Plant Biotechnology 30, 243–256 (2013)
DOI: 10.5511/plantbiotechnology.13.0904a

Review

Advances in tomato research in the post-genome era

Naama Menda, Susan R. Strickler, Lukas A. Mueller*

Boyce Thompson Institute for Plant Research, 533 Tower Road, Ithaca NY 14853, USA
* E-mail: lam87@cornell.edu Tel: +1–607-255-6557 Fax: +1–607-254-1242

Received April 2, 2013; accepted September 4, 2013 (Edited by K. Aoki)

Abstract The tomato (*Solanum lycopersicum*) draft genome, along with a draft of the wild relative, *Solanum pimpinellifolium*, were released in 2012, almost a decade after the International Tomato Genome project was initiated. Tomato is an important domesticated crop species, as well as a model organism for many aspects of plant biology such as fleshy fruit development, ripening, disease resistance, plant architecture, and compound leaf development. For these reasons, there has been a substantial effort for producing a high quality reference genome that will serve as an anchor for tomato species, and for closely related Solanaceae plants. The utility of this genome has already been demonstrated by a relatively large number of studies that have been published since the release of the sequence, covering a wide range of topics including gene expression, genetic diversity, phylogeny, comparative genomics, and epigenetics. With the availability of the potato genome, it is now possible to perform detailed comparative genomic analysis of gene families in the Solanaceae, facilitated by conservation and synteny between their genomes. A large number of ongoing efforts will result in the sequencing of hundreds of wild and domesticated tomato accessions from various populations, uncovering the breeding history of tomatoes and introducing new genomic technologies to accelerate breeding processes. In this review, we provide an overview of the origins of tomato and its position in the wider Solanaceae, and demonstrate the impact of the tomato genome sequence on Solanaceae research on the basis of recent literature that has made use of this new resource.

Key words: Tomato genome, sequencing, Solanaceae, reference genome, model organism.

The economic value of tomato to agriculture has made it the centerpiece of the Solanaceae family (also known as the Nightshades), as well as a model organism for the study of plant development, fruit ripening, and disease resistance. As such, numerous genetic and molecular tools have been developed for the species over the last decades, including high-density genetic maps, molecular markers, BAC and EST libraries, mapping populations, introgressions lines, mutant collections, microarray chips, and a high-quality draft genome sequence released in 2012 (Tomato Genome Consortium 2012). Since tomatoes are diploids, with a modest genome size (~900 Mb), have homozygous inbreds available, and share a high level of genome conservation at the macro and micro synteny levels with other Solanaceae (Fulton 2002; Ku et al. 2000), the sequenced genome can serve as a reference for wild tomatoes and other species in the wider Solanaceae family. The wild tomato relatives, most of which can be crossed to tomato, have also retained allelic diversity required for adaptations to various environments and habitats (Labate et al. 2009; Xia et al. 2010).

In this review, we discuss the existing resources and tools that have emerged based on the tomato genome sequence, how the new technologies and findings facilitate research and breeding of tomatoes and related species, and current research that will advance studies of the Solanaceae family.

Tomato domestication

Early domestication of edible wild tomato berries was commenced by South American indigenous populations in the South American Andes, where the majority of the 12 native wild tomato species of the Lycopersicum subgenus have evolved. It is likely these early populations also transported the plants to central America and Mexico, followed by the Spaniards who brought tomatoes and also potatoes to Europe at the beginning of the 16th century (Blanca et al. 2012). At first they were perceived as poisonous and used only for decoration, but they rapidly became staple food in local cuisine, and distributed across the globe as a food crop. Recent SNP analysis and morphological characterization of

---

Abbreviations: BAC, bacterial arteficial chromosome; EST, expressed sequence tag; FISH, fluorescence in-situ hybridization; GWAS, genome wide association study; ITAG, international tomato annotation group; SGN, sol genomics network; SolCAP, Solanaceae coordinated agricultural project; SNP, single nucleotide polymorphism; QTL, quantitative trait locus; eQTL, expression quantitative trait locus.

domesticated tomato, *S. pimpinellifolium*, and the weedy species *Solanum lycopersicum* var *cerasiforme*, suggests that the South American domestication was completed later on in Central America (Blanca et al. 2012).

Tomatoes now have worldwide distribution as the number one vegetable crop, constituting in monetary value 14% of all vegetables produced (FAOSTAT database http://faostat3.fao.org/). By the beginning of the 20th century, tomato breeding yielded dozens of inbreds, known today as heirloom lines, displaying variation mostly in fruit shape, size, and color properties. Despite the dramatic differences in fruit traits of heirloom lines, they vary only in a handful of Mendelian genes, such as the fruit shape genes *OVATE*, *FASCIATED*, *LOCULE NUMBER*, and *SUN* (Rodriguez et al. 2011), and the fruit color genes *PSY1* (Fray and Grierson 1993), *Beta-carotene* (Ronen et al. 2000), *Delta* (Ronen et al. 1999), and *HIGH-PIGMENT* (Lieberman et al. 2004).

Wild tomato species

The 12 members of the Lycopersicon clade (excluding cultivated tomato) display a high level of phenotypic and genetic variation, adaptations to a wide range of environments, and growth in a myriad of habitats, from high elevations with extreme drought and heat conditions, down to the coasts of the Pacific Ocean, and the Galápagos Islands more than 1,000 miles away from mainland South America (Peralta and Spooner 2000). The three wild species most closely related to cultivated tomato are the red-fruited *Solanum pimpinellifolium*, and two orange-fruited species, native to the Galápagos Islands, *Solanum cheesmaniae* and *Solanum galapagense*. Phylogenetic and molecular evidence shows that the putative closest wild tomato relative is *S. pimpinellifolium*, (Tomato Genome Consortium 2012).

Wild tomatoes have an important role as a reservoir of genetic diversity and are important for breeding purposes, for example, many alleles contributing to biotic and abiotic stress have been introduced into commercial inbreds from *S. chilense* (Xia et al. 2010; Verlaan et al. 2011), *S. peruvianum* (Rick 1986; Virginia Sanchez-Puerta and Williams Masuelli 2011), *S. habrochaites* (Finkers et al. 2007), *S. pimpinellifolium* (Duan et al. 2012, Lin and Martin 2007; Merk et al. 2012), and *S. pennellii* (Semel et al. 2007; Sharlach et al. 2013; Yu et al. 2010). Moreover, their rich environmental adaptations, and incomplete reproductive barriers, make tomatoes an excellent group for studying evolutionary processes such as adaptation, divergence, and speciation (Moyle 2008). Current resequencing efforts of multiple wild tomato accessions, such as the 150 Tomato Genomes Project, will further clarify the intra and interspecific relationships within the tomato sub clade. Additionally, hundreds of wild accessions that have been collected for decades in South America (Rick 1976) make up an invaluable resource for breeding, molecular, phylogenetic, and genomic research.

Tomato breeding

During extensive breeding of tomato, selection was focused on traits desirable for either processing tomatoes or fresh market tomatoes, splitting tomato cultivation into two major industries. Processing tomatoes are usually grown in open fields, need to have high sugar and total soluble solids content, and have simultaneous fruit ripening to allow lower costs associated with cultivation, and machinery harvesting. Fresh market tomatoes are bred for traits associated with consumer preference, such as large fruit size, uniform fruit shape (usually round), uniform red color, and also long shelf life and fruit firmness (Rick 1978). A single recessive gene, *SELF PRUNING* (Pnueli et al. 1998) regulates determinate plant habit, having the plant set all fruits and reach maximum size concurrently, while in indeterminate plants fruits of all growth stages develop on the same vine, which allow for a longer period of harvest. Yet the differences between the various tomato classifications are genome-wide, as demonstrated by genome analysis of SNPs among fresh market, processing, and heirloom tomatoes, showing distinct signatures for each group, which can define each as a separate subpopulation (Sim et al. 2012b).

