

Chapter 1

Tomato Genome Sequence

Shusei Sato and Satoshi Tabata

1.1 Introduction

The Solanaceae is a large family consisting of approximately 100 genera and 2500 species that grow in all habitats from rainforests to deserts (Knapp 2002). The Solanaceae family includes several plants of agronomic importance, including potato, eggplant, pepper, and tobacco, as well as tomato (*Solanum lycopersicum*). As well as its economic importance, tomato is considered to be a useful model plant species and has been the subject of extensive research, including genetic characterization. Tomato was consequently chosen as a target for genome sequencing. *S. lycopersicum* has a diploid genome of simple architecture that is approximately 900 Mb in size and is distributed across 12 chromosomes (Michaelson et al. 1991). Many Solanaceae species have highly syntenic genomes, each also with 12 chromosomes, and the reference genome sequence of the tomato thus provides a framework for the genomic analysis of Solanaceae plants in general and is a source of important information for molecular breeding.

In November 2003, the International Solanaceae Project (SOL; <http://solgenomics.net/solanaceae-project/index.pl>), a consortium initially involving researchers from ten countries, launched the tomato genome-sequencing project. The initial aim was to sequence gene-rich regions of the 12 chromosomes through high-quality sequencing of bacterial artificial chromosomes (BACs) that were

S. Sato (✉)

Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan

e-mail: shuseis@ige.tohoku.ac.jp

S. Tabata

Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

e-mail: tabata@kazusa.or.jp

© Springer-Verlag Berlin Heidelberg 2016

H. Ezura et al. (eds.), *Functional Genomics and Biotechnology in Solanaceae and*

Cucurbitaceae Crops, Biotechnology in Agriculture and Forestry 70,

DOI 10.1007/978-3-662-48535-4_1

selected based on DNA markers mapped on the genome and accumulated end-sequence information (Mueller et al. 2005, and <http://sgn.cornell.edu/>). In 2008, a whole-genome-sequencing strategy was also adopted, with the aim of covering the entire genome. Alongside the sequencing efforts, DNA markers were evaluated and high-density genetic linkage maps were constructed to assist assembly of the whole-genome structure (Fulton et al. 2002; Frary et al. 2005; Shirasawa et al. 2010). The chloroplast and mitochondrial genomes were sequenced independently of the nuclear genome (Kahlau et al. 2006). The International Tomato Annotation Group (ITAG) subjected the obtained sequences to assembly and performed further analyses. Ultimately, researchers from 14 countries contributed to the project, and the results were published in 2012 (The Tomato Genome Consortium 2012).

In this chapter, we summarize the process of tomato genome sequencing and the features of the tomato genome revealed by the obtained sequence information.

1.2 Tomato Genome Sequencing

The tomato genome, estimated to be 900 Mb long, has rather simple architecture composed of pericentromeric heterochromatin and distal euchromatin. Pericentromeric heterochromatin, rich in repetitive sequences, is estimated to occupy three-quarters of the tomato genome. The remaining one-quarter (220 Mb) of the tomato genome consists of distal euchromatic segments; these regions were thought to contain more than 90 % of the genes prior to the project. Therefore, the strategy of the initial phase of the tomato genome-sequencing project was to sequence the euchromatic portions of the 12 chromosomes using a BAC-by-BAC sequencing approach. The tomato variety used for the sequencing project was the ‘Heinz 1706’ cultivar, provided by the Heinz Corporation (Pittsburgh, PA, USA). ‘Heinz 1706’ was chosen because the well-characterized *Hind*III BAC library available at the time of project inception was constructed from this cultivar (Budiman et al. 2000). Two *Eco*RI and *Mbo*I BAC libraries were also constructed, and end sequences of all three BAC libraries were analyzed. In this approach, molecular genetic markers were used to anchor seed BAC clones. The tiling path was generated by walking from seed clones in both directions using the analyzed BAC end-sequencing data. This BAC-by-BAC approach resulted in the sequencing of 117 Mb of tomato euchromatic regions with high accuracy (Mueller et al. 2005).

