Review

Functional genomics of tomato: Opportunities and challenges in post-genome NGS era

RAHUL KUMAR[1],* and ASHIMA KHURANA[2]

[1]*Repository of Tomato Genomics Resources, Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India*

[2]*Zakir Husain Delhi College, Botany Department, University of Delhi, New Delhi 110 002, India*

**Corresponding author (Fax, 91-40-24119430; Email, rksl@uohyd.ernet.in)*

The Tomato Genome Sequencing Project represented a landmark venture in the history of sequencing projects where both Sanger's and next-generation sequencing (NGS) technologies were employed, and a highly accurate and one of the best assembled plant genomes along with a draft of the wild relative, *Solanum pimpinellifolium*, were released in 2012. However, the functional potential of the major portion of this newly generated resource is still undefined. The very first challenge before scientists working on tomato functional biology is to exploit this high-quality reference sequence for tapping of the wealth of genetic variants for improving agronomic traits in cultivated tomatoes. The sequence data generated recently by 150 Tomato Genome Consortium would further uncover the natural alleles present in different tomato genotypes. Therefore, we found it relevant to have a fresh outlook on tomato functional genomics in the context of application of NGS technologies in its post-genome sequencing phase. Herein, we provide an overview how NGS technologies vis-à-vis available reference sequence have assisted each other for their mutual improvement and how their combined use could further facilitate the development of other 'omics' tools, required to propel the Solanaceae research. Additionally, we highlight the challenges associated with the application of these cutting-edge technologies.

[Kumar R and Khurana A 2014 Functional genomics of tomato: Opportunities and challenges in post-genome NGS era. *J. Biosci.* **39** 917–929] **DOI** 10.1007/s12038-014-9480-6

1. Introduction

Tomato (*Solanum lycopersicum*) has been a model of choice for Solanaceae family, which includes more than 3000 species that can thrive in diverse habitats from deserts to rain forests (Knapp 2002). Due to their significance in our diet, Solanaceae, the nightshade family, represents one of the three most economically important groups of plants, besides grasses and legumes, to humankind. Genomes of many Solanaceae species, including tomato, potato, eggplant, pepper, petunia and tobacco share high conservation among them at both micro- as well as macro-syntenic levels. Therefore, requirement of genomic information of a reference genome, in this case tomato, was long awaited for molecular breeding in Solanaceae crops. In 2003, an International Solanaceae Genome Project (SOL, *http://solgenomics.net/solanaceae-project/index.pl*), involving more than 10 countries, decided to find answers for the huge diversity present in Solanaceae species by sequencing a reference genome from this family. As a result, the Tomato Genome Sequencing Project, which aimed to sequence only gene-rich regions, was launched. Initially, BAC-by-BAC approach to sequence only 220 Mb sequence covering most of the euchromatic region was adopted (IRGSP 2005); however, next-generation sequencing (NGS) technologies were also included at later phase to sequence the complete genome. Although it took almost a decade, sequencing of tomato genome was completed and a highly accurate and well-annotated reference sequence with other findings was published in May 2012 (TGC 2012). The published tomato

Keywords. Genome; omics; RNA-Seq; tomato; transcriptome

Published online: 20 October 2014

genome is one of the best assembled and annotated genome till date, and thus, remains a reliable basis for the further genomic studies. This genome information is expected to provide us deeper molecular insights into the evolution of solanaceous genomes. Additionally, it will help unravel characteristics for the observed plant diversity generated from a common set of genes. However, several questions associated with the sequenced genome still remain unanswered. Some of the key aspects which require further attention are: validation of all the predicted gene models, assigning functions to all the encoded genes, and revelation and exploitation of the genomic diversity present in *Solanum* genus, in general, and tomatoes in particular.

The available quality reference genome sequence along with cost-effective methods such as NGS would accelerate the ongoing as well as future scientific efforts towards exploiting this information for comparative genetic and genomic studies and in-depth sequence analysis in Solanaceae. For instance, re-sequencing of *S. lycopersicum* varieties would facilitate linking morphological variations to their DNA sequence variation. Indeed, such attempts have already been undertaken and information of the re-sequenced tomato genomes by 150 Tomato Genome Consortium was recently released. The other advantage offered by accurately annotated tomato genome is its utility in identification of yet to be identified proteins by linking the less explored proteome to the highly explored genome (Mueller *et al.* 2009). Thus, availability of a reference genome has opened up new avenues in study of the frontier research areas such as transcriptome, smRNAome, methylome, proteome, metabolome, interactome, etc. in Solanaceae. In this review, the newly emerged opportunities associated with the tomato genome sequence and the current challenges posed to us, especially those related to bioinformatics tools and data analysis are discussed. Further, emphasis is given on their potential implications to promote research and breeding of tomatoes and related species. .

2. Overview of the tomato genome and its sequencing project

Due to its position at the crossroads of Sanger's sequencing and NGS methods, tomato genome sequencing project is the milestone project in the history of sequencing projects. Collectively, both technologies were used to sequence complete genome of tomato cultivar 'Heinz 1706'. Tomato consists of 12 chromosomes that have distinct gene rich euchromatin at distal ends whereas gene poor heterochromatin is distributed towards centromeres (Peterson *et al*. 1996). The predicted size of the tomato genome is approximately 900 megabases (Mb) (Arumuganathan and Earle 1991). The base accuracy of the final assembled sequence is approximately one substitution error per 29.4 kilobases (kb) and one indel error per 6.4 kb. A total of 34,727 protein-coding genes have been predicted in tomato genome and of these, and the presence of 30,855 has further been validated by RNA sequencing, henceforth called RNA-Seq, data (Mueller *et al*. 2009; TGC 2012). A striking observation into the evolution of tomato genome was achieved by comparative genomic analysis of tomato, grapes and *Arabidopsis* genomes, which clearly indicated occurrence of two consecutive genome triplication events during its evolution. The first triplication event took place when the rosid and euasterid lineages diverged, whereas the second event occurred during the diversion of euasterid I and euasterid II lineages. Eventually, these two triplication events set the stage for evolution of genes and added new gene family members that subsequently acquired important fruit-specific functions and played a crucial role in the evolution of fleshy fruits. Some of the important genes of this group include transcription factors such as Ripening-Inhibitor (RIN; Vrebalov *et al.* 2002) and Colorless non-ripening (CNR; Manning *et al.* 2006), ethylene biosynthesis and signalling components such as ACC synthase (ACS; Nakatsuka *et al.* 1998) and ethylene receptor (ETR; Klee and Giovannoni 2011), red light photoreceptors associated with fruit quality such as phytochromes (PHYB1/PHYB2; Pratt *et al.* 1995; Gupta *et al*. 2014), enzymes involved in lycopene biosynthesis such as phytoene synthase (PSY1/PSY2; Giorio *et al.* 2008) and cell wall maintenance such as xyloglucan endotransglucosylase/ hydrolases (XTH10). Contrary to that, some of the gene families such as cytochrome P450, which are associated with toxic alkaloid biosynthesis, were found to be either contracted or completely eliminated in tomato (TGC 2012). Apart from sequencing of tomato genome, many dedicated Web resources such as Solanaceae Genome Network (SGN; *http://solgenomics.net/*), Tomato Functional Genomics Database (TFGD; *http://ted.bti.cornell.edu.*), Kazusa Tomato Genomics Database (KaTomicsDB; *http://www.kazusa.or.jp/tomato/),* EU-SOL *(http://www.eu-sol.net/),* the Variant browser *(http://www.tomatogenome.net/VariantBrowser)* and many others have also been developed by SOL community which are currently helping in comprehensive integration of the various 'omics' data sets (figure 1). TFGD, in particular, is a comprehensive database for storage, query, mining, analysis, visualizing and integrating large-scale tomato functional genomics data sets and enables users to easily retrieve biologically important information through a set of efficient query interfaces and analysis tools (Fei *et al.* 2011).

