6.1	5.5	6.0	6.1	5.8	5.8	5.7	6.0	6.2	5.5	5.4	5.3	6.0	5.2	4.8	5.8	5.6	6.1	5.1	5.0	5.1
SPME1420	Branched chain amino acid	2-Methyl-1-propanol	3.4	4.3	4.0	4.1	3.6	3.8	3.4	3.4	4.2	3.1	3.9	3.5	3.4	4.3	3.4	3.6	2.8	4.0	4.1	3.1	3.3	4.6
SPME1657	Branched chain amino acid	3-Methylbutanal	4.9	4.8	3.9	4.5	4.6	3.9	4.0	3.9	4.5	4.0	4.4	3.6	3.5	4.5	3.2	3.7	3.6	4.5	4.5	3.9	4.2	4.7
SPME1753	Branched chain amino acid	2-Methylbutanal	4.5	4.6	3.8	4.6	4.3	3.8	4.0	4.0	4.5	4.1	4.4	4.0	3.6	4.8	3.9	4.0	3.6	4.5	5.0	3.9	4.0	4.7

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
SPME2832	Branched chain amino acid	3-Methylbutanenitrile	5.3	5.0	4.1	4.7	5.2	4.7	4.2	4.2	4.9	3.8	4.9	4.0	4.6	4.8	4.0	4.1	4.3	5.0	4.8	4.2	3.6	4.7
SPME2885	Branched chain amino acid	3-Methyl-1-butanol	5.9	5.6	4.8	5.4	5.5	4.8	4.8	4.9	5.5	5.1	5.3	4.7	5.0	5.3	4.7	4.7	4.5	5.4	5.7	4.9	5.2	5.4
SPME2963	Branched chain amino acid	2-Methyl-1-butanol	5.5	5.6	4.8	5.3	5.2	4.9	4.9	4.8	5.4	5.1	5.2	5.0	4.8	5.5	5.2	4.9	4.4	5.3	6.1	4.8	4.9	5.4
SPME5478	Branched chain amino acid	3-Methylbutanoic acid - Isovaleric acid-	5.5	5.8	4.5	5.4	5.5	5.1	4.8	4.6	5.5	5.5	5.6	4.6	5.3	5.3	4.7	5.2	4.6	5.5	5.4	5.1	5.0	5.7
SPME5637	Branched chain amino acid	2-Methylbutanoic acid	4.2	4.7	4.1	4.4	4.3	3.9	4.0	3.9	4.6	4.5	4.6	4.1	4.1	4.7	4.1	4.2	3.8	4.1	4.8	4.4	4.2	2.7
SPME7167	Branched chain amino acid	2-Methyl-1-butanol, acetate	4.2	3.9	2.9	3.8	3.5	3.4	3.7	3.4	4.0	4.7	3.6	2.7	3.6	4.0	3.9	3.0	3.6	4.1	4.6	3.1	3.5	4.0
SPME7968	Branched chain amino acid	Propanal, 3-(methylthio) - Methional-	4.3	4.7	4.7	4.8	4.6	4.4	4.7	4.7	4.8	4.1	4.9	4.4	4.3	5.1	4.2	4.4	3.8	5.0	4.0	4.4	4.3	5.1
SPME3224	Branched chain amino acid	Dimethyl disulfide	3.9	4.4	4.6	4.6	4.8	4.6	4.7	4.5	4.7	4.5	4.9	4.8	4.2	4.8	3.8	4.7	3.6	4.9	5.0	4.8	4.5	4.9
SPME10125	Carotenoid	6-Methyl-5-hepten-2-one	7.0	6.8	6.8	6.8	6.8	6.9	6.8	6.9	6.9	6.9	6.9	6.7	7.0	7.1	7.0	6.4	7.0	7.1	5.9	6.8	5.4	5.7
SPME12763	Carotenoid	Isophorone	4.7	4.7	4.7	4.6	4.7	4.7	4.7	4.7	4.6	4.8	4.5	4.7	4.6	4.6	4.7	4.3	4.7	4.8	4.4	4.7	4.5	4.6
SPME12941	Carotenoid	Acetophenone	4.8	4.8	4.8	4.8	4.8	4.7	4.8	4.8	4.7	4.8	4.7	4.7	4.7	4.8	4.8	4.6	4.8	4.8	4.8	4.8	4.8	4.8
SPME15662	Carotenoid	4-Ketoisophorone	4.8	5.1	5.1	5.0	5.1	5.1	5.3	4.3	4.7	5.1	5.2	5.1	4.8	5.1	5.0	5.1	4.8	4.9	4.8	5.1	4.3	4.5
SPME16488	Carotenoid	Acetophenone, 2'-hydroxy	4.8	4.1	4.2	4.3	4.1	4.5	4.2	5.7	4.3	4.7	4.9	4.5	4.1	4.8	4.5	5.8	4.9	4.7	5.2	5.0	4.6	5.3
SPME17111	Carotenoid	p-Methylacetophenone	4.8	4.7	4.6	4.9	4.6	4.7	4.7	4.9	4.6	4.9	4.7	4.6	4.9	4.8	4.8	4.3	5.0	4.9	4.9	4.6	4.8	4.6
SPME18347	Carotenoid	beta-Cyclocitral	5.2	5.3	5.1	5.1	5.2	5.2	5.3	5.4	5.1	5.3	4.9	5.1	5.2	5.1	5.2	4.5	5.3	5.5	4.3	5.1	4.4	4.9
SPME22247	Carotenoid	3,4-Dehydroionene	3.5	3.5	3.5	3.3	3.5	3.6	3.6	3.6	3.6	3.7	3.4	3.5	3.3	3.4	3.4	3.1	3.7	3.6	3.4	3.6	3.2	3.5
SPME22684	Carotenoid	beta-Damascenone, trans	5.5	5.7	5.8	5.2	5.4	5.7	5.8	5.1	5.3	5.8	5.3	5.3	5.2	5.3	5.0	5.0	5.3	5.3	5.4	5.3	5.3	5.1
SPME25278	Carotenoid	beta-Ionone, (E)	5.6	5.8	5.7	5.4	5.7	5.7	5.8	5.9	5.3	5.8	5.3	5.6	5.5	5.5	5.6	4.9	5.6	5.9	4.7	5.5	4.8	5.1
SPME26704	Carotenoid	trans-Pseudoionone	5.0	4.6	4.7	4.6	4.6	4.9	4.7	5.0	4.7	4.8	4.7	4.4	4.7	5.0	4.8	4.2	5.1	5.1	2.6	4.7	2.7	2.6
SPME8687	Carotenoid	beta-Ocimene, (Z)	3.9	4.2	4.3	4.3	4.3	4.4	4.5	4.7	4.7	4.1	4.5	4.4	4.5	4.3	4.9	4.5	4.2	4.3	4.5	4.4	4.9	4.6
SPME1243	Ester	Ethyl acetate	4.7	4.8	4.6	4.6	4.6	4.4	4.6	4.5	5.5	4.6	4.5	4.4	4.7	4.6	4.7	4.4	4.5	4.5	5.5	4.6	4.7	4.6
SPME1711	Ester	Isopropyl acetate	2.6	2.7	3.2	3.6	2.7	3.1	3.2	3.0	3.9	3.4	2.6	3.1	3.4	3.9	3.3	3.0	3.3	3.5	4.1	3.1	3.5	4.0
SPME2674	Ester	Propyl acetate	3.0	3.1	2.7	3.0	2.9	2.9	3.3	2.6	4.3	3.2	2.7	2.7	3.0	3.7	2.8	2.5	2.8	2.9	4.7	2.6	3.5	3.2
SPME_4188	Lipid	3-Hexanal (Z)	4.7	4.7	4.7	4.7	4.7	4.6	4.7	4.7	4.6	4.7	4.6	4.7	4.7	4.7	4.7	4.4	4.7	4.6	4.6	4.6	4.6	4.6
SPME_4418	Lipid	Hexanal	6.0	6.1	6.1	6.1	6.2	6.1	6.2	6.1	5.9	6.2	6.1	6.1	6.2	6.1	6.2	5.7	6.2	6.3	5.6	6.1	5.7	5.9
SPME_6102	Lipid	3-Hexenol (Z)	5.1	5.4	5.4	5.4	5.4	5.3	5.4	5.4	5.4	5.3	5.6	5.4	5.5	5.5	5.7	5.3	5.3	5.4	5.3	5.4	5.3	5.4

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
SPME10333	Lipid	3-Octen-1-ol, (Z)	4.9	4.6	4.6	4.6	4.6	4.5	4.6	4.6	4.5	4.5	4.6	4.4	4.7	4.6	4.7	4.2	5.0	4.6	4.4	4.5	4.6	4.7
SPME10420	Lipid	2-Pentylfuran	5.2	5.4	5.4	5.3	5.3	5.4	5.6	5.4	5.0	5.4	5.5	5.4	5.3	5.4	5.4	5.1	5.3	5.8	4.9	5.4	4.9	5.0
SPME10470	Lipid	2-Hexenoic acid, (E)	4.4	4.3	4.5	4.5	4.4	4.3	4.5	4.5	4.5	4.3	4.5	4.4	4.5	4.5	4.6	4.4	4.4	4.4	4.3	4.5	4.2	4.3
SPME10571	Lipid	2,4-Heptadienal, (E,Z)	4.3	4.5	4.4	4.5	4.5	4.3	4.5	4.4	4.5	4.7	4.6	4.4	4.7	4.6	4.7	4.3	4.3	4.4	4.6	4.4	4.6	4.8
SPME10672	Lipid	Octanal	5.4	5.4	5.3	5.3	5.3	5.3	5.4	5.3	5.4	5.5	5.3	5.3	5.3	5.3	5.3	5.1	5.3	5.4	5.4	5.3	5.4	5.4
SPME10951	Lipid	2,4-Heptadienal, (E,E)	5.3	5.4	5.4	5.4	5.5	5.4	5.5	5.4	5.4	5.4	5.5	5.4	5.5	5.5	5.5	5.3	5.4	5.4	5.5	5.4	5.4	5.6
SPME11638	Lipid	3-Octen-2-one, (E)	4.9	4.9	4.6	4.8	4.7	4.7	4.7	4.9	4.7	4.7	4.7	4.7	5.1	4.9	4.9	4.3	4.9	5.1	4.8	4.6	4.6	4.9
SPME11881	Lipid	Cyclohexanone, 2,2,6-trimethyl	4.8	4.9	4.7	4.6	4.8	4.8	4.8	4.9	4.6	4.9	4.5	4.7	4.8	4.7	4.8	4.1	4.8	4.9	4.5	4.7	4.4	4.6
SPME12307	Lipid	Heptanoic acid	5.2	5.2	5.0	5.1	5.2	4.9	5.1	5.1	5.1	5.1	5.0	5.0	5.0	5.1	5.0	4.7	5.2	5.1	5.0	5.0	5.0	5.2
SPME12394	Lipid	2-Octenal, (E)	5.9	6.2	6.0	6.0	6.0	6.0	6.0	5.8	6.0	6.0	5.9	5.9	5.9	6.0	5.8	5.6	6.0	6.1	5.9	5.9	5.9	6.0
SPME1264	Lipid	2-Methylfuran	4.1	3.6	3.3	3.9	3.4	3.3	3.3	3.6	3.6	3.8	4.1	3.4	3.7	4.0	3.7	3.6	4.1	3.8	2.7	3.7	2.6	2.6
SPME13944	Lipid	Nonanal	5.8	5.8	5.8	5.8	5.8	5.7	5.8	5.8	5.8	5.9	5.7	5.7	5.8	5.8	5.7	5.6	5.8	5.9	5.8	5.8	6.0	5.8
SPME15306	Lipid	2,2-Dimethylocta-3,4-dienal	5.3	5.1	4.9	5.0	5.0	5.1	5.0	5.4	5.0	5.2	5.0	4.8	5.0	5.2	5.1	4.6	5.4	5.4	3.8	5.0	4.2	4.8
SPME15721	Lipid	Octanoic acid	5.8	5.9	5.6	5.7	5.8	5.5	5.7	5.6	5.6	5.7	5.5	5.6	5.6	5.6	5.5	5.3	5.7	5.7	5.6	5.5	5.6	5.7
SPME15970	Lipid	2-Nonenal, (E)	4.9	4.9	4.8	4.9	5.0	4.8	4.9	4.8	4.8	4.8	4.6	4.6	4.8	4.9	4.7	4.5	4.9	4.9	4.9	4.8	5.4	5.3
SPME17304	Lipid	Decanal	5.1	5.1	5.1	5.1	5.1	5.1	5.1	5.0	5.0	5.1	5.0	5.0	5.1	5.1	5.1	4.9	5.1	5.1	5.1	5.1	5.1	5.1
SPME17717	Lipid	2,4-Nonadienal, (E,E)	4.3	4.5	4.5	4.3	4.4	4.5	4.6	4.4	4.5	4.1	4.5	4.2	4.3	4.5	4.2	3.9	4.4	4.4	4.0	4.4	4.3	4.6
SPME19151	Lipid	2-Decenal, (E)	5.4	5.5	5.3	5.2	5.4	5.3	5.4	5.2	5.3	5.4	5.2	5.2	5.3	5.2	5.1	4.9	5.2	5.5	5.3	5.2	5.4	5.6
SPME19978	Lipid	2,4-Decadienal, (E,Z)	5.2	5.7	5.6	5.3	5.4	5.4	5.6	5.1	5.5	5.3	4.9	5.2	5.2	5.3	5.1	4.4	5.5	5.4	5.1	5.0	5.4	5.4
SPME2002	Lipid	1-Penten-3-ol	4.9	5.3	5.3	5.1	5.3	5.3	5.4	5.2	5.1	5.2	5.4	5.3	5.3	5.2	5.2	5.1	4.9	5.2	5.2	5.3	5.1	5.4
SPME20471	Lipid	1-Nitro-2-phenylethane	5.3	6.3	5.7	6.3	5.8	5.8	6.1	5.6	5.8	5.5	5.4	6.0	4.2	6.4	4.6	4.7	5.8	6.8	5.2	5.4	5.3	6.7
SPME20661	Lipid	2,4-Decadienal, (E,E)	5.1	5.8	5.8	5.3	5.5	5.7	5.9	5.2	5.6	5.4	5.3	5.3	5.3	5.4	5.1	4.5	5.4	5.5	5.3	5.3	5.5	5.5
SPME2113	Lipid	1-Penten-3-one	5.0	5.3	5.2	5.0	5.2	5.1	5.2	5.1	5.0	5.3	5.2	5.2	5.2	5.1	5.3	5.0	5.1	5.0	5.2	5.2	4.9	5.1
SPME21773	Lipid	2-Undecenal, (E)	5.0	5.1	5.0	4.9	5.1	5.1	5.2	4.9	5.0	5.1	5.0	5.0	5.0	4.9	4.9	4.8	4.9	5.3	5.0	5.0	4.8	5.2
SPME22515	Lipid	2-Dodecanone	4.2	4.2	4.2	4.2	4.2	4.3	4.3	4.3	4.1	4.2	4.3	4.3	4.2	4.2	4.1	3.9	4.2	4.2	4.1	4.2	4.1	4.2
SPME22860	Lipid	Dodecanal	4.3	4.3	4.4	4.3	4.2	4.4	4.4	4.3	4.3	4.4	4.4	4.4	4.4	4.4	4.4	4.2	4.3	4.3	4.4	4.4	4.3	4.4
SPME23134	Lipid	1-Dodecanol	4.1	4.1	4.1	4.1	4.1	4.2	4.2	4.2	4.1	4.1	4.2	4.2	4.1	4.1	4.1	3.9	4.1	4.2	4.1	4.1	4.1	4.1
SPME2329	Lipid	Pentanal	5.2	5.5	5.3	5.3	5.4	5.3	5.4	5.1	5.3	5.4	5.3	5.3	5.3	5.3	5.2	5.1	5.3	5.4	5.2	5.4	5.3	5.4
SPME23873	Lipid	Geranylacetone	6.5	6.0	6.2	6.0	6.1	6.4	6.2	6.4	5.8	6.3	6.4	6.1	5.9	6.3	6.2	5.6	6.5	6.3	4.3	6.1	5.0	5.1
SPME2476	Lipid	Furan, 2-ethyl-	4.9	5.1	5.0	4.9	4.9	5.0	5.1	5.2	4.7	5.1	5.2	4.9	5.1	5.0	5.4	4.9	4.8	5.1	4.9	5.0	4.6	4.7
SPME25389	Lipid	2-Undecen-1-ol, (E)	3.8	3.8	3.9	3.8	3.7	3.8	3.8	3.8	3.8	3.8	3.8	3.9	3.8	3.9	3.8	3.7	3.8	3.8	3.9	3.9	3.8	3.8
SPME3314	Lipid	2-Pentenal, (E)	5.1	5.3	5.2	5.2	5.3	5.2	5.3	5.1	5.1	5.4	5.2	5.3	5.2	5.2	5.3	5.1	5.2	5.3	5.2	5.2	5.0	5.0