In 2008, the sequencing consortium adopted the selected BAC mixture (SBM) approach with the aim of accelerating progress (The Tomato Genome Consortium 2012). A total of 30,800 BAC clones were selected having considered the BAC end-sequence data accumulated in the initial phase of the project and the removal of BACs that had repetitive elements at their ends. The chosen BAC clones were pooled, and shotgun sequencing was performed using the Sanger sequencing method. A total of 4.2 million reads corresponding to 3.1 Gb were produced, and these sequences were assembled into contigs that covered 540 Mb of the genome

and encompassed >80 % of the previously registered tomato ESTs (<http://www.kazusa.or.jp/tomato/>). The success of the shotgun approach prepared the way for a next-generation sequencing (NGS) approach.

In 2009, the sequencing consortium decided to take advantage of the emerging NGS platforms and increase the scope of the project from euchromatic regions only to the whole tomato genome. Three NGS platforms, Roche/454, SOLiD, and Illumina, were used to generate 21 Gb, 64 Gb, and 82 Gb of sequence data, respectively. A de novo assembly of the ‘Heinz 1706’ genome was subsequently performed using the Sanger data (3.3 Gb, including ~200,000 BAC and fosmid paired-end sequences and 4.2 million SBM reads) and 454 data (21 Gb). Two programs, Newbler and CABOG, were used to generate independent assemblies; these were subsequently integrated. The structural accuracy of the de novo assembly was confirmed by mapping to paired-end sequences of the BAC and fosmid clones. The high coverage Illumina and SOLiD reads were used to improve overall base accuracy. As a result of read-mapping and error-base correction, high-base accuracy was achieved, resulting in only one base calling error per 29.4 kb and one indel error per 6.4 kb. Contig gaps were filled by integrating 117 Mb of BAC-clone Sanger sequences from the initial phase of the project. The resulting high-quality scaffolds were linked with two BAC-based physical maps and anchored using a high-density genetic map (Shirasawa et al. 2010), introgression-line mapping, and genome-wide BAC fluorescence in situ hybridization (FISH). The final tomato genome assembly consisted of 91 scaffolds covering 760 Mb. The scaffolds were then aligned with the 12 chromosomes, and most of the gaps were found to be restricted to pericentromeric regions (Table 1.1). The 21 Mb of sequences

Table 1.1 Status of tomato genome sequence (Assembly SL2.40)

Chromosome	Number of scaffolds	Cumulative scaffold length (bp)	Average GC %
chr1	9	90,303,444	33.7
chr2	7	49,917,694	33.6
chr3	13	64,839,514	34.0
chr4	6	64,063,812	33.7
chr5	3	65,021,238	34.0
chr6	8	46,040,936	34.0
chr7	4	65,268,321	34.1
chr8	9	63,031,857	34.1
chr9	10	67,661,191	34.1
chr10	6	64,833,805	34.0
chr11	6	53,385,525	34.1
chr12	10	65,485,353	34.2
Subtotal anchored scaffolds	91	759,852,690	34.0
Unanchored scaffolds (chr0)	3132	21,492,721	37.8
Total	3223	781,345,411	34.1

contained in the 3132 unanchored scaffolds were designated as chr0 and were primarily repetitive sequences (Table 1.1).

1.3 Features of Tomato Genome

1.3.1 Organization of Tomato Genome

Detailed analysis of the cytogenetic and genetic features of tomato genome organization was carried out based on the obtained tomato genome sequences anchored on the 12 chromosomes (pseudomolecules). By comparing the BAC-clone FISH results and the physical locations of these clones on pseudomolecules, it became clear that tomato pachytene chromosomes consist of prominent pericentromeric heterochromatin with 4–10× more DNA per unit length than distal euchromatin (The Tomato Genome Consortium 2012). FISH analysis using Cot 100 DNA (including most repeats) as a probe demonstrated that the repeats are concentrated around centromeres and telomeres and within chromomeres. Using the positional information from FISH BAC probes, recombination nodule locations derived from cytological mapping were compared with the physical locations on the pseudomolecules. This revealed a much higher recombination frequency in distal euchromatin than in pericentromeric heterochromatin. This distribution was confirmed by the comparison of genetic distance and physical distance using molecular genetic markers.