Importantly, one of the first steps of newly procured NGS raw data, after removal of low quality sequencing reads and filtering the adaptor sequences, is its mapping to the reference genome. In the case of 'orphan species' where no reference genome sequence is available, highly accurate *de novo* assembly remains an astronomical challenge.

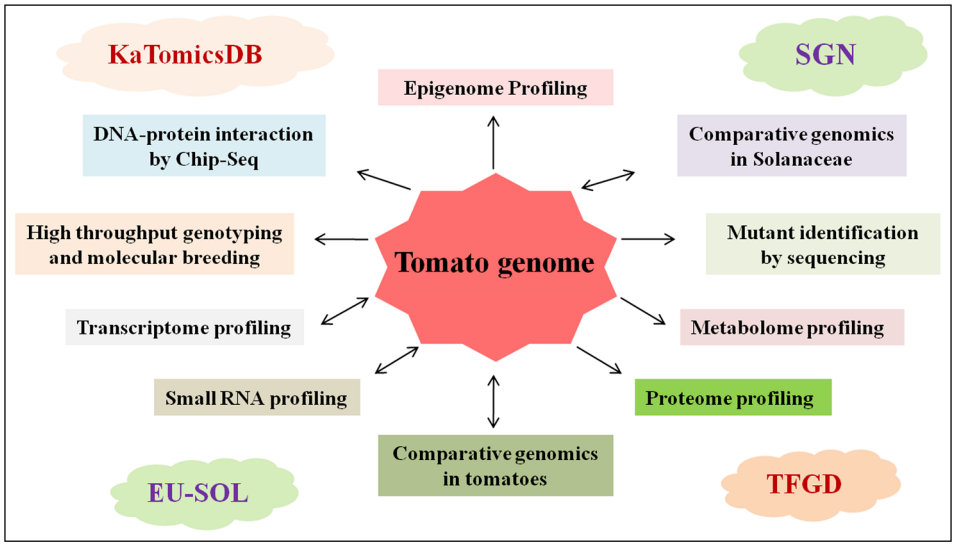

Figure 1. Impact of the availability of reference tomato genome on various aspects of its functional genomics. Normal arrows denote that the reference sequence has enabled development of other 'omics' tools, whereas bidirectional arrows denote that both the 'omics' tool as well as reference genome is mutually benefiting each other. Further, dedicated Web resources such as SGN, EU-SOL, TFGD and KaTomicsDB are helping in integration of all the data generated by these tools. SGN, Solanaceae Genome Network; EU-SOL, European SOL group; TFGD, Tomato Functional Genomics Database; KaTomicsDB, Kazusa Tomato Genomics Database.

Besides the information provided on trends associated with evolution of tomatoes, the tomato reference genome has enabled many other research topics which have been covered in an excellent review published recently (Menda *et al.* 2013). Therefore, the topics covered in this review mainly focus on the role of NGS technologies in driving the tomato functional genomics in the post-genome era.

3. Application of NGS technologies in tomato functional genomics

3.1 *Gene expression profiling in tomato*

RNA-Seq is a powerful technique which is employed for profiling and quantification of RNA transcripts in an organism. Another advantage offered by this technique is discovery of novel spliced transcripts for expressed genes. In this technology, principles of NGS are applied to the complementary DNAs (cDNA) derived from transcript populations (Wang *et al.* 2009). RNA-Seq is superior to other similar methods as it provides precise measurement of the levels of transcripts and their isoforms. Due to this property, digital expression profile of RNAs by this method has been an ideal replacement for the microarray-based expression profiling. Since the ease of its execution and eventual output is affected by the presence of a reference genome, the reference sequence of tomato genome has brought transcriptome analysis by NGS to a fast lane and currently this technology is being utilized to study transcriptomes of different tissues/stages/conditions in tomato in a more comprehensive manner (Aoki *et al.* 2013).

RNA-Seq profiling has led to identification of tissue-specific expression of many genes in tomato. Tomato Genome Consortium itself employed Illumina-based RNA-Seq method for the expression profiling of 10 tissues, including vegetative tissues such as root and leaf and reproductive tissues such as bud, flower and fruit tissues, including 1–3 cm mature green, breaker and red stages from the 'Heinz 1706' cultivar and provided evidence for the expression of 30,855 genes. Additionally, the 454 platform was also used to examine tissue-specific gene expression profiles from 'Moneymaker' cultivar. In this experiment, seven tissues, including root, stem, leaf, flower, mature green, breaker and ripe fruits, were used. Likewise, both Illumina and 454 platforms were used to examine gene expression profiles of *S. pimpinellifolium* tissues (TGC 2012). RNA-Seq has been employed for profiling of transcriptomes of the five peripheral tissues of the developing tomato pericarp and identified approximately 21,000 genes with half of them expressing ubiquitously (Matas *et al.* 2011). Huang *et al.* (2013) applied this technique to study expression pattern of members of SUN, OFP and GABBY transcription factors gene families and determined that some may exhibit tissue-specific expression in tomato. Besides, genomic-scale profiling of transcript accumulation, another striking contribution of its application, is in studying the evolutionary trends

associated with genomes. Indeed, RNA-Seq technique has already been successfully used to define both gene sequence and expression divergence between cultivated tomato and its five related wild species and elucidated selection pressure-related patterns in gene expression in tomato (Koenig *et al.* 2013). Currently, this technique is being extensively used in defining transcriptional dynamics in different tissues/conditions in tomato. A few of such important and already published studies have been listed in table 1.

The available genome sequence and concomitant upsurge of RNA-Seq studies in tomato have immensely benefited the assembly and annotation of other already sequenced or ongoing Solanaceae genome sequencing projects. Such improvements have further catalysed RNA-Seq-based transcriptome studies in these species (Kim *et al*. 2014). Those studies/approaches which are still missing in case of this reference plant but available in some other Solanaceae plants can now be easily performed in tomato and parallels can be drawn to identify both conserved and diverged mechanisms which have played significant role in the evolution of this group of plants. For example, similar approaches – which have already been successfully employed in other Solanaceae species such as to study leaf transcriptome in response to salt stress in *Petunia hybrida* (Villarino *et al*. 2014), to get insights into organ-specific pathogen defense response and potato tuber-*Phytophthora infestans* interactions (Gao *et al*. 2013), effect of ecological factors, including climate factors, soil factors, and tillage factors on growth and development of tobacco (*Nicotiana tabacum*) leaves (Lei *et al*. 2014) – can also be employed to have better understanding of similar aspects in tomato.

RNA-Seq has enabled improvement in genome annotation as it provides insights into the spliced forms of genes as well as presence of alternative transcription start sites (TSS) in full-length cDNAs (figure 1) (Martin *et al.* 2013; Zouine *et al.* 2014). The other potential implications of this technique to the tomato functional genomics include its likely use in molecular breeding for marker development, in the field of proteomics for protein identification, identification of novel non-coding RNA, network data analysis and generation of evolutionary transcriptome landscapes in tomato. However, several limitations of this technique remain and some of the challenges RNA-Seq faces include generation of large relatively shorter reads for a particular transcript and its subsequent computational analysis (Schliesky *et al.* 2012), incorporation of biases during RNA fragmentation step prior to library construction, and enrichment of reads mapping the 3′-end of the transcripts upon cDNA fragmentation (Martin *et al.* 2013).