As mentioned, an important advantage for tomato breeding is the fact that domesticated tomatoes can relatively easily outcross with wild tomato species. This property has been exploited for designing genetic markers and mapping populations (Eshed and Zamir 1995; Frary et al. 2005; Lippman et al. 2007), which have been the main driving force in marker-assisted selection for desirable traits such as fruit size (Frary et al. 2000), ripening (Moore et al. 2002), shape (Rodriguez et al. 2011), sugar content (Fridman et al. 2002), and resistance to biotic and abiotic stress (Labate and Robertson 2012; Verlaan et al. 2013). Since virtually all heirloom tomatoes are disease susceptible, and have lost genetic diversity allowing for disease resistance (Bai and Lindhout 2007; Ranc et al. 2012; Robbins et al. 2011), modern plant breeding practices require introgressing the lost traits back from the wild relatives (Grandillo et al. 2011), for example the bacterial resistance gene *Pto* was introgressed into tomato cultivars in the 1930s from *Solanum pimpinellifolium* (Lin and Martin 2007). A century of plant breeding and selection resulted in numerous commercial hybrids and cultivars which contain wild genome regions with favorable alleles, but also introgressed genome fragments, referred to as cryptic introgressions, that can negatively affect desirable traits (Labate and Robertson 2012).

Decades of breeding for visually appealing traits and beneficial properties for large scale agricultural business has neglected the properties of flavor, resulting in the common complaint concerning the loss of the typical tomato flavor in fresh market tomatoes. The poor flavor may be attributed to the dramatic loss in genetic diversity in the commercial varieties (Klee and Tieman 2013; Mathieu et al. 2009). Today's breeders face a challenge of re-introducing the complex trait of flavor, which is the result of a combination of multiple alleles. Concomitantly, commercial hybrids have to be continuously bred for novel disease resistance genes, as the average effectiveness of a new resistance allele is projected to last for no more than five years until the pathogen overcomes the resistance. While heirloom lines of *S. lycopersicum* still have many positive flavor-related alleles, it is difficult to cherry-pick those qualities to avoid the negative disease susceptible alleles, thus the wild species are the most valuable source for reintroducing variation for useful lost traits.

Tomato genome sequencing

The tomato sequencing project began in 2004 by an international consortium of participants from 14 countries. The accession chosen for the genome sequencing was 'Heinz 1706', a processing inbred cultivar, which has a number of known introgressions from wild relatives, including resistance genes for Fusarium and Verticillium wilt (Ozminkowski 2004). Initially the approach was sequencing a BAC tiling path of the euchromatin, which contains more than 90% of the genes, but spans less than 25% of the 900 Mb genome (Mueller et al. 2009). While the BAC sequencing approach yielded high quality reads anchored to chromosomes using a FISH map and high-density genetic map, it could not resolve the repeat-rich heterochromatic regions. In 2009, with more than 1,200 BACs sequenced, whole-genome 454 sequencing was used to the BAC sequences, and provid higher coverage for assembling the entire genome. The genome was annotated by the International Tomato Annotation Group (ITAG), providing predicted gene models, proteins, cDNA alignments, CDs, and repeats. Both the genome assembly and the annotations are versioned and can be accessed at the sol genomics network website (Bombarely et al. 2011; Tomato Genome Consortium 2012).

The tomato sequencing consortium also released a draft genome for *S. pimpinellifolium*, and determined an estimated divergence of 0.6% compared to the 'Heinz 1706' reference genome, as well as several putative *S. pimpinellifolium* introgressed genome regions in the 'Heinz 1706' genome, showing its breeding history involved crossing with the wild relative. This study is a starting point for facilitating breeding programs using introgression analysis and introducing beneficial alleles from tomato wild species into commercial germplasm.

Published research enabled by tomato genome

The availability of the high quality *S. lycopersicum* 'Heinz 1706' genome has greatly promoted the development of tomato research in many areas. It has been utilized in many studies since its public release, predominantly in screening and identifying candidate genes related mostly to fruit development and ripening processes, databases and bioinformatics studies based on the tomato sequence, and serving as a reference genome for other Solanaceae species. These studies (Table 1, Figure 1) provide a proof of concept for the capacity of the tomato genome to serve as a true reference genome for closely-related Solanaceae and to further enable the research of orthologous genes and gene families, gene mapping, and genome evolution both within tomato and also amongst more divergent plants. Also, the rapid development in sequencing technology is shifting modern breeding, from multi-generation marker-assisted selection to genome-wide approaches for improving crops, as well as moving basic research towards systems approach studies of gene networks, transcriptomics, proteomics, and epigenomics. The following will discuss specific work enabled by the tomato genome sequence and insight gained from these studies.

Assembly guidance and benchmarks

Next generation sequencing produces short reads that are difficult to assemble without the availability of a quality reference genome from a closely related species. The tomato genome was first used for reference-guided assembly with *S. pimpinellifolium* Illumina reads, allowing for SNP detection between the two genomes and information concerning divergence of the presumed closest relative to cultivated tomato (Tomato Genome Consortium 2012). A number of other genomes are being assembled using this approach including *S. lycopersicum* lines and close relatives, such as the Tomato 150 genomes project (Finkers and van Heusden 2013), and the SOL-100 project (http://solgenomics.net/organism/sol100/view) (see Future Research section).

The utility of the tomato genome assembly is also demonstrated by its use as a benchmark in other genome projects and as a test set for bioinformatics program development. The number of predicted genes in tomato has been used as a guideline in assemblies of sequences from other related species, such as in a transcriptome assembly of *Centaurea solstitialis* (Dlugosch et al. 2013) to determine the completeness of the assembly. To compare benchmarks of cloud computing using

Table 1. Research topics enabled by the tomato reference genome, and the primary outcomes.

Topic	Outcomes	References
Assembly guidance and benchmarks	Genome and transcriptome assembly Gene prediction	Dlugosch et al. 2013, Cohen et al. 2013, Asamizu et al. 2012
Gene annotation	Gene location, structure, and function. RNaseq annotation	Moummou et al. 2012, Yarmolinsky et al. 2013, Kachanovsky et al. 2012, Chalivendra et al. 2013, Fujisawa et al. 2013, Liu et al. 2013
Epigenetics and expression	Genomic methylation miRNA and transcript identification Tissue specific expression Gene and networks prediction Protein expression	Zhong et al. 2013, Karlova et al. 2013, Van Vu et al. 2012, Ruzicka et al. 2012, Huang et al. 2013, Gupta et al. 2013, Kumar et al. 2012, Tzfadia et al. 2012, Quadrana et al. 2013, Sablok et al. 2013, Cigliano et al. 2013, Osorio et al. 2013, Hendelman et al. 2013
Phenotype to genotype	Trait-specific marker development Gene mapping and expression QTL analysis SNP location, linking with gene function	Sharlach et al. 2013, Sim et al. 2012b, Kadirvel et al. 2013, MacAlister et al. 2012, Chibon et al. 2012, Asamizu et al. 2012
Gene families	Gene family prediction Genome distribution Phylogenetic analysis Database framework	Huang et al. 2013, Sakamoto et al. 2012, Pirrello et al. 2012, Villagarcia et al. 2012, Sanseverino et al. 2010
Comparative genomics	Genome polymorphism Candidate gene prediction Resequencing Gene and sequence conservation Comparative mapping Orthologs mapping	Tomato Genome Consortium 2012, Asamizu et al. 2012, Di Matteo et al. 2013, Andolfo et al. 2013, Góngora-Castillo et al. 2012, Shirasawa et al. 2010, Sugita et al. 2013, Wei et al. 2012, Wang et al. 2013, Shahin et al. 2012, Goulet et al. 2012, De Smet et al. 2013

Figure 1. Word cloud generated from titles and keywords of peer-reviewed papers citing the tomato genome. Size of each word is proportional to its rate of occurrence. The cartoon was generated by http://www.tagxedo.com.