Early RFLP mapping of random genomic clones led to the estimation that a large proportion of the tomato genome consists of low-copy, noncoding DNA (Zamir and Tanksley 1988). This is supported by DNA renaturation kinetics, which are consistent with predominantly low-copy DNA, despite the substantial proportion of the genome that is heterochromatic (Peterson et al. 1998). The number of repetitive sequences in the obtained tomato reference genome is far fewer than in the smaller, 740 Mb, sorghum genome, with ~4000 intact long terminal repeat (LTR) retrotransposons identified as opposed to the ~11,000 identified in sorghum (Patterson et al. 2009). The average insertion age (as estimated by base substitutions in LTR sequences) of the tomato LTR retrotransposons was older than that of sorghum [2.8 versus 0.8 million years (Myr) ago]. In addition, no high-copy full-length LTR retrotransposons were identified in tomato. The largest cluster contained just 581 members, with all the other clusters containing <100 members. Features of repetitive sequences in the tomato genome were also revealed by *k*-mer frequency analysis. *k*-mer frequencies are a repeat-library-independent, and thus unbiased, method for accessing the repetitive portion of a genome. When the frequencies of each 16-mer in the tomato genome sequence were calculated, only 24 % of the genome was found to be composed of 16-mers with frequencies that occur ≥10 times. This indicates that tomato has a distinctly lower repetitive element content

than the smaller sorghum genome, in which 41 % of the genome is composed of 16-mers with frequencies ≥ 10 . These characteristics of the repeated portions of the tomato genome facilitated the creation of long scaffolds and the assignment of scaffold sequences to specific chromosomes.

1.3.2 Gene Structure

The tomato genome was annotated by the iTAG consortium. An integrated gene prediction pipeline based on EuGene (Foissac et al. 2008) and RNA-seq data was used which produced a consensus annotation of 34,727 protein-coding genes in tomato (iTAG v2.3: http://solgenomics.net/organism/Solanum_lycopersicum/genome). As a large amount of the RNA-seq data were accumulated by using NGS platforms, most of the predicted protein-coding genes (30,855) are supported by transcribed sequence information. More than 90 % of the predicted genes (31,741, with e-value $< 1e-3$) are homologous to *A. thaliana* genes (TAIR10). Functional descriptions were putatively assigned to 78 % of the tomato proteins, and the remaining 22 % received a designation of “unknown protein.” Small RNA data from three tomato libraries supported the prediction of 96 known miRNA genes in tomato, which is consistent with the copy number found in other model and non-model plant species investigated to date.

In order to survey conserved features in protein-coding genes, gene family clusters among different plant species were defined using OrthoMCL software (Li et al. 2003). The protein-coding genes of tomato, potato, *Arabidopsis*, rice, and *Vitis vinifera* (grape) were included in the analysis, and a total of 154,880 gene sequences from these five species were grouped into 23,208 gene groups (“families,” each with at least two members). Of the 34,727 protein-coding genes predicted on the reference tomato genome, 25,885 were clustered in a total of 18,783 gene groups. Of these 18,783 gene groups, 8615 are common to all five genomes, 1727 are confined to eudicots (tomato, potato, grape, *Arabidopsis*), and 727 to plants with fleshy fruits (tomato, potato, grape) (The Tomato Genome Consortium 2012). A total of 5165 gene groups were identified as Solanaceae specific, while 562 were tomato specific and 679 were potato specific. Such genes provide candidates for further validation and exploration of diverse roles in species-specific traits, including fruit and tuber biogenesis.

