Full Length Research Paper

'Omics' approaches in tomato aimed at identifying candidate genes for ascorbic acid accumulation in the fruit

Adriana Sacco, Valentino Ruggieri, Marcella Molisso and Amalia Barone*

Department of Agricultural Sciences, University of Napoli Federico II, Via Università 100, 80055 Portici, Italy.

Tomato (*Solanum lycopersicum*) is one of the most important vegetables in the world with significant importance for human health and nutrition. This species has long served as model system for plant genetics, development, physiology, pathology, and fleshy fruit ripening, resulting in the accumulation of many genetic and genomic resources. In addition, the tremendous development of high-throughput technologies, such as transcriptomics, metabolomics and proteomics, collectively denoted as 'omics' technologies, has led to a huge collection of data and platforms today available on the net. Nowadays, identifying all the components of a single biological system is within our means; however, assigning function to genes, proteins and metabolites is still a daunting task. Major challenges include interpretation and integration of large datasets to understand the principles underlying the regulation of genes, metabolites and proteins, and how their combined interactions associate with variation in phenotype. In this review, we will focus on the role of the different high-throughput technologies in enhancing tomato breeding particularly for fruit quality traits. We also describe how two 'omics' approaches could be combined in order to identify candidate genes for the genetic control of ascorbic acid accumulation in tomato fruit. We report the example of transcriptomic and genomic approaches established on the use of different high-throughput platforms available for tomato.

Key words: Tomato, introgression lines, quality trait, genomics, transcriptomics, candidate gene, single nucleotide polymorphism (SNPs).

INTRODUCTION

Nowadays, by taking the advantages of development of new sequencing technologies, the genome sequences of several plant species have been revealed. Among the almost 30 plant genomes publicly available, there is the tomato (*Solanum lycopersicum*) genome, whose sequencing started at the end of 2004 in the framework of the International Solanaceae Genome Project (SOL) and completed at the end of the year 2011 (Tomato Genome Consortium, 2012). The tomato is one of the three most important vegetables in the world with significant importance for human health and nutrition. In the last years, its global production has increased approximately 10% since for many countries it is a significant source of vitamins and minerals (Giovannucci et al., 2002). Moreover, tomato has been always considered as model species for fleshy fruit development and ripening, as well as for genomics studies of other Solanaceae (Mueller et al., 2005). For these reasons many genetic and genomic resources have been

*Corresponding author. E-mail: ambarone@unina.it. Tel: +39 081 2539491. Fax: +39 081 2531718.

developed for this species, including databases for transcriptomics, metabolomics and proteomics data, which are now available on the net (Fei et al., 2011). Indeed, in the post-genomic era, high-throughput technologies as microarray, mass-spectrometry and protein chip, have led to the collection of a large amount of data developed by the scientific community. These techniques allow measuring thousands of variables (genes, metabolites, proteins) simultaneously across populations.

The data generated by these techniques are often collectively denoted as 'omics' data (Joyce and Palsson, 2006). To understand the organization of cellular functions at different levels (gene, metabolite, or protein) and link them to a particular phenotype, an integrative approach is needed and is often referred to as 'systems biology' (Kitano, 2002, 2010). Biological systems are complex and cannot be understood by focusing on any one aspect of their highly interacting components. Because the functioning of a plant as a system concerns each of its molecular constituents (DNA, RNA, proteins, metabolites, ions, etc.), the expanding development of high-throughput data generation technologies made it possible to apply a systems biology paradigm in plant science. Large sets of comprehensive and quantitative data from plant samples grown under a wide variety of conditions have been produced. Massive databases from such high-throughput data have been created (Joyard and McCormick, 2010). The goal of systems biology is to understand how all these components function to bring about the observed phenotypes, and to elucidate the complete network of causes and effects from the molecule to the ecosystem. Identifying all the components of a single biological system is now within our means; however, assigning function to genes, proteins and metabolites is still a daunting task.

Major challenges include interpretation and integration of large datasets to understand the principles underlying the regulation of genes, metabolites and proteins, and how their combined interactions associate with variation in phenotype (Kim et al., 2010; Fukushima, 2009). Several attempts have been made to integrate multiple 'omics' data sets from different species. Even if we are still far from the initial objective of fully understanding how a given system works, is undeniable that the systematic analyses of the different 'omics' levels can facilitate the discovery of new candidate genes/QTLs and/or to assign functions to unknown proteins. Networks and pathways have been reconstructed using transcriptome, genome–wide transcription factor binding, proteome and metabolome data, and subsequently used to infer functional interactions among genes, proteins and metabolites (Moreno-Risueno et al., 2010). Moreover, systematic analyses of the transcriptome and metabolome and correlation of the expression pattern of genes with the accumulation pattern of metabolites have been successful ways to deduce the functions of genes.

The guilty-by-association principle states that a set of genes (or proteins and metabolites) involved in a certain biological process is generally co-regulated and thus co-expressed under the control of a shared regulatory system (Saito and Matsuda, 2010). Therefore, if an unknown gene is co-expressed with known genes of a particular biological process, researchers assume that this unknown gene may be involved in this process. This co-occurrence principle can be extended to metabolite co-accumulation relationships with the expression pattern of genes of the particular pathway in which the metabolite is involved.

Following the comprehension of the whole biological system under investigation, it is of fundamental importance to identify the hubs that regulates it, in order to focus on these key-elements that could be successfully transferred into new varieties by breeding schemes or genetic transformation. In this review, we focus on how different genome-wide datasets have been and can be used to reconstruct biological networks in tomato flesh fruit, and to dissect the QTLs that underlie their genetic control, reporting as example the approaches established on the combined use of different platforms available for tomato.