3.2 *Small RNA (sRNA) profiling in tomato*

The advent of NGS technologies has been accompanied by an upsurge in profiling of small RNAs. It is now well established that 21–24 nt long sRNAs, generated from double-stranded RNA (dsRNA) by Dicer-like (DCL) family nucleases, play important roles in post-transcriptional regulation of gene expression (Baulcombe 2004). Prior to the completion of tomato genome sequencing project, when Sanger's sequencing was the predominant method, a few reports were published on sRNA profiling in plants. In tomato, the first such comprehensive report on tomato sRNAs appeared only in 2007. In this study, profiling of 4,018 sRNAs of mature green fruit resulted in identification of nine known and three novel miRNAs, with one of the novel miRNA having a fruit preferential accumulation. Further, identification of 12 novel sRNAs with no homologues available in *Arabidopsis* suggested that these sRNAs may have a species-specific role in this model fruit species (Pilcher *et al.* 2007). Later, additional conserved as well as unique sRNAs were identified in tomato (Itaya et *al.* 2008). Moreover, sRNA profiling in tomato enabled identification of conserved and nonconserved miRNAs and other short RNAs in tomato fruit and leaf (Moxon *et al.* 2008). Several miRNAs such as miR390, encoding receptor-like kinases, showed preferential accumulation in very small fruits than in leaves or flower buds, which showed a subsequent decline in mature fruits, indicating that miR390 has a specific role in early fruit development. Additionally, 23 or 24 nt sRNAs, thought to be produced by a RNA polymerase-IV-dependent pathway that generates heterochromatin related siRNA (Onodera *et al.* 2005), were found to be prevalent in fruits than leaves, suggesting that mechanisms of transcriptional control by sRNAs are relatively extensive in fruits (Moxon *et al.* 2008). Likewise, additional studies directed at the comprehensive analysis of sRNAs have been undertaken in different tomato cultivars (Itaya *et al.* 2008; Moxon *et al.* 2008; Mohorianu *et al.* 2011; TGC 2012; Zuo *et al.* 2012). In one such study that was based on incomplete 'Heinz 1706' genome sequence, predominance of 24 nt sRNAs in the flowering stages whereas higher representation of 21 nt forms in the late stages of fruit development suggested that sRNA accumulation is not a random event and is tightly coordinated with the different stages of fleshy fruit development (Mohorianu *et al.* 2011). The completion of tomato genome project has further expanded the horizon of such studies as it is now relatively easier to map such short sequences on the available tomato sequence (figure 1). This study has identified presence of 96 conserved miRNA genes in tomato (TGC, 2012). Further, upon examination of mapping of sRNAs onto respective promoters, the majority of sRNAs that originated from fruit tissues were 24mers, while those that originated from flower tissue were 22mers, suggesting that methylation and demethylation may play extensive role in the regulation of expression of fruit-related genes. Although the exact function of these sRNAs mapping to promoter regions is still unknown, these findings

Table 1. List of a few key published reports which demonstrate that the availability of tomato reference genome sequence has propelled the studies on tomato functional genomics through RNA-Seq and other NGS approaches

S. No.	Type of study	Tissues used in RNA-Seq and NGS experiments in tomato varieties	Broad outcome of the study	Reference	NGS platform
i	Transcriptome profiling	Epidermal layers, collenchyma, parenchyma, and vascular tissues of the fruit pericarp	DEGs associated with the five peripheral tissues of the developing tomato pericarp identified	Matas *et al*. 2011	FLX-454
ii	Transcriptome profiling	Meristem	Genes/TFs regulating meristem maturation determines inflorescence architecture in tomato	Park *et al.* 2012	Illumina
iii	Genome sequencing and transcriptome, including sRNA profiling	Roots, leaf, flower buds, open flowers and fruits	Organ/tissues/stages-specific genes identified	TGC 2012	Illumina and FLX-454
iv	Transcriptome profiling	Cotyledons, mature leaves, shoot apical meristem and hypocotyl	Quantitative genetic basis for leaf morphology revealed	Chitwood *et al*. 2013a	Illumina
v	Transcriptome profiling	Roots, aerial seedling, vegetative apices, stem between the fourth and fifth leaves and inflorescences	Distinct developmental architectures asociated with tomato development in a cultivated and wild relative revealed	Chitwood *et al*. 2013b	Illumina
vi	Transcriptome profiling	Lateral root, whole root and root tip	Cytokinin and auxin-regulated aspectes of root development identified	Gupta *et al.* 2013	Illumina
vii	Transcriptome profiling	Leaf and meristem	Genetic basis of tomato yield heterosis revealed	Jiang *et al*. 2013	Illumina
viii	Transcriptome profiling	Seedlings, stem, root, flower, leaf, vegetative meristem and fruit	Evolutionary trends associated with evolution of cultivated tomato and its five related wild species revealed	Koenig *et al*. 2013	Illumina
ix	Transcriptome profiling	Flagellin or *Pseudomonas* infiltrated tomato leaflets and tobacco leaves	Cell wall-associated kinase, *SlWAK1,* involved in plant immunity identified	Rosli *et al*. 2013	Illumina
x	Transcriptome profiling	Leaf	ABA has the potential to improve pathogen-resistance and abiotic stress tolerance in tomato	Wang *et al*. 2013	Illumina
xi	Transcriptome profiling	Leaf	Cytokinin and cytokinin regulated genes involved in compound leaf development	Shi *et al.* 2013	Illumina
xii	Transcriptome profiling	Trichomes	TFs involved in regulation of terpene biosynthesis identified	Spyropoulou *et al.* 2014	FLX-454
xiii	Transcriptome profiling	Fruit	Arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism	Zouari *et al*. 2014	Illumina
xv	Epigenoime, transcriptome and ChIP-seq	Different stages of fruit development	Changes associated with DNA methylation during fruit ripening revealed	Zhong *et al*. 2013	Illumina
xvi	Transcriptome profiling and ChIP-chip analysis	Mature green and red ripe fruits	Direct targets for FUL1 and FUL2 TFs during fruit ripening identified	Fujisawa *et al.* 2014	Illumina
xvii	sRNA profiling	Mature green, breaker and red-ripe fruits	Stage-specific small RNAs identified	Zuo *et al.* 2012	Illumina
xviii	microRNA profiling	5-DPA, Mature green, breaker and red-ripe fruits	Global identification of miRNA targets during fruit ripening	Karlova *et al.* 2013	Illumina
xix	Transcriptome and proteome profiling	Pollens	Improved annotation of transcriptional units helped in identification of more proteins	Lopez-Casado *et al*. 2012	454-generated sequences

Table 1 (continued)

S. No.	Type of study	Tissues used in RNA-Seq and NGS experiments in tomato varieties	Broad outcome of the study	Reference	NGS platform
xx	Genome sequencing for variant identification	Young leaf	SNPs present in 'Micro-Tom' genome identified	Kobayashi *et al.* 2014	Illumina and FLX-454
xxi	Structural variation browser for tomato and its wild relatives	150 Tomato Genome Sequencing Project (*http://www.tomatogenome.net/VariantBrowser*)	Sequences of 150 tomato genomes reported	uploaded on SGN on October 16, 2013	Illumina and FLX-454
xxii	TILLING-by-sequencing	Leaf	Mutation in EMS-mutagenized M2-population identified	Rigola *et al.* 2009	FLX-454
xxiii	TILLING-by-sequencing	Seedling	Mutation in EMS-mutagenized M2-population identified	Tsai *et al.* 2011	Illumina

have opened up a new avenue of research that would address the biogenesis, regulation and function of this category of sRNAs (TGC 2012). In a recent investigation, parallel analysis of RNA ends (PARE) resulted in identification of 119 target genes, including 106 new genes, in tomato (Karlova *et al.* 2013). Moreover, mapping of RNA-Seq data onto the reference genome has led to the identification and characterization of novel noncoding RNAs (ncRNAs). The ncRNAs have been predicted to function in the regulation of RNA accumulation via a protein-independent function (Aoki *et al.* 2013), albeit in comparison to the sRNAs the long ncRNAs (>200 nts) are far less explored, especially in plants, and represent an area with great potential in case of tomato functional genomics.

