
Review

Functional genomics of tomato in a post-genome-sequencing phase

Koh Aoki^{*1,2}, Yoshiyuki Ogata¹, Kaori Igarashi³, Kentaro Yano³, Hideki Nagasaki⁴, Eli Kaminuma⁴ and Atsushi Toyoda⁴

¹) Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen, Naka, Sakai, Osaka 599-8531, Japan

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

³) School of Agriculture, Meiji University, 1-1-1 Higashi-Mita, Tama, Kawasaki, Kanagawa 214-8571, Japan

⁴) National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

Completion of tomato genome sequencing project has broad impacts on genetic and genomic studies of tomato and Solanaceae plants. The reference genome sequence derived from *Solanum lycopersicum* cv ‘Heinz 1706’ serves as the firm basis for sequencing-based approaches to tomato genomics. In this article, we first present a brief summary of the genome sequencing project and a summary of the reference genome sequence. We then focus on recent progress in transcriptome sequencing and small RNA sequencing and show how the reference genome sequence makes these analyses more comprehensive than before. We discuss the potential of in-depth analysis that is based on DNA methylome sequencing and transcription start-site detection. Finally, we describe the current status of efforts to resequence *S. lycopersicum* cultivars to demonstrate how resequencing can allow the use of intraspecific genomic diversity for detailed phenotyping and breeding.

Key Words: tomato, genome, transcriptome, small RNA, comparative genomics.

Introduction

The tomato (*Solanum lycopersicum*) is regarded as a model plant that represents the Solanaceae family, which comprises 1000–2000 species that grow in all habitats from rainforests to deserts (Knapp 2002). Additionally, tomato is regarded as a model plant for the study of fruit development (Giovannoni 2004). Many Solanaceae plants—including potato, pepper, eggplant, tobacco and petunia—have highly syntenic genomes that each comprises 12 chromosomes; therefore, the reference genome sequence of the tomato was long awaited for molecular breeding of Solanaceae crops that are important for human nutrition.

The International Solanaceae Project (SOL, <http://solgenomics.net/solanaceae-project/index.pl>) launched the tomato genome sequencing project in November 2003 (Mueller *et al.* 2005). The aim of this sequencing project was to provide an information basis that could be used to link traits of Solanaceae plants to DNA sequence. This genome information is expected to lead us to the deeper understanding of plant diversity generated from a common set of genes. After an intensive collaboration of plant scientists from 14

countries, this sequencing project has been completed, and an annotated reference sequence and all findings were published in May 2012 (The Tomato Genome Consortium 2012). The published sequence is highly accurate, hence serves as a reliable basis for the further genomic studies.

With the prevalence of next generation sequencing (NGS) technology, the tomato genome sequence will facilitate a wide range of genetic and genomic studies that are based on comparative and in-depth sequence analysis. For example, resequencing of *S. lycopersicum* varieties sets the stage for linking phenotypic variation to DNA sequence variation; morphological and metabolic phenotypes of many economically important tomato cultivars—including *S. lycopersicum* varieties—have been intensively investigated; findings from such studies can be meaningfully reevaluated in the context of high-resolution sequence data. Another study is comprehensive sequencing of tomato transcripts. Owing to the versatility of the NGS technology, transcriptome analysis goes far beyond conventional gene-expression profiling, and facilitates comprehensive detection of small interfering RNA (siRNA), non-coding RNA (ncRNA) and splicing variants. Transcriptome analyses using NGS technology have led to the characterization of previously unrecognized mechanisms of gene regulation.

This review aims to summarize recent advances in sequencing-based genomics research on tomato in four parts.

Communicated by H. Ezura

Received October 9, 2012. Accepted December 18, 2012.

*Corresponding author (e-mail: kaoki@plant.osakafu-u.ac.jp)

First, we briefly describe an overview of the tomato genome sequencing project. Second, we present an overview of ‘Heinz 1706’ reference genome sequence. Third, we summarize recent progress with various types of transcriptome analysis. We also discuss the possibilities for further functional analysis that is based on DNA-methylation and transcription start-site analysis. In additional section, we present genome resequencing projects that involve *S. lycopersicum* cultivars and wild relatives of domestic tomatoes.

