Plant Biotechnology 30, 257–263 (2013)
DOI: 10.5511/plantbiotechnology.13.0707a

Review

Structural analyses of the tomato genome

Shusei Sato, Kenta Shirasawa, Satoshi Tabata*

Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan
*E-mail: tabata@kazusa.or.jp Tel: +81–438-52-3933 Fax: 81–483-52-3934

Received May 28, 2013; accepted July 7, 2013 (Edited by K. Aoki)

Abstract Over the last 10 years, considerable efforts have been undertaken to develop genomic resources in tomato, including genomic clones, physical maps, DNA markers, mapping populations, and genetic linkage maps. Such resources facilitate the investigation of genome structure and gene functions, and the identification of genes of agronomic importance. In parallel, an international project with the participation of over 90 groups from 14 countries has been utilizing these resources to proceed with the deciphering of all of the genetic information carried by tomato. With the aid of new sequencing technologies and sophisticated bioinformatics, sequencing of the whole genome of tomato was successfully completed and the results were published in May 2012. The resulting large number of DNA markers, high-density linkage maps, and information on the structure and function of almost all of the gene components in the tomato genome are expected to contribute to a wide variety of biological fields by accelerating the processes of identification, isolation, and functional assignment of genes of interest, understanding of the evolutional process of Solanaceae and other plants, and breeding of new varieties with better agronomic traits.

Key words: *Solanum lycopersicum*, tomato, DNA marker, high-density genetic linkage maps, genome sequencing.

It is no exaggeration to say that whole genome sequence information is a prerequisite for modern genetics, genomics, physiology, breeding, and other advanced research fields in biology. The recent progress in DNA sequencing technology, especially new generation sequencers (NGSs) and bioinformatics, has allowed us to obtain sequence information for the entire genomes of complex organisms in a short period of time (Lister et al. 2009). Since the attainment of a high-accuracy sequence of the whole genome of *Arabidopsis thaliana* in 2000 (Arabidopsis Genome Initiative 2000), nucleotide sequences that cover the entirety or substantial portions of genomes have been published for a number of plant species including rice (International Rice Genome Sequencing Project 2005), poplar (Tuskan et al. 2006), grapevine (The French–Italian Public Consortium for Grapevine Genome Characterization 2007), *Lotus japonicus* (Sato et al. 2008), and more. Meanwhile, draft sequences of the genomes, which are cost-effective but less accurate, have recently been accepted as an information source that is sufficient for most purposes in molecular genetics (Ming et al. 2008). Therefore, draft sequencing of the genomes of a variety of model and crop plants using NGSs is in progress.

Tomato (*Solanum lycopersicum*) belongs to the family Solanaceae, which consists of approximately 100 genera and 2,500 species. Among various Solanaceae plants, including several plants of agronomic importance such as potato, eggplant, pepper, and tobacco, tomato was chosen as a target for genome sequencing because it had been the most intensively genetically characterized. *S. lycopersicum* has a diploid genome ($n=12$) of rather simple architecture, approximately 950 Mb in size. One-quarter of the genome is presumed to be made up of gene-rich euchromatic regions (Peterson et al. 1996).

In 2003, the international Tomato Sequencing Project was launched by members from 10 countries, with an aim to sequence the gene-rich regions of 12 chromosomes through high-quality sequencing of bacterial artificial chromosomes (BAC) chosen based on DNA markers mapped on the genome (Mueller et al. 2005, The tomato genome sequencing consortium 2005, and http://sgn.cornell.edu/). Later, a whole genome sequencing strategy was also adopted, to cover the entire genome. In parallel with genome sequencing, the development of DNA markers and construction of high-density genetic linkage maps were conducted to assist in the reconstruction of the whole genome structure by mapping (Frary et al. 2005; Fulton et al. 2002). The chloroplast and mitochondrial genomes were sequenced independently from the nuclear genome (Kahlau et al. 2006). The obtained sequences were subjected to assembly and information analyses were performed by the International Tomato Annotation Group (ITAG). Ultimately, a total of 14 countries have contributed to this project, and the results were published in 2012

This article can be found at http://www.jspcmb.jp/
Published online September 12, 2013

(The Tomato Genome Consortium 2012). This review summarizes the recent studies on the genome structure of tomato, focusing on high-density genetic linkage maps and sequence information for the whole genome.