GENETIC AND GENOMIC TOMATO RESOURCES

Tomato has long served as a model system for plant genetics, development, physiology, pathology, and fleshy fruit ripening, resulting in the accumulation of substantial information regarding the biology of this economically important crop. Besides a large amount of already well-established genetic and genomic resources, today even more high-throughput datasets and different platforms have been generated. Among the tomato genetic resources, besides wild and related species publicly available at the Tomato Genetics Resource Center (TGRC) (http://tgrc.ucdavis.edu/index.cfm), different mutant collection (http://zamir.sgn.cornell.edu/mutants/ and http://tomatoma.nbrp.jp/) (Barone et al., 2009), and TILLING populations were developed in several countries (Minoia et al., 2010; Piron et al., 2010; Okabe et al., 2011). A powerful material to dissect genetic complex traits as quantitative trait loci (QTLs) is represented by the introgression line (IL) populations. These populations consist of a number of homozygous lines each containing marker defined segments from the wild genome in a uniform cultivated genetic background. They allow the same genetic stocks to be used worldwide in genetics and genomics applications for tomato breeding. Currently, different IL populations are available derived from wild tomato species such as *Solanum pennellii*, *Solanum habrochaites*, *Solanum pimpinellifolium*, *Solanum lycopersicoides*, *Solanum chmielewskii* and *Solanum sitiens* (Fernie et al., 2006). These lines have been widely used to localize QTLs on the molecular map, and to identify putative genes involved in their genetic control (Lippman et al., 2007). This has greatly helped the breeding work for these traits, which show a continuous variation and are strongly influenced by environmental conditions.

On the other hand, in addition to the recently tomato ge-

nome sequence (www.solgenomics.net, release SL2.40 January 2011), a large amount of genomic resources are now available. High-density genetic and physical maps, derived from interspecific crosses between *S. lycopersicum* and *S. pennellii*, *S. habrochaites*, *S. pimpinellifollium*, and other wild relatives have been developed (Foolad, 2007). Moreover, many EST collections (more than 330,000 EST deposited in the Tomato Gene Index database) deriving from various tomato species and tissues and different developmental stages are also publicly available. Different microarray platforms (TOM1, TOM2, combimatrix, affimetrix, and agilent) have been used to study the transcriptomic change in different tissues and at different environmental conditions (Alba et al., 2004, 2005; Di Matteo et al., 2010; Balaji et al., 2008; Lemaire-Chamley et al., 2005). A SolCAP chip (http://solcap.msu.edu) containing more than 8000 single nucleotide polymorphism (SNPs) has been made available for the tomato scientific community (Sim et al., 2012a). These SNPs were mainly discovered based on NGS-derived transcriptomic sequences obtained from six tomato accessions (Hamilton et al., 2012). Finally, very recently, a diversity array technology (DArT) platform for tomato using the *S. pennellii* ILs population has been developed and validated (Van Schalkwyk et al., 2012). These recent genomic resources add to other high-throughput genotyping platforms that are being used to explore the level of polymorphism detectable within cultivated tomato by genome-wide analysis (Sim et al., 2009, 2012b; Robbins et al., 2010; Shirasawa et al., 2010a).

Alongside of genomic resources, there are an increasing number of powerful computational pipelines for sequence analysis and genome annotation. SGN (Solanaceae Genome Network, http://solgenomics.net/) is a website that provides a virtual workbench for researchers working on the *Solanaceae* family, which hosts various sources of data and analysis tools. The Metabolome Tomato Database (MoTo DB) is an open-access metabolome database for tomato fruit. The database was developed using fruits from 96 different tomato cultivars in different ripening stages ensuring a representative fruit sample. The Tomato Functional Genomics Database (TFGD, http://ted.bti.cornell.edu/) provides a comprehensive resource to store, query mine, analyse, visualize and integrate large-scale tomato functional genomics data sets (Fei et al., 2011). A web-based system (plant MetGenMAP) has also been developed, which can com-prehensively integrate and analyze large-scale gene expression and metabolite profile data sets along with diverse biological information (Joung et al., 2009). Other web resources that collect data generated from different tomato 'omics' approaches are publicly available and reviewed in Yano et al. (2007) and Barone et al. (2008).

A transcriptomic-based approach

Thanks to the tremendous technical advances of the post-genomics era, data generation is no longer the limiting factor in advancing biological research. In addition, data integration, analysis, and interpretation have become key bottlenecks and challenges that biologists conducting genomic research face daily. In the last few years, many reports in tomato research focused on the possibility of understanding the complex of fruit ripening (Alba et al., 2005; Kok et al., 2008; Palma et al., 2011) and nutritional quality composition (Shauer et al., 2006; Moco et al., 2006, 2008) by using different 'omics' approaches. In our lab, we undertook two different 'omics' approaches to in-depth understand molecular mechanisms underlying tomato quality traits with particular attention to the ascorbic acid (AA) synthesis and accumulation in the fruits. The first approach consists on the screening of a *S. pennellii* IL population (genomic level), for the ascorbic acid content (metabolic level) to detect QTLs controlling its synthesis and accumulation into the tomato fruits, and on comparing the microarray analyses (transcriptomic level) of the ILs carrying the specific QTL. We identified two genotypes, IL12-4 and IL10-1, carrying a QTL for higher and lower AA accumulation compared to the control M82, respectively. Surprisingly, the transcriptomic analyses of these ILs revealed that the genes directly involved in the main metabolic pathway of AA synthesis and recycling / catabolism showed not differentially expression with respect to the cultivated parent. Indeed, the higher AA content in the IL12-4 was supposed to be controlled through the up-regulation of genes driving pectin degradation, thus releasing intermediates for the L-galactonic acid pathway, which is an alternative biosynthetic pathway for AA synthesis in plant (Di Matteo et al., 2010) previously detected only in strawberry (Agius et al., 2003) and grapevine (Cruz-Rus et al., 2010).