GenomeThreader, the tomato genome was used as a test dataset for gene structure prediction (Cohen et al. 2013).

Gene annotation

Gene annotation can include analysis of gene location and structure as well as work to predict gene function based on sequence similarity to genes of known function. The tomato reference assembly was useful in determining the genomic location of *SlscADH1*, an alcohol dehydrogenase involved in fruit ripening (Moummou et al. 2012). In another study, analysis of the genome revealed a single copy of the *sulfite reductase* (*SiR*) gene that is necessary for protecting against sulfite toxicity (Yarmolinsky et al. 2013). The sequence of the gene *phytoene synthase 1* (*PSY1*) from yellow-fruited mutant was compared to the 'Heinz 1706' sequence to determine there were no defects in the gene that may responsible for its phenotype (Kachanovsky et al. 2012).

Several studies have used the tomato genome data to predict gene function both within tomato and in related species. Proteins found in the stigmas and styles from key developmental stages in these tissues aimed to identify protein changes associated with the onset of reproductive barriers. Proteins in the study were identified using predicted proteins from the tomato dataset (Chalivendra et al. 2013). *RIPENING INHIBITOR* (*RIN*) target loci found in a CHiP-chip study were identified and positioned by mapping these sequences back to the tomato genome (Fujisawa et al. 2013). Additionally, genomic and protein data from tomato was used for the annotation of assembled RNA-seq transcripts from a tomato relative, *Capsicum frutescens* (Liu et al. 2013).

Epigenetics and gene expression

Both epigenetics and gene expression are currently prevalent areas of research, both of which benefit greatly from the tomato genome sequence. For example, genome resequencing was used in conjunction with bisulfite sequencing to study changes in genomic methylation during fruit ripening in tomato (Zhong et al. 2013). It was found that 1% of the genome is differentially methylated in fruit and binding sites for the ripening-inhibitor *RIN* gene are frequently found in these regions (Zhong et al. 2013). To study potential microRNAs targets in gene silencing, the annotation available for the tomato genome allowed for identification of miRNA-mRNA pairs throughout the genome (Karlova et al. 2013). It has also been advantageous to search the genome sequence to ensure there are no similar sequences that may results in off-target effects (Van Vu et al. 2012).

RNA-seq technology is now commonly used in gene expression analysis (Strickler et al. 2012), but most analysis pipelines require a high quality reference genome or transcriptome as a template for read mapping. To this end, the tomato genome has been useful in many studies. For example, reads from wild type and mutant tomatoes grown in soil where arbuscular mycorrhizal fungi were present were mapped to the genome to ensure the reads were from tomato and to identify novel transcripts (Ruzicka et al. 2012). *SUN*, *OFP*, *GABBY* transcription factor expression was analyzed by mapping reads to the tomato genome to determine that some may exhibit tissue-specific expression and there are chromosomal locations enriched with these genes (Huang et al. 2013). A number of novel genes were found to have auxin or cytokinin-induced differential expression, in root tissues of *S. lycopersicum* 'Micro-Tom' as deduced by quantifying reads mapped to the reference (Gupta et al. 2013). Some of these genes are orthologous to cytokinin or auxin-regulated genes in other plant species (Gupta et al. 2013). RNA-seq reads can also be mapped back to predicted ITAG2.3 cDNAs from the tomato genome data using next generation aligners, such as Bowtie (Kumar et al. 2012).

Insights in expression can also be gained through microarray experiments and identification of gene regulatory elements. To explore gene expression networks, microarray probes from the tomato Affymetrix chip were mapped to the tomato genome and the corresponding genes were identified for use in a pathway prediction tool called MORPH (Tzfadia et al. 2012). Regulatory elements for 47 genes involved in tocopherol synthesis were predicted based on ITAG2.3 annotations and tomato genome sequence. Expression networks were generated and as a result, several key genes were identified that may be good targets for improving the nutritional value of tomato (Quadrana et al. 2013). Lastly, to study protein expression, translation usage bias was calculated for mutaions in the varieties 'Micro-Tom' and 'Heinz 1706' (Sablok et al. 2013).

Data generated directly as part of the tomato genome project has proven useful in expression analysis for other projects. Expression data generated in the tomato genome project was used to look at expression of histone modifiers and the map position for these genes was determined for *S. pennellii* introgression lines (Cigliano et al. 2013). The tomato genome project RNA-seq data was also used to look at expression of *SlNADP-ME* genes in different stages of fruit development to better understand their role in ripening (Osorio et al. 2013), and to identify and observe expression of two tomato *Argonaute1* (*SlAGO1*) transcripts in different tomato tissues (Hendelman et al. 2013).

Phenotype to genotype studies

The ability to determine the genomic location of a DNA sequence is an asset when developing markers for use in trait mapping. The tomato physical map was used to identify positions for markers linked to yellow leaf curl (*Ty*) resistance from a wild species introgression (Kadirvel et al. 2013), and also for designing markers such as those used in fine mapping a bacterial spot resistance locus derived from a *S. pennellii* introgression in cultivated tomato (Sharlach et al. 2013). Markers were also developed based on the *S. pimpinellifolium* assembly to map *TERMINATING FLOWER* (*TMF*), a gene involved in flowering, and ITAG2.3 coding sequence annotations were used to observe expression of the mapped gene (MacAlister et al. 2012). The functionality of the QTL tool, Marker2sequence, was demonstrated using the tomato genome annotations to identify genes linked to QTL of interest (Chibon et al. 2012).

To find the physical locations of SNPs on the SolCAP tomato array, sequences flanking SNPs were mapped back to the tomato genome assembly using BLAST (Sim et al. 2012b). These SNPs were then used to look at population structure within cultivated tomato and *S. pimpinellifolium*. A similar approach to locate SNPs was used with another SNP chip dataset that included many tomato lines, hybrids, and some wild species (Asamizu et al. 2012). SNPs that occurred near genes were further analyzed to determine possible effects on gene function.

Gene families

The tomato genome sequence can be mined to find gene families, analyze gene relationships, and determine gene family distribution across the chromosomes. For example, protein homology can also be used to identify genes of interest such as in a study that identified putative Receptor-like Kinases (RLKs) found by BLAST of the ITAG2.3 protein dataset (Sakamoto et al. 2012). Putative *SUN*, *OFP*, *GABBY* transcriptions factors in tomato predicted proteins and genome assembly, were identified for phylogenetic analysis and also to determine these transcription factors often cluster within the genome (Huang et al. 2013). Phylogenies were produced for both *Ethylene Response Factor* (*ERF*) factors using the tomato genome to identify 146 putative ERF proteins (Pirrello et al. 2012), and for *ERECTA* genes, which are involved in plant architecture and found in tomato and a number of other species (Villagarcia et al. 2012).

Gene family databases can also be created based on data gained from the tomato genome. Specifically, members of the disease resistance *R* gene family found in tomato genome and other species were used to create a database for this gene type (Sanseverino et al. 2010).