Development of DNA markers and construction of genetic linkage maps

Interspecific maps

Genetic linkage maps are useful tools for molecular genetics and genomics because they facilitate gene isolation by map-based cloning strategies, comparative genomics, and anchoring genome sequences to chromosomes, among others. The genetic linkage map is established by linkage analysis of DNA markers using a mapping population, e.g., F_2 population or recombinant inbred lines (RILs). Because the frequency of DNA polymorphism, as detected by the presence or absence of DNA markers in the parental lines of the mapping population, is a key factor determining density of the DNA markers on the resultant genetic linkage map, distantly related lines have usually been selected to construct genetic linkage maps that are saturated with DNA markers.

The domestication of tomato and generation of modern varieties has resulted in population bottlenecks that have led to low genetic diversity in cultivated tomato (Rick et al. 1976). The primary genetic linkage map was therefore constructed using an F_2 population derived from an interspecific cross between *S. lycopersicum* 'LA1500' and *S. pennellii* 'LA716' (Bernatzky et al. 1986). For this map, restriction fragment length polymorphism (RFLP) markers were employed. Since then, RFLP markers have been applied to construct more than 20 interspecific genetic linkage maps in tomato (reviewed in Foolad 2007; Labate et al. 2007; Shirasawa et al. 2013), and the number of RFLP markers has exceeded 1,000 (Tanksley et al. 1992). Subsequently, several types of DNA markers have been developed, e.g., random amplified polymorphic DNA (RAPD), simple sequence repeat (SSR) or microsatellite, amplified fragment length polymorphism (AFLP), cleaved amplified polymorphic sequence (CAPS), and single nucleotide polymorphism (SNP) markers. A total of 1,088 CAPS, 1,342 RFLP, 155 SSR, and 19 SNP markers have been mapped onto an interspecific genetic linkage map based on a cross between *S. lycopersicum* 'LA925' and *S. pennellii* 'LA716'; the Tomato-EXPEN 2000 map (Frary et al. 2005; Fulton et al. 2002; Shirasawa et al. 2010a).

Along with advances in genomic studies in tomato, large amounts of sequence information, e.g., >200,000 expressed sequence tags (ESTs) and approximately 90,000 BAC-ends, have been released for tomato species from the SOL Genomics Network (SGN), which is recognized as the major research community for molecular genetics and genomics in Solanaceae (Mueller et al. 2005). SSR markers can be rapidly and easily developed from this sequence information by SSR-motif searches and primer designs based on their flanking sequences. These markers have several advantages over other types of DNA markers, including multi-allele detection ability, high transferability across species, and sufficient flexibility so that they can be used with various laboratory systems (Kalia et al. 2011). More than 20,000 SSR markers have been therefore developed, from EST and BAC-end sequences (Ohyama et al. 2009; Shirasawa et al. 2010a). These efforts have led to the mapping of 2,116 DNA markers including 1,282 SSR and 151 SNP markers onto the Tomato-EXPEN 2000 map, covering 1,503.1 cM (Shirasawa et al. 2010a). The linkage map is registered as the Kazusa F2-2000 genetic linkage map on the SGN website (http://solgenomics.net) and has been used for the genome sequencing project (details are described in the following section 'Genome sequencing').

Since the release of the tomato genome sequence (The Tomato Genome Consortium 2012), SNPs, which are the most abundant polymorphisms in the genome, have been discovered by a re-sequencing strategy using NGSs. In the re-sequencing strategy, the sequence reads obtained from the whole genome, from complexity-reduced genomes such as restriction-site associated DNA (RAD), or from transcribed sequences are mapped onto the reference genome or unigenes. In tomato, 62,576 nonredundant putative SNPs were identified from whole-transcriptome sequencing of six accessions that span cultivated market classes (Hamilton et al. 2012). Of these SNPs, 7,720 were selected for the genotyping of three interspecific mapping populations, i.e., the Tomato-EXPEN 2000 (*S. lycopersicum* 'LA925' and *S. pennellii* 'LA716'), Tomato-EXPEN 2012 (*S. lycopersicum* 'Moneymaker' and *S. pennellii* 'LA716'), and Tomato-EXPIM 2012 (*S. lycopersicum* 'Moneymaker' and *S. pimpinellifolium* 'LA121'), and 3,503, 3,687, and 4,491 SNPs were mapped on each genetic linkage map, respectively (Sim et al. 2012a). On the other hand, the diversity arrays technology (DArT; Jaccoud et al. 2001) platform for 990 loci has dissected the recombination points of the map based on introgression lines (ILs) consisting of *S. pennellii* 'LA716' introgressed into *S. lycopersicum* 'M82' (Van Schalkwyk et al. 2012), the population of which was originally developed by marker-assisted selection with 350 RFLP markers (Eshed et al. 1994).