1.3.3 Genome Triplication

The draft genome of grape (*V. vinifera*) indicated that no recent genome duplication had occurred, and this enabled the discovery of ancestral traits and features related

to the genetic organization of flowering plants (French-Italian Public Consortium for Grapevine Genome Characterization 2007). Further analysis revealed that whole-genome triplication contributed to the establishment of the grape genome and that this triplication is common to many dicot plants but is absent in monocots. To test the hypothesis that the whole-genome triplication in the rosid lineage, which includes grape and *Arabidopsis*, occurred in a common ancestor shared with tomato and other asterids (Tang et al. 2008), the tomato and grape genomes (French-Italian Public Consortium for Grapevine Genome Characterization 2007) were compared. A comparison of grape triplet chromosomes to the tomato genome inferred 1730 tomato-grape (asterid-rosid) homologous DNA segments. The distribution of synonymous nucleotide substitution rates (Ks) between corresponding gene pairs in duplicated blocks suggests that one polyploidization in tomato preceded the asterid-rosid divergence. Since each of the “triplets” of grape chromosomal segments matches optimally with a distinct homologous block 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 the tomato genome. 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 and was followed by widespread gene loss. Based on alignments of multiple tomato segments to single grape genome segments, the tomato genome can be partitioned into three nonoverlapping “subgenomes.” The smaller number of tomato-tomato (501) compared with tomato-grape (1730) homologous segments is consistent with substantial gene loss and rearrangement following this additional polyploidy. Based on the Ks of triplicated genes, the tomato triplication is estimated at 71 Myr, and therefore, the majority of post-triplication gene loss predates the ~7.3 Myr tomato-potato divergence (Wu and Tanksley 2010).

These two genome triplication events shaped the evolution of genes involved in fleshy fruit development. Most of the genes were eliminated by widespread gene loss following the triplication events, with the duplicates that remained acquiring new and distinct functions. This group of genes includes pleiotropic transcription factors that are necessary for ethylene biosynthesis [*RIN* (Vrebalov et al. 2002), *CNR* (Manning et al. 2006)], enzymes necessary for ethylene biosynthesis and signaling (*ACS* (Nakatsuka et al. 1998), *ETR* (Klee and Giovannoni 2011)), red-light photoreceptors that are associated with fruit quality [*PHYB1*/*PHYB2* (Pratt et al. 1995)], and enzymes necessary for lycopene biosynthesis [*PSY1*, *PSY2* (Giorio et al. 2008)] (Fig. 1.1).