In the IL10-1, the different expression of genes involved in carbohydrate catabolism, fatty acid biosynthesis, glyoxylate metabolism and antioxidant system were involved in a reduced AA level of tomato ripe fruit (Di Matteo et al., 2012). Therefore, in both cases, the combined use of the ILs with the microarray platform, allowed the identification of new genes candidate to the control of AA level in the tomato fruit. The identified differentially expressed transcripts were mapped onto the tomato genome available at the Solanaceae Genomics Network. The genetic positions were obtained by BLASTN (Altschul et al., 1990) searches against the entire Tomato WSG Chromosomes (SL2. 40) database (http://solgenomics.net/index.pl). Among the 20 genes of the IL12-4 model (Di Matteo et al., 2010), three mapped on the introgression 12-4, whereas in case of IL10-1, among the 17 genes which establish the model (Di Matteo et al., 2012), eight mapped to the introgression 10-1. Therefore, we can state that the transcriptomic analysis led to dissect the two QTLs for AA content that map to chromosomal regions 10-1 and 12-4 into candidate genes, whose specific function will be further investigated by other 'omics' approaches, such as the TILLING or VIGS platforms today avail-
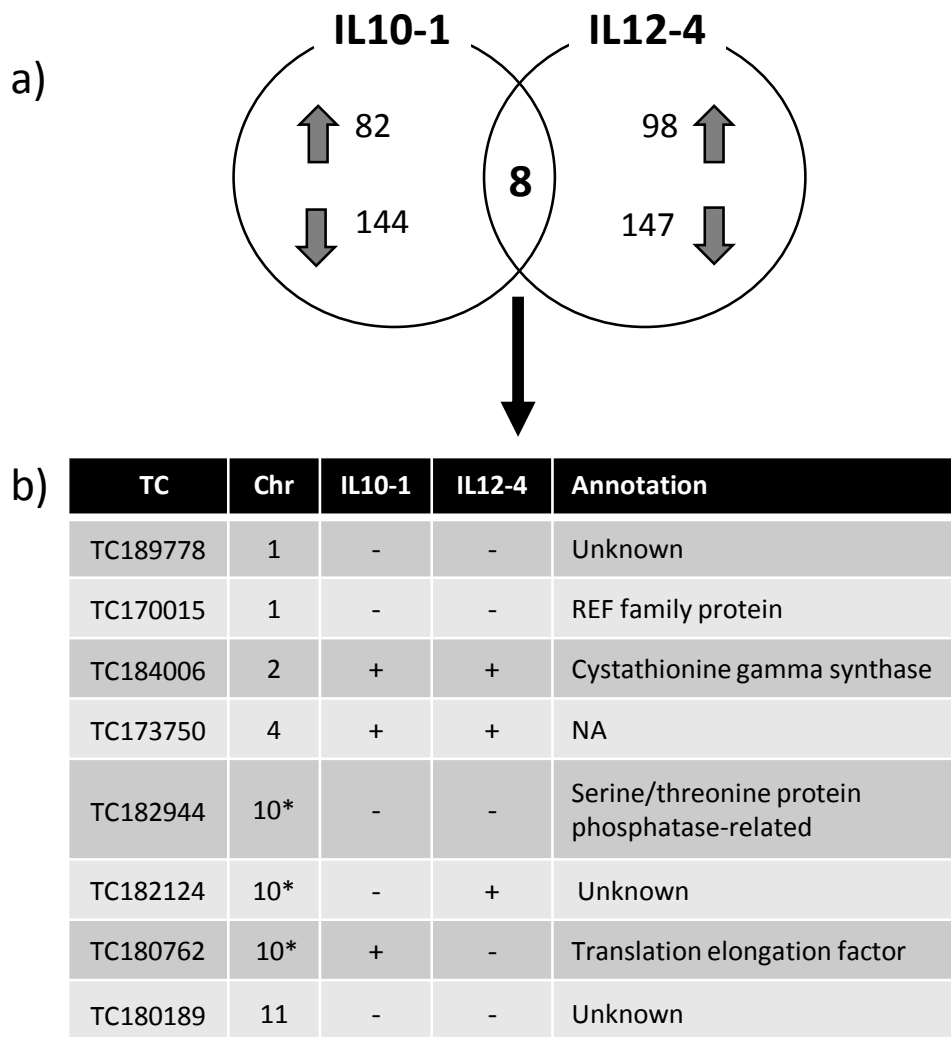
Figure 1. a) Venn diagram showing the shared differentially expressed genes between the *Solanum pennellii* introgression lines (ILs) 12-4 and 10-1; b) list of the eight shared genes between the ILs 12-4 and 10-1.

available for tomato (Minoia et al., 2010; Orzaez et al., 2009).

In addition, it is interesting to note that candidate genes for AA control in tomato fruits of ILs 10-1 and 12-4 have also been found outside of the introgression regions, giving us a more complete picture of the genes and molecular mechanisms controlling metabolic pathways and their interactions. For example, out of these genes, 10 mapped to chromosome 1, five per each IL; for some of these genes, it is possible to hypothesize an interaction of 'key-elements' mapping on chromosome 1 with genes mapping on the regions 10-1 and 12-4, in controlling AA synthesis and accumulation in tomato fruit. As a whole, comparing the differentially expressed transcripts obtained from the two different microarray experiments (Figure 1), we found that a set of 8 transcripts were shared between the IL12-4 and the IL10-1, and among them two (TC182124 and TC180762) mapping on the introgressed

region 10-1 also exhibit a contrasting expression pattern. One is annotated as a translation elongation factor p whereas the other is still not annotated. For the latter, a possible role as regulator of genes controlling AA level in tomato fruit could be hypothesized and further investigated in the future. These results highlight the powerful employment of the ILs in combination with the microarray approach.

Comprehensively, identified genes mapping within or outside the introgressed regions may represent key-control points in the mechanisms regulating the AA content in tomato fruit and so very useful in breeding program aimed to increasing nutritional quality in tomato fruits.

A genomic-based approach

The introgression lines as source of favorable alleles to transfer in the cultivated varieties have been so far used

Table 1. Characteristics of SNP assayed on the 96 tomato sample collection by the SolCAP genomic platform.