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
SPME3372	Lipid	Pentane, 1-chloro	4.6	4.8	4.7	4.7	4.8	4.8	4.9	4.5	4.6	4.8	4.6	4.7	4.4	4.7	4.5	4.3	4.7	5.1	4.4	4.6	4.8	4.7
SPME3559	Lipid	1-Pentanol	5.4	5.8	5.7	5.6	5.6	5.6	5.8	5.3	5.6	5.6	5.7	5.6	5.6	5.7	5.4	5.4	5.4	5.7	5.4	5.7	5.3	5.4
SPME3652	Lipid	2-Penten-1-ol, (E)	4.9	5.2	5.2	5.0	5.2	5.1	5.3	5.2	5.0	5.1	5.3	5.1	5.0	5.2	5.1	5.0	4.9	5.1	5.0	5.1	4.8	5.2
SPME5665	Lipid	1-Pentanol, 4-methyl	3.2	3.4	3.2	3.3	3.1	3.1	3.3	3.4	3.4	3.2	3.4	2.9	2.6	3.6	2.7	3.1	2.6	2.8	3.8	3.1	2.7	2.6
SPME5899	Lipid	2-Hexenal, (E)	5.8	5.9	5.9	5.9	5.9	5.8	5.9	5.9	5.8	5.8	6.0	6.0	6.0	6.0	6.1	5.8	5.9	5.8	5.8	5.9	5.7	5.9
SPME6840	Lipid	2-Hexen-1-ol, (Z)	3.8	3.8	3.9	4.1	3.9	3.7	3.9	3.8	4.1	3.8	4.1	3.8	3.9	4.4	3.9	4.3	3.9	3.9	4.0	4.0	3.9	4.1
SPME6908	Lipid	1-Hexanol	5.4	5.4	5.5	5.5	5.6	5.3	5.6	5.5	5.4	5.6	5.6	5.4	5.5	5.8	5.5	5.2	5.2	5.6	5.3	5.5	5.0	4.9
SPME7656	Lipid	4-Heptenal, (Z)	3.6	3.8	3.7	3.7	3.6	3.8	3.8	3.7	3.6	3.7	3.7	3.8	3.8	3.8	3.7	3.4	3.8	3.9	3.0	3.7	3.3	3.6
SPME7822	Lipid	Heptanal	4.5	5.1	5.1	5.0	5.1	5.1	5.2	5.0	5.1	5.2	5.1	5.1	5.1	5.1	5.0	4.7	5.1	5.1	5.1	5.1	5.1	5.3
SPME8073	Lipid	2,4-Hexadienal, (E,E)	5.3	5.3	5.3	5.2	5.3	5.3	5.4	5.5	5.2	5.2	5.4	5.4	5.5	5.3	5.5	5.1	5.3	5.2	5.3	5.2	5.2	5.4
SPME9019	Lipid	2-Heptenal, (E)	6.0	6.4	6.2	6.2	6.3	6.2	6.4	6.1	6.2	6.2	6.3	6.3	6.2	6.3	6.2	6.0	6.2	6.4	5.9	6.3	6.2	6.3
SPME9134	Lipid	Hexanoic acid	5.8	5.9	5.6	5.8	5.4	5.7	5.2	5.3	5.6	5.7	5.7	5.7	5.7	5.7	5.7	5.3	5.8	3.0	5.5	5.7	5.5	5.8
SPME9535	Lipid	1,5-Octadien-3-ol, (Z)	4.0	4.4	3.9	4.1	3.9	3.8	4.1	3.9	4.1	4.1	4.2	3.8	4.2	4.3	4.1	3.9	3.8	4.2	4.1	4.2	4.2	4.6
SPME9583	Lipid	2-Vinylfuran	4.1	4.0	3.9	3.9	3.9	3.8	4.1	4.1	4.0	4.0	4.0	4.0	3.9	4.0	3.9	3.8	3.9	4.1	4.1	4.1	4.0	4.2
SPME9765	Lipid	1-Propanol, 3-(methylthio)	3.2	3.9	3.8	4.1	3.7	3.6	3.6	3.8	3.5	3.0	2.7	3.2	2.7	4.4	2.7	2.5	3.0	4.8	3.1	3.0	3.2	4.1
SPME11620	Phenolic	Benzyl alcohol	5.9	5.3	5.3	5.6	5.6	5.2	5.4	5.6	5.1	5.4	5.2	5.2	5.7	5.7	5.5	4.9	5.3	6.6	5.7	5.1	5.5	6.0
SPME12116	Phenolic	Phenylacetaldehyde	5.3	6.7	6.4	6.4	6.3	6.2	6.6	6.1	6.4	6.2	6.6	6.4	5.1	6.6	5.4	6.1	5.7	6.8	5.8	6.5	5.8	7.1
SPME13486	Phenolic	Benzaldehyde, 4-methyl	4.3	4.2	4.0	4.0	4.0	4.0	4.1	4.2	4.0	4.1	4.1	4.0	4.1	4.1	4.1	3.8	4.1	4.1	4.1	4.0	4.1	4.1
SPME14806	phenolic	2-Phenylethanol	4.4	5.4	5.1	5.0	5.1	5.0	5.2	4.9	5.1	5.1	5.2	5.2	4.3	5.1	4.7	4.8	4.8	5.2	4.8	5.1	4.6	5.6
SPME15466	Phenolic	Benzyl cyanide	5.0	6.1	6.1	6.0	6.1	6.1	6.3	5.5	6.0	5.1	6.4	6.2	5.1	6.5	5.7	5.0	5.2	6.5	5.3	6.1	4.5	6.3
SPME15616	Phenolic	Benzoic acid	4.6	4.7	4.6	4.7	4.7	4.6	4.7	4.6	4.7	4.7	4.7	4.6	4.6	4.8	4.6	4.5	4.6	4.7	4.7	4.7	4.6	4.9
SPME16229	Phenolic	Benzene, 1,4-dimethoxy	3.8	3.8	3.8	3.8	3.7	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.8	3.6	3.8	3.8	3.8	3.8	3.8	3.8
SPME17609	Phenolic	Cinnamaldehyde	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.0	4.2	4.2	4.2	4.2	4.2	4.2
SPME18245	Phenolic	Benzaldehyde, 2-amino	3.8	3.9	3.2	3.5	3.8	3.1	3.7	3.5	3.4	4.0	3.9	3.1	4.2	3.6	4.4	3.7	4.0	3.6	4.7	3.5	3.9	3.2
SPME23011	Phenolic	Vanillin	3.1	3.1	3.2	3.1	3.1	3.1	3.2	3.2	3.2	3.1	3.3	3.3	3.2	3.2	3.2	3.1	3.1	3.2	3.2	3.2	3.0	3.4
SPME25497	Phenolic	Dimethyl phthalate	3.9	3.8	3.7	3.7	3.6	3.7	3.7	3.9	3.9	3.9	3.6	3.6	3.6	3.7	3.6	3.7	3.9	3.8	3.8	4.0	3.7	3.8
SPME27724	Phenolic	Benzophenone	4.6	4.5	4.2	4.2	4.2	4.3	4.3	4.5	4.2	4.5	4.1	4.1	4.1	4.2	4.2	3.8	4.6	4.3	4.1	4.2	4.2	4.3
SPME9383	Phenolic	Benzaldehyde	6.2	6.0	5.8	5.9	6.0	5.7	5.9	5.6	5.7	5.8	6.0	5.8	5.7	6.0	5.6	5.7	5.8	6.0	5.8	5.9	5.7	6.2
SPME12248	Phenylpropanoid	Salicylaldehyde	5.2	4.6	4.5	4.4	4.7	4.4	4.5	5.3	4.6	4.7	4.4	4.4	4.5	4.6	4.8	4.3	5.0	5.1	4.5	4.4	5.1	5.2
SPME13638	Phenylpropanoid	Guaiacol	6.4	4.7	5.2	5.3	5.0	5.0	5.1	7.0	5.4	5.4	5.5	5.5	5.4	5.8	5.5	6.5	6.6	6.7	5.7	5.2	5.7	6.9
SPME17474	Phenylpropanoid	Methyl salicylate	6.8	4.4	4.8	4.7	4.5	4.8	4.9	6.9	5.1	5.4	5.5	5.5	5.0	5.2	4.8	6.3	6.4	6.6	5.8	5.1	6.0	6.9
SPME21897	Phenylpropanoid	Eugenol	5.2	3.2	2.9	3.7	3.7	3.1	3.1	6.1	3.2	4.0	3.1	3.0	2.8	4.1	3.1	4.1	5.7	3.6	3.9	3.2	3.6	5.3

ID	Type	Annotation	S.ly.Moneymaker	S.ly.F6_20_010	S.pi.G1.1554	1_31_S.pi.G1.1703	2_17_S.pi.G1.1593	2_19_S.pi.G1.1594	2_25_S.pi.PI126925	2_7_S.pi.PI126933	2_9_S.pi.G1.1589	3_13_S.pi.LA1279	3_17_S.pi.LA1374	3_19_S.pi.LA1472	3_3_S.pi.LA0114	3_31_S.pi.LA1596	3_5_S.pi.LA0121	4_13_S.pi.LA1924	4_19_S.pi.LA2839	4_21_S.pi.LA2854	4_25_S.pi.LA1349	4_7_S.pi.LA1645	5_17_S.ch.LA0421	6_17_S.ch.LA1139
SPME21968	Phenylpropanoid	**Chavibetol**	4.6	2.9	3.3	3.6	3.5	3.3	3.7	5.5	3.4	3.5	3.4	3.1	3.3	3.7	3.4	3.7	5.0	3.5	3.5	3.6	3.7	4.7
SPME23199	Phenylpropanoid	**Isoeugenol, (Z)**	3.2	2.7	3.5	2.8	2.9	3.4	3.5	3.6	2.7	3.6	3.2	3.2	2.9	3.1	3.2	2.4	3.4	3.2	3.2	3.3	2.6	3.1
SPME10316	Terpene	**Sabinene**	4.1	3.7	3.8	3.8	3.7	3.8	3.8	4.2	4.2	3.3	3.9	3.8	4.0	3.5	4.3	3.8	4.1	3.7	3.8	3.8	4.3	4.1
SPME10623	Terpene	**Dehydro-1,8-cineole**	4.3	4.0	4.0	4.1	3.9	4.2	4.1	4.3	4.0	4.3	4.1	3.9	4.2	4.3	4.2	3.7	4.4	4.3	3.2	4.1	3.1	3.4
SPME10909	Terpene	**alpha-Terpinene**	3.2	2.9	3.0	4.0	3.1	3.0	3.1	2.9	2.7	3.5	3.0	3.0	3.5	3.8	2.9	2.6	4.3	3.4	3.9	3.2	3.7	3.2
SPME11024	Terpene	**alpha-Phellandrene**	3.9	3.4	3.6	4.0	3.4	3.8	3.4	4.0	3.9	4.0	3.8	3.7	4.1	4.1	4.0	3.1	4.4	4.1	4.2	3.7	3.8	3.1
SPME11486	Terpene	**p-Cymene**	4.5	4.4	4.4	4.9	4.2	4.4	4.6	4.9	4.7	5.0	4.8	4.3	4.9	4.8	4.7	4.2	5.0	4.8	5.1	4.4	5.2	4.9
SPME11666	Terpene	**Limonene**	5.1	4.7	4.4	5.2	4.8	4.5	4.6	4.9	4.7	4.8	4.5	5.1	6.0	5.4	5.7	4.2	5.1	5.8	5.8	4.5	4.9	5.1
SPME11818	Terpene	**beta-Phellandrene**	3.4	3.4	3.2	3.8	3.5	3.4	3.2	2.7	3.4	3.5	3.1	3.3	2.6	3.7	3.1	2.7	3.9	3.0	3.7	3.0	2.6	3.1
SPME13053	Terpene	**Linalool oxide A**	5.8	5.8	4.7	5.8	5.6	4.8	4.7	5.7	5.5	6.2	5.6	5.3	6.0	5.7	5.6	5.3	5.6	5.8	6.4	5.1	6.4	4.7
SPME13735	Terpene	**p-Cymenene**	4.8	4.7	4.5	5.1	4.5	4.6	4.6	5.0	4.8	5.2	5.0	4.6	5.3	5.0	5.1	4.4	5.2	5.2	5.5	4.7	5.5	4.8
SPME13819	Terpene	**Linalool**	5.5	5.1	4.6	5.1	5.0	4.6	4.6	4.8	4.8	5.2	4.5	4.7	5.0	5.1	5.2	4.3	5.3	5.2	5.0	4.7	5.4	4.3
SPME15913	Terpene	**Nerol oxide**	3.8	4.0	3.4	3.6	3.6	3.3	3.4	3.7	3.8	3.9	3.8	3.6	3.8	3.8	3.7	3.1	4.0	3.6	4.4	3.4	3.5	3.1
SPME16100	Terpene	**Camphor**	5.6	5.6	5.4	5.5	5.4	5.4	5.5	5.5	5.4	5.6	5.3	5.4	5.4	5.4	5.5	5.2	5.6	5.6	5.5	5.5	5.6	5.6
SPME16359	Terpene	**alpha-Phelladren-8-ol**	3.6	3.0	3.2	4.8	3.7	3.4	3.6	3.6	4.0	4.7	2.9	3.7	3.6	4.5	3.7	2.7	5.1	3.8	4.3	3.5	4.9	3.7
SPME16789	Terpene	**p-Cymen-9-ol**	4.0	3.9	3.9	4.4	3.9	3.9	3.9	4.2	4.1	4.4	4.0	3.9	4.1	4.2	4.0	3.6	4.6	4.0	4.1	3.9	4.6	4.3
SPME16940	Terpene	**p-Menth-1-en-9-al**	5.2	5.2	4.2	4.9	4.9	4.3	4.3	4.7	4.7	5.5	4.5	4.3	4.9	4.7	4.8	4.0	5.1	5.0	5.2	4.2	5.3	4.4
SPME16998	Terpene	**p-Cymen-8-ol**	5.1	4.9	4.7	5.3	4.8	4.9	4.8	5.2	4.9	5.4	5.1	4.7	5.3	5.1	5.0	4.6	5.4	5.3	5.5	4.8	5.7	5.1
SPME17258	Terpene	**alpha-Terpineol**	5.8	5.5	4.9	5.5	5.3	5.0	5.1	5.2	5.1	5.6	4.9	5.0	5.4	5.4	5.5	4.7	5.7	5.5	5.4	5.0	5.8	5.0
SPME18632	Terpene	**Neral**	5.6	5.4	5.4	5.5	5.4	5.7	5.5	5.7	5.5	5.6	5.5	5.2	5.6	5.7	5.7	5.1	5.8	5.8	4.2	5.4	3.7	4.4
SPME19038	Terpene	**Carvone**	3.4	3.3	3.3	3.5	2.9	3.2	3.4	3.5	3.2	3.3	3.2	3.6	4.1	3.6	3.9	2.8	3.5	3.8	4.0	3.1	3.4	3.5
SPME19448	Terpene	**Geranial**	6.3	6.0	6.0	6.2	6.1	6.2	6.1	6.3	6.2	6.2	6.1	5.9	6.3	6.4	6.3	5.6	6.5	6.5	4.8	6.0	3.8	4.3
SPME22759	Terpene	**alpha-Copaene**	4.0	4.1	4.2	3.7	3.9	4.2	4.2	3.8	4.0	4.2	3.9	3.6	3.7	3.8	3.6	3.5	3.7	3.7	3.8	3.8	3.7	3.4
SPME23316	Terpene	**p-Cymene, 2,5-dimethoxy**	3.1	2.8	3.3	2.9	2.9	3.2	3.3	3.3	3.2	3.2	3.2	3.3	3.1	3.2	3.1	2.7	3.1	2.6	3.1	3.0	3.0	2.7
SPME26156	Terpene	**delta-Cadinene**	2.7	2.9	2.7	2.7	2.6	2.7	2.9	2.9	3.0	2.9	3.1	2.6	2.9	2.7	2.7	2.9	2.6	2.7	2.7	2.7	2.9	2.7
SPME26474	Terpene	**Dihydroactinidiolide**	4.9	4.9	4.9	4.8	5.0	5.0	5.0	5.1	4.5	5.1	4.6	4.9	4.7	4.8	4.9	4.3	5.0	5.2	4.0	4.8	4.5	4.4
SPME26530	Terpene	**alpha-Calacorene**	3.2	3.0	3.8	3.7	3.4	3.7	3.8	3.6	3.9	3.4	3.8	3.5	3.8	3.7	3.6	3.6	3.1	3.9	4.0	3.7	3.8	3.3
SPME9779	Terpene	**Linaloyl oxide**	4.6	4.3	3.1	4.2	3.8	3.3	3.2	3.5	3.4	4.5	3.6	3.4	4.0	4.1	4.3	3.4	4.3	4.2	4.5	4.0	4.7	3.7
SPME9893	Terpene	**m-Cymene**	3.1	2.6	2.7	3.5	3.0	2.6	2.8	2.6	2.7	3.4	2.7	2.7	2.7	3.1	2.8	2.5	3.9	2.9	2.8	2.7	3.2	2.9
SPME10803	Terpenoid	**2-Carene**	4.0	4.0	3.7	3.9	4.0	3.8	3.9	4.1	3.5	4.0	3.6	3.8	3.7	4.0	4.0	3.4	4.1	4.2	3.3	3.9	3.8	3.5
SPME11311	Terpenoid	**alpha-Terpinene**	3.4	2.7	2.7	3.4	2.7	2.7	2.9	3.0	3.2	3.3	3.3	3.1	3.5	3.4	3.4	2.7	3.7	3.3	3.7	3.3	3.6	2.7