Comparative genomics

The availability of the tomato genome sequence is pivotal for Solanaceae and fleshy fruit comparative genomics as well as studies amongst more divergent species. Within the tomato clade, genome comparisons have been useful in determining information about variation both within cultivated and in comparison to wild tomatoes. For example, 'Micro-Tom' BAC-end sequences were mapped to the 'Heinz 1706' genome as a reference to find polymorphism and rearrangement between the two genome (Asamizu et al. 2012). As expected, more polymorphism was observed in the heterochromatin. Two putative genome rearrangements were found, an inversion on chromosome 2 and an inversion and translocation on chromosome 3. Another study predicted genes from a *S. pennellii* introgression involved in higher expression of fruit phenolics based on 'Heinz 1706' annotations (Di Matteo et al. 2013). Genome comparison can also be done through resequencing studies, i.e. short read sequencing at a low depth of coverage, both within *S. lycopersicum* and with closely related species by mapping reads to the 'Heinz 1706' reference genome. This has been demonstrated by mapping *S. pimpinellifolium* reads to 'Heinz 1706' to calculate SNPs (Tomato Genome Consortium 2012).

The tomato genome has been used in comparison with a number of other Solanaceae genomes. Potato studies have been particularly information-rich, since full genome sequence is available for this species as well (Potato Genome Sequencing Consortium et al. 2011). Identification and chromosomal distribution of tomato and potato pathogen recognition genes was performed using ITAG annotations and it was observed many of these genes are undergoing adaptive divergence (Andolfo et al. 2013). Orthologous clusters of genes involved in geminivirus response were generated using tomato, *C. annuum*, and *S. tuberosum* to find conservation of genes involved in this disease response pathway (Góngora-Castillo et al. 2012). Other studies have found the tomato genome data to be useful in generating comparative maps between species. SSR markers were created for comparative mapping between pepper and tomato (Shirasawa et al. 2010) and COS markers between pepper and tomato were anchored for linkage map construction (Sugita et al. 2013). DNA sequences from *Physalis* were used to find orthologous matches to the tomato genome allowing for the conclusion that sequence similarity between the two species is sufficient for marker transfer from tomato (Wei et al. 2012).

Interesting information can be obtained by comparing tomato to more divergent species. By searching tomato sequence to identify chromoplast proteins, putative homologs were identified and subsequently compared to five other carotenoid-rich plant species, including watermelon, carrot, orange cauliflower, red papaya, and red bell pepper (Wang et al. 2013). Statistics on potato and tomato orthologous group comparisons were useful in comparing results in a study with lily and tulip to determine that the number of orthologous groups shared in the Liliaceae may to be lower than in tomato-potato comparisons, likely due to a greater divergence time in the Liliaceae family (Shahin et al. 2012). The tomato genome was used to deduce that an insertion found in the esterase gene, *SlCXE1*, may be Solanum-specific based on sequence comparisons to species represented in the NCBI database (Goulet et al. 2012). Another study focused on single copy genes found in tomato, Chinese cabbage, and banana. These genes often exhibit higher expression and sequence conservation and may be resistant to duplication since they are often essential housekeeping genes (De Smet et al. 2013).

Future improvements to the reference genome

There is no doubt the availability of a high quality reference tomato genome has had a huge impact on Solanaceae research. Despite this observation, as with any whole genome sequencing project, there is still room for improvement. Many gaps still exist in the genome assembly and a substantial number of genes have not been placed on an actual chromosome, hence the existence of a chromosome 0 in the assembly. It is hoped that FISH data will aid in locating some of these regions. Some important genes have been lost during domestication and breeding that are known to be represented in wild tomato species, such as some parts of the *PTO* locus (Lin and Martin 2007), and the *SUN* locus which is completely missing from 'Heinz 1706', but exists in other tomato varieties (Xiao et al. 2008). Furthermore, *S. lycopersicum* exhibits reduced genetic diversity likely due to the bottlenecks it underwent during domestication (Miller and Tanksley 1990). As mentioned, introgression of wild species in breeding has helped to alleviate the lack of diversity in tomato and current and future efforts will aim to gain a better understanding of these species and the diversity they capture. While the tomato genome has been useful for reference-guided assembly of near relatives, reference genomes for more divergent wild relatives will be needed for comparative studies. It is important to note, these wild tomato species tend to be heterozygous, often an effect of obligate outcrossing, which can complicate genome assembly. Lastly, there currently exists whole genome data for a limited number of cultivated and wild species, so it will be necessary to generate more complete data concerning variation within tomato species and wild tomato accessions so that population genetics-based models can be properly applied.

Resources

In addition to sequenced genomes, a number of resources exist for the Solanaceae research community, which can aid and impact research in a variety of ways (Table 2). The sol genomics network database contains a number of tools useful both to breeders and basic researchers (Bombarely et al. 2011). Many maps and markers can be found there, as well as a genome browser for the tomato genome and annotations, pages concerning information for tomato loci, and a BLAST search tool. A centralized area for breeders, called the Breeder's Toolbox, provides an easy access point for a number of breeding tools and information. Members of the Solanaceae research community can edit some data types, such as genetic loci, and phenotyped accessions, allowing information to be easily disseminated to other researchers (Menda et al. 2008). Another part of the site consists of the SolCyc pathway databases, which contain detailed information on metabolic pathways in the Solanaceae. The sol genomics database is clade-oriented, meaning that it contains data for a number of other related species, as well as common model plant species. Furthermore, ontology terms have been developed for the Solanaceae to provide a standardized vocabulary for describing their phenotypes (Cooper et al. 2013; Jung et al. 2011; Menda et al. 2008).

A number of other databases exist that store Solanaceae genetic, phenotypic, and phylogenetic data (Table 2). Some databases contain an abundance of expression data, such as the Sequence Read Archive (SRA) housed at the NCBI (Leinonen et al. 2011), and the Tomato Functional Genomics Database (Fei et al. 2011), while others focus on the results of tomato-related projects. The SolCAP project has mined SNPs between a large number of cultivated tomato lines and some wild species and has also generated extensive phenotypic data for both tomato and potato (Hamilton et al. 2011a; Hamilton et al. 2012; Merk et al. 2012). An array has been created from the SNPs identified in this project and the SNP data can also be found in the sol genomics database. Tomato stocks and wild accessions can be found and obtained by querying the databases at the Tomato Genetic Resource Collection (http://tgrc.ucdavis.edu) as well as the Germplasm Resources Information Network (USDA database, http://www.ars-grin.gov/).

In addition to databases, several other genetic and genomic resources are available for the tomato community. Of particular use to breeding, linkage maps have been created mainly from populations derived from crosses to wild species. These studies have allowed a number of quanitative trait loci (QTL) affecting important traits, such as yield, morphology, and fruit characterisitics to be identified for use in tomato breeding (Table 2).

Future of Solanaceae research

Genome sequencing technology enables more rapid breeding for desirable traits using previously known genetic knowledge, and utilizing the high level of genome conservation in the Solanaceae family. As more accessions are re-sequenced using the tomato reference genome, it is expected to impact breeding programs, addressing current challenges in ensuring food security in a world of climate change. Currently, several projects exist that attempt to explore the limits of genetic diversity in cultivated tomato, wild tomato species, and other Solanaceae family members. The 150 Tomato Resequencing Project aims to re-sequence many tomato lines including 10 heirlooms, 43 landraces, and 30 wild accessions, in addition to three wild species, which will be sequenced at higher coverage for *de novo* genome assembly, and 60 F8 generation individuals from a *S. pimpinellifolium* RILs (http://www.tomatogenome.net/) (Finkers and van Heusden 2013). The data generated from these efforts should be publicly available in the near future. A community effort to create a common Solanaceae-based genomic framework has resulted in the creation of a project called SOL-100 for sequencing genomes encompassing the phylogenetic diversity of the group (http://solgenomics.net/organism/sol100/view).

