# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**
Open access books available

**122,000**
International authors and editors

**135M**
Downloads

Our authors are among the

**154**
Countries delivered to

**TOP 1%**
most cited scientists

**12.2%**
Contributors from top 500 universities

CLARIVATE ANALYTICS
**BOOK CITATION INDEX**
INDEXED

**WEB OF SCIENCE**™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us?
Contact book.department@intechopen.com

# Reaping the Benefits of Next-generation Sequencing Technologies for Crop Improvement — Solanaceae

Sushil Satish Chhapekar, Rashmi Gaur, Ajay Kumar and Nirala Ramchiary

Additional information is available at the end of the chapter

## Abstract

Next-generation sequencing (NGS) technologies make possible the sequencing of the whole genome of a species decoding a complete gene catalogue and transcriptome to allow the study of expression pattern of entire genes. The huge data generated through whole genome and transcriptome sequencing not only provide a basis to study variation at gene sequence (such as single-nucleotide polymorphism and InDels) and expression level but also help to understand the evolutionary relationship between different crop species. Furthermore, NGS technologies have made possible the quick correlations of phenotypes with genotypes in different crop species, thereby increasing the precision of crop improvement. The Solanaceae family represents the third most economically important family after grasses and legumes due to high nutritional components. The current advances in NGS technology and their application in Solanaceae crops made several progresses in the identification of genes responsible for economically important traits, development of molecular markers, and understanding the genome organization and evolution in Solanaceae crops. The combination of high-throughput NGS technologies with conventional crop breeding has been shown to be promising in the Solanaceae translational genomics research. As a result, NGS technologies has been seen to be adopted in a large scale to study the molecular basis of fruit and tuber development, disease resistance, and increasing quantity and quality of crop production.

**Keywords:** Solanaceae, NGS, capsicum, eggplant, tomato, potato

## 1. Introduction

In developing countries, "population" and "food security" are the two major issues. These problems get worse with the sudden climate changes that hamper production, yield, and quality of food crops. Therefore, to keep in mind the food security for billions of peoples, an initiative is required for improving the quality and yield of important crops. Several traditional

plant-breeding practices have been carried out for producing new varieties that can withstand with such changing climatic conditions besides increasing the productivity. These time-consuming practices could make considerable progress in crop improvement using selective germplasm, however, resulted in loss of biodiversity in the process. The recent advances in crop genomics, particularly the use of high throughput next-generation sequencing (NGS) technologies, look promising to identify causal genetic factors at genome by sequencing the whole genome and transcriptome of a species. As a result, the complete gene catalogue of a crop species and functional genes in different tissues could be identified besides allowing studying the genetic pathways involved in growth and development and biochemical pathways that eventually could be correlated with the crop phenotypes [1, 2]. Furthermore, the sequence data generated in vast amount provide a basis of genetic variation such as single-nucleotide polymorphisms (SNPs), which ultimately provide a relationship between genotype and phenotype in different species.

The Solanaceae family comprises approximately 2500 flowering plant species under 102 genera. The family represents the third most economically important family after grasses and legumes. Among the most important plants of this family are the potato (*Solanum tuberosum*), eggplant (*Solanum melongena*), tomato (*Solanum lycopersicun*), and capsicum or pepper (*Capsicum annuum*). They serve as important food crops and consumed worldwide due to their high nutritional components. Solanaceae crops have high nutritional value due to the presence of quality proteins, mineral salts, starch, vitamins, and antioxidants. Tomato majorly contributes to dietary nutrition globally with beneficial effects to human health mainly attributed to antioxidant compounds in the fruit such as lycopene and several other compounds such as carotenoids, zeaxanthin, and vitamin C. Capsicum fruits are rich source of metabolites that are beneficial for human health, such as carotenoids (provitamin A), vitamin C, vitamin A (which destroy free radicals), vitamin E, flavonoids, and capsaicinoids (anticancer agent). Although these compounds function as antioxidants and nutrients, they are used in traditional medicine also due to their enormous medicinal properties. Eggplant serves as an excellent source of antioxidants such as anthocyanins and several phenolics. Apart from this, it has a significant effect in reducing blood and liver cholesterol rates in humans. Worldwide, potato tubers are the principal source of starch along with proteins, vitamins, and antioxidants.

Here in this chapter, an attempt has been made to compile current research progress made based on NGS technology in four most important Solanaceae crop plants: tomato, potato, eggplant, and pepper. Furthermore, the application of NGS technology on those four crops toward translational research has been discussed.