Tomato genome sequencing: transition from Sanger sequencing era to NGS era

Within the history of genome sequencing, the tomato genome project occurred during the transition period between multi-parallel Sanger sequencing and NGS. In 2004, the project was originally launched by the SOL as consortium sequencing project, and it involved 10 countries (Korea, China, UK, India, The Netherlands, France, Japan, Spain, Italy and USA). The ‘Heinz 1706’ cultivar, which was provided by the Heinz Corporation (Pittsburgh, PA), was used for this sequencing project because the original *Hind*III BAC library was made using this cultivar. This sequencing project initially involved a BAC-by-BAC sequencing approach that had been successfully applied to precedent model plants such as *Arabidopsis thaliana* (Arabidopsis Genome Initiative 2000), *Oryza sativa* (International Rice Genome Sequencing Project 2005) and *Lotus japonicus* (Sato *et al.* 2008). Three BAC libraries—specifically an *Eco*RI, a *Mbo*I and a *Hind*III library—were constructed. In this approach, a limited number of BAC clones were anchored to the genome (Peters *et al.* 2006). To anchor BAC clones, individual clones were screened for the presence of molecular genetic markers and marker-positive BACs were linked to the respective genetic loci defined by the respective markers. A fluorescent *in situ* hybridization (FISH) approach was used to verify the chromosome map positions of individual BACs and to delineate euchromatin/heterochromatin boundaries (Peterson *et al.* 1999). The 12 chromosomes were split up between 10 participant countries for BAC-by-BAC sequencing. Concurrently, Argentina and Italy sequenced the mitochondria (<http://www.mitochondrialgenome.org/>) and chloroplast genomes (NCBI accession number: NC_007898) (Kahlau *et al.* 2006), respectively, although mitochondria genome sequence has not been completely finished yet. This BAC-by-BAC approach was used to sequence 263 Mb that include 36% of the previously registered tomato ESTs.

In 2008, the sequencing consortium adopted the Selected BAC Mixture (SBM) approach to accelerate progress (The Tomato Genome Consortium 2012). Based on sequences from the ends of BACs and the criteria that at least one such end did not have similarity to repetitive sequence, 30,800 BAC clones were selected. These selected BACs were pooled and sequenced using a Sanger-based shotgun approach; 3.1 Gb was sequenced via the SBM method and these 3.1 Gb covered 540 Mb of the genome. The SBM con-

tigs were merged with the BAC-by-BAC contigs; together, these contigs cover 81 % of the previously registered tomato ESTs (<http://www.kazusa.or.jp/tomato/>). The success of the shotgun approach prepared the way for a NGS approach.

In 2009, the shotgun approach was applied to the whole tomato genome using emerging NGS platforms. Three NGS platforms—454 (Margulies *et al.* 2005), SOLiD (McKernan *et al.* 2009) and Illumina (Harris *et al.* 2008)—were used to generate 21 Gb, 64 Gb and 82 Gb, respectively, in NGS reads (Ahmadian *et al.* 2006, Ju *et al.* 2006). A *de novo* assembly of the ‘Heinz 1706’ genome was initially based on 454 and Sanger reads. High-quality BAC end sequences and high-coverage Illumina and SOLiD datasets were used to fill gaps and to improve overall base accuracy. The resulting tomato genome consisted of 91 scaffolds that covered 760 Mb that were in turn aligned with the 12 chromosomes. A combination of Sanger and NGS technologies was used to achieve high base accuracy, with only one error per 29.4 kb and only one indel error per 6.4 kb.

As described here, the change in sequencing approach over the course of the tomato genome sequencing project coincided with the transition from Sanger sequencing technology to NGS technologies. In the initial stage of the project, the goal was to sequence the euchromatic regions (size estimate 220 Mb), and sequencing the euchromatic regions was thought to be less than twice the effort of sequencing the *Arabidopsis* genome (150 Mb) and a moderate goal for BAC-by-BAC Sanger sequencing. But ultimately, the whole genome (760 Mb) was sequenced essentially via a shotgun approach that depended on NGS technology; this achievement demonstrated that genome size is not a limiting factor. During later stages of the project, advances in bioinformatics greatly facilitated the mapping and assembly of the relatively short, but highly redundant, reads that were generated with the NGS platforms. Projects that follow publication of this highly accurate reference genome involve in-depth sequencing of RNAs.

Overview of the tomato genome

Before describing any post-genome-sequencing studies, let us have a quick overview of the ‘Heinz 1706’ reference genome sequence (The Tomato Genome Consortium 2012). Each of the 12 chromosomes consists of pericentric heterochromatin and of euchromatin at the distal ends. The recombination rates and the gene and transcript densities are higher in euchromatin than in heterochromatin. Based on ITAG Release 2.3 (http://solgenomics.net/organism/Solanum_lycopersicum/genome), there are 34,727 of these predicted genes in the reference genome; based on RNA sequencing data 30,855 genes out of these predicted genes correspond to transcribed genes. The genome is highly syntenic with other commercially important Solanaceae plants such as potato, eggplant, pepper and tobacco.