Missing value data point	Analyzed SNP (Number)	Segregating SNP (No.)		
		MAF = 0	MAF<10%	MAF>10%
MV = 0	6427	1410	3102	1915
MV<10%	1154	-	446	708
MV>10%	38	n.d.	n.d.	n.d.
Total	7619	1410	3548	2623

MV = Missing value. MAF=Minor frequency allele.

in tomato with the aim of identifying QTLs, both in conventional breeding (Lippman et al., 2007) and combined with 'omics' platforms (Schauer et al., 2006; Di Matteo et al., 2010, 2012). Despite these approaches allowed to identify many QTLs controlling very different traits, they allowed to explore reduced genetic backgrounds, generally limited to a few wild species (Fernie et al., 2006). By contrast, a wider source of genetic variation can be found among and within tomato breeding lines or cultivated varieties and ecotypes collected from different geographical regions. This is particularly true for traits important for adaptation to different environments. Most of this variation is of a quantitative nature and therefore requires specific genetic strategies for detecting QTLs, such as association mapping. In this case, large populations have to be phenotyped for the trait under study and genotyped through molecular markers uniformly distributed all over the genome, to perform a wide-genome approach of association mapping aimed at identifying new genes controlling the trait.

As an alternative, a candidate-gene approach could also be used, where variation at genomic level is specifically investigated for a number of genes already known to be involved in determining the phenotypic trait under study. In both cases, a high number of markers and/or genes already mapped on chromosomes are required. For tomato, a high-density map including different markers and genes is already available (Foolad, 2007; Shirasawa et al., 2010a, 2010b). Moreover, recently a genomic platform for SNPs detection has been built in the framework of the Solanaceae Coordinated Agricultural Project (SolCAP) from NIFA/USDA, based on the ILLUMINA Infinium Technology. The SolCAP tomato panel initially included around 8000 SNPs. These consist of Sanger-based eSNPs from genome sequences of two processing tomato lines (TA496 and Heinz 1706), besides those identified from ILLUMINA transcriptomic sequencing of three fresh-market lines, one processing line, one cherry tomato and one accession of the wild species *S. pimpinellifolium*. Therefore, depending on the germplasm assayed by the SolCAP genomic platform, various levels of polymorphism could be detected. A germplasm panel that consists of 489 accessions has been so far (September 27th, 2011) genotyped in the framework of the SolCAP activities (Sim et al., 2012b).

This panel includes 141 accessions for processing, 122 accessions for fresh-market, 88 vintage tomatoes, 103 accessions belonging to various wild species and 35 accessions of miscellaneous materials (hybrids, F1 etc).

Data from this experiment are available for the scientific community (http://solcap.msu.edu) in order to compare the level of polymorphism detected in this population representative of genetic variability available among tomatoes with that highlighted in other specific tomato collections belonging to specific institutions, as already reported by Hamilton et al. (2012). In our laboratory, an association mapping approach by candidate gene has been undertaken with the aim of identifying among 96 different genotypes new alleles in genes that could increase the level of antioxidants in the fresh and processed fruit. In this context, a collection of *S. lycopersicum* accessions is being investigated for fruit quality traits in a 2-years trial in order to measure physiological and metabolic characters correlated with antioxidant synthesis and storage. In the meanwhile, the collection has been genotyped using the SolCAP platform for high-throughput genomic analysis. The collection under study mainly consists of Italian ecotypes (39 accessions), Latin American cultivars (29 accessions), some vintage cultivars and modern varieties coming from different geographical regions (that is, from Spain, China, Africa, USA). The variability exhibited by this collection, as evidenced by SNP analysis on the SolCAP genomic platform, resem-bles in the percentage that highlighted by the SolCAP experiment on the panel of 489 genotypes. Indeed, most of SNPs analyzed (98.2%, Table 1) had a missing value <10%, that means they lacked in less than 86 genotypes out of 96, and this value is similar to the 97.6% observed in the larger tomato population of 489 genotypes. The 38 SNPs that showed a missing value (MV, Table 1) higher that 10% were excluded by subsequent analyses.

Among others, a threshold value of minor frequency allele (MAF) of 10% was established. Overall, 1410 SNPs out of 7619 (18.5%) did not segregate among the 96 genotypes, since they only exhibited a single allele in all genotypes and no minor alleles were detected (MAF = 0). The distribution of all 7619 SNPs on the 12 tomato chromosomes was obtained by physically mapping them on the macromolecules obtained from the complete sequenced tomato genome (version 2.40 available at the

Table 2. Distribution over the 12 tomato chromosomes of SNP (with MV = 0 and MAF>10%) suitable for association mapping studies.

Allelic frequency	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	Total
10%<AF<15%	28	12	43	30	179	13	11	15	9	3	66	7	416
15%<AF<20%	10	28	39	179	207	27	12	14	8	3	111	19	657
20%<AF<25%	16	18	20	31	4	3	8	12	36	3	26	2	179
25%<AF<30%	22	25	33	19	5	9	5	9	20	9	65	8	229
30%<AF<35%	6	15	17	9	7	10	4	8	13	14	3	5	111
35%<AF<40%	7	7	6	20	6	2	10	15	10	7	5	8	103
40%<AF<45%	11	6	28	5	8	2	3	7	12	3	6	11	102
45%<AF<50%	30	4	9	8	5	2	9	10	9	1	4	14	105
Total	130	115	195	301	421	68	62	90	117	43	286	74	1902