## 2. Next-generation sequencing technologies

Knowing the genome sequence of a species has an advantage in crop breeding. This became possible with the revolution of DNA sequencing technologies. The Sanger method [3] was the first-generation sequencing method based on DNA chain termination method of the single-pass sequencing of one clone at a time. With the advent of NGS technologies, the sequencing

of complete genome or transcriptome of a species/genotype has become possible within a few hours. Utilizing various NGS platforms that are based on diverse chemistry and detection methods, several crop genomes, including major Solanaceae crops have been sequenced [4–7]. Among the various NGS technologies, three widely utilized platforms are Roche/454, Illumina Genome Analyzer (GA), and ABI SOLiD. The Roche/454 GSFLX chemistry is based on pyrosequencing and can produce up to 1 million reads of 600 bp to 1 kb [8]. The ABI SOLiD chemistry is based on emulsion polymerase chain reaction and sequencing by ligation technology, which can sequence up to 100 million reads of 50 bp in size [9]. The Illumina/Solexa GA based on sequencing by synthesis method produces 320 to 640 million reads of 100–150 bp [10].

The third- and the fourth-generation sequencing technologies are being developed, the majority of which allow the detection of single molecules with real-time sequencing. The popular third-generation sequencing platforms are Ion Torrents/Life Technologies, Heli-Scope™/Helicos Biosciences, and PacBio RS/Pacific Biosciences. The fourth generation is nanopore sequencing technology (Roche/IBM and Oxford). Ion Torrent company introduced a very different approach in 2010 as "Personal Genomic Machine," which was later commercialized by Life Technology. The chemistry is based on the real-time detection of the pH change (release of hydrogen ions), with the incorporation of a nucleotide into a growing DNA strand by a silicon detector [11]. The technology provides an average read length of ∼ 200 bp. The HeliScope introduced by Helicos BioSciences was the first commercially available single-molecule sequencing (SMS) platform [12]. The technology is based on highly sensitive fluorescence detection system with the incorporation of each nucleotide carrying fluorescent dye in the growing strand. The read length obtained ranges from 30 to 35 bp. PacBio RS, a single-molecule real-time (SMRT) sequencing technology, is based on the DNA sequencing by synthesis method and contains the provision of the real-time imaging of fluorescently tagged nucleotides for studying the sequence and structure of nucleic acid [13]. This technology not only can produce a comparatively longer DNA sequence (average read lengths of 5500–8500 bp) but also has wider application in epigenetics research as the technology is able to detect DNA methylation such as 4-methylctosine (mC), 5-mC, and 6-methyladenine (mA) [14].

The development of nanopore sequencing technology [15] begins an era of fourth-generation sequencing technology and has promised a cheap and fast method of sequencing. The principle involves threading a single-stranded DNA/RNA molecule electrophoretically through a nanopore that causes altering the pore's electrical properties and thereby modulating the ionic current through the nanopore. Braha et al. [16] designed a biosensor using "α-hemolysin," a toxin isolated from *Staphylococcus aureus*. The first commercial sequencing device was announced by Oxford Nanopore Technologies in 2012. Later, the technology was adapted and commercialized by other companies like Roche with IBM, Electronic BioSciences, and NABsys [17, 18]. This technology has advantage as sample preparation is not needed and the transduction and recognition occur in real time, on a molecule-by-molecule basis. The technology produces very long reads (up to 10 kb), which could be are capable of inexpensive *de novo* sequencing.