Comparison of the reference tomato genome with those of plants in the euasterids (*Mimulus*, *Lactuca* and

Table 1. Publicly available RNA-Seq and sRNA-Seq datasets from tomato (September, 2012)

Submission ID or Accession ID	NGS platform	Strategy	Samples	Reference
SRA049915	Illumina HiSeq2000	RNA-Seq	<i>S. lycopersicum</i> cv ‘Heinz 1706’: 1 cm fruit, 2 cm fruit, 3 cm fruit, MG ^a , B ^b , B10 ^c , bud, flower, leaf, root. <i>S. pimpinellifolium</i> : IMG ^d , B ^b , B5 ^c , leaf	The Tomato Genome Consortium (2012)
SRA047925	454 GS FLX Titanium	RNA-Seq	<i>S. lycopersicum</i> cv ‘MoneyMaker’: root, stem, leaf, flower, MG ^a , B ^b , R ^e <i>S. pimpinellifolium</i> : leaf, R ^e	The Tomato Genome Consortium (2012)
SRA050797	AB SOLiD System 3.0	RNA-Seq	<i>S. lycopersicum</i> cv ‘Heinz 1706’: young leaves, old leaves, roots, stems, flowers, fruits	The Tomato Genome Consortium (2012)
SRA027382	454 GS FLX	RNA-Seq	<i>S. lycopersicum</i> cv ‘Ailsa Craig’: 1 cm fruit, MG ^a , B ^b , B7 ^c , B7 ^c (<i>rin</i>), B7 ^c (<i>nor</i>), B7 ^c (<i>Nr</i>), B7 ^c (<i>hp1</i>), B7 ^c (<i>apricot</i>), B7 ^c (TAGL1-RNAi) <i>S. lycopersicum</i> cv ‘M82’: pollen, unpollinated style, pollinated style <i>S. lycopersicum</i> M82 × M82: pollinated style <i>S. pennellii</i> : pollen <i>S. pennellii</i> LA716: pollen <i>S. pennellii</i> introgression line IL2-2: B7 <i>S. pennellii</i> LA716 × LA716: pollinated style <i>S. pennellii</i> LA716 × M82: pollinated style <i>S. pennellii</i> M82 × LA716: pollinated style	Lopez-Casado <i>et al.</i> (2012)
GSE12081	454 ^f	Small RNA-Seq	<i>S. lycopersicum</i> cv ‘Micro-Tom’: leaf, 1-15mm green fruits	Moxon <i>et al.</i> (2008)
GSE18110	Illumina Genome Analyzer	Small RNA-Seq	<i>S. lycopersicum</i> cv ‘Micro-Tom’: bud, flower, 1–3 mm fruit, 5–7 mm fruit, 11–14 mm fruit, MG ^a , B ^b , B3 ^c , B5 ^c , B7 ^c	Mohorianu <i>et al.</i> (2011)
GSE32470	Illumina Genome Analyzer II	Small RNA-Seq	<i>S. lycopersicum</i> cv ‘Heinz 1706’: leaves, flowers, fruit	The Tomato Genome Consortium (2012)

^a MG, mature green fruit; ^b B, breaker fruit; ^c Bn, breaker +n day fruit; ^d IMG, immature green fruit; ^e R, ripe fruit. ^f Model was not identified in the record.

Helianthus) or rosid (*Vitis* and *Arabidopsis*) family revealed that two consecutive genome triplication events occurred, the first when the rosid and euasterid lineages diverged approximately 130 million years ago and the next when the euasterid I and euasterid II lineages diverged approximately 60 million years ago. These two genome triplication events set the stage for evolution of genes involved in fleshy fruit development; duplicated genes acquired new and distinct functions. This group of genes includes transcription factors (*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)). Conversely, cytochrome P450 gene subfamilies that are involved in biosynthesis of toxic glycoalkaloid show contraction or complete loss in tomato.

Transcriptome analyses

Gene expression profiling

The reference sequence of the tomato genome has paved a fast lane for transcriptome analysis. NGS technology is also used for transcriptome sequencing. A comprehensive way to measure transcriptome composition is by direct high-throughput sequencing of cDNA, or, namely, by RNA-Seq (Nagalakshmi *et al.* 2008). If enough reads are collected from a sample, normalized read counts can be used to estimate gene expression level (Mortazavi *et al.* 2008). We have listed the publicly available RNA-Seq datasets and small RNA (sRNA)-Seq datasets in Table 1.