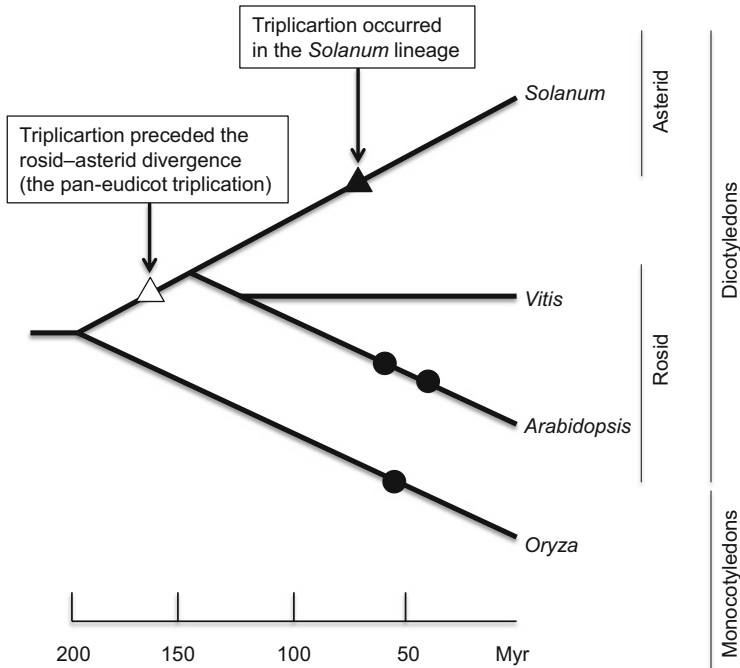


Fig. 1.1 Two triplication events in the *Solanum* genome. Reported polyploidization events in monocotyledon and eudicotyledon genomes. A *white triangle* indicates occurrence of a triplication event after divergence of dicotyledons from monocotyledons and before divergence of rosid and asterid (pan-eudicot triplication). The triplication event identified in the *Solanum* lineage (tomato and potato) is shown with a *black triangle*. *Black circles* indicate genome duplication reported in previous publications

1.4 Comparative Genomics of the Tomato Genome

1.4.1 Comparative Genome Analysis Against Potato

In the potato (*S. tuberosum*) genome-sequencing project (Potato Genome Sequencing Consortium 2011), which was published prior to the tomato genome, a homozygous doubled-monoploid potato clone was used for sequencing in order to overcome the highly heterozygous nature of most potato cultivars. A whole-genome shotgun sequencing approach was applied using different NGS platforms, primarily Illumina technology. A final assembly of 727 Mb was compiled from 96.6 Gb of raw sequences (Potato Genome Sequencing Consortium 2011).

Tomato and potato are estimated to have diverged ~ 7.3 Myr (Wu and Tanksley 2010). Sequence alignment of 71 Mb of euchromatic regions from the tomato reference genome to their counterparts in potato revealed 8.7 % nucleotide divergence with an average of one indel per 110 bp. The intergenic and repeat-rich heterochromatic sequences generally showed nucleotide divergence of >30 %

between the two species, consistent with the high-sequence diversity in these regions among different potato genotypes (Potato Genome Sequencing Consortium 2011). The chromosome pseudomolecules of the potato genome were updated by anchoring the scaffolds on the integrated genetic and physical reference map comprising nearly 2500 markers (Sharma et al. 2013). The dot plot alignments between the updated pseudomolecules of the potato genome and those of the tomato genome revealed 19 paracentric inversions including eight large inversions that were previously known from cytological studies.

In order to carry out a precise comparison between protein-coding genes of tomato and potato, 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 is 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 (The Tomato Genome Consortium 2012). A total of 138 (0.75 %) gene pairs had significantly higher than average non-synonymous (K_a) vs. synonymous (K_s) nucleotide substitution rate ratios (ω), indicating diversifying selection, and many high ω -group genes were found to encode proteins that regulate biological processes, such as transcription factors. Conversely, 147 gene pairs (0.80 %) had significantly lower than average ω , indicating purifying selection, and most low- ω genes were found to be structural genes such as histone superfamily proteins and ribosomal proteins.

Comparison of the predicted genes also revealed genes conserved only in tomato or potato. Cytochrome P450 provides an example; several cytochrome P450 sub-families show complete loss in tomato with respect to potato. Some of these losses, such as *CYP80N1* and *CYP82E4*, may be ecologically significant. Their absence may limit the biosynthesis of toxic glycoalkaloid and thus promote the development of a nutritionally attractive fruit that, in turn, enhances seed dispersal by animals (Cipollini and Levey 1997; Chakrabarti et al. 2007).