Intraspecific maps

In contrast to the great advances in the interspecific genetic linkage maps, intraspecific maps, which are generated from crosses within *S. lycopersicum*, have progressed much more slowly, due to the low genetic diversity in cultivated tomato. Intraspecific maps are

considered more useful than interspecific ones for breeding and genetics, in which quantitative trait loci (QTLs) for agronomically important traits, e.g., fruit quality and yield, are targeted. The first intraspecific map was constructed with 132 RFLP, 33 RAPD, and 211 AFLP markers using RILs derived from a cross between two inbred lines of *S. lycopersicum*, 'Levovil' and var. *cerasiforme* 'Cervil' (Saliba-Colombani et al. 2000).

SNPs, the most abundant source of genomic variation, are the most promising source of polymorphisms for closely related intraspecific lines. SNPs have therefore been identified from the EST sequences of several tomato cultivars (Aoki et al. 2010; Yamamoto et al. 2005). Two mapping populations, derived from crosses between *S. lycopersicum* 'Micro-Tom' and either *S. lycopersicum* 'Ailsa Craig' or *S. lycopersicum* 'M82,' were subjected to linkage analysis of their SNP markers. A total of 1,137 markers, including 793 SNPs, along with 344 SSR and intronic polymorphism markers, were mapped onto two genetic linkage maps, which covered 1,467.8 and 1,422.7 cM, respectively (Shirasawa et al. 2010b). Comparative analysis of these two intraspecific maps with the Tomato-EXPEN 2000 indicates that the intraspecific maps cover the whole tomato genome, and that the marker order is mostly conserved in the three maps (Shirasawa et al. 2010b). Because 'Micro-Tom,' a miniature dwarf cultivar originally bred for home gardening purposes (Scott et al. 1989), is regarded as a model tomato line (Meissner et al. 1997), various genomic and genetic resources have been developed from this cultivar (Ariizumi et al. 2011). These intraspecific genetic linkage maps will also support genomic and genetic studies in tomato in general.

The tomato genome sequence aids in the identification of the positions of SNPs obtained from re-sequencing data for tomato lines. The 62,576 nonredundant putative SNPs reported by Hamilton et al. (2012) will be a source of SNPs for cultivated tomato lines of *S. lycopersicum* (Sim et al. 2012b). In addition, 171,792 SNP, insertion, and deletion candidates were also identified from the mapping analysis of 'Micro-Tom' BAC-end sequences compared with the reference genome of *S. lycopersicum* 'Heinz 1706' (Asamizu et al. 2012). The positions of the polymorphisms on the genetic linkage maps can be speculated from the relationship between the genetic and physical positions within each chromosome (The Tomato Genome Consortium 2012; Sim et al. 2012a). Therefore, the minimum subset of SNP markers can be easily selected for construction of a genetic linkage map covering both whole genome and local regions for attractive QTLs, and even for an intraspecific genetic linkage map.

Genome sequencing

International project for tomato genome sequencing

The tomato genome, estimated to be 950 Mb long, is composed of pericentromeric heterochromatin and distal euchromatin, rich in repetitive sequences and genes, respectively. The genome of the inbred tomato cultivar 'Heinz 1706' was sequenced using a combination of the conventional Sanger method and next-generation sequencing methods. A total of 21 Gb of Roche/454 Titanium shotgun and mate-pair reads, and 3.3 Gb of Sanger paired-end reads, including ~200,000 BAC and fosmid paired-end sequences, were assembled using both Newbler and CABOG (Celera Assembler with the Best Overlap Graph) and then integrated into a single assembly. The structural correctness of the de novo assembly was confirmed by mapping paired-end sequences of the BAC and fosmid clones. By using Illumina GA (82 Gb) and SOLiD (Single Responsibility, Open-Closed, Liskov Substitution, Interface Segregation and Dependency Inversion) (112 Gb) reads, base accuracy was improved to less than one substitution error per 29.4 kb and one indel error per 6.4 kb. Contig gaps were filled by integrating 117 Mb of BAC clone sequences completed by the Sanger method, and the resulting high-quality scaffolds were linked with two BAC-based physical maps and anchored using a high-density genetic map (Shirasawa et al. 2010a), introgression line mapping, and genome-wide BAC fluorescence in situ hybridization (FISH). The genome size predicted by both assemblers was ~900 Mb. Of these, ~760 Mb was assembled into 91 scaffolds aligned on the 12 tomato chromosomes, with most gaps restricted to pericentromeric regions. The 21 Mb of sequence contained in 3,132 unanchored scaffolds was designated as chr0, which is composed primarily of repetitive sequences.