Table 2. Major resources of genomic, genetic, and phenotypic tools for tomato and close relatives.

Topic	Resource	Data types	References
Databases	sol genomics network (SGN)	Solanaceae genomics, genetic, phenotypic, taxonomic information and tools.	Bombarely et al. 2011
	Solanaceae Source	Phylogenetic and taxonomic information	PBI Solanum Project. 2013
	Tomato Functional Genomics Database	Expression and metabolites	Fei et al. 2011
	SolCyc	Solanaceae biochemical pathways	Bombarely et al. 2011
	NCBI Sequence Read Archive (SRA)	580 Solanaceae experiments (RNA, DNA, and whole genome reads).	Leinonen et al. 2011
	TGRC	Tomato monogenic mutants (>800). Wild tomato accessions (>1,000). Tomato cultivars (>200)	http://tgrc.ucdavis.edu/
	GRIN	Thousand of accessions for tomato, potato, pepper, eggplant, petunia, and tobacco	http://www.ars-grin.gov/
	EU-SOL	Core collection of >7,000 tomato accessions from public and private collections	https://www.eu-sol.wur.nl/
	SolCAP	~8,000 verified SNPs for tomato and potato	Sim et al. 2012b, Hamilton et al. 2011b
	Tomato M82 mutants	EMS and Fast-neutron monogenic mutants	Menda et al. 2004
	LycoTILL	TILLING of a tomato EMS population	Minoia et al. 2010
	UCD TILLING core	*Solanum lycopersicum*, cv. VFNT Cherry mutant populations	http://tilling.ucdavis.edu/index.php/Tomato_Tilling
	UTiLLdb	Tomato and pepper EMS mutants	Dalmais et al. 2008
	TOMATOMA	Tomato 'Micro-Tom' variety population	Okabe et al. 2011
	MiBase	EST and unigene dataset from 'Micro-Tom'	Yano et al. 2006
	KafTom	The Kazusa Micro-Tom full-length cDNA Site	http://www.pgb.kazusa.or.jp/kaftom/
	Kappa-View 4 SOL	Metabolic pathway maps of Solanaceae and related species	http://kpv.kazusa.or.jp/kpv4-sol
	MassBase	A mass spectral tag archive for metabolomics.	http://webs2.kazusa.or.jp/massbase/
	Nijmegen Experimental Garden and Genebank	ex situ plant collection of non-tuberous Solanaceae species in the world	http://www.ru.nl/bgard/
Genomes	*S. lycopersicum*	The reference tomato genome of Heinz1706	Tomato Genome Consortium 2012
	S. pimpinellifolium	De novo 40X draft genome	Tomato Genome Consortium 2012
	Solanum tuberosum group phureja	The reference potato genome	Potato Genome Sequencing Consortium et al. 2011
	Nicotiana benthamiana	63X draft genome	Bombarely et al. 2012
	MicroTom tomato	Resequencing of a model tomato cultivar	http://trace.ddbj.nig.ac.jp/DRASearch/study?acc=DRP000312
Introgression lines	*S. pennellii*	76 introgression lines, M82 background	Eshed and Zamir 1995
	S. habrochaites LA1777	99 NILs and BCRILs in E6203 background	Monforte and Tanksley 2000
	S. habrochaites LYC4	30 introgression lines, Money Maker background	Finkers et al. 2007
	S. habrochaites LA407	64 IBC lines	Francis et al. 2001
	S. lycopersicoides LA2951	90 introgression lines, VF36 background	Canady et al. 2005
Linkage maps	Eggplant	F2 mapping populations	Doganlar et al. 2002b, Doganlar et al. 2002a, Brand et al. 2012, Wu et al. 2009b, Wu et al. 2006
	Pepper	F2 mapping populations	Wu et al. 2006, Wu et al. 2009a
	Petunia	F2 mapping populations	http://solgenomics.net
	Potato	BC1 map, meta-QTL consensus map	Tanksley et al. 1992, Danan et al. 2011
	Tobacco	F2 mapping populations	Bindler et al. 2007
	Tomato	F2, BC1, and BC2 populations, introgresion lines, QTL maps, SolCAP markers, FISH, FCP, physical BAC map, AGP, pachytene chromosomes, ITAG.	Fulton et al. 2002, Eshed and Zamir 1995, Shirasawa et al. 2010, Sim et al. 2012a, Ashrafi et al. 2009, Stack et al. 2009, Doganlar et al. 2002c, Bernacchi and Tanksley 1997, Jimenez-Gomez et al. 2007, van der Knaap and Tanksley 2003, Van Der Knaap et al. 2002, Tanksley et al. 1996, Grandillo and Tanksley 1996
QTLs	SolQTL	In silico analysis of QTLs,	Tecle et al. 2010
	Real Time QTL	QTLs scored for S. pennellii introgression lines	Gur et al. 2004

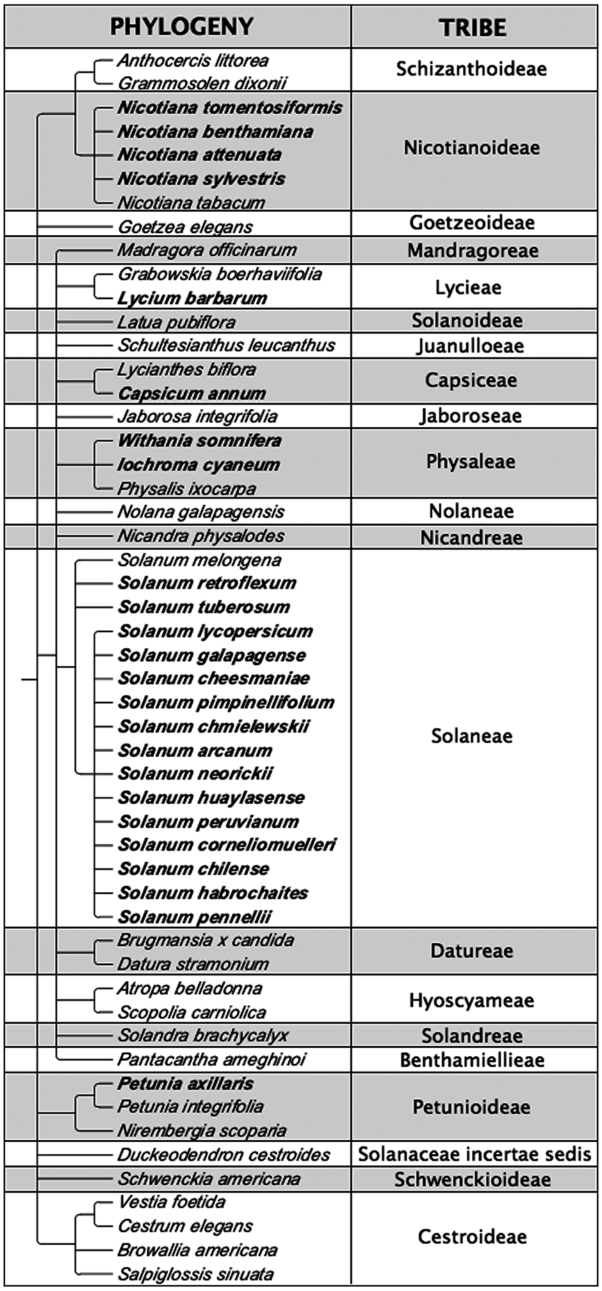

Figure 2. Cladogram representing the Solanaceae comprising the SOL-100. Species with a sequencing project are shown in bold. The tree was generated with data from NCBI Taxonomy (Federhen 2012).