1.4.2 Comparative Genome Analysis of Tomato and Wild Relatives

The reference tomato genome sequence was obtained from ‘Heinz 1706’, a cultivated variety. To explore variation between cultivated tomato and the nearest wild tomato species, the tomato genome-sequencing consortium sequenced the *S. pimpinellifolium* genome (accession LA1589) using a whole-genome shotgun approach with Illumina technology (The Tomato Genome Consortium 2012). A final assembly of 739 Mb was generated from 39.3 Gb quality-trimmed sequences (43.7-fold coverage). Mapping the *S. pimpinellifolium* reads to the *S. lycopersicum* pseudomolecules revealed a nucleotide divergence of only 0.6 % (5.4 million SNPs), indicating a remarkably high level of genomic similarity between the two species. Correspondingly, no large structural variation was detected in gene-rich

euchromatic regions; however, a *k*-mer-based mapping strategy revealed that several pericentromeric regions containing coding sequences are absent in *S. pimpinellifolium*. The chromosome 1 indel contains a putative self-incompatibility locus, while the indel on chromosome 10 is segregated in the broader *S. pimpinellifolium* germplasm, suggesting the existence of an even greater reservoir of genetic variation among other isolates.

More than 90 % (32,955) of the predicted genes in the *S. lycopersicum* genome are present in the genome of *S. pimpinellifolium*. As expected from the pedigree of ‘Heinz 1706’, which has *S. pimpinellifolium* as one of its ancestors, putative *S. pimpinellifolium* introgressions were detected. Examination of the variation between the two species for 32,955 (92 %) of the iTAG annotated genes revealed 6659 identical genes and 3730 genes with 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, 1550 genes either gained or lost a stop codon in *S. pimpinellifolium*. Since the identified SNPs can be used as markers for the whole *S. pimpinellifolium* genome, it will be possible to explore the biological relevance of this variation and its relationship to domestication and crop improvement. Within cultivated germplasms, particularly among the small-fruited cherry tomatoes, several chromosomal segments are more closely related to *S. pimpinellifolium* than to ‘Heinz 1706’, supporting previous observations on the recent admixture of these gene pools as a consequence of breeding (Ranc et al. 2008). ‘Heinz 1706’ itself has been reported to carry introgressions from *S. pimpinellifolium* (Ozminkowski 2004). Genomic regions with low divergence between *S. pimpinellifolium* and ‘Heinz 1706’ but with high divergence among domesticated cultivars were regarded as *S. pimpinellifolium* introgressions. 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).

1.5 Continuing Sequencing Efforts and Future Perspectives

NGS allowed the tomato genome-sequencing project, which began by using clone-by-clone Sanger technology of selected regions, to progress to the sequencing and assembly of the whole genome. The comprehensive datasets, which include large amounts of NGS data and BAC/cosmid end Sanger reads, alongside scrupulous attention-to-error correction, produced one of the highest-quality genome sequences to date (Assembly SL2.40). Nevertheless, the Tomato Genome Sequence Consortium is pursuing efforts to further improve the genome and reach “gold standard.” These endeavors are currently focused upon gap closure and scaffold validation. A large number (~2000) of additional BAC clones have been sequenced

using NGS platforms with the aim of closing gaps within and between scaffolds. For smaller gaps of up to 1000 bp, an additional high-throughput method was developed using 454 technology and applied to gap closure. Scaffold validation was enhanced by adding >600 BAC clones to the tomato FISH map (SOL Newsletter April 2013, Issue 35: <http://solgenomics.net/>). The locations of the BAC clones were used both for estimating gap size between scaffolds and for validation and adjustment of the order and orientation of the scaffolds. Many of the localized BAC clones were selected from chr0 scaffolds (unanchored scaffolds), and the obtained FISH map data allowed these scaffolds to be mapped to pseudomolecules. The accumulated new data will be incorporated and the updated reference tomato genome information will be released as SL2.50 (Lucas Mueller, personal communication).