AF = Allelic frequency; AF = 10% includes segregation ratios = 86:10; AF = 50% includes segregation ratios = 48:48; MV = missing value; MAF = minor frequence allele.
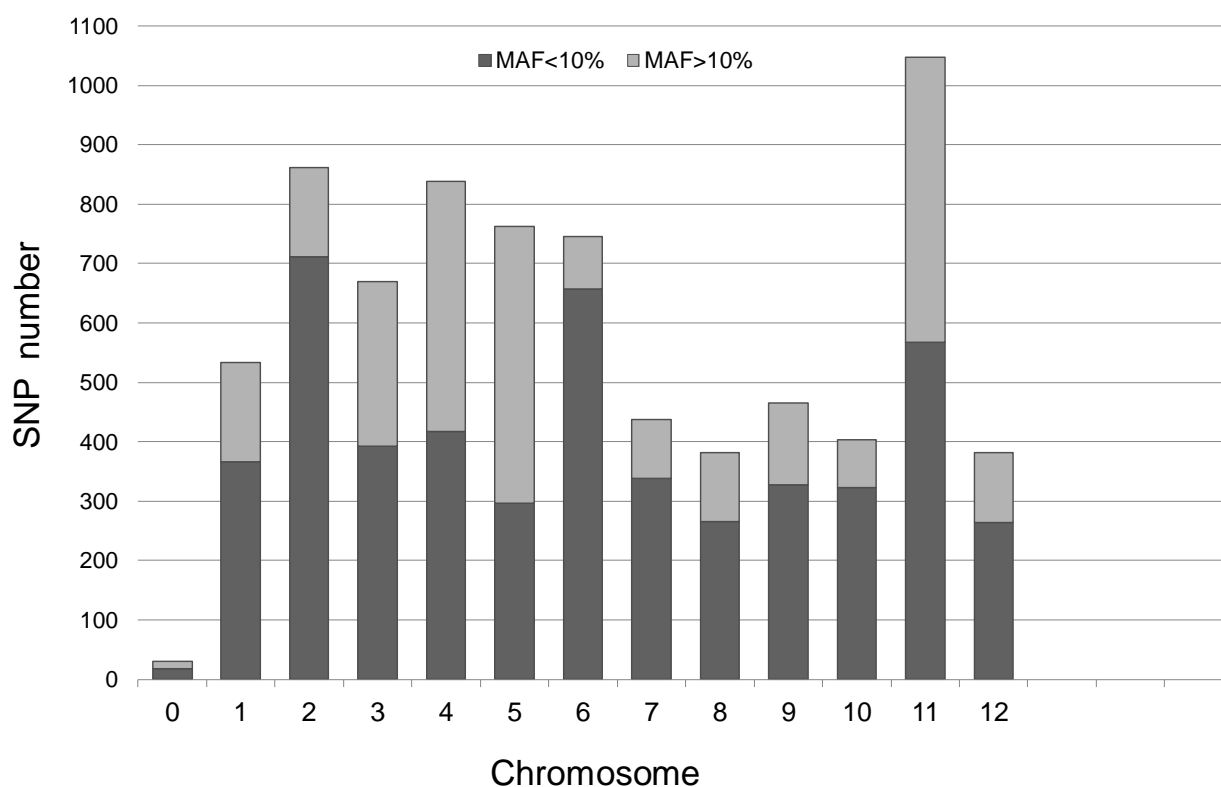
Figure 2. Distribution of 7619 SNPs on the 12 tomato chromosomes. MAF = Minor frequency allele.

http://solgenomics.net). The highest number of SNPs map on chromosome 11 (Figure 2), with an equal distribution between SNPs with MAF>10% and < 10%. A lower number of SNPs map on chromosomes from 7 to 12, whereas this number is higher for chromosomes from 1 to 6. In most cases, except than chromosomes 5 and 11, on each chromosome SNPs with MAF<10% are prevalent. Since many SNPs (3548 corresponding to 46.6%) fall within the group with MAF<10%, they should be mostly considered rare alleles (present only in one or two geno-types), and not useful for association mapping analysis. This high frequency of rare alleles, estimated considering the threshold value of 10%, confirms the one evidenced by Labate et al. (2009) who analyzed 30 tomato land-races by different types of markers. In our collection, the 2623 SNPs showing MV<10% and MAF>10% were the only suitable for the linkage disequilibrium analysis re-quired to associate gene polymorphism with phenotype variation. They were located on the tomato molecular map (Figure 3) and it is evident that some chromosomes are better covered and that markers mainly cluster in specific areas of each chromosome, generally in distal ones. This would suggest extending our analysis to a greater number of genotypes and of markers, even though
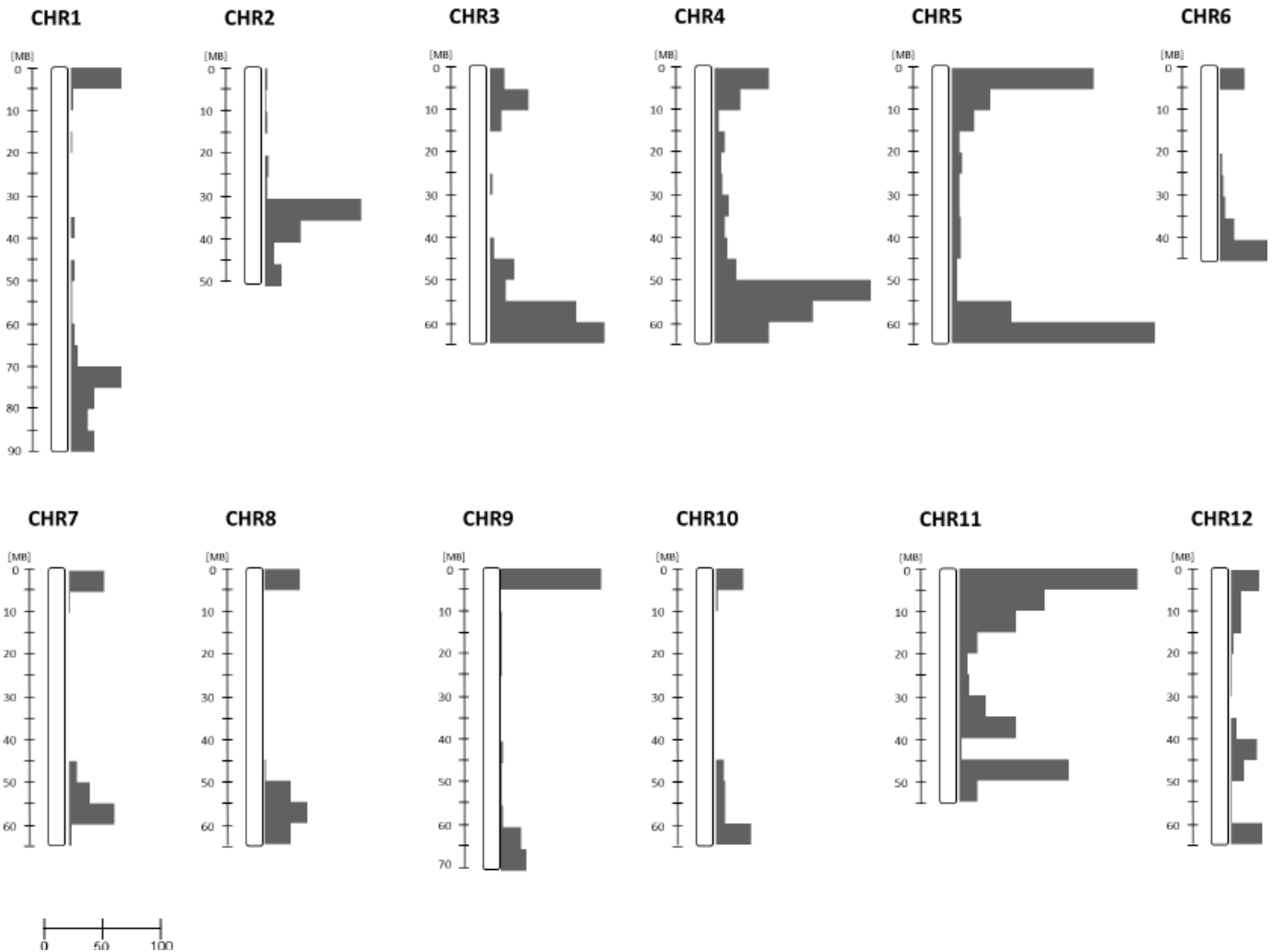
Figure 3. Mapping on tomato chromosomes of 2623 SNPs showing MV ≤ 10% and MAF > 10%. MV = missing value; MAF = minor allele frequency.