Chapter 6

General Discussion

Likely brought to Europe by Hernán Cortés and to Asia by Fernando de Magallanes, tomatoes conquered plates and palates around the world. In every country tomatoes are incorporated into the agricultural and culinary traditions. Yet, while expanding and conquering the world, they left behind most of the genetic diversity in its wild relatives back in the Andean regions. Tomatoes hold tight to some survival characteristics such as self-pollination, but a lot of the variability decreased in cultivated varieties.

For the past decades, scientists have been trying to recover the missing diversity with more knowledge and less fear to the poisonous features of the nightshades. For this reason they evaluate wild germplasm that might improve the genetic diversity within cultivated tomatoes.

Tomato germplasm (dis)similarities screening

In 2012 a high quality sequence of the tomato genome was published (Tomato Genome Consortium, 2012). This historical event expanded the frontiers of the tomato -omics approaches and the research within the Solanaceae.

Two years before the publication of the genome sequence, we were working with a custom made SNP array (Chapter 2) and the combination of different approaches and technologies allowed us to screen tomato and its wild relatives. For the array, the SNP search was performed through sequencing of a few selected *S. lycopersicum* cherry and round cultivars with contrasting phenotypes and in a number of introgression free old cultivars. The 6000 selected polymorphisms were enough to screen across cultivars and wild relatives. Later the SNP positions could be linked to the published reference tomato genome. Overall we compared different germplasm, determined the variation among cultivars, populations and accessions of different tomato wild relatives.

One finding drawing our attention and not addressed before was that most accessions of *S. cheesmaniae* and *S. galapagense* were genetically similar although they were phenotypically quite different. Screening 4072 SNPs, in 26 accessions of *S. cheesmaniae* and 15 accessions of *S. galapagense* gave no clear polymorphisms. The few differences were mostly due to two *S. cheesmaniae* accessions: G1.1516 and LA3124. They were quite different compared to the other *S. cheesmaniae* accessions. When we grew the accessions in the greenhouse, *S. galapagense* accession LA1137 had a *S. cheesmaniae* phenotype and even for the only two markers that we considered as species specific SNPs for*S. galapagense, S. galapagense* LA1137 was an exception to the rule in that it contained the *S. cheesmaniae* alleles for those two SNPs. So, this accession

might be misclassified. In general, we found more genetic differences between cherry and round cultivars than genetic differences between the two species *S. cheesmaniae* and *S. galapagense,* but that might be due to the fact that we specifically screened for differences between round and cherry. The two species clustered together in a Neighbour Joining analysis (NJ, Fig. 1). Yet, the observed phenotypical differences in leaf structure and trichomes are very clear (Darwin *et al.* 2003). An example of the leaf morphology was shown in Table 1 of the introduction chapter of this thesis. Apparently only a few genetic differences can lead to these very distinct morphological differences. Aflitos *et al.* (2014) compared whole genome sequences from different tomato species and also found that *S. cheesmaniae* and *S. galapagense* clustered together. However, when we grew 11 *S. cheesmaniae* and 14 *S. galapagense* accessions in a controlled climate chamber and performed metabolic profiling we could clearly distinguish the two species. It was already known that metabolites such as acyl sugars were characteristic of the *S. galapagense* accessions (Lucatti *et al.*, 2013). Minor genetic differences can lead to significant changes in metabolomic profiles and morphology. We think it should be reconsidered that *S. cheesmaniae* and *S. galapagense* are really different species as proposed by Darwin *et al.* (2003) or the boundaries to define different species should be better specified.

In Figure 1, we show that some of our 68 *S. pimpinellifolium* accessions intermingled with the *S. lycopersicum* cultivars. We suppose that most of those accessions are admixtures between species. Sim *et al.* (2012) showed a clear differentiation of the clades with the SolCap array except for one accession which they called a wild cherry. Lin *et al.* (2014) made an analysis based on 20111 SNPs and 331 accessions from the red fruited tomato clade and 10 accessions of wild relatives. In that study, the *S. lycopersicum* var. *cerasiforme* accessions ended up in the two main clusters and some even clustered with the group of 166 big-fruited *S. lycopersicum* accessions. Furthermore, Blanca *et al.* (2015) made a genetic diversity analysis with 7720 SNPs including 530 tomato accessions of *S. lycopersicum* var. *lycopersicum*, 316 *S. lycopersicum* var. *cerasiforme* (Blanca *et al.* support the idea that *cerasiforme* is a true phylogenetic tomato group), 145 accessions of *S. pimpinellifolium* and 17 other wild relatives. They called everything in between species admixtures. When performing a NJ analysis, some *S. pimpinellifolium* accessions were closer to the *S. lycopersicum* clade than others. These studies confirm that the choice of accessions might influence the conclusions. Researchers as Moyle (2008) proposed strict similarities. But since there is a lot of admixture between the Lycopersicon species, the use of too few accessions can result in biased results. The *S. pimpinellifolium* accessions we found among the *S. lycopersicum* clade may need a re-classification.

Also, morphological characteristics can be misleading. For example, Harlan (1971) suggested that biloculed domesticated forms found in south Mexico and Guatemala are the oldest cultivated types. The revised key for the *Lycopersicon* and related *Solanum* species formulated by Rick *et al.* (1990) state that wild relatives are biloculed. But, in our study we found *S. galapagense* accessions with three locules (Fig. 2).

To be or not to be a cherry tomato

We distinguished higher levels of polymorphisms on most of the chromosomes 4, 5 and 12 between cherry and round tomatoes. Additionally, smaller more polymorphic regions were found on chromosomes 1 and 2 (Chapter 2). We believe that those chromosomes and regions are the main drivers in making tomatoes cherry or round. Lin *et al.* (2014) made an inventory of the known QTL regions along some of these chromosomes. Several of them were related to fruit weight (*fw1.1*, *fw2*.1, *fw2.2*, *fw2.3, fw5.2, fw12.1*), locule number (*lcn2.1*, *lcn 2.2*, *lcn12.1*), soluble solids content (*ssc5.1*, *ssc5.2*, *ssc5.3*) and firmness (*fir5.1*). In those QTL regions are genes causing the separation between small and big fruited tomatoes. Even though Lin *et al.* (2014) also found highly divergent SNPs on chromosome 4, they did not draw any conclusions from it. Nevertheless, Xu *et al.* (2013) mentioned in another mapping study, several QTLs related to fresh weight, firmness and sugar content also on chromosome 4.

For chromosome 1, Lin *et al.* (2014) depicted introgressions and sweeps from inbreeding lines, fresh market hybrids and processing hybrids. Along this thesis we saw a clear pattern. The SNPs in the region on chromosome 1 between 5.4 and 83.4 Mb were random and didn't distinguish cherry from round/beef tomatoes (Fig. 5B, Chapter 2); the genetic map between *S. lycopersicum* and *S. pimpinellifolium* had a gap on chromosome 1 due to lack of differences between round tomatoes and cherries (Chapter 3). Nevertheless, we could retrieve more markers for this QTL hotspot region (Chapter 4) and finally we saw clear metabolomic differences between *S. lycopersicum* genotypes and *S. pimpinellifolium* accessions that matched the metabolic patterns observed in the RILs (Chapter 5). The last ones specially related to changes in the phenylpropanoid and alkaloid pathways. All these data have led us to believe that there has always been a selection in this region of chromosome 1. The reason for this might have been the selection against the bitter flavour of alkaloids. With that selection pressure, there was an automatic selection towards a different phenylpropanoid composition.

Combinations of genotyping technologies

Our SNP array provided generally applicable markers in tomato and its wild relatives. The possibility of aligning reads to a reference sequence opened up possibilities for people to consistently refer to specific physical regions among genotypes.

Our genetic map proved to be useful for the improvement of the alignment of the tomato genome. The clearest example was the one on chromosome 12, where we observed a clear discrepancy between our genetic and physical maps, pointing at a likely misplaced scaffold along the tomato genome (Chapter 2, Additional file 6 and Chapter 3 Figure 2). Sim *et al.* (2012b) also got a disruption between genetic and physical positions on chromosome 12 with genetic maps from crosses to *S. pennellii* (EXPEN) and *S. pimpinellifolium* (EXPIM), but they did not discuss it further. This clearly shows the benefits of comparing genetic and physical maps.

When comparing the SNP array data to the reference Heinz 1706 genome we found genome regions, for example on chromosome 1 at 86.8 Mb, where SNPs known to be derived from *S. pimpinellifolium*, unexpectedly, were equal to Heinz, but different from Moneymaker. These events are most likely caused by the fact that Heinz 1706 has a *S. pimpinellifolium* ancestor (Ozminkowski, 2004). *S. pimpinellifolium* introgressions in Heinz were already reported on chromosomes 4, 9, 11 and 12 (Tomato Genome Consortium, 2012).

In Chapter 3 and 4, we demonstrated the advantages of combining different genotyping platforms. After our first screening with SNPs, we were able to perform further *in silico* mapping of a TYLCV resistance gene and metabolite related QTL. After the *in silico* improvement, regions were better delimited and QTLs could be better defined. This possibility of retrieving more information about regions of interest can be very cost-effective.

Genomics and metabolomics, towards QTL

Populations screened with QTL mapping methods are key to elucidate simple and complex traits interacting between wild relatives and commercial tomatoes. In our project, we applied several metabolomic approaches in leaves (Chapter 3) and fruits (Chapter 4) of a recombinant inbred line (RIL) population derived from a cross between *S. lycopersicum* and *S. pimpinellifolium*. We chose different analytical platforms in order to target compounds belonging to diverse pathways in order to obtain insight in the variation within the metabolomes of the two species. At the moment, there is no analysis platform that can target all the metabolites of a sample at once.