Extensive molecular marker analysis revealed that, as a result of domestication, genetic diversity in the cultivated tomato is much lower than in its wild relatives. The availability of a high-quality genome from the domesticated cultivar ‘Heinz 1706’ is facilitating the sequencing of additional cultivated and wild tomato ecotypes, with the aim of analyzing genetic variations and improving the data available for marker-associated breeding. One large-scale example of these ongoing projects is the “150 tomato genomes project” (<http://www.tomatogenome.net>). In this project, 84 ecotypes including 10 old varieties, 43 cultivated lines, and 30 wild accessions have been selected for sequencing. Moreover, some 60 F8 individuals of *S. pimpinellifolium* recombinant inbred lines (RILs) will also be sequenced with the aim of identifying recombination breakpoints at the sequence level. Popular cultivars in tomato experimental studies such as ‘Ailsa Craig’, ‘Rutgers’, ‘M82’, and ‘Micro-Tom’ will also be sequenced (Aoki et al. 2013; <http://solgenomics.net/organism/1/view>). Although most of these datasets are not currently publicly available, they will serve as excellent information resources for developing SNP markers and intraspecific maps.

In addition to cultivated and wild tomato ecotypes, hundreds of Solanaceae species will be sequenced using NGS technologies to create a common Solanaceae-based genomic framework that includes sequences and phenotypes of 100 genomes encompassing the phylogenetic diversity of Solanaceae group. This clade-oriented project, called “SOL-100,” involves sequencing 100 different Solanaceae genomes and linking these sequences to the reference tomato sequence. The ultimate aim of this project is to explore key issues of plant biodiversity, genome conservation, and phenotypic diversification, and more information is available at the SOL Genomics Network (SGN) site (<http://solgenomics.net/organism/sol100/view>). At the time of writing (August 2013), genome-sequencing projects involving 25 Solanaceae species are ongoing (Table 1.2; <http://solgenomics.net/organism/1/view>), and the obtained results are beginning to emerge (Bombarely et al. 2012; Sierro et al. 2013).

The highly accurate ‘Heinz 1706’ reference genome sequence will, alongside genome sequences of *S. pimpinellifolium* and potato, pave the way for comparative and functional studies and for genomics-assisted breeding in Solanaceae. Additional sequencing and bioinformatics resources are currently being devoted to

Table 1.2 List of Solanaceae species analyzed in the SOL-100 project

<i>Nicotiana tomentosiformis</i>	<i>Solanum chilense</i>
<i>Nicotiana benthamiana</i>	<i>Solanum neorickii</i>
<i>Nicotiana attenuata</i>	<i>Solanum galapagense</i>
<i>Nicotiana sylvestris</i>	<i>Solanum pimpinellifolium</i>
<i>Petunia axillaris</i>	<i>Solanum pennellii</i>
<i>Lycium barbarum</i>	<i>Solanum huaylasense</i>
<i>Capsicum annuum</i>	<i>Solanum corneliomuelleri</i>
<i>Withania somnifera</i>	<i>Solanum chmielewski</i>
<i>Iochroma cyaneum</i>	<i>Solanum peruvianum</i>
<i>Solanum tuberosum</i>	<i>Solanum cheesmaniae</i>
<i>Solanum retroflexum</i>	<i>Solanum arcanum</i>
<i>Solanum melongena</i>	<i>Solanum habrochaites</i>
	<i>Solanum lycopersicum</i>

expand the Heinz 1706 sequence into a “gold standard.” Extensive sequencing efforts on cultivated and wild tomato accessions will provide marker and gene pools of sufficient depth for crop improvement. Moreover, the SOL community aims to sequence and analyze 100 additional Solanaceae genomes (SOL100) and develop the needed translational tools. Along with the systematic development of material and information resources, genomic studies of cross-Solanaceae species analyses will bear considerable fruit in coming years.