Figure 2 depicts a subset of the species representing the major sub-clades of the family. The Solanaceae family is one of the most morphologically diverse plant families, with more than 3,000 described species (Knapp et al. 2004; PBI Solanum Project 2013), and worldwide distribution. In addition to tomato, the family includes many plants of high economic value, such as food crops (potato, eggplant, and pepper), ornamentals (Petunia and Schizanthus), and plants with unique biochemical metabolites, many of which are medicinal alkaloids, such as nicotine, atropine, and scopolamine. The goals of the project are to study how a common set of genes gave rise to such phenotypically diverse species in the taxon, and to understand the genetic basis for plant biodiversity as well as to generate sequence information for the linking of both genetic and physical maps of sequenced species. This will be used in the genetic mapping of traits especially to compare information gained from phenotyping studies in other species to a lesser-studied species. Sequence for several SOL-100 species, such as *Nicotiana benthamiana* (Bombarely et al. 2012) and *S. pimpinellifolium* (Tomato Genome Consortium 2012), is already available, with several other projects currently underway (Table 2). By uploading data to the SOL-100 project page, researchers preparing to sequence a Solanaceae species can inform other researchers of their plans.

As more Solanaceae genome sequences become available, will be of outmost importance the generation of high quality annotations for these genomes. Also, improvements can continue to be made to the currently available genomes. For the tomato genome, the community has contributed improvements to the genome based on new research. For example, some genes were annotated with incorrect exon-intron boundaries, which users have reported, and these updates will be incorporated in future genome releases. Identification of gene regulatory regions, such as promoter locations, would also be a useful information for the improvement of gene annotation, understanding levels of control on gene expression, and also for interesting studies of regulatory element evolution. Additionally, knock-out lines, such as the SALK lines available for *Arabidopsis* (Alonso et al. 2003), would prove immensely useful in determining gene function and phenotypic effect.

An important step in tomato research will be linking genotype and phenotype information in a form that will be useful to breeding programs. Genome-wide association mapping (GWAS) is a commonly used method to link phenotype data to genotype, and a large amount of SNP data has been generated as a result of the SolCAP project providing a valuable resource (Sim et al. 2012a). Although the low variation in cultivated tomato and relatively higher linkage disequilibrium may reduce the resolution of GWAS, some recent work has shown the potential of genome wide association mapping in tomato (Ranc et al. 2012; Sim et al. 2012c; Xu et al. 2013).

Only one year after the publication of the genome, it has already had a tremendous impact on tomato research. Now that the tomato genome is available, additional data is rapidly being generated. Data such as expression, eQTLS, metabolomic data, and epigenetics all interplay to result in phenotype and it will be critical to develop methods that link these various data types so that we can gain a greater understanding of tomato and how it may be improved, especially in terms of flavor, disease resistance, and adaptations to climate change.

Acknowledgements

We thank Aureliano Bombarely and Noe Fernandez for critical review of the manuscript.

References

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301: 653–657

Andolfo G, Sanseverino W, Rombauts S, Peer Y, Bradeen J, Carputo D, Frusciante L, Ercolano M (2013) Overview of tomato (*Solanum lycopersicum*) candidate pathogen recognition genes reveals important Solanum R locus dynamics. *New Phytol* 197: 223–237

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

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

Bai Y, Lindhout P (2007) Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Ann Bot* (*Lond*) 100: 1085–1094

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

Bindler G, van der Hoeven R, Gunduz I, Plieske J, Ganal M, Rossi L, Gadani F, Donini P (2007) A microsatellite marker based linkage map of tobacco. *Theor Appl Genet* 114: 341–349

Blanca J, Canizares J, Cordero L, Pascual L, Jose Diez M, Nuez F (2012) Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS ONE* 7: e48198

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

Bombarely A, Menda N, Tecle IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA (2011) The Sol Genomics Network (solgenomics.net): Growing tomatoes using Perl. *Nucleic Acids Res* 39 (Database issue): D1149–D1155

Canady MA, Meglic V, Chetelat RT (2005) A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. *Genome* 48: 685–697

Chalivendra SC, Lopez-Casado G, Kumar A, Kassenbrock AR, Royer S, Tovar-Mèndez A, Covey PA, Dempsey LA, Randle AM, Stack SM, et al. (2013) Developmental onset of reproductive barriers and associated proteome changes in stigma/styles of Solanum pennellii. *J Exp Bot* 64: 265–279

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

Cigliano RA, Sanseverino W, Cremona G, Ercolano MR, Conicella C, Consiglio FM (2013) Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles. *BMC Genomics* 14: 57

Clarkson JJ, Knapp S, Garcia VF, Olmstead RG, Leitch AR, Chase MW (2004) Phylogenetic relationships in Nicotiana (Solanaceae) inferred from multiple plastid DNA regions. *Mol Phylogenet Evol* 33: 75–90

Cohen J, Filippis I, Woodbridge M, Bauer D, Hong NC, Jackson M, Butcher S, Colling D, Darlington J, Fuchs B, et al. (2012) RAPPORT: running scientific high-performance computing applications on the cloud. *Philos Trans A Math Phys Eng Sci* 371: 20120073

Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, Preece J, Athreya B, Mungall CJ, Rensing S, et al. (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol* 54: e1

Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savois V, Aubert G, Brunaud V, de Oliveira Y, Guichard C, et al. (2008) UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol* 9: R43

Danan S, Veyrieras JB, Lefebvre V (2011) Construction of a potato consensus map and QTL meta-analysis offer new insights into the genetic architecture of late blight resistance and plant maturity traits. *BMC Plant Biol* 11: 16

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

De Smet R, Adams KL, Vandepoele K, Van Montagu MC, Maere S, Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. *Proc Natl Acad Sci USA* 110: 2898–2903

Di Matteo A, Ruggieri V, Sacco A, Rigano MM, Carriero F, Bolger A, Fernie AR, Frusciante L, Barone A (2013) Identification of candidate genes for phenolic accumulation in tomato fruit. *Plant Sci* 205–206: 87–96

Dlugosch KM, Lai Z, Bonin A, Hierro J, Rieseberg LH (2013) Allele Identification for Transcriptome-Based Population Genomics in the Invasive Plant *Centaurea solstitialis*. *G3 (Bethesda)* 3: 359–367

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) Conservation of gene function in the solanaceae as revealed by comparative mapping of domestication traits in eggplant. *Genetics* 161: 1713–1726

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics* 161: 1697–1711

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

Duan H, Zhu Y, Qi D, Li W, Hua X, Liu Y, Deng X (2012) Comparative study on the expression of genes involved in carotenoid and ABA biosynthetic pathway in response to salt stress in tomato. *J Integr Agric* 11: 1093–1102

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

Federhen S (2012) The NCBI Taxonomy database. *Nucleic Acids Res* 40 (Database issue): D136–D143

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

(Database issue): D1156–D1163

Finkers R, van Heusden AW, Meijer-Dekens F, van Kan JA, Maris P, Lindhout P (2007) The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to *Botrytis cinerea*. *Theor Appl Genet* 114: 1071–1080

Finkers R, van Heusden S (2013) The 150+ Tomato Genome (re-) Sequence project; lessons learned and potential applications. *Tomato Breeder's Roundtable*