Organization of the tomato genome

The obtained pseudomolecules complement the cytogenetic and genetic understanding of tomato genome organization. Tomato pachytene chromosomes consist of prominent pericentric heterochromatin with 4–10×more DNA per unit length than distal euchromatin. FISH demonstrates that repeats are concentrated around centromeres, in chromomeres, and at telomeres. The distribution of recombination nodules and genetic markers indicates a much higher frequency of recombination in distal euchromatin than in pericentric heterochromatin. Gene-encoded transcripts map to distal euchromatin, while miRNA genes are evenly distributed along the chromosomes. Small RNAs also map to distal euchromatin, a situation completely different from that of *Arabidopsis*, in which they map preferentially to

pericentric heterochromatin.

Tomato repetitive DNA is more ancient, with family members more diverged than in most angiosperms. Early RFLP mapping of random genomic clones showed the tomato genome to be largely comprised of low-copy, non-coding DNA (Zamir et al., 1988), consistent with the predominantly low-copy DNA renaturation kinetics, despite a substantial portion of the genome being heterochromatic (Peterson et al. 1996). Compared with the smaller genome of sorghum (740 Mb) (Paterson et al. 2009), tomato has fewer intact long terminal repeat (LTR) retrotransposons (~4,000, versus 11,000 for sorghum) with older average insertion ages (2.8 versus 0.8 million years ago (mya)), and no high-copy full length LTR retrotransposons (the largest cluster having only 581 members, with all other clusters <100). K-mer frequencies are a repeat library-independent and thus unbiased method to access the repetitive portion of a genome. When the frequency for each 16mer in the tomato genome sequences was calculated, only 24% of the genome are composed of 16mers with frequencies ≥10 times. This indicates that the tomato genome has a distinctly lower repetitive content compared to Sorghum genome of similar size, where 41% of the genome are composed of 16mers with frequencies ≥10. These characteristics of the repeat component of the tomato genome facilitated the assignment of assembled repetitive sequences to specific chromosomes.

Gene structure

An integrated gene prediction pipeline based on EuGene (Foissac et al. 2008) and RNA-seq data were used to annotate the genome. The annotation predicted 34,727 protein-coding genes, most of which (30,855) are supported by RNA-seq data and show homology to *Arabidopsis thaliana* proteins (31,741 with e-value $<1e^{-3}$). Functional descriptions could be assigned to 78% of tomato proteins, while 22% received a description of "unknown protein." Small RNA data from three tomato libraries supported the prediction of 96 known miRNA genes in tomato, consistent with the copy number in other model and non-model plant species investigated to date.

The protein-coding genes of tomato, potato, *Arabidopsis* (TAIR9), rice (RAP2), and *Vitis vinifera* (grape) (version 1.0) have been grouped into 23,208 gene-groups ("families"; each with at least two members) using OrthoMCL (Li et al. 2003). Of the 34,727 protein-coding genes predicted for tomato, 25,885 were clustered in a total of 18,783 gene-groups. From the 18,783 gene-groups, 8,615 are common to all five genomes, 1,727 are confined to eudicots (tomato-potato-grape-*Arabidopsis*), and 727 to plants with fleshy fruits (tomato-potato-grape). A total of 5,165 gene-groups were identified as Solanaceae-specific, while 562 are specific to tomato and 679 to potato. Such genes provide candidates for further validation and exploration of possible roles in species-specific traits, including fruit and tuber biogenesis.