NGS platforms were used to examine the tissue-specific expression profiles of many tomato genes. Specifically, 10 tissues from the ‘Heinz 1706’ cultivar—root, leaf, bud, flower, 1 cm fruit, 2 cm fruit, 3 cm fruit, mature green fruit, breaker fruit and breaker + 10 day fruit (red fruit)—were subjected to RNA-Seq analysis using the Illumina platform (Table 1). The average number of reads per replicate sample was 10 ± 1.6 million. Similarly, gene expression profiles of *S. pimpinellifolium* tissues—including leaf, immature

green fruit, breaker fruit, breaker + 5 day fruit—were also subjected to RNA-Seq analysis using the Illumina platform (Table 1). The 454 platform was also used to examine tissue-specific gene expression profiles; RNA from seven tissues of *S. lycopersicum* cv ‘MoneyMaker’—root, stem, leaf, flower, mature green fruit, breaker fruit and ripe fruit, and 2 tissues of *S. pimpinellifolium*; leaf and ripe fruit—were used in these analyses (Table 1).

Mapping of the RNA-Seq data onto the reference genome sequence demonstrates that some transcripts originate in genomic regions that do not contain protein-coding genes; these transcripts may include non-coding RNAs and may function in the regulation of RNA accumulation via a protein-independent mechanism.

In the ‘omics’ framework, transcriptome sequencing can provide firm support for protein identification in proteomics analysis. Proteomic profiling that derives from accurate mass spectrometry depends heavily on the availability of a DNA reference database. Thus, the capacity for protein identification is limited in non-model organisms due to a lack of high-quality reference databases. However, RNA sequencing on NGS platforms may be useful for generating reliable reference databases at low cost and such databases should facilitate efficient matching of peptides masses to corresponding gene sequences.

This concept was tested by comparing the efficiency of protein identification using a custom RNA-Seq-based transcript database versus using a public database of Sol Genomics Network (<http://solgenomics.net/>) tomato unigene build (version released in June 2009) (Lopez-Casado *et al.* 2012). To construct custom unigene databases, RNA-Seq data from 454 sequencing of mature pollen, style, leaf and fruit of *S. lycopersicum* cv ‘M82’ and two wild relatives (*S. pennellii* and *S. habrochaites*) was used to assemble transcript sequences (Lopez-Casado *et al.* 2012). For comparison, another set of tomato unigene database, the version released in June 2009 from the SOL genomics network (SGN), was used. Quantitative proteomic analysis of pollen samples was conducted using the isobaric tag for relative and absolute quantitation (iTRAQ) method (Wiese *et al.* 2007); in the iTRAQ method, quantitative information is represented by isotope-encoded ‘reporter ions’ that are observed only in MS/MS spectra (Ross *et al.* 2004). To evaluate the potential of a custom RNA-Seq database for protein identification, peptide searches were performed using the custom RNA-Seq database or the SGN unigene database, and numbers of proteins that were identified with each database were compared. The results demonstrated that the number of proteins identified with the custom RNA-Seq database was greater than that with the SGN unigene database, yet the percentages of identified mass spectra were similar. More importantly, the percentages of matched amino acids in a peptide were comparable using the two databases. These results indicate that a custom RNA-Seq database can be used as a reliable reference database for proteomics analysis and, therefore, valuable for proteomics of non-model plants.

Small RNA profiling

The availability of a reference genome sequence has expanded the field of exploration of small RNA function to tomato. Recent studies established that 21–24 nt small RNAs (sRNA), which are generated from double stranded RNA (dsRNA) by Dicer-like (DCL) family nuclease, are involved in the control of gene expression (Phillips *et al.* 2007). These dsRNA can be formed by two different mechanisms; specifically, micro RNA (miRNA) is generated from a precursor RNA that has a short hairpin structure, whereas small interfering RNA (siRNA) is produced from long dsRNA, formation of which is dependent on the activity of RNA-dependent RNA polymerase (Brodersen and Voinnet 2006).