Profiling of sRNA have also been conducted in other Solanaceae species such as potato, pepper *etc*. For example, analysis of miRNAome from leaf, stem, root and four early developmental stages of tuberization resulted in identification of 89 conserved miRNAs, 147 potato-specific miRNAs and 112 candidate potato-specific miRNAs in potato (Lakhotia *et al*. 2014). Similarly, sRNAs were analyzed from leaf and stolon tissues and 28 conserved miRNA families were reported and potato-specific miRNAs were identified (Zhang *et al*. 2013). Large scale sequence analysis of *Potato spindle tuber viroid*-specific small RNAs revealed that their accumulation is accompanied by specific changes in gene expression in potato spindle tuber viroid (PSTVd)-infected tomato plants (Wang *et al*. 2011). Further, genome-wide analysis of sRNAs in pepper has identified 177 microRNAs corresponding to 37 microRNA families. It was revealed that their distribution correlated well with gene density in the hot pepper genome and is similar to that already reported in tomato but is in contrast to what is observed in *Arabidopsis thaliana* (TGC 2012; Kim *et al*. 2014). These observations strengthen the micro- and macro-syntenic level conservation between Solanaceae genomes, even at distribution of sRNAs encoding loci, which is much different from that in *Arabidopsis*. This observation indicates that besides conservation in the position of genes, their regulation might also be conserved among these genomes.

3.3 *Profiling of DNA-protein interaction*

The transcriptional networks driving plant development are mostly under the tight regulation of transcription factors (TFs). These TFs impart their influence on regulation of gene expression by binding directly to their target genes. NGS-based technologies have become a thrust in this area, and one of the most popular methods to elucidate DNA-protein interaction remains chromatin immunoprecipitation coupled with high-throughput sequencing of enriched fraction (ChIP-seq) (Kaufmann *et al.* 2009). Several reports have been published where ChIP-seq along with ChIP-chip

methods were employed to screen genomes for locations associated with binding of several TFs such as RIPENING-INHIBITOR (RIN) and FRUITFULL homologs (FUL1/FUL2) (Fujisawa *et al.* 2011, 2014; Zhong *et al.* 2013), IS sigma factor PSPTO_1203 and Fur (Ferric uptake regulator) regulon of *P. syringae* pv. tomato (Butcher *et al.* 2011; Markel *et al.* 2011). Examination of genome-wide targets for master regulators of fruit ripening, including RIN, FUL1 and FUL2, by combining RNA-Seq with ChIP-chip assay, a total of 292, 860 and 878, respectively, target ripening-associated genes were identified in tomato (Zhong *et al.* 2013; Fujisawa *et al.* 2014). Therefore, ChIP-seq and RNA-Seq with ChIP-chip are powerful tools that can improve mechanistic understanding of transcriptional networks underlying tomato development; however presence of close paralogous TFs and availability of accurate ChIP-seq algorithms slightly curtails overall efficacy of this technique.

3.4 *Mutation identification by sequencing in tomato*

The information of tomato genome sequence has further accelerated the functional genomic research in Solanaceae. Importantly, one of the non-transgenic 'reverse genetics' approaches, viz. Target Induced Local Lesions In Genomes (TILLING; McCallum *et al.* 2000), has been immensely benefited by the availability of genome sequence and advancement of sequencing technology. TILLING is a functional genomics tool to discover novel and rare mutations in populations. Although effective, some of the challenges associated with this tool retain its labour-intensive nature, ineffectiveness for pools deeper than eight individuals, and use of sequencing to characterize the mutations. An alternative approach which addresses these shortcomings is TILLING-by-sequencing. Mutation identification by re-sequencing of individual genomes is possible, albeit it is not yet feasible for large populations. Several lines of evidence have demonstrated that re-sequenceing of PCR products by combining NGS technologies and specific computational pipeline coupled with multiplexing of sequencing samples represents a powerful approach to detect mutations in a subset of candidate genes (Rigola *et al.* 2009; Reddy *et al.* 2012). Illumina-based method for sequencing target genes amplified from multidimensional pooled templates has been successfully applied to identify and determine novel mutations and their density in diploid and polyploid populations in rice, tomato and *Arabidopsis* (Tsai *et al.* 2011, 2013). We are also adopting similar approach to identify mutations in the EMS-mutagenized tomato population developed at RTGR, University of Hyderabad. Since *in silico* prediction of mutation hot regions in genome via CODDLE (an online computational tool to predict mutations) requires pre-information of the genic sequences, availability of a reference sequence becomes an indispensable source for designing PCR strategies for amplification of candidate genes (figure 1). High-density single nucleotide polymorphism (SNP) array (7,720 SNP)-based genotyping of a collection of 426 tomato accessions has revealed that over 97% of the markers are polymorphic in the entire collection (Sim *et al*. 2012a, b). Though such method cannot be employed to identify rare mutations in mutagenized populations in tomato till date, nonetheless, Genotyping-by-Sequencing (GBS) approach can be employed, as in the case of potato (Uitdewilligen *et al*. 2013), to identify such mutations in a segregating population.

3.5 *Proteome analysis in tomato*

One of the biggest challenges faced by researchers working in the field of plant proteomics, besides the extraction, isolation and detection related issues, is lack of comprehensive DNA sequence information and even if it is available for some plant species, highly accurate and wide-spread annotation and high-quality GO classification of the annotated genes remain inadequate. Though, with the available genomic sequences the primary bottleneck to the widespread use of high-throughput proteomics is partially alleviated, however, the ongoing efforts to have superior quality definition of transcriptional units are only going to help further in higher resolution proteomic studies in tomato. This hypothesis was interrogated by performing a quantitative analysis of the proteomes of pollen from domesticated tomato (*S. lycopersicum*) and its two wild relatives (*S. pennellii* and *S. habrochaites*) against a custom RNA-Seq transcript database (454 db1) versus a public database of Sol Genomics Network (*http://solgenomics.net/*) tomato unigene build (version released in June 2009; SGN db) and a combined database having both SGN db and 454 db1 data (454 db2) (Lopez-Casado *et al.* 2012). While ESTScan method generated a collection of 39,967 predicted proteins for SGN db, a total of 263,600 and 81,991 proteins were predicted for 454 db1 and 454 db2, respectively. For protein identification, the mass spectra derived from pollen proteins, which was obtained through isobaric tag for relative and absolute quantitation (iTRAQ) method (Wiese *et al.* 2007), was searched against all three databases. The results demonstrated that the 454 db1 enabled identification of maximum number of proteins in comparison of the two other databases, even though the percentages of identified mass spectra were similar and percentages of matched amino acids in a peptide remained comparable. The outcome further suggested that custom RNA-Seq database can be used as a reliable reference database for proteomics analysis and hence, can be used as an asset for proteomics of non-model plants (Lopez-Casado *et al.* 2012). Based on these preliminary studies, a collaborative effort under a 'tomato fruit transcript expression