Genome triplication

Comparison of the *S. lycopersicum* and *Vitis vinifera* genomes (Jaillon et al. 2007), involving 1,730 tomato-grape (asterid-rosid) homologous DNA segments, supports the hypothesis that a whole genome triplication affecting the rosid lineage occurred in a common eudicot ancestor (Tang et al. 2008). The distribution of synonymous (*Ks*) nucleotide substitutions between corresponding gene pairs in duplicated blocks suggests one polyploidization in tomato preceded the asterid-rosid divergence. Since each of the 'triplets' of grape chromosomal segments has different best-matching homologous blocks in tomato, it can be inferred that tomato-grape genome structural divergence followed this triplication.

Comparison with the grape genome also reveals a more recent triplication in tomato. While few individual tomato genes remain triplicated, about 73% of tomato gene models are in blocks that are orthologous to one grape region, collectively covering 84% of the grape gene space. Among grape genomic regions, 22.5% have one orthologous region in tomato, 39.9% have two, and 21.6% have three. The most parsimonious explanation is that a whole genome triplication occurred in the tomato lineage, followed by widespread gene loss (Fig. 1). Based on alignments of multiple tomato segments to single grape genome segments, the tomato genome can be partitioned into three non-overlapping 'subgenomes.' The smaller number of tomato–tomato (501) compared with tomato–grape (1,730) homologous segments is consistent with substantial gene loss and rearrangement following this additional polyploidy. The tomato triplication is estimated at 71 mya based on the Ks of triplicated genes, and therefore the vast majority of post-triplication gene loss predates the ~7.3 mya tomato–potato divergence (Wu et al. 2010).

Comparative genome analysis against potato

It is estimated that tomato and potato diverged ~7.3 mya (Wu et al. 2010). Sequence alignment of 71 Mb of euchromatic regions from the *S. lycopersicum* genome to their counterparts in *S. tuberosum* (Potato Genome Sequencing Consortium 2011) revealed 8.7% nucleotide divergence with an average of one indel every 110 bp. The intergenic and repeat-rich heterochromatic sequences generally showed nucleotide divergence of more than 30% between the two species, consistent with the high sequence diversity in these regions between different potato genotypes (Potato Genome Sequencing Consortium 2011). Alignment of tomato–potato orthologous regions confirmed eight large inversions

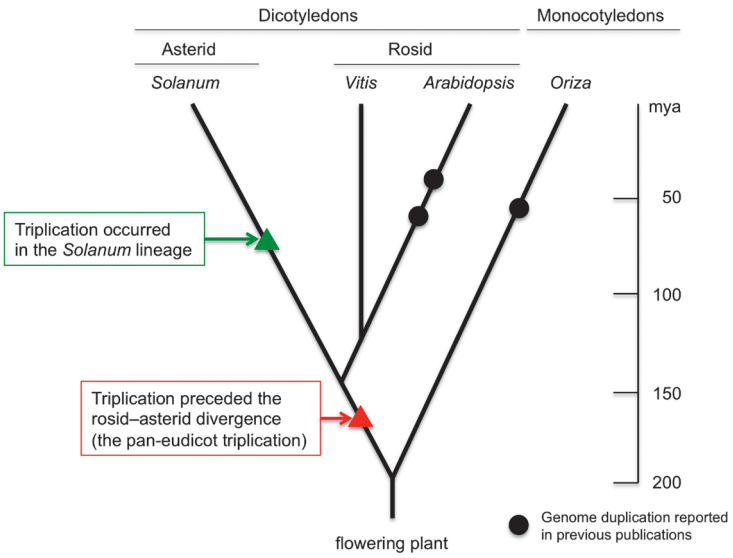

Figure 1. Two triplication events in the lineage of the *Solanum* genome.

known from cytological studies.

To facilitate interspecies comparison, the potato genome was re-annotated using the same pipeline as that used for tomato annotation. The annotation predicted 35,004 genes for potato, which was comparable to the number of genes (34,727) predicted for the tomato genome. By comparing the predicted genes in the tomato and potato genomes, 18,320 clearly orthologous tomato–potato gene pairs were identified. A total of 138 (0.75%) gene pairs had significantly higher than average non-synonymous (Ka) vs. synonymous (Ks) nucleotide substitution rate ratios (ω), indicating diversifying selection, whereas 147 (0.80%) had significantly lower than average ω. The proportions of the high and low ω group between sorghum and maize, the lineages of which diverged at least 11.9 mya (Swigonova et al. 2004), are 0.70% and 1.19%, respectively, suggesting that diversifying selection may have been more rapid in tomato–potato than in sorghum-maize.