Before the tomato genome was completed, Sanger sequencing was used to profile tomato sRNA. The first comprehensive sRNA profiling was reported in 2007 (Pilcher *et al.* 2007); 4,108 sRNA were cloned from mature green fruit and nine known and three novel miRNAs were identified. None of these 12 miRNAs had homology to *Arabidopsis* miRNAs; this finding indicated that the 12 miRNAs each have a species-specific role in tomato. Itaya and coworkers (Itaya *et al.* 2008) and Zhang and coworkers (Zhang *et al.* 2008) reported similar results. Deep sequencing using NGS platform was then reported in 2008 (Moxon *et al.* 2008). Sequencing libraries were produced from leaf, bud and green fruit (1 to 15 mm diameter) of the dwarf cultivar ‘Micro-Tom’ (Meissner *et al.* 1997). This group used the 454 platform to generate 721,874 reads that yielded 225,000 and 102,000 unique sequences from fruits and leaves, respectively (Table 1). From these reads, 20 sequences matched known miRNAs (miR156, miR159, miR160, miR162, miR164, miR165, miR165, miR166, miR167, miR168, miR169, miR170, miR171, miR172, miR319, miR390, miR393, miR396, miR399, miR894); when 2-nt mismatches were allowed, 10 additional sequences match known miRNAs (miR394, miR395, miR397, miR398, miR408, miR472, miR482, miR828, miR858, miR1151). Tissue-dependent expression levels were examined on northern blots. Interestingly, the expression of some individual miRNAs differed at different stages of fruit development. For example, the accumulation of miR390, which may regulate genes that encode receptor-like kinases, was much higher in very small fruits than in leaves or flower buds, but miR390 accumulation was very low in mature fruits. This finding indicates that miR390 has a specific role in early fruit formation. The size distribution of non-redundant sRNA had a peak at 21 nt in leaf, but in fruits, there were more 23 or 24 nt sRNAs than 21 nt sRNAs. The 23- or 24-nt sRNAs are thought to be generated via a RNA polymerase IV-dependent pathway that produces heterochromatin-related siRNA (Onodera *et al.* 2005). Thus, this result suggested a more extensive control at the transcriptional level by DNA methylation triggered by 23- or 24-nt sRNAs in fruit tissues than in leaf.

More detailed sRNA profiling, including profiling of

10 time points from closed buds to red-ripe fruit of the ‘Micro-Tom’ cultivar, was reported in 2011 (Mohorianu *et al.* 2011); this work was based on the preliminary (~43% complete) ‘Heinz 1706’ genome (Table 1). Preliminary genome sequence facilitated the profiling of not just miRNAs, but also of many other sRNAs. When sRNA reads were mapped to the preliminary genome, 43,336 sRNA-producing loci were identified. Analysis of the sRNA expression profiles revealed that 24-nt sRNAs predominate in the flowering stages, but that representation of 21-nt forms increases in the late stages of fruit development. This result clearly demonstrates that sRNA expression is not random but is timed to coincide with the stages of fleshy fruit development. Most of the sRNAs that did not match to known miRNAs were differentially expressed during fruit development. Expression profiles of 43,336 sRNAs were classified into 63 co-expression clusters with respect to the similarity in the developmental expression pattern. One of the intriguing findings is that many clusters showed dominance of a single sRNA length. For example, two clusters that had similar expression profiles both of which show a remarkable drop with the onset of fruit development (namely, Cluster C consisted of 41 sRNAs and Cluster D consisted of 13 sRNAs), differed in size-class composition; with a clear dominance of 24-nt class in Cluster C and a dominance of 22-nt class in Cluster D. This suggests that different sRNA biogenesis mechanisms are specifically and independently regulated throughout fruit developmental process.

With the completion of the ‘Heinz 1706’ genome, mapping of NGS reads to the reference genome revealed the presence of 96 conserved miRNA genes in tomato (The Tomato Genome Consortium 2012). Among the 34 miRNA families identified, 10 are highly conserved in plants. Interestingly, the sRNAs specifically mapped to short regions, typically a 100–200 bp region within a promoter that produces a significant amount of sRNAs. The other interesting feature of sRNAs that map to the promoters of protein-coding genes is the dynamic expression profile during fruit development (The Tomato Genome Consortium 2012). Notably, the majority of these sRNAs that map to the promoters of protein-coding genes are 24-nt RNAs, and such RNAs are known to mediate methylation or de-methylation of DNA. Therefore, the sRNA that map to promoters may control gene expression at the transcriptional level. The biogenesis, regulation and function of these sRNAs that map to the promoters remain to be elucidated.