preliminary information could be derived for realizing a genome-wide association mapping approach.

Indeed, in order to explore the level of variability available in our collection, the segregation ratios observed for 1902 markers with MV = 0 and MAF>10%, distributed for each chromosome, are reported in Table 2, excluding those that still map on chromosome 0 or are not assigned to any chromosome (http://Solgenomics.net). Segregation ratios were grouped in eight different intervals, ranging from 10% (86:10) to 50% (48:48) of the allelic frequencies. Each interval includes five segregation ratios corresponding approximately to 5% of the allelic frequencies. Two intervals (10 to 15% and 15 to 20%) include about 50% of the markers (1073 SNPs of 1902) while the other ones include about 100 SNPs, except than the interval 25 to 30% (229 SNPs). These segregation ratios suggest that a higher variability is evident in our tomato collection compared to that exhibited by the collection of 30 accessions analyzed by Labate et al. (2009). More-

over, distribution of the 1902 SNP markers with MAF >10% on chromosomes shows that most of all map on chromosome 5 (421 SNPs) but also on chromosomes 4 (301 SNPs) and 11 (286 SNPs) while only 43 SNPs map on chromosome 10.

Consequently, in order to find new genes associated to AA variation by a genome-wide AM approach, genotypes should be added to our collection with the aim of further increasing the genetic variability to be explored. Finally, as for the candidate gene AM approach, based on the gene annotation actually available for the sequenced tomato genome, among the 7720 SNPs represented on the SolCAP platform, 28 genes belong to the metabolic pathways that lead to AA biosynthesis and accumulation (Ioannidi et al., 2009; Zou et al., 2006).

Among these, 25 segregate in our population of 96 genotypes, 17 as minor alleles (MAF<10%) whereas the other eight show a high level of variability and map on chromosomes 3, 4, 5 and 10. These polymorphisms will
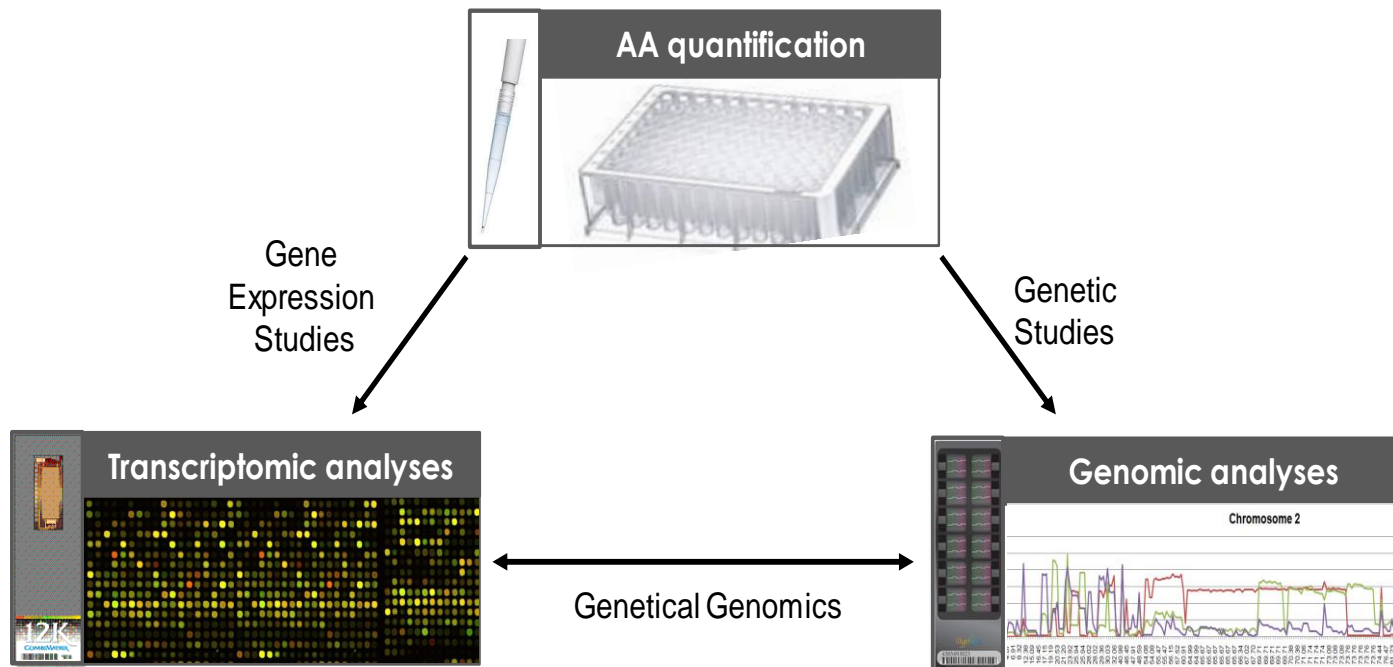
Figure 4. Integration of 'omics' data to improve and accelerate the production of new elite lines: combined transcriptomic and genomic analyses of different tomato lines will facilitate the identification of candidate genes for the trait of interest and will enhanced tomato breeding efficiency.