Francis DM, Kabelka E, Bell J, Franchino B, Clair DS (2001) Resistance to bacterial canker in tomato (*Lycopersicon hirsutum* LA407) and its progeny derived from crosses to L-esculentum. *Plant Dis* 85: 1171–1176

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

Frary A, Nesbitt TC, Frary A, Grandillo S, Van Der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB (2000) fw2. 2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289: 85–88

Frary A, Doganlar S, Daunay MC, Tanksley SD (2003) QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor Appl Genet* 107: 359–370

Fray R, Grierson D (1993) Identification and genetic-analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. *Plant Mol Biol* 22: 589–602

Fridman E, Liu Y, Carmel-Goren L, Gur A, Shoresh M, Pleban T, Eshed Y, Zamir D (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. *Mol Genet Genomics* 266: 821–826

Fujisawa M, Nakano T, Shima Y, Ito Y (2013) A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. *Plant Cell* 25: 371–386

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

Góngora-Castillo E, Ibarra-Laclette E, Trejo-Saavedra DL, Rivera-Bustamante RF (2012) Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). *Virol J* 9: 1–16

Goulet C, Mageroy MH, Lam NB, Floystad A, Tieman DM, Klee HJ (2012) Role of an esterase in flavor volatile variation within the tomato clade. *Proc Natl Acad Sci USA* 109: 19009–19014

Grandillo S, Tanksley SD (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species Lycopersicon pimpinellifolium. *Theor Appl Genet* 92: 935–951

Gur A, Semel Y, Cahaner A, Zamit D (2004) Real Time QTL of complex phenotypes in tomato interspecific introgression lines. *Trends Plant Sci* 9: 107–109

Hamilton J, Hansey C, Whitty B, Stoffel K, Massa A, Van Deynze A, De Jong W, Douches D, Buell CR (2011) Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12: 302

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

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genom* 12: 302

Hendelman A, Kravchik M, Stav R, Zik M, Lugassi N, Arazi T (2013) The developmental outcomes of P0-mediated ARGONAUTE destabilization in tomato. *Planta* 237: 363–377

Huang Z, Van Houten J, Gonzalez G, Xiao H, van der Knaap E (2013) Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Mol Genet Genom* 288: 1111–1129

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

Jung S, Menda N, Redmond S, Buels RM, Friesen M, Bendana Y, Sanderson L, Lapp H, Lee T, MacCallum B, et al. (2011) The Chado Natural Diversity module: a new generic database schema for large-scale phenotyping and genotyping data. *Database* (*Oxford*) 2011: bar051

Kachanovsky DE, Filler S, Isaacson T, Hirschberg J (2012) Epistasis in tomato color mutations involves regulation of phytoene synthase 1 expression by *cis*-carotenoids. *Proc Natl Acad Sci USA* 109: 19021–19026

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

Karlova R, van Haarst JC, Maliepaard C, van de Geest H, Bovy AG, Lammers M, Angenent GC, de Maagd RA (2013) Identification of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis. *J Exp Bot* 64: 1863–1878

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

Knapp S, Bohs L, Nee M, Spooner DM (2004) Solanaceae—a model for linking genomics with biodiversity. *Comp Funct Genomics* 5: 285–291

Knapp S (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J Exp Bot* 53: 2001–2022

Knapp S (2010) Four new vining species of solanum (Dulcamaroid clade) from Montane habitats in tropical America. *PLoS ONE* 5: e10502

Ku H, Vision T, Liu J, Tanksley S (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny. *Proc Natl Acad Sci USA* 97: 9121–9126

Kumar R, Tyagi AK, Sharma AK (2011) Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development. *Mol Genet Genom* 285: 245–260

Kumar R, Ichihashi Y, Kimura S, Chitwood DH, Headland LR, Peng J, Maloof JN, Sinha NR (2012) A high-throughput method for Illumina RNA-Seq library preparation. *Front Plant Sci* 3: 202

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

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

Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration (2011) The sequence read archive. *Nucleic Acids Res* 39 (Database issue): D19–D21

Li W, Chetelat RT (2010) A pollen factor linking inter- and intraspecific pollen rejection in tomato. *Science* 330: 1827–1830

Lieberman M, Segev O, Gilboa N, Lalazar A, Levin I (2004) The tomato homolog of the gene encoding UV-damaged DNA binding protein 1 (DDB1) underlined as the gene that causes the high pigment-1 mutant phenotype. *Theor Appl Genet* 108: 1574–1581

Lin N, Martin GB (2007) Pto- and Prf-mediated recognition of AvrPto and AvrPtoB restricts the ability of diverse Pseudomonas syringae pathovars to infect tomato. *Mol Plant Microbe Interact* 20: 806–815

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

Liu S, Li W, Wu Y, Chen C, Lei J (2013) De novo transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. *PLoS ONE* 8: e48156

MacAlister CA, Park SJ, Jiang K, Marcel F, Bendahmane A, Izkovich Y, Eshed Y, Lippman ZB (2012) Synchronization of the flowering transition by the tomato TERMINATING FLOWER gene. *Nat Genet* 44: 1393–1398

Mathieu S, Cin VD, Fei Z, Li H, Bliss P, Taylor MG, Klee HJ, Tieman DM (2009) Flavour compounds in tomato fruits: identification of loci and potential pathways affecting volatile composition. *J Exp Bot* 60: 325–337

Mazourek M, Pujar A, Borovsky Y, Paran I, Mueller L, Jahn MM (2009) A dynamic interface for capsaicinoid systems biology. *Plant Physiol* 150: 1806–1821

Menda N, Buels RM, Tecle I, Mueller LA (2008) A community-based annotation framework for linking solanaceae genomes with phenomes. *Plant Physiol* 147: 1788–1799

Menda N, Semel Y, Peled D, Eshed Y, Zamir D (2004) In silico screening of a saturated mutation library of tomato. *Plant J* 38: 861–872

Merk HL, Yarnes SC, Van Deynze A, Tong N, Menda N, Mueller LA, Mutschler MA, Loewen SA, Myers JR, Francis DM (2012) Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato germplasm. *J Am Soc Hortic Sci* 137: 427–437

Miller J, Tanksley S (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theor Appl Genet* 80: 437–448

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

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. *Genome* 43: 803–813

Moore S, Vrebalov J, Payton P, Giovannoni J (2002) Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *J Exp Bot* 53: 2023–2030

Moummou H, Tonfack LB, Chervin C, Benichou M, Youmbi E, Ginies C, Latché A, Pech J, Van der Rest B (2012) Functional characterization of SlscADH1, a fruit-ripening-associated short-chain alcohol dehydrogenase of tomato. *J Plant Physiol* 169: 1435–1444

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

Mueller LA, Lankhorst RK, Tanksley SD, Giovannoni JJ, White R, Vrebalov J, Fei Z, van Eck J, Buels R, Mills AA, et al. (2009) A snapshot of the emerging tomato genome sequence. *The Plant Genome* 2: 78–92

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

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer J (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In: Nee M, Symon D, Lester R, Jessop J (eds) *Solanaceae IV: advances in biology and utilization*. Kew: The Royal Botanic Gardens, Kew, pp 111–137

Osorio S, Vallarino JG, Szecowka M, Ufaz S, Tzin V, Angelovici R, Galili G, Fernie AR (2013) Alteration of the interconversion of pyruvate and malate in the plastid or cytosol of ripening tomato fruit invokes diverse consequences on sugar but similar effects on cellular organic acid, metabolism, and transitory starch accumulation. *Plant Physiol* 161: 628–643