References

- Aoki K, Ogata Y, Igarashi K, Yano K, Nagasaki H, Kaminuma E, Toyoda A (2013) Functional genomics of tomato in a post-genome-sequencing phase. *Breed Sci* 63:14–20
- Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB (2012) A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Mol Plant Microbe Interact* 25:1523–1530
- Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res* 10:129–136
- Chakrabarti M, Meekins KM, Gavilano LB, Siminszky B (2007) Inactivation of the cytochrome P450 gene CYP82E2 by degenerative mutations was a key event in the evolution of the alkaloid profile of modern tobacco. *New Phytol* 175:565–574
- Cipollini ML, Levey DJ (1997) Secondary metabolites of fleshy vertebrate-dispersed fruits: adaptive hypotheses and implications for seed dispersal. *Am Nat* 150:346–372
- 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. *Curr Bioinforma* 3:87–97
- 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

- French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467
- 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
- Giorio G, Stigliani AL, D'Ambrosio C (2008) Phytoene synthase genes in tomato (*Solanum lycopersicum* L.)—new data on the structures, the deduced amino acid sequences and the expression patterns. *FEBS J* 275:527–535
- 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
- Klee HJ, Giovannoni JJ (2011) Genetics and control of tomato fruit ripening and quality attributes. *Annu Rev Genet* 45:41–59
- Knapp S (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J Exp Bot* 53:2001–2022
- Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189
- Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet* 38:948–952
- Michaelson MJ, Price HJ, Ellison JR, Johnston JS (1991) Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. *Am J Bot* 78:183–188
- Mueller LA, Tanksley SD, Giovannoni JJ, van Eck J, Stack S, Choi D, Kim BD, Chen M, Cheng Z, Li C, Ling H, Xue Y, Seymour G, Bishop G, Bryan G, Sharma R, Khurana J, Tyagi A, Chattopadhyay D, Singh NK, Stiekema W, Lindhout P, Jesse T, Lankhorst RK, Bouzayen M, Shibata D, Tabata S, Granell A, Botella MA, Giuliano G, Frusciante L, Causse M, Zamir D (2005) The tomato sequencing project, the first cornerstone of the International Solanaceae Project (SOL). *Comp Funct Genomics* 6:153–158
- Nakatsuka A, Murachi S, Okunishi H, Shiomi S, Nakano R, Kubo Y, Inaba A (1998) Differential expression and internal feedback regulation of 1-aminocyclopropane-1-carboxylate synthase, 1-aminocyclopropane-1-carboxylate oxidase, and ethylene receptor genes in tomato fruit during development and ripening. *Plant Physiol* 118:1295–1305
- 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, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556
- Peterson DG, Pearson WR, Stack SM (1998) Characterization of the tomato (*Lycopersicon esculentum*) genome using in vitro and in situ DNA reassociation. *Genome* 41:346–356
- Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195
- Pratt LH, Cordonnier-Pratt MM, Hauser B, Caboche M (1995) Tomato contains two differentially expressed genes encoding B-type phytochromes, neither of which can be considered an ortholog of Arabidopsis phytochrome B. *Planta* 197:203–206
- Ranc N, Muñoz 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
- Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W, Carboni MF, D'Ambrosio JM, de la Cruz G, Di Genova A, Douches DS, Eguiluz M, Guo X, Guzman F, Hackett CA, Hamilton JP, Li G, Li Y, Lozano R, Maass A, Marshall D, Martinez D, McLean K, Mejía N, Milne L,

- Munive S, Nagy I, Ponce O, Ramirez M, Simon R, Thomson SJ, Torres Y, Waugh R, Zhang Z, Huang S, Visser RG, Bachem CW, Sagredo B, Feingold SE, Orjeda G, Veilleux RE, Bonierbale M, Jacobs JM, Milbourne D, Martin DM, Bryan GJ (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3 (Bethesda)* 3:2031–2047
- 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 (2010) An interspecific linkage map of SSR and intronic polymorphism markers in tomato. *Theor Appl Genet* 121:731–739
- Sierro N, Batten JN, Ouadi S, Bovet L, Goepfert S, Bakaher N, Peitsch MC, Ivanov NV (2013) Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol* 14:R60
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* 320:486–488
- The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641
- Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (*rin*) locus. *Science* 296:343–346
- Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. *BMC Genomics* 11:182
- Zamir D, Tanksley SD (1988) Tomato genome is comprised largely of fast-evolving, low copy-number sequences. *Mol Gen Genet* 213:254–261