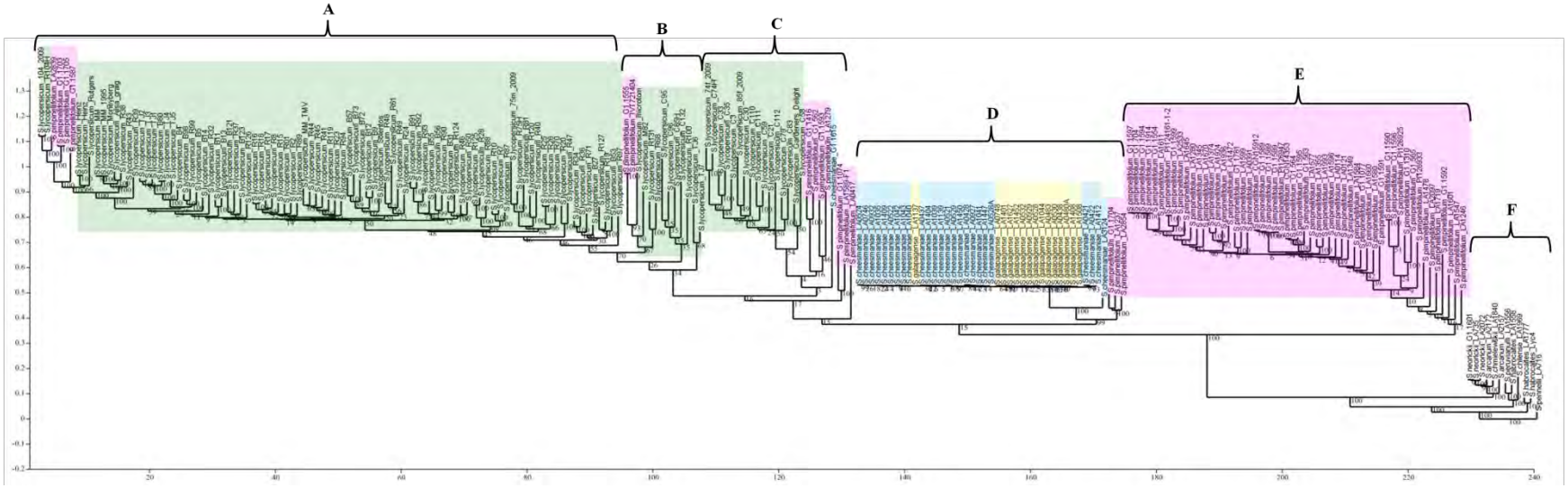


Figure 1. Neighbour joining clustering analysis of 237 tomatoes based on 4072 SNPs distributed over the genome with the Manhattan similarity measure, 100 bootstraps and S. *pennellii* as outgroup. Clusters: **A)** 89 *S. lycopersicum* round and beef cultivars in green (including Heinz, Moneymaker, Rutgers, Solentos and Ailsa Craig) and 4 *S. pimpinellifolium* accessions in pink (LA2839, G1.1587, G1.170, G1.1705). **B)** 11 *S. lycopersicum* round and cherry cultivars (including M82 and microtom) and 2 *S. pimpinellifolium* accessions (G1.1555, IVT721404). **C)** 16 *S. lycopersicum* cherry cultivars (including Gardeners Delight), 7 *S. pimpinellifolium* accessions (G1.1416, G1.562, G1.1593, LA1279, G1.1615, G1.1074, LA1589-F1, LA0417) and 1 *S. cheesmaniae* accession (G1.1615). **D)** 25 *S. cheesmaniae* in light blue (LA0421, LA0422, LA0428, LA0437, LA0521, LA0522, LA0524, LA0528A, LA0529, LA0746, LA0927, LA0932, LA1035, LA1039, LA1040, LA1041, LA1042, LA1043, LA1139, LA1404, LA1409, LA1412, LA1447, LA1450, LA3124), 15 *S. galapagense* in yellow (LA0438, LA0480A, LA0483, LA0528, LA0530, LA0532, LA0748, LA0929, LA1044, LA1137, LA1401, LA1408, LA1452, LA1508, LA1627) and 3 *S. pimpinellifolium* (G1.1704, LA1237, LA2854) accessions. **E)** 52 *S. pimpinellifolium,* accessions (CGN14353, CGN15912, G1.1077, G1.1310, G1.1554, G1.1589, G1.1589, G1.1589, G1.1590, G1.1591, G1.1592, G1.1594, G1.1595, G1.1596, G1.1597, G1.1599, G1.1781, G1.1914, G1.563, G1.564, G1.565, G1.704, LA0114, PI126925, PI126933, PI12694, PI124161-1-2, LA0121, LA0391, LA1246, LA1280, LA1345, LA1349, LA1355, LA1374, LA1472, LA1478, LA1547, LA1577, LA1580, LA1584, LA1596, LA1599, LA1601, LA1611, LA1645, LA1660, LA1670, LA1719, LA1924, LA1936, LA1993, LA2097, LA2533) and 1 *S. lycopersicum* round. **F)** Outgroup: 2 *S. arcanum* , 1 *S. chilense*, 1 *S. chmielewskii,* 2 *S. habrochaites,* 3 *S. neorickii,* 1 *S. peruvianum* and 1 *S. pennellii* accessions.

Figure 2. Horizontal cut of *S. galapagense* fruits with two or more locules (red arrows are pointing at locule boundaries).

Polar primary metabolites were profiled with gas chromatography-time of flight-mass spectrometry (GC-TOF; Lisec *et al.*, 2006). For both leaves and fruits only few QTLs were found for compounds detected with this platform. This suggests that the primary metabolism between the two genotypes is very similar. Earlier studies of the primary metabolism of introgression lines with less related species like *S. pennellii* have shown greater differences (Bermúdez *et al.*, 2008; Toubiana *et al.*, 2012). With *S. pimpinellifolium*, far more differences were found within the secondary metabolism and the volatile organic compounds. Semi-polar secondary metabolites were profiled with liquid chromatography-quadrupole time of flight-mass spectrometry (LC-QTOF; de Vos *et al.*, 2007) and volatile organic compounds were profiled with solid-phase micro extraction (SPME)-GC (Tikunov *et al.*, 2005). With these metabolomic approaches we covered a broad range of metabolites that were later linked to different traits.

The metabolic profiling carried out for leaves of the RILs was useful to correlate metabolites with a possible role in the resistance mechanism to TYLCV. The TYLCV resistance of *S. pimpinellifolium* is very complex. Based on co-localisation and correlation analysis, the QTL found on chromosome 3 seemed to be related to the presence of flavonoid glycosides and the one on chromosome 11 to hexoses (Chapter 3). Those metabolites seemed to play an important role in conferring resistance likely through the use of different transport mechanisms to fight pathogens (Sade *et al.*, 2014). Khan *et al.* (2012) also evaluated the same RIL population for seed quality related traits. They discovered certain epistatic effects among chromosomes to cope with different types of stresses. Chromosomes 3 and 11 showed an interaction under cold stress conditions. But they stated that cold may affect the water content of the cells and then causes an osmotic stress. The QTL we found was also related to chromosomes 3 and 11 and could be related to an osmotic/transport response under stress conditions. When performing a Random Forest analysis, one of the most significant markers for the TYLCV resistance on chromosome 3 was related to a glycogen synthase kinase gene and this type of genes play a role in signalling, transport and stress responses (Jonak and Hirt, 2002). Therefore, this gene and mechanisms related to transport (perhaps related to phosphates) could be interesting to target for further elucidation of the resistance mechanism.

Other examples of interesting QTLs were the ones related to the powerful antioxidant chlorogenic acid. QTLs for this compound were found in both leaves and fruits. Interestingly, for fruits, the strongest QTL of chlorogenic acid was on chromosome 1 while for leaves it was found on the top of chromosome 10. Those two hotspot regions caused the greatest differences between the species and may have pleiotropic effects on several other traits

between *S. lycopersicum* and *S. pimpinellifolium*. When Causse *et al.* (2001) evaluated sensory traits in a RIL population derived from a cross between *S. lycopersicum* cherry and round, they only found a QTL for sourness on chromosome 1 and no QTLs on chromosome 10. Within our analysis, those two QTL regions seem to play an important role in determining differences in secondary metabolite production.

Xu *et al.* (2013) stated that *S. pimpinellifolium* might be a valuable allele source for quality improvement related to firmness, sugars and acids. But as mentioned before, we observed at least two cases, on chromosomes 2 and 4, where fruit weight QTL and sugar QTL are in the same region of the chromosome but originating from different parents. For example, on chromosome 2 the *fw2.2* gene that regulates cell division (Cong and Tanksley, 2006) is in the same chromosome region as a QTL for sugar content (as the one reported in Chapter 4). There could be a dilution effect of sugars due to the size of the fruits (Prudent *et al.*, 2010), but that doesn´t mean that the sugar content cannot be improved within a certain range or that hybrid effects cannot be of profit. In the case of breeding for homozygous plants, the extra costs in breeding must be compared to the added value. Sometimes it could be better to target other sugar allele sources that are not linked to fruit weight, like the *S. pennellii* apoplastic invertase (*lin5*) on top of chromosome 9 (Fridman *et al.*, 2000).

Our RIL population showed different quality related QTLs that could be beneficial in tomato breeding. This thesis presents data for breeders which allow them to target specific regions and to improve traits. Fine mapping and positional cloning could be next steps to get more insight in the mechanisms behind the changes.

S. pimpinellifolium for tomato quality improvement

Beyond the differences and the interactions found among the RILs, we also found a large variation of the metabolic phenotypes in the other *S. pimpinellifolium* accessions (Chapter 5). In a genetic diversity plot comparing the RILs and the accessions, we observed additional variation in the *S. pimpinellifolium* accessions (Fig. 3). As expected, there was no genetic variation between the *S. galapagense* and the *S. cheesmaniae* accessions. But the combination of genetic and metabolic variation between *S. lycopersicum* and *S. pimpinellifolium* could lead to association analyses in which more QTLs can be found and compared. Ranc *et al.* (2012) and Xu *et al.* (2013) already showed that association mapping is a possibility in tomato. This type of approach can lead to find more stable QTL across populations. However, the needed

presence of a minimum allele frequency could result in neglecting specific exceptional genotypes.

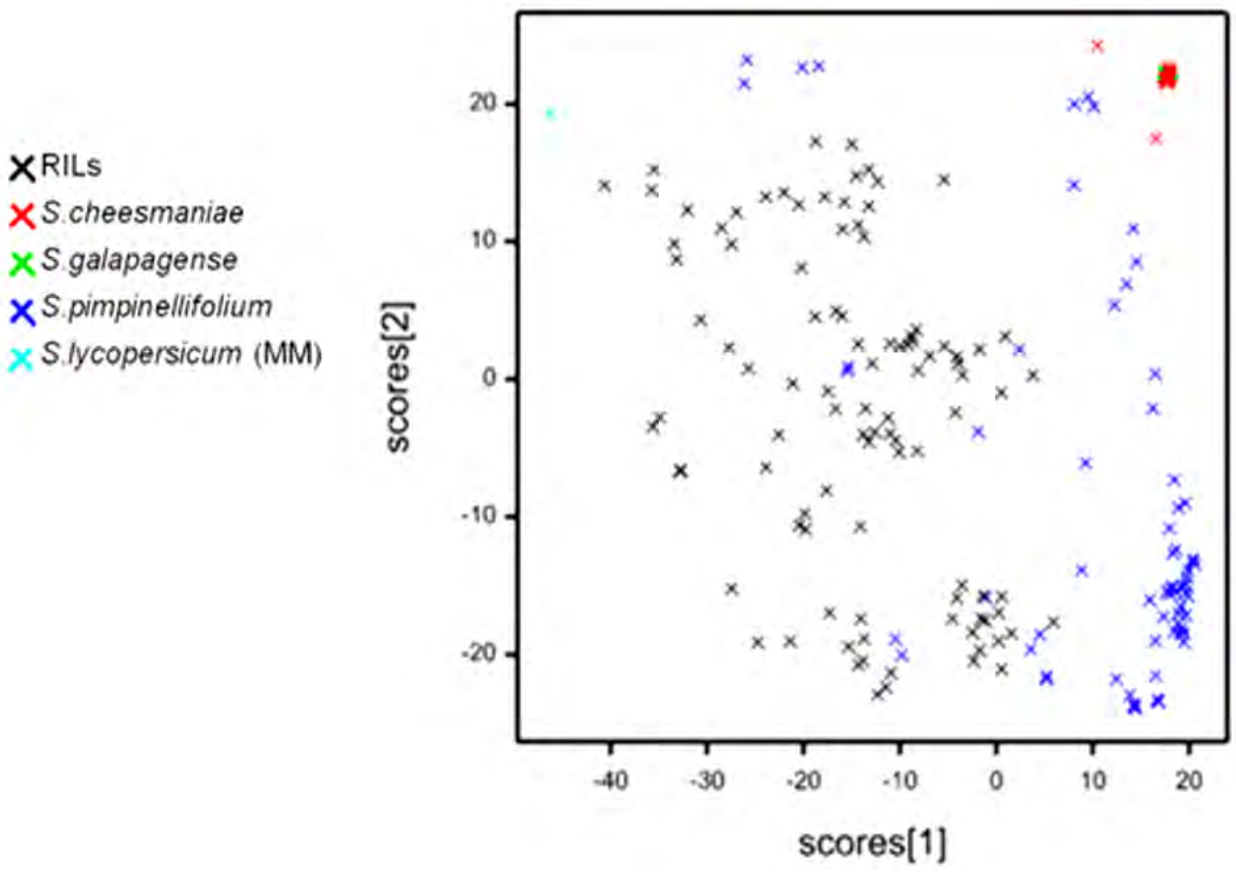


Figure 3. Genetic diversity plot with SNPs distributed homogenously along the genome for the 100 RILs (black), 68 *S. pimpinellifolium* accessions (blue) and *26 S. cheesmaniae* (red) and 15 *S. galapagense* (in green, behind the red cluster of genotypes). Scores stand for calculated distances between the genotypes according to the SNP differences.

In general, *S. pimpinellifolium* accessions had higher levels of phenylpropanoids/flavonoids. Therefore, *S. pimpinellifolium* can be a good source for improvement of flavonoid content of tomatoes. Other aroma related compounds, such as phenylethanol, could be also targeted for improvement with *S. pimpinellifolium* as donor. An association panel consisting of different accessions of *S. pimpinellifolium* could help in further exploration of metabolic pathways leading to the accumulation of different metabolites. Our method of metabolic profiling can be followed by more specific targeted analyses when needed.

Our study presents the physical positions of QTL on the tomato genome. This can help breeders to target specific QTL regions and compare these regions with their own germplasm. Tools like the variant browser developed by Wageningen UR (available at:

http://www.tomatogenome.net) are of great help for breeders and researchers in countries with limited bioinformatics resources; this type of resources allow people to compare and explore the variation within the tomato genome.

Further challenges of quality parameters

As acknowledged in the beginning of this thesis, *quality* is to be defined by parameters based on personal or cultural prejudices and expectations. Flavour is one of the traits that could exemplify the best the complexity lying behind the perception of quality. Because, not only personal perceptions are crucial to target, but there is a combination of several factors along the production chain. In this chain, growers demand high yield, distributors (supermarkets) demand long shelf life and consumers demand better appearance in combination with better flavour. Even the fact of making the fruits ripe off-the-vine and then apply different post-harvest handling can have huge effects on flavour (Baldwin *et al.*, 2011). Klee and Tieman (2013) proposed to target flavour improvement in tomato by identifying the compounds contributing to consumer preferences, the genes controlling those compounds and the most important alleles giving differences on the role of those genes. Nevertheless, the identification of compounds contributing to consumer preferences is a major complex challenge by itself. That is also partly dependent on the evaluated human population. For example, pink tomatoes are quite popular in certain Asiatic countries, while not in Europe. Pink tomatoes are related to the transcription factor *MYB12* and the lack of accumulation of naringenin chalcone in the fruit peel (Ballester *et al.*, 2010). This shows that preferences can be related to alterations of complete metabolic pathways. Another not as tangible possible quality attribute in tomatoes could be the increase of antioxidant and free-radical scavenging activity by flavonoids; though there are different debates according to the bioavailability and effects of the flavonoids (Roos and Kasum, 2002).

Breeding efforts could be ignored by societies not willing to incorporate certain *quality* improvements. There is a great challenge for scientists, breeders, politicians, communicators and even marketing specialists to keep on creating research and dialectics. The collection and distribution of information will continue to be a key to better define *quality* parameters related to the real impact of health related compounds and the perception of flavour.

References

Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, Yativ M, Domínguez E, Wang Z, de Vos RCH, Jetter R, Schreiber L, Heredia A, Rogachev I, Aharoni, A: **Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network**. PLoS Genetics **2009**, 5(12):e1000777.

Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L, Bakker F, Dirks R, Breit T, Gravendeel B, Huits H, Struss D, Swanson-Wagner R, van Leeuwen H, van Ham RCHJ, Fito L, Guignier L, Sevilla M, Ellul P, Ganko E, Kapur A, Reclus E, de Geus B, van de Geest H, Hekkert BtL, van Haarst J, Smits L, Koops A, Sanchez-Perez G, van Heusden AW, Visser R, Quan Z, Min J, Liao L, Wang X, Wang G, Yue Z, Yang X, Xu N, Schranz E, Smets E, Vos R, Rauwerda J, Ursem R, Schuit C, Kerns M, van den Berg J, Vriezen W, Janssen A, Datema E, Jahrman T, Moquet F, Bonnet J, Peters S: **Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon)* clade by whole-genome sequencing.** The Plant Journal **2014,** 80(1):136-148.