profiling initiative' program at the Cornell University, USA, has already been initiated. It aims to integrate the proteome and transcriptome of tomato fruit, especially related to cell wall proteome analysis (Ruiz-May and Rose 2013) in different genotypes, including both wild type ripening fruit and those of the ripening impaired mutants *ripening inhibitor* (*rin*), *non-ripening* (*Nor*) and *never ripe* (*Nr*) to derive estimates of secretome genes under transcriptional and/or post-transcriptional control (*http://solgenomics.net/secretom/detail/profiling*). Additionally, comparative and quantitative proteomics approach has also been undertaken to create a more comprehensive catalog of the tomato cell wall proteome focusing principally on tomato leaves at different stages of infection by the oomycete of *Phytophthora infestans* as well as proteins that are regulated by abiotic stresses (*http://solgenomics.net/secretom/detail/proteomics*). Moreover, comparative protein profiling has already provided important insights into the proteome changes during fruit ripening (Rocco *et al*. 2006; Faurobert *et al*. 2007; Barsan *et al*. 2010; Yeats *et al*. 2010; Konozy *et al*. 2013) and under various abiotic stresses, including salinity (Chen *et al*. 2009; Manaa *et al*. 2013); aluminium (Zhou *et al*. 2009) and heat (Zhou *et al*. 2011) in tomato. Similarly, comparative proteomics has also resulted in identification of a chromoplast-specific carotenoid-associated (CHRC) protein which seems to be important for enhanced accumulation of carotenoids in *high-pigment1* (*hp1*) tomato fruits (Kilambi *et al*. 2013). Proteomic analysis of ripening tomato fruit infected by *Botrytis cinerea* has led to identification of many defense related proteins which are accumulated at higher level at mature green stage than in red fruits (Shah *et al*. 2012). Further, integrated transcriptomic and proteomic analysis of a compatible tomato-aphid interaction has revealed a predominant role of salicylic acid in plant response against the aphid infection (Coppola *et al*. 2013). The number of identified proteins in most of these studies is not very high, nonetheless with the upsurge of other tomato genome re-sequencing and transcriptome projects and continuously evolving computational tools, it will further help improve upon the existing genome annotation; which in turn is expected to revolutionize the higher resolution proteomic studies in tomato.

3.6 *Epigenome analysis in tomato*

One of the most benefited research communities, working on plants, by the arrival of NGS methods remains the epigenetic community, which immediately capitalized on it in exploring the role of DNA methylation in the regulation of gene expression. The first study on high-resolution DNA methylome in *Arabidopsis* was based on immunoprecipitation of methylcytosine-enriched fraction (Zhang *et al.* 2006); however, the first single-base resolution epigenome maps in *Arabidopsis* were based on sequencing-by-synthesis NGS method (Lister *et al.* 2008). Importantly, the latter study also provided the initial insights into the multiple layers of epigenetic regulation controlling transcription in plants. A direct link between the location of sRNAs and DNA methylation and their combined influence on the transcription of genes was also established (Lister *et al.* 2008).

Recently, the first ever single-base resolution methylomes of tomato fruits revealed that fruit epigenome is not static and methylation status of the genome changes continuously during different phases of fruit development. The whole-genome bisulphite sequencing covering four stages of fruit development led to identification of 52,095 differentially methylated regions in the 90% of the genome covered in this analysis in wild-type tomato. Comparison of fruit methylomes of two non-ripening mutants, viz. *ripening-inhibitor* (*rin*) and *Colorless nonripening* (*Cnr*), demonstrated that while the average methylation level in the 5′-ends of genes gradually declines during wild-type fruit development, it remains high in these two mutant genotypes. The results further suggested that the DNA methylation profile is not only altered in different tissues but developmental transitions, in this case green fruit to red fruit, and brings about notable variations even in the same tissue. Another striking feature of this study was presence of most of the RIN-binding sites in the demethylated regions of the promoters of numerous ripening related genes, though their sRNA density remains unchanged. Overall, this study suggested that dynamic epigenome ensures the fidelity of developmental processes such as ripening (Zhong *et al.* 2013).

Further, mining RNA-Seq data in search of transcription start site (TSS) variation has also improved gene structure annotation (Martin *et al.* 2013). With the distinct identity of TSSs in genome, it has become lot easier to map sRNAs into the different bins corresponding to 5′-UTR, gene-body and 3′-UTR and establishing direct association between methylation status of a particular bin with the expression of that gene. Such correlation was supported by a recent evidence where good correlation was observed between region-specific accumulation of sRNA and hypermethylation of cytosines and suppression of corresponding target genes in F_1 progeny of *S. lycopersicum* cv 'M82' × *S. pennellii* LA716 (Shivaprasad *et al.* 2012).

3.7 *Network analysis in tomato*

The integration of high throughput transcriptomic, proteomic and metabolomic data can provide precise information of the conditional changes associated with metabolic activity of an organism. Even though detailed insights into the gene structure of tomato genome has been achieved, deriving meaningful biological information from the sequenced genome remains an uphill task. The already published studies on

transcriptome, proteome and metabolome have helped in redefining the earlier hypotheses, nonetheless it remains essential to construct networks of co-expressed genes and use gene ontology (GO) information for highlighting the important gene candidates as critical components of functional networks (Martin *et al.* 2013). Many such attempts have already established a direct correlation between transcriptomics and metabolomics datasets, for the elucidation of gene-to-metabolite networks, in tomato (Fait *et al.* 2008; Lee *et al.* 2012; Osorio *et al.* 2012; Etalo *et al.* 2013). However, the information of transcriptome used in such efforts was based on microarray experiments, which represents only about 25% of the total genes encoded by tomato (Alba *et al.* 2005), and still has scope for further improvement. Such analysis has successfully been employed in potato where RNA-Seq data was used for co-expression network analysis and 18 gene modules representing tissue-specific transcriptional networks of major potato organs and developmental stages were identified (Massa *et al*. 2011). Information of gene structure and isoforms can also be useful for assigning valid functions to proteins identified in genome projects. Hence, application of RNA-Seq method along with information of genome structure can result in rapid improvement in construction of such molecular networks (figure 1). Secondly, proteome component is still missing in these studies and once included, the network information would become more holistic which eventually can act as a catalyst to help elucidate unknown functions of yet to be characterized genes. To the best of our knowledge, no such report has been published so far in the context of plants, however recent evidence demonstrate that pre-information of gene structure along with application of computation-guided strategies could be a powerful approach as this combination enabled discovery of a new enzyme and its metabolic step in bacteria (Zhao *et al.* 2013).