It will be intriguing to combine sRNA sequencing with other types of sequencing, such as DNA methylome sequencing (Lister *et al.* 2008, Zhang *et al.* 2006) or CAP analysis of gene expression (CAGE), which captures sequences containing transcription start sites (Shiraki *et al.* 2003). Reportedly, region-specific accumulation of sRNA and hypermethylation of cytosines, which are both associated with DNA methylation-mediated gene regulation, correlate with suppression of corresponding target genes in F1 progeny of *S. lycopersicum* cv ‘M82’ × *S. pennellii* LA716

(Shivaprasad *et al.* 2012). Additionally, some recent findings indicate that DNA methylation is associated with gene expression regulation during tomato fruit development (Teyssier *et al.* 2008).

Interspecific and intraspecific comparison of tomato genome sequences

The tomato reference genome was derived from a cultivar of *S. lycopersicum*, designated ‘Heinz 1706’. According to the conserved nature of Solanacea genomes, availability of the reference genome is facilitating the sequencing of varieties belonging to *S. lycopersicum* and wild relatives.

A comparison of the ‘Heinz 1706’ reference genome with the genome of *S. pimpinellifolium* (accession LA1589), which is thought to be a wild ancestor of *S. lycopersicum*, has been reported (The Tomato Genome Consortium 2012). Based on the *de novo* assembly of *S. pimpinellifolium* genome sequence, the divergence between the two genomes was estimated to be 0.6 %, or 5.4 million single nucleotide polymorphisms (SNPs). As expected from the pedigree of ‘Heinz 1706’, which has *S. pimpinellifolium* as one of its ancestors, putative *S. pimpinellifolium* introgressions were detected. Genomic regions with low divergence between *S. pimpinellifolium* and ‘Heinz 1706’ but with high divergence within domesticated cultivars were regarded as *S. pimpinellifolium* introgression. Based on these criteria, 40 regions that were considered to be introgressed from *S. pimpinellifolium* were detected. Interestingly, there appear to be large introgressions on chromosome 9 and 11 and each introgression is implicated in the breeding of disease resistance loci into ‘Heinz 1706’ using *S. pimpinellifolium* germplasm.

Genome project of domesticated cultivars includes ‘Micro-Tom’, a dwarf cultivar that is regarded as one of the model systems in studies of tomato (Meissner *et al.* 1997). Systematic bioresources of ‘Micro-Tom’ including EMS-mutagenized lines, gamma ray-mutagenized lines and full-length cDNAs have been developed and are publicly available (Aoki *et al.* 2010, Saito *et al.* 2011); these resources makes ‘Micro-Tom’ a good system for tomato genomics. Reportedly, ‘Micro-Tom’ has relatively large number of loci that are polymorphic when compared the respective loci in other cultivated tomatoes (Shirasawa *et al.* 2010). We conducted genome sequencing of ‘Micro-Tom’ (accession: DRX000482, DRX000454, DRX000455, DRX000627 and DRX000628) and identified approximately 1,230,000 SNPs and 190,000 indels in comparison of the “Micro-Tom” sequence and the ‘Heinz 1706’ reference genome sequence (unpublished data). This means that there is one nucleotide difference between the two genomes in every 700 bases. This frequency appears to be higher than that observed in intra-specific comparison of rice cultivars, where one SNP was identified in every 2,890 bases (Arai-Kichise *et al.* 2011). This result is consistent with of the finding that there are many polymorphic loci in ‘Micro-Tom’ (Shirasawa *et al.* 2010).

A comprehensive clade-oriented genome sequencing project is ongoing as a collective effort of the Solanaceae research community; this collaboration is called the SOL-100 project (<http://solgenomics.net/organism/sol100/view>). In the SOL-100 framework, 17 genome sequencing projects of *S. lycopersicum* cultivars are currently registered (September 2012), including ‘Ailsa Craig’, ‘Rutgers’, ‘M82’ and ‘Micro-Tom’ which are popular cultivars in tomato experimental studies (<http://solgenomics.net/organism/1/view>). Although most these datasets are not currently publicly available, they will serve as excellent information resources for developing SNP markers and intra-specific maps (Saliba-Colombani *et al.* 2000, Shirasawa *et al.* 2010).

Conclusion

Highly accurate ‘Heinz 1706’ reference genome sequence paves the road for sequencing-based functional genomics of tomato and of its wild relatives. In this review, we presented transcriptome analyses as one field that will benefit from this reference genome. The mapping of NGS reads onto the reference genome facilitates quantitative estimation of the expression levels of any transcripts—including those derived from annotated genes and non-annotated transcription units such as ncRNAs. Additionally, sRNA sequencing may accelerate the discovery of novel mechanisms of transcriptional and post-transcriptional regulation of tomato genes during fruit development.