Agarwal M, Shrivastava N, Padh H: **Advances in molecular marker techniques and their applications in plant sciences**. Plant Cell Rep **2008**, 27:617–631.

Alba JM, Monserrat M, Fernandez-Muñoz R: **Resistance to the two-spotted spider mite (*Tetranychus urticae*) by acylsucroses of wild tomato (*Solanum pimpinellifolium*) trichomes studied in a recombinant inbred line population.** Experimental and Applied Acarology **2009**, 47(1):35-47.

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B: **Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon***. Theor Appl Genet **2001**, 103:1283–1292.

Anbinder I, Reuveni M, Azari R, Paran I, Nahon S, Shlomo H, Chen L, Lapidot M, Levin I: **Molecular dissection of Tomato leaf curl virus resistance in tomato line TY172 derived from *Solanum peruvianum***. Theoretical and Applied Genetics **2009**, 119(3):519-530.

Anithakumari AM, Tang J, van Eck HJ, Visser RGF, Leunissen JAM, Vosman B and van der Linden CG: **A pipeline for high throughput detection and mapping of SNPs from EST databases**. Molecular Breeding **2010**, 26(1):65-75.

Appleby N, Edwards D, Batley J: **New technologies for ultra-high throughput genotyping in plants**. Plant Genomics **2009**, 513:19-39.

Arens P, Mansilla C, Deinum D, Cavellini L, Moretti A, Rolland S, van der Schoot H, Calvache D, Ponz F, Collonnier C, Mathis R, Smilde D, Caranta C, Vosman B: **Development and evaluation of robust molecular markers linked to disease resistance in tomato for distinctness, uniformity and stability testing**. Theoretical and Applied Genetics **2010**, 120(3):655-64.

Ashrafi H, Kinkade M, Foolad MR: **A new genetic linkage map of tomato based on a *Solanum lycopersicum* × *S. pimpinellifolium* RIL population displaying locations of candidate pathogen response genes**. Genome **2009,** 52(11):935-956.

Bai Y, Lindhout P: **Domestication and breeding of tomatoes: what have we gained and what can we gain in the future?** Annals of Botany **2007**, 100(5):1085-1094.

Baldwin E, Goodner K, Plotto A: **Interaction of volatiles, sugars, and acids on perception of tomato aroma and flavor descriptors**. Journal of food science **2008**, 73(6):S294-S307.

Baldwin E, Plotto A, Narciso J, Bai J: **Effect of 1-methylcyclopropene on tomato flavour components, shelf life and decay as influenced by harvest maturity and storage temperature**. Journal of the Science of Food and Agriculture **2011**, 91(6):969-980.

Baldwin EA, Scott JW, Shewmaker CK, Schuch W: **Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components**. In: HortScience: American Society for Horticultural Science, **2000**: 1013-1022.

Baldwin, E: **Fruit flavor, volatile metabolism and consumer perceptions**. Fruit Quality and its Biological Basis. CRC Press, FL **2002**: 89-106.

Ballester AR, Molthoff J, de Vos R, te Lintel Hekkert B, Orzaez D, Fernández-Moreno JP, Tripodi P, Grandillo S, Martin C, Heldens J, Ykema M, Granell A, Bovy A: **Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color**. Plant physiology **2010**, 152(1), 71-84.

Banerjee MK, Kalloo MK: **Sources and inheritance of resistance to leaf curl virus in Lycopersicon**. Theoretical and Applied Genetics **1987**, 73(5):707-710.

Bartoshuk LM, Klee HJ: **Better fruits and vegetables through sensory analysis**. Current Biology **2013**, 23(9):R374-R378.

Bartoshuk LM, Blandon A, Clark DG, Colquhoun TA, Hudson S, Klee HJ, Moskowitz HK, Sims CA, Snyder DJ, Tieman DM: **Retronasal Olfaction/Taste (Volatiles/Sweetness) Interactions in Tomatoes**. Chemical Senses **2013**, 38:259-372.

Bartoshuk LM, Colquhoun TA, Clark DG, Schwieterman M, Sims CA, Whitaker V, Klee HJ, Tieman D, Mclntyre L: **Compositions and methods for modifying perception of sweet taste.** In.: US Patents US20130280400 A1; **2013b**.

Baxter CJ, Sabar M, Quick WP, Sweetlove LJ: **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**. Journal of Experimental Botany **2005**, 56(416):1591-1604.

Bennett, AB: **Taste: Unraveling Tomato Flavor**. Current Biology **2012**, 22(11):R443-R444.

Bergougnoux V: **The history of tomato: From domestication to biopharming**. Biotechnology advances **2014,** 32(1):170-189.

Bermudez L, Urias U, Milstein D, Kamenetzky L, Asis R, Fernie A, Van Sluys M, Carrari F, Rossi M: **A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit**. Journal of Experimental Botany **2008**, 59(10):2875-2890.

Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D, Tanksley S: **Advanced backcross QTL analysis of tomato. II. Evaluation of near-isogenic lines carrying single-donor introgressions for desirable wild QTL-alleles derived from *Lycopersicon hirsutum* and *L. pimpinellifolium***. Theoretical and Applied Genetics **1998**, 97: 170-180.

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard J, Uhlig J, Zamir D and Tanksley S: **Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from *Lycopersicon hirsutum***. Theoretical and Applied Genetics **1998**, 97(3):381-397.

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH: **Potential of metabolomics as a functional genomics tool**. Trends in plant science **2004**, 9:418-425.

Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, Nuez F: **Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato**. PloS One **2012**, 7:e48198.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, Knaap E, Cañizares, J: **Genomic variation in tomato, from wild ancestors to contemporary breeding accessions**. BMC genomics **2015**, 16(1), 257.

Bol J, Linthorst H, Cornelissen B: **Plant pathogenesis-related proteins induced by virus infection**. Annual Review of Phytopathology **1990,** 28(1):113-138.

Bovy A, Schijlen E, Hall RD: **Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics**. Metabolomics **2007**, 3(3):399-412.

Bredemeijer GMM, Arens P, Wouters D, Visser D, Vosman B: **The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification**. Theoretical and Applied Genetics **1998**, 97:584-590.

Bredemeijer GMM, Cooke RJ, Ganal MW, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Roder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B: **Construction and testing of a microsatellite database containing more than 500 tomato varieties.** Theor Appl Genet **2002**, 105:1019-1026.

Bretó MP, Asins MJ, Carbonell EA: **Genetic variability in *Lycopersicon* species and their genetic relationships**. Theoretical and Applied Genetics **1993**, 86:113–120.

Broman KW: **The genomes of recombinant inbred lines**. Genetics **2005**, 169(2):1133-1146.

Butterbach P, Verlaan MG, Dullemans AM, Lohuis D, Visser RGF, Bai Y, Kormelink R: **Tomato yellow leaf curl virus resistance by *Ty-1* involves increased cytosine methylation of viral genomes and is compromised by cucumber mosaic virus infection.** Proceedings of the National Academy of Sciences **2014**, 111(35):12942-12947.

Carli P, Barone A, Fogliano V, Frusciante L, Ercolano MR: **Dissection of genetic and environmental factors involved in tomato organoleptic quality**. BMC plant biology 2011, 11(1):58.

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester H, Visser RG, Keurentjes JJ: **Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality**. Plant physiology **2012**, 158(3):1306-1318.

Causse M, Chaïb J, Lecomte L, Buret M: **Both additivity and epistasis control the genetic variation for fruit quality traits in tomato**. Theoretical and Applied Genetics **2007**, 115(3):429-442.

Causse M, Desplat N, Pascual L, Le Paslier MC, Sauvage C, Bauchet G, Bérard A, Bounon R, Tchoumakov M, Brunel D, Bouchet J: **Whole genome resequencing in tomato reveals variation associated with introgression and breeding events**. BMC Genomics **2013**, 14:791.

Causse M, Duffe P, Gomez M, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C: **A genetic map of candidate genes and QTLs involved in tomato fruit size and composition**. Journal of Experimental Botany **2004**, 55(403):1671-1685.

Causse M, Saliba-Colombani V, Lesschaeve I, Buret M: **Genetic analysis of organoleptic quality in fresh market tomato. 2. Mapping QTLs for sensory attributes.** Theoretical and Applied Genetics **2001**, 102(2-3):273-283.

Chagué V, Mercier J, Guenard M, De Courcel A, Vedel F: **Identification of RAPD markers linked to a locus involved in quantitative resistance to TYLCV in tomato by bulked segregant analysis**. Theoretical and applied genetics **1997,** 95(4):671-677.

Chen FQ, Foolad MR: **A molecular linkage map of tomato based on a cross between Lycopersicon esculentum and L. pimpinellifolium and its comparison with other molecular maps of tomato**. Genome **1999**, 42(1):94-103.

Chibon P, Schoof H, Visser RG, Finkers R: **Marker2sequence, mine your QTL regions for candidate genes**. Bioinformatics **2012**, 28(14):1921-1922.

Chibon P-Y, Voorrips RE, Visser RG, Finkers R: **MQ2: visualizing multi-trait mapped QTL results.** Molecular breeding **2013**, 32(4):981-985.

Cho A-S, Jeon S-M, Kim M-J, Yeo J, Seo K-I, Choi M-S, Lee M-K: **Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice**. Food and Chemical Toxicology **2010**, 48(3):937-943.

Cooke RJ: **Modern methods for cultivar verification and the transgenic plant challenge.** Seed Science Technology **1999**, 27:669–680.

Cong B, Tanksley SD: ***FW2. 2* and cell cycle control in developing tomato fruit: a possible example of gene co-option in the evolution of a novel organ.** Plant molecular biology **2006**, 62(6):867-880.

Crick F: **Central dogma of molecular biology**. Nature **1970**, 227:561-563.

Czosnek H, Eybishtz A, Sade D, Gorovits R, Sobol I, Bejarano E, Rosas-Díaz T, Lozano-Durán R: **Discovering host genes involved in the infection by the tomato yellow leaf curl virus complex and in the establishment of resistance to the virus using Tobacco Rattle Virus-based post transcriptional gene silencing**. Viruses **2013**, 5(3):998-1022.

Darwin SC, Knapp S and Peralta IE: **Taxonomy of tomatoes in the Galápagos Islands: native and introduced species of *Solanum* section *Lycopersicon* (Solanaceae)**. Systematics and Biodiversity **2003**, 1(1):29–53.

DeCandolle A: **Origin of cultivated plants**. Cambridge University press (reprint 2011) **1886**, 1-468.

De Vos RC, Hall RD, Moing A: **Metabolomics of a model fruit: tomato**. Annual Plant Reviews, Biology of Plant Metabolomics **2011**, 43:109-155.

De Vos RC, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD: **Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry**. Nature Protocols **2007**, 2(4):778-791.

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MS, Wang L: **The phenylpropanoid pathway and plant defence: a genomics perspective**. Molecular Plant Pathology **2002**, 3(5):371-390.

Doganlar S, Frary A, Ku HM, Tanksley SD: **Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589)**. Genome **2002,** 45(6):1189-1202.

Eybishtz A, Peretz Y, Sade D, Akad F, Czosnek H: **Silencing of a single gene in tomato plants resistant to Tomato yellow leaf curl virus renders them susceptible to the virus**. Plant Molecular Biology **2009**, 71(1-2):157-171.

Eybishtz A, Peretz Y, Sade D, Gorovits R, Czosnek H: **Tomato yellow leaf curl virus infection of a resistant tomato line with a silenced sucrose transporter gene LeHT1 results in inhibition of growth, enhanced virus spread, and necrosis**. Planta **2010**, 231(3):537-548.

Fernie AR, Tadmor Y, Zamir D: **Natural genetic variation for improving crop quality**. Current opinion in plant biology **2006**, 9(2):196-202.

Finkers R: **BreedDB: A database supporting quantitative aspects of Plant Breeding**. Stockholm, **2009**.

Food and Agriculture Organization of the United Nations (FAO): **Tomato production**. http://faostat3.fao.org/, **2014**.

Fridman E, Tzili P, Zamir D: **A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene**. Proceedings of the National Academy of Sciences 2000, 97(9):4718-4723.

Friedman M: **Anticarcinogenic, cardioprotective, and other health benefits of tomato compounds lycopene, α-tomatine, and tomatidine in pure form and in fresh and processed tomatoes**. Journal of agricultural and food chemistry **2013**, 61(40):9534-9550.

Friedman M: **Tomato glycoalkaloids: role in the plant and in the diet**. Journal of Agricultural and Food Chemistry **2002**, 50(21):5751-5780.

Friedman M, Lapidot M, Cohen S, Pilowsky M: **A novel source of resistance to tomato yellow leaf curl virus exhibiting a symptomless reaction to viral infection**. Journal-American Society for Horticultural Science **1998,** 123:1004-1007.

Galtier N, Foyer CH, Huber J, Voelker TA, Huber SC: **Effects of elevated sucrose-phosphate synthase activity on photosynthesis, assimilate partitioning, and growth in tomato (*Lycopersicon esculentum* var UC82B)**. Plant Physiology **1993,** 101(2):535-543.

Gascuel, O: **BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data**. Molecular Biology and Evololution **1997**, 14:685-695.

Goff SA, Klee HJ: **Plant volatile compounds: sensory cues for health and nutritional value.** Science **2006**, 311(5762):815-819.

Gorovits R, Moshe A, Ghanim M, Czosnek H: **Recruitment of the host plant heat shock protein 70 by tomato yellow leaf curl virus coat protein is required for virus infection**. PloS one **2013**, 8(7):e70280.

Grandillo S, Chetelat R, Knapp S, Spooner D, Peralta I, Cammareri M, Perez O, Termolino P, Tripodi P, Chiusano ML, Ercolano MR, Frusciante L, Monti L, Pignone D: **Solanum sect. *Lycopersicon***. In: **Wild Crop Relatives: Genomic and Breeding Resources, Vegetables,** C. Kole (ed.). Springer-Verlag **2011,** 129-215.

Grandillo S, Tanksley SD: **Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium***. Theoretical and Applied Genetics **1996**, 92(8):957-965.

Gross BL, Olsen KM: **Genetic perspectives on crop domestication**. Trends in plant science **2010**, 15(9):529-537.

Gunderson KL, Steemers FJ, Ren H, Ng P, Zhou L, Tsan C, Chang W, Bullis D, Musmacker J, King C, Lebruska LL, Barker D, Oliphant A, Kuhn KM, Shen R: **Whole-genome genotyping**. Methods in Enzymology **2006**, 410:359-76.

Gunderson KL: **Whole-genome genotyping on bead arrays**. Methods Molecular Biology **2009**, 529:197-213.

Hammer Ø, Harper DAT, Ryan PD: **PAST: Paleontological Statistics software package for education and data analysis**. Palaeontol Electron **2001**, 4(1):1-9.

Harlan JR: **The dynamics of domestication**. Crops & Man, American Society of Agronomy, Crop Science Society of America **1992,** 115-133.

Harvey M, Quilley S, Beynon H: **Exploring the tomato: transformations of nature, society and economy**. Bookcraft, Bath, UK. **2002**, 1-304.

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: **High-throughput genotyping by whole-genome resequencing**. Genome Research **2009,** 19(6):1068-1076.

Husain SE, James C, Shields R, Foyer CH: **Manipulation of fruit sugar composition but not content in *Lycopersicon esculentum* fruit by introgression of an acid invertase gene from *Lycopersicon pimpinellifolium***. New Phytologist **2001**, 150(1):65-72.