3.8 *Comparative genomics in Solanaceae*

The tomato genome shares high sequence homology with genomes of other Solanaceae species. Therefore, the sequenced genome of tomato 'Heinz 1706' will potentially serve as a useful reference for the sequencing of other varieties belonging to *S. lycopersicum* and wild relatives and will be used for comparative genomic studies to elucidate conserved and distinct evolutionary trends associated with each species. In addition to the sequencing and annotation of 'Heinz 1706', Tomato Genome Consortium also sequenced genome of *S. pimpinellifolium* (accession LA1589), considered to be a wild ancestor of *S. lycopersicum,* to 5× coverage and assembled it into a 739 Mb draft genome (TGC 2012). Comparison of reference sequence of 'Heinz 1706' with the *de novo* assembly of *S. pimpinellifolium* genome sequence demonstrated that the two genomes diverge by 0.6 %, or 5.4 million single nucleotide polymorphisms (SNPs). Comparison of the coding sequences, encoded gene by the two genomes, showed presence of a large number (>12,500) of gene carrying non-synonymous changes, including gains and losses of stop codons with potential consequences for gene function. Additionally, 40 regions of *S. pimpinellifolium* origin were detected in 'Heinz 1706', indicating that these regions were introgressed in the genome of latter during tomato breeding. Importantly, these introgressions were found to be associated preferentially with chromosomes 4, 9, 11 and 12 (TGC, 2012). Recent evidence indicates that 'Micro-Tom' genome sequence differs with that of the 'Heinz 1706' by approximately 1,230,000 SNPs and 190,000 indels, indicating a SNP frequency of 1 per 700 bp between the two genomes (Aoki *et al.* 2013). These results support the earlier findings, where relatively large number of polymorphic loci was reported in 'Micro-Tom' in comparison to the cultivated tomatoes (Shirasawa *et al.* 2010). We also investigated the natural structural variation associated with members of two tomato gene families involved in regulation of auxin responses, and identified 29,064 and 17,845 sequence variants, including both indels and SNPs, for 22 tomato auxin response factors (covering ~166 kb genomic region) and 26 Aux/IAA (covering ~79 kb region of tomato genome) genes in two inbred lines and 11 wild relatives. Based on our results, we observed that *S. galapagense* is a closer wild relative of domesticated tomato than *S. pimpinellifolium* (data unpublished).

Recently, comparison of tomato and hot pepper genomes revealed that gene repertories related to fruit ripening are well conserved in the two species (Kim *et al*. 2014). Whereas tomato is a climacteric fruit type and pepper is a non-climacteric species. This observation suggests that a gene regulatory mechanism likely causes differentiation in fruit ripening. Further investigation of orthologous regulatory genes, previously identified in tomato ripening, revealed a significant conservation in the expression of several transcription factor genes such as *RIN* and *NOR* and genes involved in ethylene signaling pathways (*ETR3*, *ETR4*, *EIN2* and *EILs*) during fruit ripening. In contrast, other transcription factor genes such as *CNR*, *Uniform* (*Golden-like 2*) and *Homeobox protein 1* (*HB-1*) and genes involved in ethylene biosynthesis (*ACS2*, *ACS4* and *ACO1*) showed distinct expression patterns in hot pepper and tomato. Based on these observations, it was inferred that the conservation and divergence of the transcription of these genes and their interactions may lead to qualitative and quantitative differences in the physiological phenomena underlying ripening (Kim *et al*. 2014). Further, in order to establish a common Solanaceae-based genomic framework, Solanaceae research community has already started a concerted effort to sequence at least one phylogenetically diverse genome of solanaceous

species from each clade; under the nomenclature of SOL-100 sequencing project (*http://solgenomics.net/organism/sol100/view*) (Menda *et al*. 2013). Currently, sequencing of a total of 25 genomes, including 15 genomes from *Solanum* taxon are in the sequencing pipeline and several of them are either completed or is nearing to their completion. Additionally, 17 genome sequencing projects of popular *S. lycopersicum* cultivars such as 'Ailsa Craig', 'Rutgers', 'M82' and 'Micro-Tom' have been registered (*http://solgenomics.net/organism/1/view*). Recently, sequence data of many tomato genotypes generated by 150 Tomato Genome Consortium was made public (*http://www.tomatogenome.net/VariantBrowser*). However, majority of data-sets are still not public and once available, these will serve as excellent information centers for developing SNP markers and intra-specific maps. Collectively, it will help in accelerating functional genomics research in Solanaceae in general and tomato in particular (Saliba-Colombani *et al.* 2000; Shirasawa *et al.* 2010). Moreover, SOL-100 project is a remarkable project with a huge potential in facilitating comparative genomics in Solanaceae, nonetheless the comparison of hundreds of sequenced genomes together on a common platform is going to be a huge challenge before the SOL bioinformatics team.

4. Conclusion and future perspectives

Completion of the Tomato Genome Sequencing Project has opened up a Pandora's Box for the comparative studies in Solanaceae. The high-quality 'Heinz 1706' reference genome sequence is a high-potential resource which can be used to tap the natural variation hidden in its wild relatives and improve important agronomic traits such as disease resistance, fruit quality and shelf-life, etc., in current tomato varieties. Besides the other 'omics' tools, transcriptome analysis by NGS is one such field which has been greatly propelled by the availability of tomato genome sequence. The NGS reads generated during RNA-Seq are easily mapped to the reference genome and their mapping has led to identification of novel transcripts and gene isoforms across the tomato genome. The assembled genome has enabled direct links between accumulated sRNA and methylation status and their combined impact on the expression of genes. Such high-throughput analyses have accelerated the discovery of novel regulatory mechanisms underlying various developmental aspects in tomato. Further, the tomato genome sequence would enable the correct genome assemblies of the ongoing solanaceaeous genome sequencing projects. The inception of SOL-100 and SOL-150 tomato genome projects is another milestone effort of the SOL community which will help unravel conserved and non-conserved evolutionary mechanisms associated with the tomatoes and other Solanaceae species. Therefore, the reference sequence of tomato is going to have far-reaching impact on the functional genomics studies in Solanaceae and the generated genomic as well as transcriptomic information would further speed up the molecular breeding efforts in this taxon. The genetic variants identified in comparative genomics studies will, undoubtedly, help in interpretation of their functional consequences on the genetic basis of important agronomic traits. Collectively, all these efforts would lead to transfer of important agronomic traits from one species to other and help in crop improvement, especially in Solanaceae.

Acknowledgements

The research work in RK's lab is supported by the Department of Science and Technology, Government of India. RK thanks Prof RP Sharma and Dr Y Sreelakshmi, Department of Plant Sciences, University of Hyderabad, for their support in carrying out the research work.

References

Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD and Giovannoni JJ 2005 Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* **17** 2954–2965

Aoki K, Ogata Y, Igarashi K, Yano K, Nagasaki H, Kaminuma E and Toyoda A 2013 Functional genomics of tomato in a post-genome-sequencing phase. *Breed. Sci.* **63** 14–20

Arumuganathan K and Earle ED 1991 Estimation of nuclear DNA content of plants by flow cytometry. *Plant Mol. Biol. Report.* **9** 229–241

Barsan C, Sanchez-Bel P, Rombaldi C, Egea I, Rossignol M, Kuntz M, Zouine M, Latche A, *et al*. 2010 Characteristics of the tomato chromoplast revealed by proteomic analysis. *J. Exp. Bot.* **61** 2413–2431

Baulcombe D 2004 RNA silencing in plants. *Nature* **431** 356–363

Butcher BG, Bronstein PA, Myers CR, Stodghill PV, Bolton JJ, Markel EJ, Filiatrault MJ, Swingle B, *et al*. 2011 Characterization of the *Fur* regulon in *Pseudomonas syringae* pv. tomato DC3000. *J. Bacteriol.* **193** 4598–4611

Chen S, Gollop N and Heuer B 2009 Proteomic analysis of salt-stressed tomato (*Solanum lycopersicum*) seedlings: effect of genotype and exogenous application of glycinebetaine. *J. Exp. Bot.* **6** 2005–2019