be further investigated referring to their potential association with different levels of AA content in tomato fruit. In the future, new alleles will also be searched for those candidate genes involved in AA level regulation that were identified by the transcriptomic approach here described. This high integration among 'omics' platforms will be extremely powerful in detecting genes or alleles up until now unexplored for the improvement of the nutritional quality of tomato fruit. It could be also successfully applied to other traits under the complex control of many genes/QTLs, thus allowing to identify genes/hubs and to realize breeding by precision approach in tomato (Collard and Mackill, 2008).

CONCLUSION

Our comprehension of complex molecular networks that underlie biological processes has grown dramatically in the last few years. In this review, we have highlighted the use of some 'omics' platforms and tools today available for tomato researches to better understand molecular mechanisms controlling fruit AA synthesis and storage. Our goal was to unveil new genes and their relations (transcriptomic-based approach) or new alleles (genomic-based approach) involved in modulating AA accumulation in tomato fruit in order to manipulate them for enhancing fruit antioxidant content. Comprehensively, this would lead to an approach of genetical genomics, as proposed by Kirst and You (2007), which incorporates the transcription level information at integration of those from geno-

typing and phenotyping to identify candidate genes for complex traits (Figure 4). The application of post-genomics tools should accelerate the selection process and the combined use of different 'omics' strategies and will considerably shorten the time required for the production of elite lines. Indeed, as genome sequencing becomes less costly and the development of the most recently technique, such as RNAseq (Wang et al., 2009), protein–DNA binding microarrays (Badis et al., 2009) and genome-wide profiling of histone modifications and DNA methylation (Lister et al., 2008; Zhang et al., 2009) is increasing, the comprehension of complex biological phenomena will certainly improve.

ACKNOWLEDGEMENTS

REFERENCES

Agius F, Gonza´lez-Lamothe R, Caballero JL, Mun˜oz-Blanco J, Botella MA, Valpuesta V (2003). Engineering increased vitamin C levels in plants by overexpressing of a D-galacturonic acid reductase. Nat. Biotechnol. 21:177–181.
Alba R, Fei Z, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose, JKC, Martin G, Tanksley SD, Bouzayen M, Jahn MM, Giovannoni J (2004). ESTs, cDNA microarrays, and gene expression profiling: Tools for dissecting plant physiology and development. Plant J. 39:697–714.

Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni JJ (2005). Transcriptome and Selected Metabolite Analyses Reveal Multiple Points of Ethylene Control during Tomato Fruit Development. Plant Cell. 17(11):2954–2965.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, Kuznetsov H, Wang CF, Coburn D, Newburger DE, Morris Q, Hughes TR, Bulyk ML (2009). Diversity and complexity in DNA recognition by transcription factors. Science. 324:1720–1723.

Balaji V, Mayrose M, Sherf O, Jacob-Hirsch J, Eichenlaub R, Iraki N, Manulis-Sasson S, Rechavi G, Barash I, Sessa G (2008). Tomato Transcriptional Changes in Response toClavibacter michiganensis subsp michiganensis Reveal a Role for Ethylene in Disease Development. Plant Physiol. 146:1797–1809.

Barone A, Chiusano ML, Ercolano MR, Giuliano G, Grandillo S, Frusciante L (2008). Structural and functional genomics of tomato. Int. J. Plant Genomics. doi:10 1155/2008/820274.

Barone A, Di Matteo A, Sacco A, Carputo D, Frusciante L (2012). High-Troughput Genomics Enhances Tomato Breeding Effieciency Current Genomics 10: 1-9.

Collard BCY, Mackill DJ (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil .Trans. R. Soc B. 363:557–572.

Cruz-Rus E, Botella MA, Valpuesta V, Gomez-Jimenez MC (2010). Analysis of genes involved in L-ascorbic acid biosynthesis during growth and ripening of grape berries. J. Plant Physiol. 167:739–748.

Di Matteo A, Sacco A, Anacleria M, Pezzotti M, Delledonne M, Ferrarini A, Frusciante L, Barone A (2010). The ascorbic acid content of tomato fruits is associated with the expression of genes involved in pectin degradation. BMC Plant Biol. 10:163.

Di Matteo A, Sacco A, De Stefano R, Frusciante L, Barone A (2012). Comparative transcriptomic profiling of two tomato lines with different ascorbate content in the fruit. Biochem. Genet. 50: 908-921.

Fei Z, Joung JG, Tang X, Zheng Y, Huang M, Lee JM, McQuinn R, Tieman DM, Alba R, Klee HJ, Giovannoni JJ (2011). Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. Nucleic Acids Res. 39:1–8.

Fernie AR, Tadmore Y, Zamir D (2006). Natural genetic variation for improving crop quality. Curr. Opin. Plant Biol. 9:196–202.

Foolad MR (2007). Genome mapping and molecular breeding of tomato. Int. J. Plant Genomics. Article ID 64358 52 pages doi:10 1155/2007/64358.

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2009). Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach. Curr. Opin. Chem. Biol. 13:532–538.

Giovannucci E, Rimm EB, Liu Y, Stampfer MJ, Willett WC (2002). A prospective study of tomato products lycopene and prostate cancer risk. J. Natl. Cancer Inst. 94:391–398.

Hamilton JP, Sim SC, Stoffel K, Van Deynze A, Buell CR, Francis DM (2012). Single Nucleotide Polymorphism Discovery in Cultivated Tomato via Sequencing by Synthesis. The Plant Genome. 5:17–29.

Ioannidi E, Kalamaki MS, Engineer C, Pateraki I, Alexandrou D, Mellidou I, Giovannoni J, Kanellis AK (2009). Expression profiling of ascorbic acid-related genes during tomato fruit development and ripening and in response to stress conditions. J. Exp. Bot. 60:663–678.