Hutton SF: **Fine-Mapping and Cloning of Ty-1 and Ty-3**; and **Mapping of a New TYLCV Resistance Locus, "Ty-6" [abstract]**. Tomato Breeders Round Table **2013**: http://tgc.ifas.ufl.edu/2013/abstracts/SamOrchardAbstract%20TBRT%202013.pdf

Illumina®: **Infinium® HD Assay: Ultra Protocol Guide**. Catalog #WG-901-4007 ©Illumina, Inc. California, USA. **2009**:1-224.

Iovene M, Wielgus SM, Simon PW, Buell CR, Jiang J: **Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato**. Genetics **2008,** 180(3):1307-1317.

Jenkins JA: **The Origin of the Cultivated Tomato**. Economic Botany **1948**, 2:379-392.

Ji Y, Scott JW, Hanson P, Graham E, Maxwell DP: **Sources of resistance, inheritance, and location of genetic loci conferring resistance to members of the tomato-infecting begomoviruses**. In Tomato Yellow Leaf Curl Virus Disease. Springer; **2007**: 343-362.

Jinap S, Hajeb P: **Glutamate. Its applications in food and contribution to health**. Appetite **2010**, 55(1):1-10.

Jonak C, Hirt H: **Glycogen synthase kinase 3/SHAGGY-like kinases in plants: an emerging family with novel functions**. Trends in plant science **2002**, 7(10): 457-461.

Kabelka E, Franchino B, Francis DM: **Two loci from Lycopersicon hirsutum LA407 confer resistance to strains of Clavibacter michiganensis subsp. michiganensis**. Phytopathology **2002**, 92(5):504-510.

Kadirvel P, de la Peña R, Schafleitner R, Huang S, Geethanjali S, Kenyon L, Tsai W, Hanson P: **Mapping of QTLs in tomato line FLA456 associated with resistance to a virus causing tomato yellow leaf curl disease**. Euphytica **2013**, 190(2):297-308.

Kamenetzky L, Asís R, Bassi S, de Godoy F, Bermudez L, Fernie AR, Van Sluys M-A, Vrebalov J, Giovannoni JJ, Rossi M: **Genomic analysis of wild tomato introgressions determining metabolism-and yield-associated traits**. Plant physiology **2010**, 152(4):1772-1786.

Kasrawi MA, Suwwan MA, Mansour A: **Sources of resistance to tomato-yellow-leaf-curl-virus (TYLCV) in Lycopersicon species**. Euphytica **1988**, 37(1):61-64.

Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M: **Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population**. Genetics **2007**, 175(2):891-905.

Khan N, Kazmi RH, Willems LAJ, Van Heusden AW, Ligterink W, Hilhorst HWM: **Exploring the natural variation for seedling traits and their link with seed dimensions in tomato**. PloS ONE **2012**, 7(8):e43991.

Klee HJ, Tieman DM: **Genetic challenges of flavor improvement in tomato**. Trends in Genetics **2013**, 29(4):257-262.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L: **VarScan: variant detection in massively parallel sequencing of individual and pooled samples**. Bioinformatics **2009**, 25(17):2283-5.

Labate JA and Robertson LD: **Evidence of cryptic introgression in tomato (*Solanum lycopersicum* L.) based on wild tomato species alleles**. BMC Plant Biology **2012**, 12:133.

Labate JA, Baldo AM: **Tomato SNP discovery by EST mining and resequencing**. Molecular Breeding **2005**, 16:343–349

Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** Genome Biology **2009**, 10(3):R25.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. Bioinformatics **2009**, 25(16):2078-2079.

Lin T, Zhu G, Zhang J, Xu X, Yu Q, Zheng Z, Zhang Z, Lun Y, Li S, Wang X, Huang Z, Li J, Zhang C, Wang T, Zhang Y, Wang A, Zhang Y, Lin K, Li C, Xiong G, Xue Y, Mazzucato A, Causse M, Fei Z, Giovannoni JJ, Chetelat RT, Zamir D, Stadler T, Li J, Ye Z, Du Y, Huang S: **Genomic analyses provide insights into the history of tomato breeding**. Nature Genetics **2014,** 46: 1220-1226.

Lippman Z, Tanksley SD: **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 **2001**, 158(1):413-422.

Lippman ZB, Semel Y, Zamir D: **An integrated view of quantitative trait variation using tomato interspecific introgression lines**. Current opinion in genetics & development **2007**, 17(6):545-552.

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR: **Gas chromatography mass spectrometry-based metabolite profiling in plants**. Nature Protocols **2006**, 1(1):387-396.

Lommen A: **Data (pre-) processing of nominal and accurate mass LC or GC data using MetAlign**. Plant Metabolomics. Humana Press **2012,** 229-253.

Long, J: **De tomates y jitomates en el siglo XVI**. Estudios de Cultura Náhuatl **1995**, 25: 239-252.

Lucatti AF, van Heusden AW, de Vos RC, Visser RG, Vosman B: **Differences in insect resistance between tomato species endemic to the Galapagos Islands**. BMC evolutionary biology **2013**, 13(1): 175.

Lunn D, Phan TD, Tucker GA, Lycett GW: **Cell wall composition of tomato fruit changes during development and inhibition of vesicle trafficking is associated with reduced pectin levels and reduced softening**. Plant Physiology and Biochemistry **2013**, 66:91-97.

Marshall JA, Knapp S, Davey MR, Power JB, Cocking EC, Bennett MD, Cox AV: **Molecular systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) using the nuclear ITS rDNA region**. Theoretical and Applied Genetics **2001**, 103:1216–1222.

McClean PE, Hanson MR: **Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species**. Genetics 1986, **112**:649–667.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M DePristo MA: **The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data**. Genome Research **2010**, 20(9):1297-1303.

Mézard C: **Meiotic recombination hotspots in plants**. Biochemical Society Transactions **2006**, 34(4):531-534.

Miller JC, Tanksley SD: **RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon**. Theoretical and applied genetics **1990**, 80(4):437-448.

Moco S, Bino RJ, Vorst O, Verhoeven HA, De Groot J, Van Beek TA, Vervoort J, Ric De Vos CH: **A liquid chromatography-mass spectrometry-based metabolome database for tomato**. Plant Physiology **2006**, 141(4):1205-1218.

Monforte AJ, Friedman E, Zamir D, Tanksley SD: **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 **2010**, 102(4): 572–590.

Morin PA, Luikart G, Wayne RK, and the SNP workshop group: **SNPs in ecology, evolution and conservation**. Trends in Ecology & Evolution **2004**, 19: 208-216.

Mouritsen, OG, Styrbæk K. **Umami: Unlocking the secrets of the fifth taste**. Columbia University Press **2014**.

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

Moyle LC, Muir CD: **Reciprocal insights into adaptation from agricultural and evolutionary studies in tomato**. Evolutionary Applications **2010**, 3:409-421.

Muir CD, Moyle LC: **Antagonistic epistasis for ecophysiological trait differences between *Solanum* species**. New Phytologist **2009,** 183:789-802.

Muños S, Ranc N, Botton E, Bérard A, Rolland S, Duffé P, Carretero Y, Le Paslier M-C, Delalande C, Bouzayen M: **Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL**. Plant Physiology **2011**, 156(4):2244-2254.

Nakazato T, Bogonovich M, Moyle LC: **Environmental factors predict adaptive phenotypic differentiation within and between two wild Andean tomatoes**. Evolution **2008**, 62:774-792.

Nakazato T, Warren DL, Moyle LC: **Ecological and geographic modes of species divergence in wild tomatoes**. American Journal of Botany **2010**, 97:680-693.

Nesbitt TC and Tanksley SD: **Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes**. Genetics **2002**, 162(1):365-379.

Nijveen H, van Kaauwen M, Esselink D, Hoegen B, Vosman B: **QualitySNPng: a user-friendly SNP detection and visualization tool**. Nucleic Acids Research **2013**, gkt333: 1-4.

Osorio S, Tohge T, Fernie A: **Application of metabolomic profiling for identifying valuable traits in tomato**. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources **2009**, 4(024):1-9.

Ozminkowski, R. **Pedigree of variety Heinz 1706**. Tomato Genetics Cooperative Report **2004**, 54:26.

Palmer JD, Zamir D: **Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon***. Proceedings of the National Academy of Sciences **1982**, 79:5006–5010.

Paran I, Goldman I, Tanksley SD, Zamir D: **Recombinant inbred lines for genetic mapping in tomato**. Theoretical and Applied Genetics **1995**, 90(3-4):542-548.

Peralta IE, Knapp S, Spooner DM: **Nomenclature for wild and cultivated tomatoes**. TGC Report **2006**, 56:6-12.

Peralta IE, Knapp S, Spooner DM, Lammers TG: **New species of wild tomatoes (*Solanum* section *Lycopersicon*: Solanaceae) from Northern Peru**. Systematic Botany **2005**, 30:424-434.

Peralta IE, Spooner DM, Knapp S: **Taxonomy of wild tomatoes and their relatives (*Solanum* sections *Lycopersicoides*, *Juglandifolia*, *Lycopersicon*; Solanaceae)**. Systematic Botany Monograhs **2008**, 84:1–186.

Peralta IE, Spooner DM: **Morphological characterization and relationships of wild tomatoes (*Solanum* L. Sect. *Lycopersicon*)**. Monographs In Systematic Botany **2005**, 104:227.

Pereira da Costa JH, Rodríguez GR, Pratta GR, Picardi LA, Zorzoli R: **QTL detection for fruit shelf life and quality traits across segregating populations of tomato**. Scientia Horticulturae **2013**, 156:47-53.

Pérez de Castro A, Díez MJ, Nuez F: **Inheritance of Tomato yellow leaf curl virus resistance derived from *Solanum pimpinellifolium* UPV16991**. Plant Disease **2007**, 91(7):879-885.

Perez-Fons L, Wells T, Corol DI, Ward JL, Gerrish C, Beale MH, Seymour GB, Bramley PM, Fraser PD: **A genome-wide metabolomic resource for tomato fruit from *Solanum pennellii***. Scientific reports **2014**, 4:1-8.

Petro-Turza M: **Flavor of tomato and tomato products**. Food Reviews International **1986**, 2(3):309-351.

Picó B, Ferriol M, Diez MJ, Vinals FN: **Agroinoculation methods to screen wild *Lycopersicon* for resistance to Tomato yellow leaf curl virus**. Journal Plant Pathology **2001**, 83(3):215-220.

Picó B, Sifres A, Elia M, Díez MJ, Nuez F: **Searching for new resistance sources to tomato yellow leaf curl virus within a highly variable wild Lycopersicon genetic pool**. Acta Physiologiae Plantarum **2000**, 22(3):344-350.

Pilowsky M, Cohen S: **Screening additional wild tomatoes for resistance to the whitefly-borne tomato yellow leaf curl virus**. Acta Physiologiae Plantarum **2000**, 22(3):351-353.

Prudent M, Bertin N, Génard M, Muños S, Rolland S, Garcia V, Petit J, Baldet P, Rothan C, Causse M: **Genotype-dependent response to carbon availability in growing tomato fruit**. Plant, cell & environment **2010**, 33(7):1186-1204.

Ranc N, Muños S, Santoni S, Causse M: **A clarified position for *Solanum lycopersicum* var. cerasiforme in the evolutionary history of tomatoes (solanaceae).** BMC Plant Biology **2008**, 8:130.

Ranc N, Muños S, Xu J, Le Paslier M-C, Chauveau A, Bounon R, Rolland S, Bouchet J-P, Brunel D, Causse M: **Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of Solanum lycopersicum var. cerasiforme**. G3: Genes | Genomes | Genetics **2012**, 2(8):853-864.

Ravi I, Baunthiyal M, Saxena J: **Advances in Biotechnology**. Springer, **2014**. ISBN: 978-81-322-1554-7.

Rick C, Fobes J: **Association of an allozyme with nematode resistance.** Tomato Genetics Cooperative Report **1974**, 24:25.

Rick CM: **Potential genetic resources in tomato species: clues from observations in native habitats**. Basic life sciences **1973**, 2:255-269.

Rick CM, Laterrot H, Philouze J: **A revised key for *Lycopersicon* species**. Tomato Genetics Cooperative Report **1990**, 40: 31

Rodriguez F, Wu F, Ané C, Tanksley S, Spooner DM**: Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history?.** BMC Evol Biol **2009,** 9(1):191.

Rodríguez GR, Muños S, Anderson C, Sim S-C, Michel A, Causse M, Gardener BBM, Francis D, van der Knaap E: **Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity**. Plant physiology **2011**, 156(1):275-285.

Ross JA, Kasum CH: **Dietary flavonoids: bioavailability, metabolic effects, and safety**. Annual review of Nutrition **2002**, 22(1):19-34.

Sade D, Brotman Y, Eybishtz A, Cuadros-Inostroza Á, Fernie AR, Willmitzer L, Czosnek H: **Involvement of the hexose transporter gene LeHT1 and of sugars in resistance of tomato to tomato yellow leaf curl virus**. Molecular Plant **2013**, 6(5):1707-1710.

Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie A, Czosnek H, Brotman Y: **Comparative metabolomics and transcriptomics of plant response to Tomato yellow leaf curl virus infection in resistant and susceptible tomato cultivars**. Metabolomics **2015**, 11(1):81-97.

Saliba-Colombani V, Causse M, Langlois D, Philouze J, Buret M: **Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits**. Theoretical and Applied Genetics **2001**, 102(2-3):259-272.

Salinas M, Capel C, Alba JM, Mora B, Cuartero J, Fernández-Muñoz R, Lozano R, Capel J: **Genetic mapping of two QTL from the wild tomato *Solanum pimpinellifolium* L. controlling resistance against two-spotted spider mite (Tetranychus urticae Koch)**. Theoretical and Applied Genetics **2013**, 126(1):83-92.

Sato Y, Itagaki S, Kurokawa T, Ogura J, Kobayashi M, Hirano T, Sugawara M, Iseki K: **In vitro and in vivo antioxidant properties of chlorogenic acid and caffeic acid**. International Journal of Pharmaceutics **2011**, 403(1–2):136-138.

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J: **Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement**. Nature biotechnology **2006**, 24(4):447-454.

Schauer N, Zamir D, Fernie AR: **Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex**. Journal of Experimental Botany **2005**, 56(410):297-307.

Schijlen EG, Ric de Vos C, van Tunen AJ, Bovy AG: **Modification of flavonoid biosynthesis in crop plants**. Phytochemistry **2004**, 65(19):2631-2648.

Schwahn K, de Souza LP, Fernie AR, Tohge T: **Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade**. Journal of integrative plant biology **2014**, 56(9):864-875.

Service XS: **Project Report: Illumina iSelect Genotyping**. Private report produce by the company: 2333 BZ Leiden. **2010**: 1-28.

Sharma A, Zhang L, Nio-Liu D, Ashrafi H, Foolad MR: **A *Solanum lycopersicum x Solanum pimpinellifolium* linkage map of tomato displaying genomic locations of R-genes, RGAs, and candidate resistance/defense-response ESTs**. International Journal of Plant Genomics **2008**, 1-18.

Sherman JD, Stack SM: **Two-dimensional spreads of synaptonemal complexes from solanaceous plants. V. Tomato (*Lycopersicon esculentum*) karyotype and idiogram**. Genome **1992**, 35(2): 354-359

Shirasawa K, Isobe S, Hirakawa H, Asamizu E, Fukuoka H, Just D, Rothan C, Sasamoto S, Fujishiro T, Kishida Y, Kohara M, Tsuruoka H, Wada T, Nakamura Y, Sato S, Tabata S: **SNP discovery and linkage map construction in cultivated tomato**. DNA Research **2010,** 17(6):381-391.

Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM: **Development of a Large SNP Genotyping Array and Generation of High-Density Genetic Maps in Tomato**. PLoS ONE **2012**, 7(7):,1-11.

Sim S-C, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MG, Chetelat RT, Hutton SF, Scott JW, Gardner RG, Panthee DR, Mutschler M, Myers JR, Francis DM: **High-Density SNP Genotyping of Tomato (*Solanum lycopersicum* L.) Reveals Patterns of Genetic Variation Due to Breeding**. PLoS ONE **2012b**, 7(9):1-18.

Simms D, Cizdziel PE and Chomcynski P: **TRIZOL: A new reagent for optimal single-step isolation of RNA**. Focus **1993**, 15:99–102.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH: **JBrowse: A next-generation genome browser**. Genome Research **2009**, 19(9):1630-1638.

Small DM, Voss J, Mak, YE, Simmons KB, Parrish T, Gitelman D. **Experience-dependent neural integration of taste and smell in the human brain**. Journal of Neurophysiology **2004**, 92(3):1892-1903.

Spooner DM, Peralta IE, Knapp S: **Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]**. Taxon **2005**:43-61.

Spooner DM, Anderson GJ, Jansen RK: **Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (*Solanaceae*)**. American Journal Botany **1993**, 80:676–688.

Stack SM, Royer SM, Shearer LA, Chang SB, Giovannoni JJ, Westfall DH, White RA, Anderson LK: **Role of fluorescence in situ hybridization in sequencing the tomato genome**. Cytogenetic and Genome Research **2009**, 124:339–350.

Steemers FJ and Gunderson KL: **Whole genome genotyping technologies on theBeadArray™ platform**. Biotechnology Journal **2007**, 2:41–49.

Steward CN, Via LE: **A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications**. Biotechniques **1993**, 14:748–750.

Street SL, Kyes RC, Grant R, Ferguson B: **Single nucleotide polymorphisms (SNPs) are highly conserved in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques**. BMC Genomics **2007**, 8:480.

Tadmor T, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E. **Identification of *malodorous*, a wild species allele affecting tomato aroma that was selected against during domestication**. Journal of Agricultural and Food Chemistry **2002**, 50(7): 2005-2009.

Tam SM, Hays JB, Chetelat RT: **Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato**. Theoretical and Applied Genetics **2011**, 123(8):1445-1458.

Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JAM: **QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species**. BMC Bioinformatics **2006**, 7:438.

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T: **Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium***. Theor Appl Genet **1996**, 92(2):213-224.

Taylor IB: **Biosystematics of the tomato.** In: J. Atherton and J. Rudich (Eds.), The Tomato Crop. Springer Netherlands **1986**, 1-34.

Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, Clark TG: **A genotype calling algorithm for the Illumina BeadArrayplatform**. Bioinformatics **2007**, 23:2741-6.

Tieman DM, Bliss P, McIntyre LM, Blandon-Ubeda A, Bies D, Odabasi AZ, Rodríguez 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: **The chemical interactions underlying tomato flavor preferences**. Current Biology **2012**, 22(11):1035-1039.

Tieman DM, Taylor M, Schauer N, Fernie AR, Hanson AD, Klee HJ: **Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde**. Proceedings of the National Academy of Sciences **2006**, 103(21):8287-8292.

Tieman DM., Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ: **Identification of loci affecting flavour volatile emissions in tomato fruits**. Journal of Experimental Botany **2006b**, 57(4):887-896.

Tikunov Y, Lommen A, De Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG: **A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles**. Plant Physiology **2005**, 139(3):1125-1137.

Tikunov YM, de Vos RC, Paramás AMG, Hall RD, Bovy AG: **A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach**. Plant physiology **2010**, 152(1):55-70.

Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH: **MSClust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data**. Metabolomics **2012**, 8(4):714-718.

Tikunov YM, Molthoff J, de Vos RC, Beekwilder J, van Houwelingen A, van der Hooft JJ, Nijenhuis-de Vries M, Labrie CW, Verkerke W, van de Geest H, Bovy AG: **NON-smoky glycosyltransferase1 prevents the release of smoky aroma from tomato fruit**. The Plant Cell Online **2013**, 25(8):3067-3078.

Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, Fait A: **Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations**. PLoS genetics **2012,** 8(3): e1002612.

Tomato Genome Consortium. **The tomato genome sequence provides insights into fleshy fruit evolution**. Nature **2012**, 485:635-641.

UPOV: **International convention for the protection of new varieties of plants** of December 2, 1961, as Revised at Geneva on November 10, 1972, on October 23, 1978, and on March 19, **1991**: http://www.upov.int/en/publications/conventions/1991/act1991.htm.

Van Berloo R, van Heusden S, Bovy A, Meijer-Dekens F, Lindhout P, van Eeuwijk F: **Genetic research in a public-private research consortium: prospects for indirect use of Elite breeding germplasm in academic research**. Euphytica **2008**, 161:293-300.

Van Eeuwijk FA, Law JR: **Statistical aspects of essential derivation, with illustrations based on lettuce and barley**. Euphytica **2004**, 137:129-137.

Van Ooijen J: **MapQTL® 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species.** Kyazma BV: Wageningen, Netherlands **2009**.

Van Ooijen JW: **Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species**. Genetics Research **2011**, 93:343-349.

Van Tienderen PH, de Haan AA, van der Linden CG, Vosman B: **Biodiversity assessment using markers for ecologically important traits**. Trends in Ecology and Evolution **2002**, 17(12):577-582.

Verlaan MG, Hutton SF, Ibrahem RM, Kormelink R, Visser RG, Scott JW, Edwards JD, Bai Y: **The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 are Allelic and code for DFDGD-Class RNA–Dependent RNA Polymerases**. PLoS genetics **2013**, 9(3):e1003399.

Verlaan MG, Szinay D, Hutton SF, de Jong H, Kormelink R, Visser RG, Scott JW, Bai Y: **Chromosomal rearrangements between tomato and Solanum chilense hamper mapping and breeding of the TYLCV resistance gene Ty-1**. The Plant Journal **2011**, 68(6):1093-1103.

Villareal RL: **Tomatoes in the tropics**. Bowker Publishing Company, **1980:** 1-174. ISBN: 0-89158-989-9.

Víquez-Zamora M, Vosman B, van de Geest H, Bovy A, Visser RGF, Finkers R, van Heusden AW: **Tomato breeding in the genomics era: Insights from a SNP array**. BMC Genomics **2013**, 14:354.

Víquez-Zamora M, Caro M, Finkers R, Tikunov Y, Bovy A, Visser RG, Bai Y, van Heusden S: **Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in *Solanum pimpinellifolium***. BMC genomics **2014**, 15(1):1152.

Vogel JT, Tieman DM, Sims CA, Odabasi AZ, Clark DG, Klee HJ: **Carotenoid content impacts flavor acceptability in tomato (*Solanum lycopersicum*)**. Journal of the Science of Food and Agriculture **2010**, 90(13):2233-2240.

Voorrips RE, Verkerke W, Finkers R, Jongerius R, Kanne J: **Inheritance of taste components in tomato**. Acta Physiologiae Plantarum **2000**, 22(3):259-261.

Voorrips RE: Mapchart: **Software for the graphical presentation of linkage maps and QTLs**. Journal of Heredity **2002**, 93(1):77-78.

Vosman B, Visser D, Rouppe J, van der Voort, Smulders MJM, van Eeuwijk F: **Detection of mutants and the establishment of essential derivation among rose varieties using AFLP**. Theoretical and Applied Genetics **2004**, 109:1718–1725.

Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG: **The Influence of Fruit Ripening on the Accumulation of Health-Related Metabolites in Five Different Capsicum Accessions**. In: Wahyuni, Breeding for pepper fruit quality: a genetical metabolomics approach. PhD thesis, Wageningen University, Wageningen, Netherlands. **2014,** 93-120**.**

Xu J, Ranc N, Muños S, Rolland S, Bouchet J-P, Desplat N, Le Paslier M-C, Liang Y, Brunel D, Causse M: **Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species**. Theoretical and Applied Genetics **2013**, 126:567-581.

Xu J-G, Hu Q-P, Liu Y: **Antioxidant and DNA-protective activities of chlorogenic acid isomers**. Journal of agricultural and food chemistry **2012**, 60(46):11625-11630.

Yang X, Caro M, Hutton SF, Scott JW, Guo Y, Wang X, Rashid MH, Szinay D, de Jong H, Visser RGF, Bai Y, Du Y: **Fine mapping of the tomato yellow leaf curl virus resistance gene Ty-2 on chromosome 11 of tomato**. Molecular Breeding **2014**:1-12.

Zanor MI, Rambla J-L, Chaïb J, Steppa A, Medina A, Granell A, Fernie AR, Causse M: **Metabolic characterization of loci affecting sensory attributes in tomato allows an assessment of the influence of the levels of primary metabolites and volatile organic contents**. Journal of Experimental Botany **2009**, 60(7):2139-2154.

Summary

Summary

Tomatoes are part of the Solanaceae family which comprises over 3000 species. Around half of this family is represented by the genus *Solanum* which includes amongst others tomato, aubergine and potato. Resources to study tomato plants are developed worldwide to elucidate the relationship between genotypes and phenotypes and this made tomato a suitable model to study crop plants.

Since a few years, genotyping is no longer perceived as a major bottle neck in plant molecular breeding and this also holds for tomato. Genome (re)sequencing projects are delivering large numbers of informative markers such as single nucleotide polymorphisms (SNP). We used a set of 5528 SNPs to evaluate the tomato germplasm. Genotyping different tomato samples allowed the evaluation of the level of heterozygosity and the number of introgressions in commercial varieties. We found relatively more differences between cherry and round/beef tomatoes on chromosomes 4, 5 and 12. Furthermore, we identified a set of markers suitable to differentiate *S. lycopersicum* var. Moneymaker from all other wild relatives we evaluated. The SNPs can be used for genotyping, identification of varieties, comparison of genetic and physical maps and to confirm (dis)similarities (Chapter 2).

A part of tomato research aims at determining how to expand genetic diversity in the existing tomato crop; this can be done by incorporating useful traits found in wild germplasm. In this thesis we focused on exploring the variation between accessions of wild relatives of the species in the subsection Lycopersicon. Especially, on the species *Solanum pimpinellifolium* since it represents a good source to explore variation for quality traits that can be incorporated into cultivated tomatoes. A recombinant inbred line (RIL) population was developed from a cross between an accession of this wild relative and *S. lycopersicum* var. Moneymaker. All the lines were genotyped with our SNP array and 1974 SNPs made it possible to construct a linkage map based on 715 genetic loci. In this way we could compare genetic linkage and physical positions. Additionally, a subset of the lines was genotyped by sequencing (GBS). We identified two QTLs for resistance to Tomato Yellow Mosaic Virus (TYCV) and the sequence information was used to saturate the Quantitative Trait Loci with more markers. We found that the resistance to TYLCV was associated to a region on chromosome 11 close to the region of *qTy-p11* (~51.3 Mb) and to another region on chromosome 3 near *qTy-p3 (*~46.5 Mb) (Chapter 3). We also used this genotyping approach to target mQTL hotspots for fruit related metabolites.

Three different metabolomics platforms were used to phenotype the metabolome of the RIL population. Liquid chromatography coupled with mass spectrometry (LS) was used to detect semi-polar compounds such as flavonoids, alkaloids, phenylpropanoids, saponins, phenolic

acids, polyamines and products thereof. Gas chromatography (GC) coupled with electron impact time of flight (TOF) was used for detection of primary metabolites and solid phase microextraction (SPME)-GC for the analysis of volatiles. We performed QTL analysis on leaf and fruit samples of the RIL population. The TYLCV resistance mechanism is likely associated with sucrose and flavonoid glycosides related regions on chromosomes 11 and 3, respectively. With the combination of different ~omics platforms we provided a valuable insight into the genetics behind *S. pimpinellifolium*-derived TYLCV resistance.

For fruits, we found clear metabolite QTL-hotspots on chromosomes 1 and 10. Our results show that to increase the antioxidant properties of tomato, the region between 71-87 Mb on chromosome 1 has to originate from Moneymaker while other regions on chromosome 6 (35-44 Mb), chromosome 10 (~44.3 Mb) and chromosome 12 (~48 Mb) have to be of *S. pimpinellifolium* origin. The above-mentioned region on chromosome 6 also affects the concentration of malic acid in the fruits. Sugars can be increased by combining the wild alleles on chromosome 2 (~41.7 Mb) for sucrose and chromosome 10 (1.7 Mb) for fructose with the Moneymaker alleles on the hotspot region of chromosome 1 and chromosome 4 (~55 Mb) for fructose and glucose respectively. Off flavour regions that should be avoided in crosses with *S. pimpinellifolium* are the ones at the top of chromosome 1 and on chromosome 9 around 65 Mb where we found loci associated with the concentration of the compounds putrescine and dimethyl disulfide. An aromatic boost to the fruits can be given by the introgression of parts of the wild chromosome 8; this will increase the concentration of phenolic VOCs (Chapter 4).

S. pimpinellifolium certainly harbours characteristics that could be (re-) introduced in tomato. Therefore, we explored the metabolome of several accessions of *S. pimpinellifolium* during ripening. Clear metabolic profile differences were identified between species, especially related to the phenylpropanoid pathway. *S. pimpinellifolium* is a potential source to improve the flavonoid content of tomatoes and several other fruit aromas. Certain accessions looked even more promising than the RIL parent as a source for quality traits. All this helped us to better understand particular differences with wild relatives or even between genotypes (Chapter 5).

In general, our results give an insight in the physical positions of metabolite related QTLs that could be used by breeders that would like to exploit *S. pimpinellifolium* to improve tomato quality.

Acknowledgements

My most deep and sincere acknowledgements go to Dr. Adriaan W. (Sjaak) van Heusden. Thanks for teaching me that there is no need to call a person "Doctor" to respect, admire and appreciate him/her deeply; Thanks for teaching me that there is a moment where you should just ´throw it away´. Thanks for teaching me the value behind one ´I don't believe you´ in order to build confidence and prove arguments. Thanks Sjaak for opening a door when I was not strong enough to knock and finally, thanks for keep on beleving in me!. Voor het verleden, het heden en de toekomst: Hartelijk Bedankt!.

To Dr. Richard Visser: Thanks for all your support within Plant Breeding along with the ideas for collaboration with other departments or institutions. Thanks for believing I could make this project "happen", thanks for giving me the opportunity of representing young plant breeders from WageningenUR in Turkey and thank you for your input in every step of the way.

To Dr. Arnaud Bovy: Thank you for supporting my project with no strings attached. Thanks for all the discussions, the meetings, the advices, your understanding and even sometimes for not understanding so that I could try harder in explaining.

To Dr. Richard Finkers: Thank you Richard for essential pieces that made this thesis come true. Thank you for your willingness to find solutions to bioinformatic problems. Thank you for showing always understanding and support in this scientific roller-coaster of result's production!

To Dr. Yuling Bai and Dr. Ben Vosman: Thanks for supporting my project and for your valuable input in some of this thesis chapters.

To the Department of Plant Breeding: Thanks for the support, the shared knowledge, the smiles, the greetings in the hall, the talks near the coffee machines, the feeling of belonging. I am truly thankful for the opportunity to be a part of "Plant Breeding".

To the great people of Unifarm, particularly to Andre Maassen, Maarten Peters, Geurt Versteeg and Teus van den Brink. Thanks for the willingness to help, the support, the green fingers and the care to the plants within all the experiments.