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

PBI Solanum Project (2013) Solanaceae Source. http://www.nhm.ac.uk/solanaceaesource/

Peralta IE, Spooner DM (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (Solanum L. section Lycopersicon [Mill.] Wettst. subsection Lycopersicon). *Am J Bot* 88: 1888–1902

Peralta IE, Spooner DM (2000) Classification of wild tomatoes: A review. *Kurtziana* 28: 45–54

Pirrello J, Prasad BC, Zhang W, Chen K, Mila I, Zouine M, Latche A, Pech JC, Ohme-Takagi M, Regad F, et al. (2012) Functional analysis and binding affinity of tomato ethylene response factors provide insight on the molecular bases of plant differential responses to ethylene. *BMC Plant Biol* 12: 190

Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganal M, Zamir D, Lifschitz E (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* 125: 1979–1989

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

Quadrana L, Almeida J, Otaiza SN, Duffy T, Corrêa da Silva JV, de Godoy F, Asís R, Bermúdez L, Fernie AR, Carrari F, et al. (2013) Transcriptional regulation of tocopherol biosynthesis in tomato. *Plant Mol Biol* 81: 309–325

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

Rick CM (1978) The tomato. *Sci Am* 239: 76–87

Rick CM (1976) Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. *Genet Agr* 30: 249–259

Rick CM (1986) Reproductive isolation in the *Lycopersicon peruvianum* complex. *Solanaceae: Biology and Systematics* 477–495

Robbins MD, Sim S, Yang W, Van Deynze A, van der Knaap E, Joobeur T, Francis DM (2011) Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. *J Exp Bot* 62: 1831–1845

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

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to beta-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato. *Proc Natl Acad Sci USA* 97: 11102–11107

Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J* 17: 341–351

Ruzicka D, Chamala S, Barrios-Masias FH, Martin F, Smith S, Jackson LE, Barbazuk WB, Schachtman DP (2012) Inside arbuscular mycorrhizal roots–Molecular probes to understand the symbiosis. *Plant Genome* doi: 10.3835/plantgenome2012.06.0007

Sablok G, Wu X, Kuo J, Nayak KC, Baev V, Varotto C, Zhou F (2013) Combinational effect of mutational bias and translational selection for translation efficiency in tomato (*Solanum lycopersicum*) cv. Micro-Tom. *Genomics* 101: 290–295

Sakamoto T, Deguchi M, Brustolini OJ, Santos AA, Silva FF, Fontes EP (2012) The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense. *BMC Plant Biol* 12: 229

Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR (2010) PRGdb: a bioinformatics platform for plant resistance gene analysis. *Nucleic Acids Res* 38 (Database issue): D814–D821

Semel Y, Schauer N, Roessner U, Zamir D, Fernie AR (2007) Metabolite analysis for the comparison of irrigated and non-irrigated field grown tomato of varying genotype. *Metabolomics* 3: 289–295

Shahin A, van Kaauwen M, Esselink D, Bargsten JW, van Tuyl JM, Visser RG, Arens P (2012) Generation and analysis of expressed sequence tags in the extreme large genomes Lilium and Tulipa. *BMC Genom* 13: 640

Sharlach M, Dahlbeck D, Liu L, Chiu J, Jiménez-Gómez JM, Kimura S, Koenig D, Maloof JN, Sinha N, Minsavage GV, et al. (2013) Fine genetic mapping of RXopJ4, a bacterial spot disease resistance locus from Solanum pennellii LA716. *Theor Appl Genet* 126: 1–9

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

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, et al. (2012) 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. (2012) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS ONE* 7: e45520

Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes. *Taxon* 54: 43–61, *Solanum* L. section *Lycopersicon*, Mill., Wettst.

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

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

Strickler SR, Bombarely A, Mueller LA (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *Am J Bot* 99: 257–266

Sugita T, Semi Y, Sawada H, Utoyama Y, Hosomi Y, Yoshimoto E, Maehata Y, Fukuoka H, Nagata R, Ohyama A (2013) Development of simple sequence repeat markers and construction of a high-density linkage map of *Capsicum annuum*. *Mol Breed* 31: 909–920

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

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

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

Tzfadia O, Amar D, Bradbury LM, Wurtzel ET, Shamir R (2012) The MORPH algorithm: ranking candidate genes for membership in Arabidopsis and tomato pathways. *Plant Cell* 24: 4389–4406

USDA A National Genetic Resources Program *Germplasm Resources Information Network-(GRIN)*. [Online Database] National Germplasm Resources Laboratory, Beltsville, Maryland. Available: http://www.ars-grin.gov/

Van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theor Appl Genet* 107: 139–147

Van Der Knaap E, Lippman ZB, Tanksley SD (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. *Theor Appl Genet* 104: 241–247

Van Vu T, Roy Choudhury N, Mukherjee SK (2012) Transgenic tomato plants expressing artificial microRNAs for silencing the pre-coat and coat proteins of a begomovirus, Tomato leaf curl New Delhi virus, show tolerance to virus infection. *Virus Res* 172: 35–45

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

Verlaan M, Hutton S, Ibrahem R, Kormelink R, Visser R, Scott JW, Edwards JD, Bai Y (2013) The Tomato Yellow Leaf Curl Virus Resistance Genes *Ty-1* and *Ty-3* are allelic and code for DFDGD-class RNA–dependent RNA polymerases. *PLoS Genet* 9: e1003399

Villagarcia H, Morin A, Shpak ED, Khodakovskaya MV (2012) Modification of tomato growth by expression of truncated ERECTA protein from *Arabidopsis thaliana*. *J Exp Bot* 63:

6493–6504

Virginia Sanchez-Puerta M, Williams Masuelli R (2011) Evolution of nematode-resistant Mi-1 gene homologs in three species of Solanum. *Mol Genet Genomics* 285: 207–218

Wang Y, Yang Y, Fei Z, Yuan H, Fish T, Thannhauser TW, Mazourek M, Kochian LV, Wang X, Li L (2013) Proteomic analysis of chromoplasts from six crop species reveals insights into chromoplast function and development. *J Exp Bot* 64: 949–961

Wei J, Hu X, Yang J, Yang W (2012) Identification of single-copy orthologous genes between physalis and *Solanum lycopersicum* and analysis of genetic diversity in physalis using molecular markers. *PLoS ONE* 7: e50164

Wu J, Peng Z, Liu S, He Y, Cheng L, Kong F, Wang J, Lu G (2012) Genome-wide analysis of Aux/IAA gene family in Solanaceae species using tomato as a model. *Mol Genet Genom* 287: 295–311

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus Capsicum. *Theor Appl Genet* 118: 1279–1293

Wu F, Eannetta NT, Xu Y, Tanksley SD (2009) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor Appl Genet* 118: 927–935

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics* 174: 1407–1420

Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang Z (2010) Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Mol Ecol* 19: 4144–4154

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527–1530

Xu J, Ranc N, Muños S, Rolland S, Bouchet J, Desplat N, Le Paslier M, Liang Y, Brunel D, Causse M (2013) Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theor Appl Genet* 126: 1–15

Yarmolinsky D, Brychkova G, Fluhr R, Sagi M (2013) Sulfite reductase protects plants against sulfite toxicity. *Plant Physiol* 161: 725–743

Yu Q, Wang B, Liu L, Yang S, Yang T, Li J, Hou X (2010) Identification and effect analysis of QTLs conferring salt tolerance during seed germination in tomato. *Acta Botanica Boreali-Occidentalia Sinica* 9: 014

Zhong S, Fei Z, Chen YR, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, et al. (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol* 31: 154–159