Comparative genome analysis against the wild progenitor of domesticated tomato

To explore variation between cultivated tomato and the nearest wild tomato species, the *S. pimpinellifolium* genome (accession LA1589) was also sequenced and assembled using Illumina short reads. A total of 39.3 billion quality-trimmed base pairs (43.7-fold coverage) yielded a de novo assembly of 739 Mb. Mapping the *S. pimpinellifolium* reads to the *S. lycopersicum* pseudo-molecules revealed a nucleotide divergence of only 0.6% (5.4 million SNPs), indicating a remarkably high level of genomic similarity between the two species. Consistent with this, no large structural variation was detected in gene-rich euchromatic regions; however, a k-mer-based mapping strategy revealed the absence in *S. pimpinellifollium* of several pericentric regions that contained coding sequences. The chromosome 1 indel contains a putative self-incompatibility locus, while that on chromosome 10 is segregated in the broader *S. pimpinellifolium* germplasm, suggesting the existence of an even greater reservoir of genetic variation there.

Examination of the variation between the two species for 32,955 (92%) of the iTAG annotated genes revealed 6,659 identical genes and 3,730 genes showing only synonymous changes. Despite this high genic similarity, 68,683 SNPs from 22,888 genes are potentially disruptive to gene function, including non-synonymous changes, gain or loss of stop codons or essential splice sites, and indels causing frameshifts. In addition, 1,550 genes either gained or lost a stop codon in *S. pimpinellifolium*. With the availability of an extensive *S. pimpinellifolium* marker database, it will be possible to explore the biological relevance of this variation as it relates to domestication and crop improvement. Within cultivated germplasm, particularly among the small-fruited cherry tomatoes, several chromosomal segments are more closely related to *S. pimpinellifolium* than to Heinz, supporting previous observations on the recent admixture of these gene pools due to breeding (Ranc et al. 2008). 'Heinz 1706' itself has been reported to carry introgressions from *S. pimpinellifolium* (Ozminkowski 2004). Large introgressions were detected on both chromosomes 9 and 11, and both chromosomes have been implicated in the breeding of disease resistance loci into 'Heinz 1706' using *S. pimpinellifolium* germplasm (Ozminkowski 2004).

SOL-100 project

In the next few years, hundreds of Solanaceae will be sequenced using next-generation sequencing methods to

create a common Solanaceae-based genomic framework that includes sequences and phenotypes of 100 genomes encompassing the phylogenetic diversity of this group. This project, called "SOL-100," involves sequencing 100 different Solanaceae genomes and linking these sequences to the reference tomato sequence with the ultimate aim of exploring key issues of plant biodiversity, genome conservation, and phenotypic diversification. There is a page on the SGN with detailed information about the SOL-100 project (http://solgenomics.net/organism/sol100/view). In addition, there is a link with accompanying information on how to submit a SOL-100 genome. Progress on various genomes is also available on the SGN website.

The genome sequences of tomato, *S. pimpinellifolium*, and potato provide a starting point for comparative and functional studies and for genomics-assisted breeding. Additional sequencing and bioinformatics resources are currently being devoted to expanding the Heinz 1706 sequence into a "gold standard." Moreover, the SOL community aims to sequence 100 additional Solanaceae genomes (SOL100) and develop the needed translational tools.

Acknowledgements

This work was supported by the Kazusa DNA Research Institute Foundation and the Genomics for Agricultural Innovation Foundation (DD-4010/SGE-1001), Ministry of Agriculture, Forestry, and Fisheries (MAFF), Japan.