We also described the sequencing of genomes of cultivated tomatoes and wild relatives of tomatoes. Combining detailed phenotyping of cultivated tomatoes (for example, <http://www.phenome-networks.com/home>; <http://solgenomics.net/search/phenotypes>) with genome sequencing facilitates association of DNA sequences to agronomically important traits. Systematic development of genomics bioresources (Ariizumi *et al.* 2011, Bombarely *et al.* 2011, Carvalho *et al.* 2011) also helps us exploit the wealth of the Solanaceae genome sequence.

Acknowledgements

This work was supported, in part, by a grant from Core Facility Upgrading Program of National BioResource Project to KA. Sequencing of the ‘Micro-Tom’ genome was supported by a grant from the Genome Information Upgrading Program in 2010 to KA and AT.

Literature Cited

- Ahmadian, A., M. Ehn and S. Hober (2006) Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta* 363: 83–94.
- Aoki, K., K. Yano, A. Suzuki, S. Kawamura, N. Sakurai, K. Suda, A. Kurabayashi, T. Suzuki, T. Tsugane, M. Watanabe *et al.* (2010) Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics* 11: 210.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Arai-Kichise, Y., Y. Shiwa, H. Nagasaki, K. Ebana, H. Yoshikawa, M. Yano and K. Wakasa (2011) Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. *Plant Cell Physiol.* 52: 274–282.
- Ariizumi, T., K. Aoki and H. Ezura (2011) Systematic development of tomato bioresources in Japan. *Interdisciplinary Bio Central* 3: 1–6.
- Bombarely, A., N. Menda, I. Y. Teclé, R. M. Buels, S. Strickler, T. Fischer-York, A. Pujar, J. Leto, J. Gosselin and L. A. Mueller (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.* 39: D1149–1155.
- Brodersen, P. and O. Voinnet (2006) The diversity of RNA silencing pathways in plants. *Trends Genet.* 22: 268–280.
- Carvalho, R. F., M. L. Campos, L. E. Pino, S. L. Crestana, A. Zsogon, J. E. Lima, V. A. Benedito and L. E. Peres (2011) Convergence of developmental mutants into a single tomato model system: ‘Micro-Tom’ as an effective toolkit for plant development research. *Plant Methods* 7: 18.
- Giorio, G., A. L. Stigliani and C. D’Ambrosio (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.
- Giovannoni, J. J. (2004) Genetic regulation of fruit development and ripening. *Plant Cell* 16 (Suppl): S170–180.
- Harris, T. D., P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. DiMeo, J. W. Efcavitch *et al.* (2008) *Science* 320: 106–109.
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Itaya, A., R. Bundschuh, A. J. Archual, J. G. Joung, Z. Fei, X. Dai, P. X. Zhao, Y. Tang, R. S. Nelson and B. Ding (2008) Small RNAs in tomato fruit and leaf development. *Biochim. Biophys. Acta* 1779: 99–107.
- Ju, J., D. H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M. S. Marma, S. Shi, J. Wu *et al.* (2006) Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* 103: 19635–19640.
- Kahlau, S., S. Aspinall, J. C. Gray and R. Bock (2006) Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J. Mol. Evol.* 63: 194–207.
- Knapp, S. (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *J. Exp. Bot.* 53: 2001–2022.
- Klee, H. J. and J. J. Giovannoni (2011) Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* 45: 41–59.
- Lister, R., R. C. O’Malley, J. Tonti-Filippini, B. D. Gregory, C. C. Berry, A. H. Millar and J. R. Ecker (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523–536.
- Lopez-Casado, G., P. A. Covey, P. A. Bedinger, L. A. Mueller, T. W. Thannhauser, S. Zhang, Z. Fei, J. J. Giovannoni and J. K. Rose (2012) Enabling proteomic studies with RNA-Seq: The proteome of tomato pollen as a test case. *Proteomics* 12: 761–774.
- Manning, K., M. Tör, M. Poole, Y. Hong, A. J. Thompson, G. J. King, J. J. Giovannoni and G. B. Seymour (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38: 948–952.
- Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bembien, J. Berka, M. S. Braveman, Y. J. Chen, Z. Chen *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.