To my "Breeding for Quality Group": Marine Paupiere, Yury Tikunov, Maria Martorana, Yuni Wahyuni, Vanessa Stahl-Hermes, Jan Schaart, Fien Meijer-Dekens, Jos Molthoff, Raana Roohanitaziani, Marjolein Hunck, Jordi Petit Pedró and all the ones that joined the group in different periods: Thanks for all the moments shared and thanks for all your support. Thanks for the teas, the creative ideas coming together and the laughs.

To the government of Costa Rica for their support in certain stage of my thesis.

To the Department of Biosciences: Thanks for the collaboration, the willingness to help, the teaching, the patiente, the shared knowledge and the essential contribution for the metabolic analysis along this thesis.

To Ric de Vos: Deep thanks for all your support and time. You are one of those rare persons that are so smart and so nice at the same time that whenever we come to you with a question we left your room feeling that our knowledge increased at least a bit. Thanks for all your explanations and thanks for inspiring so much passion for knowledge and science.

To Roland Mum and Bert Schipper: Thanks for showing me your passion for metabolomics, and thanks for always having the best disposition to help.

To Henriëtte van Eekelen: Thank you dear Henriëtte for letting me learn with you some of the many tricks about performing metabolic analyses. Thanks for your support, your patiente and your friendship.

To the Department of Biometris, especially to Joao Paolo, Marcos Malosetti and Fred van Eeuwijk. Thanks for all the support, willingness to collaborate, shared statistical knowledge and beyond.

To the administrative-heart-and-soul of Plant Breeding: Janneke van Deursen, Nicole Trefflich, Letty Dijker and Annie Marchal: Thank you very very very much for making things easier. Thanks for inspiring happiness and comfort with willingness to support and help. Thank you for all your smiles and laughs making us feel closer to home!.

To Marine: Thank you dear Marinecita for walking with me in this PhD path, thanks for sharing ups and downs, thanks for helping me so much in this learning together to do metabolic analyses, thanks for all your support. I will always treasure your friendship in my heart!.

To Myluska: Thank you Mylu for the support, the sharing of ideas and the fused efforts to work together for a common goal. Thank you for the laughs, the shared-confusion moments and your friendship.

To Rutger Rottenhatter: Thanks for the support of most of these years at the office. Thanks for all the tea-times, the 'dutch lessons' and the 'boy-scouts' stories (even to learn from the dangerous ones).

To Yury Tikunov: Thanks for opening my mind into different ways of thinking. Thanks for showing me the Russian way to make efficient analysis. Thanks for all your scientific, professional and friendly support.

To Yuling Bai, Luisa Trindade and Joao Paolo: Thanks for being a proof that women can be successful researchers even when they are in foreign countries. Thanks for your great example and for being an inspiration!.

To my students/assistants also my friends: Jarno Sinnige, Jordi Petit Pedró and Marjolein Hunck. Working with you guys was/is a pleasure!. Thanks for your time, hard work and patience. Thanks for challenging me into explaining things as clear as possible. Thanks for the complicity, the sharing of frustration and success, the laughs and beyond.

To all my friends in Plant Breeding: Marine, Narges, Peter (Quy Dinh), Madhuri, Myluska, Mario, Kaile, Anne, Izan, Mas, Christos, Behzad, Xuan, Raana, Tim (& Sifra), Ningwen, Johan, Colette, Sara,, Cristina, Elma, Patrick, Koen, Gerard B., Gerard vd L., Eric, Olga, Roeland, Luisa and many more people lighting up any grey day.

To my now Doctor friends (lets use the title!): Dr. Anitha Kumari, Dr. Paula Hurtado, Dr. Wahuyuni, Dr. Ningwen Zhang, Dr. Nasim Irani, Dr. Diana Londoño, Dr. Alejandro Lucatti, Dr. Andrés Torres, Dr. Juan Du, Dr. Dongli Gao, Dr. Natalia Carreño, Dr. Awang , Dr. Pierre-Yves Chibon, Dr. Erica Renaud and many more to come!. Thanks for all the inspiration and moments of support!

To more of all the friends WageningeUR broad into my life: Pavlos, Laurita, Okello, Etalo, Iliana, Jordi, Rene, Paula, César, Carito, Roland, Natalia, Bjorn, Sabine, Deborah, Joachim, Annie, Alvin, Daniela, Marcelo, Cris, María A., Saulo, Nato, Reinier and many more. Thanks for all the shared moments, I hope we can keep on finding moments to share.

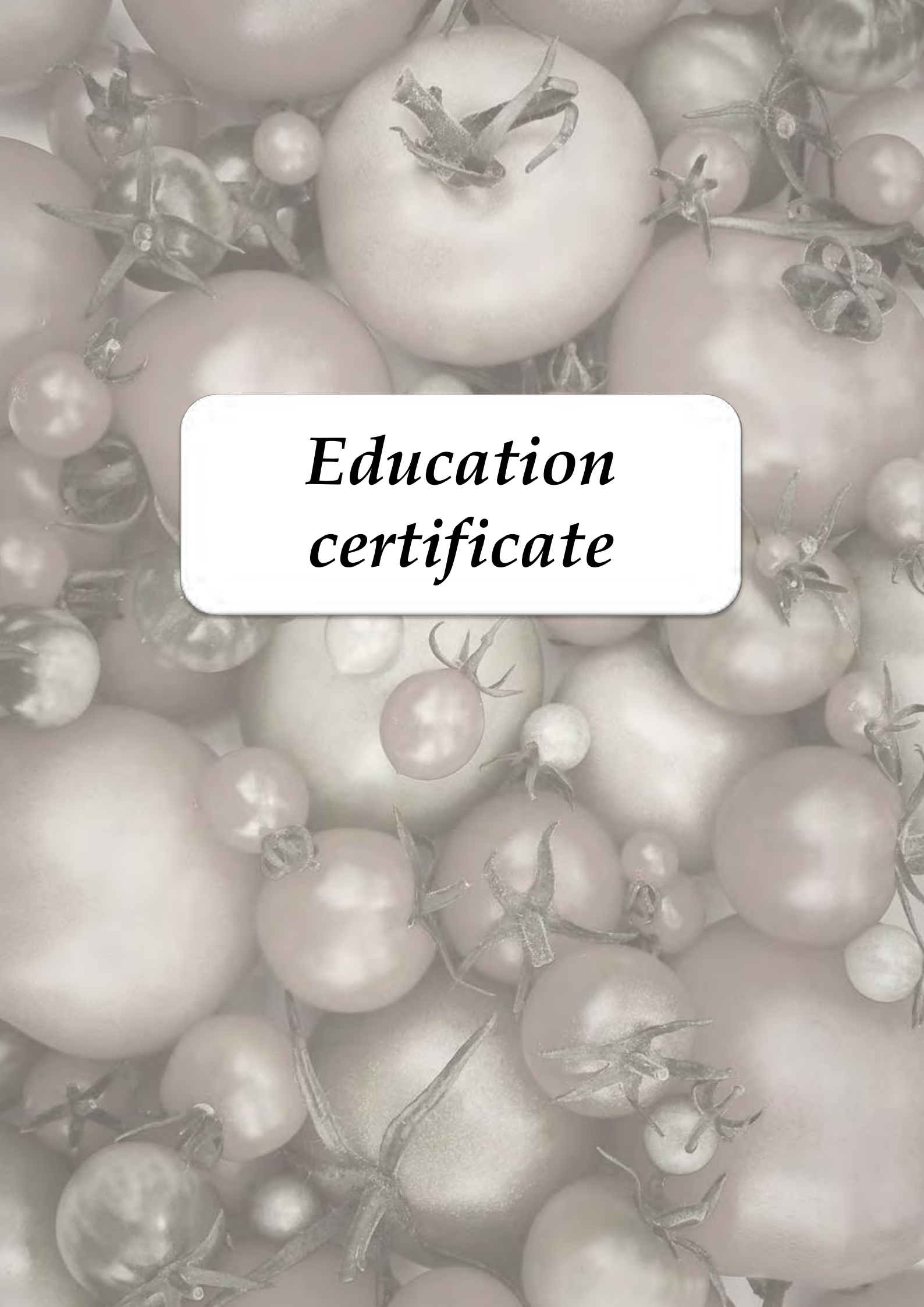
To my Solynta colleagues: Thank you guys for all the support in the final stage of delivering my thesis. Looking forward to make diploid potato breeding happen together with you!.

Thanks to my latin-family in Wageningen: Sergio, Made, Irene, Carlos Iván, Carlita, Vicky, Pieter, Marce, Fran. Chicos, gracias por el cariño, apoyo moral constante, los cuidados, los favores, los paseos, las saliditas, las bicicleteadas, las conversadas y resolvidas de mundo. Gracias por "ser y estar" siempre. Irene: Gracias lindísima por estar ahí desde 'el inicio de los tiempos', por todo el cariño, el apoyo, las regañadas, las terapias, las tardes de té y demás. Sergio: Gracias por incitarme siempre a buscarle un lado más positivo a las cosas. Made: Gracias por ser la inspiración de la dulzura, por todas las aventuras compartidas, por todo el cariño y el amor. ¡Los quiero muchísimo!.

To many friends in Costa Rica (&+): Nathalia y Rolo, Marlencita y Eddy, Andrea, Mariajosé, Giancarlo y AnaVic, Ethel, Ericka, Mariana, Diana M., las Natalias, Diana S. Gaby, Manri. Tenerlos en mi vida es una bendición. Ustedes son familia del alma y del corazón.

Por supuesto, gracias a mi familia: A los Chan, gracias por todo el apoyo incondicional a través de los años y por siempre hacernos sentir queridos. A mis abuelitos que ya no nos acompañan, gracias por todas las lecciones de vida y por todo su legado. A mi Abuelita Salomé, gracias por toda la ternura a través de los años y por todas mis mitocondrias. A mi mamá y mi papá: gracias por forjarme para ser una persona perseverante. Gracias por su amor y apoyo infinito. A GJ y Gis: gracias por su apoyo, amistad y guía a través de los años. Gracias por Fabián (con quien adquirí el hermoso compromiso de ser Madrina), por Diego y por Paula. Ricardo y Stef: Gracias por su apoyo, compañía y mil gracias por Matías e Ignacio. Gracias por todas las fotos, los videos y los mensajes. Mis sobrinos me hacen querer ser una mejor persona y me alegran el corazón más allá del tiempo y el espacio. Son realmente un regalo. A JoseB: gracias infinitas por ser más allá que mi hermanito menor, gracias por ser uno de mis mejores amigos, gracias por la complicidad, por el apoyo, por hacer siempre ver que no hay distancia ni horarios que afecten la comunicación, gracias por siempre ser y estar incondicionalmente presente. ¡Los adoro!.

Y finalmente, a la persona más importante de mi vida: ¡Gracias Rafa! No hay suficientes palabras que puedan expresar todos los agradecimientos, reconocimientos, cariños y amor... pero espero que el resto de la vida me brinde todas los oportunidades que se puedan para decirte y demostrarte todo lo que agradezco que estemos juntos en éste proyecto de vida y todo lo que ¡te amo!.



Education certificate

Education Statement of the Graduate School
Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Ana Marcela Víquez Zamora
Date: 4 September 2015
Group: Laboratory of Plant Breeding
University: Wageningen University & Research Centre

1) Start-up phase	*date*
► **First presentation of your project**	
Illumina Infinium Beadarray and Tomato Wild Relatives Research	Mar 01, 2011
► **Writing or rewriting a project proposal**	
Exploiting wild relatives of S. lycopersicum for tomato quality improvement	Sep 19, 2012
► **Writing a review or book chapter**	
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD student day, Amsterdam University	Nov 30, 2012
EPS PhD student day, Leiden University	Nov 29, 2013
► **EPS theme symposia**	
EPS Theme 4 'Genome Biology', Wageningen University	Dec 09, 2011
EPS Theme 4 'Genome Biology', Wageningen University	Dec 03, 2014
► **NWO Lunteren days and other National Platforms**	
Plant Breeding-Research Day	Mar 08, 2011
Plant Breeding-Research Day	Feb 28, 2012
100 Years Plant Breeding: Open Day	Aug 31, 2012
Plant Breeding-Research Day	Sep 24, 2014
► **Seminars (series), workshops and symposia**	
CBSG Tomato Clustermeeting	Oct 31, 2011
Plant Science Seminar by Young PSG: Speek, Present and Debate	Jul 24, 2012
CBSG Tomato Clustermeeting	Nov 20, 2012
CBSG Summit 2012	Feb 29-Mar 01, 2012
CBSG Summit and Tomato Clustermeeting	Feb 11-13, 2013
Multiple omics integration & exploring genome functionality using NGS techniques	Dec 12-13, 2013
► **Seminar plus**	
► **International symposia and congresses**	
Next Generation Plant Breeding Conference	Nov 11-14, 2012
Cost Action FA1106 - Quality Fruit 2013 Conference, Greece	Sep 22-25, 2013
Plant Breeders' Day, Turkey	May 11, 2014
► **Presentations**	
Hortifair 2011 seminar "More crop per drop" (Talk)	Nov 03, 2011
CBSG Summit 2012 (Poster)	Feb 29-Mar 01, 2012
Large scale genotyping of CBSG tomato populations - CBSG Summit 2012 (Talk)	Feb 29, 2012
Large-scale genotyping in tomato and its wild relatives - CBSG Tomato Clustermeeting (Talk)	Nov 20, 2012
CBSG Summit 2013 (Poster)	Feb 11, 2013
Workshop "Utilización práctica de la genómica", Costa Rica (Talk)	Apr 29, 2013
Cost Action FA1106 - Quality Fruit 2013 Conference (Talk)	Sep 24, 2013
Plant Breeders Day", Turkey (Talk)	May 11, 2014
Symposium "All inclusive Breeding: Integrating highthroughput science" (Poster)	Oct 16, 2014
► **IAB interview**	
Meeting with a member of the International Advisory Board of EPS	Sep 29, 2014
► **Excursions**	
Visit Rijk Zwaan and Monsanto	Oct 18, 2012
Tomato World Excursion	Nov 06, 2012
Subtotal Scientific Exposure	*18.0 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Course Genome Mining	Oct 27, 2011
Current Trends in Phylogenetics	Oct 22-26, 2012
Systems Biology course "Statistical analysis of ~omics data"	Dec 10-14, 2012
Course "Mixed model based genetic analysis in GenStat: from QTL mapping and association mapping t	Sep 02-04, 2013
Course "Introduction to R for statistical analysis"	Oct 21-22, 2013
► **Journal club**	
► **Individual research training**	
Subtotal In-Depth Studies	*4.9 credits**
4) Personal development	*date*
► **Skill training courses**	
Scientific writing	Jan- Feb, 2013
Workshop "Last Stretch of the PhD Programme"	Sep 20, 2013
Voice Matters - Course	Oct, 2013
Mini-symposium: How to Write a World-class Paper	Oct 17, 2013
The Essentials of Scientific Writing and Presenting	May, 2014
PhD Workshop Carousel	Jul 02, 2014
Social Dutch for employees	Oct 2013-Jan 2014
Entrepreneurship in and outside science	Sep-Oct 2014
► **Organisation of PhD students day, course or conference**	
Monday Mornings Seminars for the Plant Breeding Department in 2013	Jan 2013-2014
Symposium "All inclusive Breeding: Integrating highthroughput science"	Oct 2014
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*6.8 credits**
TOTAL NUMBER OF CREDIT POINTS*	**37.2**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*