Chitwood DH, Kumar R, Headland LR, Ranjan A, Covington MF, Ichihashi Y, Fulop D, Jimenez-Gomez JM, *et al*. 2013a A quantitative genetic basis for leaf morphology in a set of precisely defined tomato introgression lines. *Plant Cell* **25** 2465–2481

Chitwood DH, Maloof JN and Sinha NR 2013b Dynamic transcriptomic profiles between tomato and a wild relative

reflect distinct developmental architectures. *Plant Physiol.* **162** 537–552

Coppola V, Coppola M, Rocco M, Digilio MC, D'Ambrosio C, Renzone G, Martinelli R, Scaloni A, *et al*. 2013 Transcriptomic and proteomic analysis of a compatible tomato-aphid interaction reveals a predominant salicylic acid-dependent plant response. *BMC Genomics* **14** 515

Etalo DW, Stulemeijer IJ, van Esse HP, de Vos RC, Bouwmeester HJ and Joosten MH 2013 System-wide hypersensitive response-associated transcriptome and metabolome reprogramming in tomato. *Plant Physiol.* **162** 1599–1617

Fait A, Hanhineva K, Beleggia R, Dai N, Rogachev I, Nikiforova VJ, Fernie AR and Aharoni A 2008 Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development. *Plant Physiol.* **148** 730–750

Faurobert M, Mihr C, Bertin N, Pawlowski T, Negroni L, Sommerer N and Causse M 2007 Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiol.* **143** 1327–1346

Fei Z, Joung JG, Tang X, Zheng Y, Huang M, Lee JM, McQuinn R, Tieman DM, *et al*. 2011 Tomato functional genomics database: a comprehensive resource and analysis package for tomato functional genomics. *Nucleic Acids Res.* **39** D1156–D1163

Fujisawa M, Nakano T and Ito Y 2011 Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin immunoprecipitation. *BMC Plant Biol.* **11** 26

Fujisawa M, Shima Y, Nakagawa H, Kitagawa M, Kimbara J, Nakano T, Kasumi T and Ito Y 2014 Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. *Plant Cell* doi:10.1105/tpc.113

Gao L, Tu ZJ, Millett BP and Bradeen JM 2013 Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-*Phytophthora infestans* interactions. *BMC Genomics* **14** 340

Giorio G, Stigliani AL and D'Ambrosio C 2008 Phytoene synthase genes in tomato (*Solanum lycopersicum* L.) - new data on the structures, the deduced amino acid sequences and the expression patterns. *FEBS J.* **275** 527–535

Gupta SK, Sharma S, Santisree P, Kilambi HV, Appenroth K, Sreelakshmi Y and Sharma R 2014 Complex and shifting interactions of phytochromes regulate fruit development in tomato. *Plant Cell Environ.* doi:10.1111/pce.12279

Gupta S, Shi X, Lindquist IE, Devitt N, Mudge J and Rashotte AM 2013 Transcriptome profiling of cytokinin and auxin regulation in tomato root. *J. Exp. Bot.* **64** 695–704

Huang Z, Van Houten J, Gonzalez G, Xiao H and van der Knaap E 2013 Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. *Mol. Genet. Genomics* **288** 111–129

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

Itaya A, Bundschuh R, Archual AJ, Joung JG, Fei Z, Dai X, Zhao PX, Tang Y, *et al*. 2008 Small RNAs in tomato fruit and leaf development. *Biochim. Biophys. Acta* **1779** 99–107

Jiang K, Liberatore KL, Park SJ, Alvarez JP and Lippman ZB 2013 Tomato yield heterosis is triggered by a dosage sensitivity of the florigen pathway that fine-tunes shoot architecture. *PLoS Genet.* **9** e1004043

Karlova R, van Haarst JC, Maliepaard C, van de Geest H, Bovy AG, Lammers M, Angenent GC and 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

Kaufmann K, Muino JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P and Angenent GC 2009 Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol.* **7** e1000090

Kilambi HV, Kumar R, Sharma R and Sreelakshmi Y 2013 Chromoplast-specific carotenoid-associated protein appears to be important for enhanced accumulation of carotenoids in *hp1* tomato fruits. *Plant Physiol.* **161** 2085–2101

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, *et al*. 2014 Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **46** 270–278

Klee HJ and Giovannoni JJ 2011 Genetics and control of tomato fruit ripening and quality attributes. *Annu. Rev. Genet.* **45** 41–59

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

Kobayashi M, Nagasaki H, Garcia V, Just D, Bres C, Mauxion JP, Le Paslier MC, Brunel D, *et al*. 2014 Genome-wide analysis of intraspecific DNA polymorphism in 'Micro-Tom', a model cultivar of tomato (*Solanum lycopersicum*). *Plant Cell Physiol.* doi:10.1093/pcp/pct181

Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, *et al*. 2013 Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* **110** E2655–E2662

Konozy EH, Rogniaux H, Causse M and Faurobert M 2013 Proteomic analysis of tomato (*Solanum lycopersicum*) secretome. *J. Plant Res.* **126** 251–266

Lakhotia N, Joshi G, Bhardwaj AR, Katiyar-Agarwal S, Agarwal M, Jagannath A, Goel S and Kumar A 2014 Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (Solanum tuberosum L.) by high-throughput sequencing. *BMC Plant Biol.* **14** 6

Lee JM, Joung JG, McQuinn R, Chung MY, Fei Z, Tieman D, Klee H and Giovannoni J 2012 Combined transcriptome, genetic diversity and metabolite profiling in tomato fruit reveals that the ethylene response factor SlERF6 plays an important role in ripening and carotenoid accumulation. *Plant J.* **70** 191–204

Lei B, Lu K, Ding F, Zhang K, Chen Y, Zhao H, Zhang L, Ren Z, *et al*. 2014 RNA sequencing analysis reveals transcriptomic variations in tobacco (*Nicotiana tabacum*) leaves affected by climate, soil, and tillage factors. *Int. J. Mol. Sci.* **15** 6137–6160

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

Lopez-Casado G, Covey PA, Bedinger PA, Mueller LA, Thannhauser TW, Zhang S, Fei Z, Giovannoni JJ, *et al*. 2012 Enabling proteomic studies with RNA-Seq: the proteome of tomato pollen as a test case. *Proteomics* **12** 761–774

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ and Seymour GB 2006 A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* **38** 948–952

Manaa A, Mimouni H, Wasti S, Gharbi E, Aschi-Smiti S, Faurobert M and Ahmed H 2013 Comparative proteomic analysis of tomato (*Solanum lycopersicum*) leaves under salinity stress. *P.O.J.* **6** 268–277

Markel E, Maciak C, Butcher BG, Myers CR, Stodghill P, Bao Z, Cartinhour S and Swingle B 2011 An extracytoplasmic function sigma factor-mediated cell surface signaling system in *Pseudomonas syringae* pv. tomato DC3000 regulates gene expression in response to heterologous siderophores. *J. Bacteriol.* **193** 5775–5783

Martin LB, Fei Z, Giovannoni JJ and Rose JK 2013 Catalyzing plant science research with RNA-seq. *Front. Plant Sci.* **4** 66

Massa AN, Childs KL, Lin H, Bryan GJ, Giuliano G and Buell CR 2011 The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. *PLoS One* **6** e26801

Matas AJ, Yeats TH, Buda GJ, Zheng Y, Chatterjee S, Tohge T, Ponnala L, Adato A, *et al*. 2011 Tissue- and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *Plant Cell* **23** 3893–3910

McCallum CM, Comai L, Greene EA and Henikoff S 2000 Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol.* **123** 439–442