- McKernan, K.J., H.E. Peckham, G.L. Costa, S.F. McLaughlin, Y. Fu, E.F. Tsung, C.R. Clouser, C. Duncan, J.K. Ichikawa, C.C. Lee *et al.* (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* 19: 1527–1541.
- Meissner, R., Y. Jacobson, S. Melamed, S. Levyatuv, G. Shalev, A. Ashri, Y. Elkind and A.A. Levy (1997) A new model system for tomato genetics. *Plant J.* 12: 1465–1472.
- Mohorianu, I., F. Schwach, R. Jing, S. Lopez-Gomollon, S. Moxon, G. Szitty, K. Sorefan, V. Moulton and T. Dalmay (2011) Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *Plant J.* 67: 232–246.
- Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer and B. Wold (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621–628.
- Moxon, S., R. Jing, G. Szitty, F. Schwach, R.L. Rusholme Pilcher, V. Moulton and T. Dalmay (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* 18: 1602–1609.
- Mueller, L.A., S.D. Tanksley, J.J. Giovannoni, J. van Eck, S. Stack, D. Choi, B.D. Kim, M. Chen, Z. Cheng, C. Li *et al.* (2005) The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). *Comp. Funct. Genomics* 6: 153–158.
- Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein and M. Snyder (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344–1349.
- Nakatsuka, A., S. Murachi, H. Okunishi, S. Shiomi, R. Nakano, Y. Kubo and A. Inaba (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.
- Ondera, Y., J.R. Haag, T. Ream, P. Costa Nunes, O. Pontes and C.S. Pikaard (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120: 613–622.
- Peters, S.A., J.C. van Haarst, T.P. Jesse, D. Woltinge, K. Jansen, T. Hesselink, M.J. van Staveren, M.H. Abma-Henkens and R.M. Klein-Lankhorst (2006) TOPAAS, a tomato and potato assembly assistance system for selection and finishing of bacterial artificial chromosomes. *Plant Physiol.* 140: 805–817.
- Peterson, D.G., N.L. Lapitan and S.M. Stack (1999) Localization of single- and low-copy sequences on tomato synaptonemal complex spreads using fluorescence in situ hybridization (FISH). *Genetics* 152: 427–439.
- Phillips, J.R., T. Dalmay and D. Bartels (2007) The role of small RNAs in abiotic stress. *FEBS Lett.* 581: 3592–3597.
- Pilcher, R.L., S. Moxon, N. Pakseresht, V. Moulton, K. Manning, G. Seymour and T. Dalmay (2007) Identification of novel small RNAs in tomato (*Solanum lycopersicum*). *Planta* 226: 709–717.
- Pratt, L.H., M.M. Cordonnier-Pratt, B. Hauser and M. Caboche (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.
- Ross, P.L., Y.N. Huang, J.N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels *et al.* (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* 3: 1154–1169.
- Saito, T., T. Ariizumi, Y. Okabe, E. Asamizu, K. Hiwasa-Tanase, N. Fukuda, T. Mizoguchi, Y. Yamazaki, K. Aoki and H. Ezura (2011) TOMATOMA: a novel tomato mutant database distributing Micro-Tom mutant collections. *Plant Cell Physiol.* 52: 283–296.
- Saliba-Colombani, V., M. Causse, L. Gervais and J. Philouze (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43: 29–40.
- Sato, S., Y. Nakamura, T. Kaneko, E. Asamizu, T. Kato, M. Nakao, S. Sasamoto, A. Watanabe, A. Ono, K. Kawashima *et al.* (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15: 227–239.
- Shiraki, T., S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, R. Kodzius, A. Watahiki, M. Nakamura, T. Arakawa *et al.* (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. USA* 100: 15776–15781.
- Shirasawa, K., S. Isobe, H. Hirakawa, E. Asamizu, H. Fukuoka, D. Just, C. Rothan, S. Sasamoto, T. Fujishiro, Y. Kishida *et al.* (2010) SNP discovery and linkage map construction in cultivated tomato. *DNA Res.* 17: 381–391.
- Shivaprasad, P.V., R.M. Dunn, B.A. Santos, A. Bassett and D.C. Baulcombe (2012) Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.* 31: 257–266.
- Teyssier, E., G. Bernacchia, S. Maury, A. How Kit, L. Stammiti-Bert, D. Rolin and P. Gallusci (2008) Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. *Planta* 228: 391–399.
- The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485: 635–641.
- Vrebalov, J., D. Ruezinsky, V. Padmanabhan, R. White, D. Medrano, R. Drake, W. Schuch and J. Giovannoni (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* 296: 343–346.
- Wiese, S., K.A. Reidegeld, H.E. Meyer and B. Warscheid (2007) Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7: 340–350.
- Zhang, J., R. Zeng, J. Chen, X. Liu and Q. Liao (2008) Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. *Gene* 423: 1–7.
- Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S.W. Chan, H. Chen, I.R. Henderson, P. Shinn, M. Pellegrini, S.E. Jacobsen *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* 126: 1189–1201.