References

Aoki K, Yano K, Suzuki A, Kawamura S, Sakurai N, Suda K, Kurabayashi A, Suzuki T, Tsugane T, Watanabe M, et al. (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics* 11: 210

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815

Ariizumi T, Aoki K, Ezura H (2011) Systematic development of tomato bioresources in Japan. *Interdisciplinary Bio Central* 3: 1–7

Asamizu E, Shirasawa K, Hirakawa H, Sato S, Tabata S, Yano K, Ariizumi T, Shibata D, Ezura H (2012) Mapping of Micro-Tom BAC-end sequences to the reference tomato genome reveals possible genome rearrangements and polymorphisms. *Int J Plant Genomics* 2012: 437026

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112: 887–898

Eshed Y, Zamir D (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79: 175–179

Foissac S, Gouzy JP, Rombauts S, Mathé C, Amselem J, Sterck L, Van de Peer Y, Rouzé P, Schiex T (2008) Genome Annotation in Plants and Fungi: EuGene as a Model Platform. *Current Bioinformatics* 3: 87–97

Foolad MR (2007) Genome mapping and molecular breeding of tomato. *Int J Plant Genomics* 2007: 64358

Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley S (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. *Theor Appl Genet* 111: 291–312

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* 14: 1457–1467

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

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 29: e25

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al.; The French–Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467

Kahlau S, Aspinall S, Gray JC, Bock R (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol* 63: 194–207

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 117: 309–334

Labate JA, Grandillo S, Fulton T, Muños S, Caicedo AL, Peralta I, Ji Y, Chetelat RT, Scott JW, Gonzalo MJ, et al. (2007) Tomato. In: Kole C (ed) *Genome Mapping and Molecular Breeding in Plants, Volume 5*. Springer, New York, pp 1–125

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13: 2178–2189

Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol* 12: 107–118

Meissner R, Jacobson Y, Melamed S, Levyatuv S, Shalev G, Ashri A, Elkind Y, Levy A (1997) A new model system for tomato genetics. *Plant J* 12: 1465–1472

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). *Nature* 452: 991–996

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, et al. (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol* 138: 1310–1317

Ohyama A, Asamizu E, Negoro S, Miyatake K, Yamaguchi H, Tabata S, Fukuoka H (2009) Characterization of tomato SSR markers developed using BAC-end and cDNA sequences from genome databases. *Mol Breed* 23: 685–691

Ozminkowski R (2004) Pedigree of variety Heinz 1706. *Rep Tomato Genet Coop* 54: 26

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551–556

Peterson DG, Stack SM, Price HJ, Johnston JS (1996) DNA content of heterochromatin and euchromatin in tomato (*Lycopersicon*

esculentum) pachytene chromosomes. *Genome* 39: 77–82

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

Rick CM, Kesicki E, Fobes JF, Holle M (1976) Genetic and biosystematic studies on two new sibling species of *Lycopersicon* from interandean Peru. *Theor Appl Genet* 47: 55–68

Saliba-Colombani V, Causse M, Gervais L, Philouze J (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43: 29–40

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, et al. (2008) Genome structure of the legume, *Lotus japonicus. DNA Res* 15: 227–239

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al.; The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635–641

Scott JW, Harbaugh BK (1989) Micro-Tom - a miniature dwarf tomato. *Florida Agr Expt Sta Circ* 370: 1–6

Sequencing Project IRG; International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800

Shirasawa K, Asamizu E, Fukuoka H, Ohyama A, Sato S, Nakamura Y, Tabata S, Sasamoto S, Wada T, Kishida Y, et al. (2010a) An interspecific linkage map of SSR and intronic polymorphism markers in tomato. *Theor Appl Genet* 121: 731–739

Shirasawa K, Isobe S, Hirakawa H, Asamizu E, Fukuoka H, Just D, Rothan C, Sasamoto S, Fujishiro T, Kishida Y, et al. (2010b) SNP discovery and linkage map construction in cultivated tomato. *DNA Res* 17: 381–391

Shirasawa K, Hirakawa H (2013) DNA marker applications to molecular genetics and genomics in tomato. *Breed Sci* 63: 21–30

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

Sim SC, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, Chetelat RT, Hutton SF, Scott JW, Gardner RG, et al. (2012b) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS ONE* 7: e45520

Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) On the tetraploid origin of the maize genome. *Comp Funct Genomics* 5: 281–284

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* 320: 486–488

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, et al. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141–1160

The Tomato Genome Sequencing Consortium (2005) The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). *Comp Funct Genomics* 6: 153-158

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* 313: 1596–1604

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

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11: 182

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al.; Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195

Yamamoto N, Tsugane T, Watanabe M, Yano K, Maeda F, Kuwata C, Torki M, Ban Y, Nishimura S, Shibata D (2005) Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars. *Gene* 15: 127–134