Joung JG, Corbett AM, Moore Fellman S, Tieman DM, Klee HJ, Giovannoni JJ, Fei Z (2009). Plant MetGenMAP: An Integrative Analysis System for Plant Systems Biology. Plant Physiol. 151:1758–1768.

Joyard J, McCormick S, Guest Editors (2010). Plant Systems Biology. Plant Physiol. 152:401.

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

Kim TY, Kim HU, Lee SY (2010). Data integration and analysis of biological networks. Curr. Opin. Biotechnol. 21:78–84.

Kirst M, Yu Q (2007). Genetical genomics: successes and prospects in plants in: Varshney RK, Tuberosa R, (Eds) Genomics-assisted crop improvement. Volume 1 Genomics Approaches and Platforms. Springer Dordrecht. pp.245-266.

Kitano H (2002). Systems Biology: A Brief Overview. Science. 295: 1662 –1664.

Kitano H (2010). Grand challenges in systems physiology. Front. Physiol. 1:3.

Kok EJ, Lehesranta SJ, van Dijk JP, Helsdingen JR, Dijksma WTP, Van Hoef AMA, Koistinen KM, Karenlampi SO, Kuioer HA, Keijer J (2008). Change in gene and protein expression during tomato ripening - consequences for the safety assessment of new crop plant varietis. Food Sci. Technol. Internatl. 14: 503-518.

Labate JA, Robertson LD, Baldo AM (2009). Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (Solanum lycopersicum L.). Heredity. 103:257–267.

Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C (2005). Changes in Transcriptional Profiles Are Associated with Early Fruit Tissue Specialization in Tomato. Plant Physiol. 139:750–769.

Lippman ZB, Semel Y, Zamir D (2007). An integrated view of quantitative trait variation using tomato interspecific introgression lines. Curr. Opin. Genet. Dev. 17:545–552.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 133:523–536.

Minoia S, Petrozza A, D'Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bandahmane A, Carriero F (2010). A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC Research Notes. 3:69.

Moco S, Bino RJ, Vorst O, Verhoeven HA, de Groot J, van Beek TA, Vervoort J, de Vos CHR (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141:1205–1218.

Moco S, Forshed J, De Vos RCH, Bino RJ, Vervoort J (2008). Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography-mass spectrometry and nuclear magnetic resonance. Metabolomics. 4:202–215.

Moreno-Risueno MA, Busch W, Benfey PN (2010). Omics meet networks using systems approaches to infer regulatory networks in plants. Curr. Opin. Plant Biol. 13:126–131.

Mueller LA, Solow T, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright M, Ahrens R, Wang Y, Herbst E, Keyder E, Menda N, Zamir D, Tanksley SD (2005). The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond. Plant Physiol. 138:1310-1317.

Okabe Y, Asamizu E, Saito T, Matsukura C, Ariizumi T, Bres C, Rothan C, Mizoguchi T, Ezura H (2011). Tomato TILLING technology: development of a reverse genetics tool for the efficient isolation of mutants from Micro-Tom mutant libraries. Plant & Cell Physiol. 52:1994–2005.

Orzaez D, Medina A, Torre S, Fernández-Moreno JP, Rambla JL, Fernández-Del-Carmen A, Butelli E, Martin C, Granell A (2009). A visual reporter system for virus-induced gene silencing in tomato fruit based on anthocyanin accumulation. Plant Physiol. 150:1122–1134.

Palma JM, Corpas FJ, del Río LA (2011). Proteomics as an approach to the understanding of the molecular physiology of fruit development and ripening. J. Proteomics. 74:1230–1243.

Piron F, Nicolai M, Minoia S, Piednoir E, Moretti A, Salgues A, Zamir D, Caranta C, Bendahmane A (2010). An induced mutation in tomato eIF4E leads to immunity to two potyviruses. PLoS ONE. 5:e11313.

Robbins MD, Sim SC, Yang W, van Deynze A, van der Knapp S, Joobeur T, Francis DM (2010). Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. J. Exp. Bot. 62:1831–1845.

Saito K, Matsuda F (2010). Metabolomics for Functional Genomics Systems Biology and Biotechnology. Annu. Rev. Plant Bio. 61:463–489.

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. 24:447–454.

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 (2010a). SNP discovery and linkage map construction in cultivated tomato. DNA Res. 17:381–

391.

Shirasawa K, Asamizu E, Fukuoka H, Ohyama A, Sato S, Nakamura Y, Tabata S, Sasamoto S, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Isobe S (2010b). An interspecific linkage map of SSR and intronic polymorphism markers in tomato. Theor. Appl. Genet. 121:731-739.

Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM (2009). Oligonucleotide array discovery of polymorphisms in cultivated tomato (Solanum lycopersicum L.) reveals patterns of SNP variation associated with breeding. BMC Genomics. 10:466.

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM (2012a). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS One. 7(7): e40563.

Sim SC, Van Deynze A, Stoffel K, Ganal M, Surstewitz G, Douches D, Zarka D, Hamilton J, Robin Buell C, Francis DM (2012b). Genome-wide genotyping of cultivated tomato (Solanum lycopersicum L.) using a high-density SNP array for breeding and genetics. XX Plant and Animal Genome Conference. San Diego CA. January 15-18.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 485(7400):635–641.

Van Schalkwyk A, Wenzl P, Lopez-Cobollo SSR, Kilian A, Bishop G, Hefer C, Berger DK (2012). Bin mapping of tomato diversity array (DArT) markers to genomic regions of Solanum lycopersicum X Solanum pennellii introgression lines. Theor. Appl. Genet. 124:947–956.

Wang Z, Gerstein M, Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

Yano K, Aoki K, Shibata D (2007). Genomic databases for tomato. Plant Biotechnol. 24:17–25.

Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009). Genome-wide analysis of mono-di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. 10:R62.

Zou L, Li H, Ouyang B, Zhang J, Ye Z (2006). Cloning and mapping of genes involved in tomato ascorbic acid biosynthesis and metabolism. Plant Sci. 170:120–127.