Mohorianu I, Schwach F, Jing R, Lopez-Gomollon S, Moxon S, Szittya G, Sorefan K, Moulton V, *et al*. 2011 Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. *Plant J.* **67** 232–246

Moxon S, Jing R, Szittya G, Schwach F, Rusholme Pilcher RL, Moulton V and Dalmay T 2008 Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res.* **18** 1602–1609

Menda N, Stricler SR and Mueller LA 2013 Advances in tomato research in the post-genome era. *Plant Biotechnol.* **30** 243–256

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

Nakatsuka A, Murachi S, Okunishi H, Shiomi S, Nakano R, Kubo Y and Inaba A 1998 Differential expression and internal feedback regulation of 1-aminocyclopropane-1-carboxylate synthase, 1-aminocyclopropane-1-carboxylate oxidase, and ethylene receptor genes in tomato fruit during development and ripening. *Plant Physiol.* **118** 1295–1305

Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O and Pikaard CS 2005 Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120** 613–622

Osorio S, Do PT and Fernie AR 2012 Profiling primary metabolites of tomato fruit with gas chromatography/mass spectrometry. *Methods Mol. Biol.* **860** 101–109

Park SJ, Jiang K, Schatz MC and Lippman ZB 2012 Rate of meristem maturation determines inflorescence architecture in tomato. *Proc. Natl. Acad. Sci. USA* **109** 639–644

Peterson DG, Stack SM, Price HJ and Johnston JS 1996 DNA content of heterochromatin and euchromatin in tomato (*Lycopersicon esculentum*) pachytene chromosomes. *Genome* **39** 77–82

Pilcher RL, Moxon S, Pakseresht N, Moulton V, Manning K, Seymour G and Dalmay T 2007 Identification of novel small RNAs in tomato (*Solanum lycopersicum*). *Planta* **226** 709–717

Pratt LH, Cordonnier-Pratt MM, Hauser B and Caboche M 1995 Tomato contains two differentially expressed genes encoding B-type phytochromes, neither of which can be considered an ortholog of *Arabidopsis* phytochrome B. *Planta* **197** 203–206

Reddy TV, Dwivedi V and Sharma NK 2012 Development of TILLING by sequencing platform towards enhanced leaf yield in tobacco. *Ind. Crop. Prod.* **40** 324–355

Rigola D, van Oeveren J, Janssen A, Bonne A, Schneiders H, van der Poel HJ, van Orsouw NJ, Hogers RC, *et al*. 2009 High-throughput detection of induced mutations and natural variation using KeyPoint technology. *PLoS One* **4** e4761

Rocco M, D'Ambrosio C, Arena S, Faurobert M, Scaloni A and Marra M 2006 Proteomic analysis of tomato fruits from two ecotypes during ripening. *Proteomics* **6** 3781–3791

Rosli HG, Zheng Y, Pombo MA, Zhong S, Bombarely A, Fei Z, Collmer A and Martin GB 2013 Transcriptomics-based screen for genes induced by flagellin and repressed by pathogen effectors identifies a cell wall-associated kinase involved in plant immunity. *Genome Biol.* **14** R139

Ruiz-May E and Rose JK 2013 Progress toward the tomato fruit cell wall proteome. *Front. Plant Sci.* **4** 159

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

Schliesky S, Gowik U, Weber AP and Brautigam A 2012 RNA-Seq assembly - are we there yet? *Front. Plant Sci.* **3** 220

Shah P, Powell AL, Orlando R, Bergmann C and Gutierrez-Sanchez G 2012 Proteomic analysis of ripening tomato fruit infected by *Botrytis cinerea*. *J. Proteome Res.* **11** 2178–2192

Shi X, Gupta S, Lindquist IE, Cameron CT, Mudge J and Rashotte AM 2013 Transcriptome analysis of cytokinin response in tomato leaves. *PLoS One* **8** e55090

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

Shivaprasad PV, Dunn RM, Santos BA, Bassett A and Baulcombe DC 2012 Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *EMBO J.* **31** 257–266

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

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

Spyropoulou EA, Haring MA and Schuurink RC 2014 RNA sequencing on *Solanum lycopersicum* trichomes identifies

transcription factors that activate terpene synthase promoters. *BMC Genomics* **15** 402

TGC 2012 The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485** 635–641

Tsai H, Howell T, Nitcher R, Missirian V, Watson B, Ngo KJ, Lieberman M, Fass J, *et al*. 2011 Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol.* **156** 1257–1268

Tsai H, Missirian V, Ngo KJ, Tran RK, Chan SR, Sundaresan V and Comai L 2013 Production of a high-efficiency TILLING population through polyploidization. *Plant Physiol.* **161** 1604–1614

Uitdewilligen JG, Wolters AM, D'Hoop BB, Borm TJ, Visser RG and van Eck HJ 2013 A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS One* **8** e62355

Villarino GH, Bombarely A, Giovannoni JJ, Scanlon MJ and Mattson NS 2014 Transcriptomic analysis of *Petunia hybrida* in response to salt stress using high throughput RNA sequencing. *PLoS One* **9** e94651

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W and Giovannoni JJ 2002 A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. *Science* **296** 343–346

Wang Y, Shibuya M, Taneda A, Kurauchi T, Senda M, Owens RA and Sano T 2011 Accumulation of potato spindle tuber viroid-specific small RNAs is accompanied by specific changes in gene expression in two tomato cultivars. *Virology* **413** 72–83

Wang Y, Tao X, Tang XM, Xiao L, Sun JL, Yan XF, Li D, Deng HY, *et al*. 2013 Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. *BMC Genomics* **14** 841

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

Wiese S, Reidegeld KA, Meyer HE and Warscheid B 2007 Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7** 340–350

Yeats TH, Howe KJ, Matas AJ, Buda GJ, Thannhauser TW and Rose JK 2010 Mining the surface proteome of tomato (*Solanum lycopersicum*) fruit for proteins associated with cuticle biogenesis. *J. Exp. Bot.* **61** 3759–3771

Zhang R, Marshall D, Bryan GJ and Hornyik C 2013 Identification and characterization of miRNA transcriptome in potato by high-throughput sequencing. *PLoS One* **8** e57233

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, *et al*. 2006 Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126** 1189–1201

Zhao S, Kumar R, Sakai A, Vetting MW, Wood BM, Brown S, Bonanno JB, Hillerich BS, *et al*. 2013 Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature* **502** 698–702

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

Zhou S, Sauve R and Thannhauser TW 2009 Proteome changes induced by aluminium stress in tomato roots. *J. Exp. Bot.* **60** 1849–1857

Zhou S, Sauve R, Liu Z, Reddy S, Bhatti S, Hucko S, Yong Y, Fish T, *et al*. 2011 Heat-induced proteome changes in tomato leaves. *J. Am. Soc. Hortic. Sci.* **136** 219–226

Zouine M, Fu Y, Chateigner-Boutin AL, Mila I, Frasse P, Wang H, Audran C, Roustan JP, *et al*. 2014 Characterization of the tomato ARF gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS One* **9** e84203

Zouari I, Salvioli A, Chialva M, Novero M, Miozzi L, Tenore GC, Bagnaresi P and Bonfante P 2014 From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. *BMC Genomics* **15** 221

Zuo J, Zhu B, Fu D, Zhu Y, Ma Y, Chi L, Ju Z, Wang Y, *et al*. 2012 Sculpting the maturation, softening and ethylene pathway: the influences of microRNAs on tomato fruits. *BMC Genomics* **13** 7

MS received 21 February 2014; accepted 15 September 2014

Corresponding editor: RAJEEV KUMAR VARSHNEY