## 3. Application of NGS technology in Solanaceae genetics and genomics studies

NGS technologies have numerous potential applications in plant genetics and genomics, which include generation of genomic resources, complete decoding of a species genome, differential gene expression studies, whole genome association studies (WGAS), genomics assisted breeding (GAB), etc. (Figure 1).
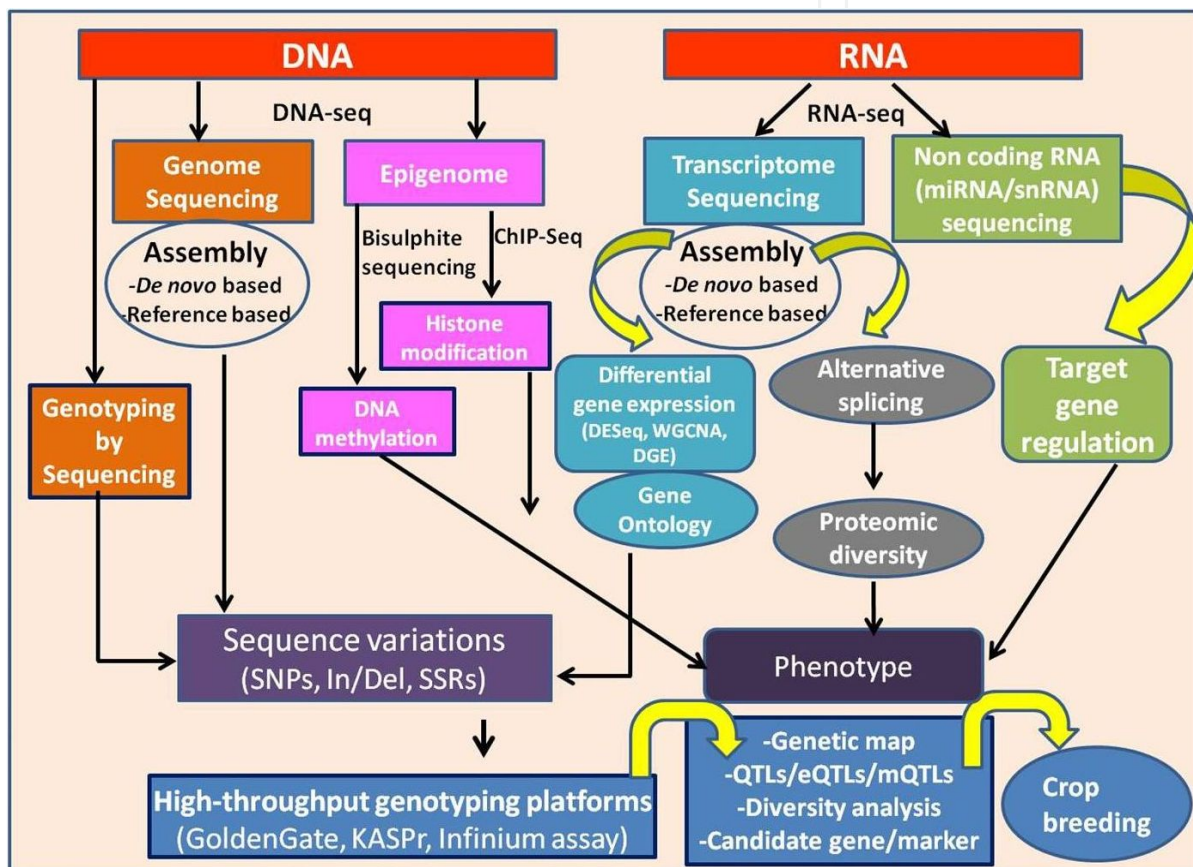


**Figure 1.** Overview of NGS applications in plant genetics and genomics

### 3.1. Transcriptome profiling of Solanaceae

Transcriptome sequencing of a species is the first step to access the functionally active genes. The transcriptome sequencing either by first-generation Sanger sequencing or by high throughput NGS approaches provides an insight into the expression of genes in a particular tissue/or different developmental stages of a species. The vast amount of sequencing data serve as a useful resource for the identification of sequence variations for the development of various markers, which would enable the mapping of candidate genes/QTLs for important traits. These applications have been discussed below in four important Solanaceae crops.

### 3.1.1. Potato

Potato (*S. tuberosum*) is the world's fourth largest crop after maize, rice, and wheat. It has a number of ploidy levels ranging from diploid ($2n = 24$) to triploids, tetraploids, pentaploids, and hexaploids. Most of the cultivated varieties are autotetraploid ($4n = 48$). Potato is the world's most important food crops that have edible tuber produced from stolons under favorable environmental conditions. It is accepted worldwide as a cheap source of dietary starch, protein, vitamins, and antioxidants, especially to feed large populations in developing countries. To date, only 4,20,074 ESTs are available in NCBI database (http://www.ncbi.nlm.nih.gov/nucest/?term=potato) that served as a valuable resource in various studies of gene discovery and expression analysis in potato germplasm [19–22]. In 2011, Massa et al. [23] reported a transcriptome sequence of *S. tuberosum* group Phureja clone DM1-3 516R44 using Illumina GAII platform. In this study, a total of 22,704 transcripts were identified, and 83% of these were of known function. The expression analysis was performed in a set of 32 tissues at various developmental stages and revealed that more than twenty thousand genes were found to be expressed in normal potato tissue and of these, some showed tissue-specific expression. In another study, using the weighted gene correlation network analysis (WGCNA), 18 gene co-expression modules were identified that comprised of a total of 5400 genes [24]. These modules were classified according to the high correlated expression profiles of genes in particular developmental stages. Two modules contained mainly transcription factors that showed co-expression in fruit development (e.g., Leafy Cotyledon 1 and transcriptional factor B3 domains) and tuber-tissue-specific expression (e.g., APETALA and WRKY). In another study, using digital gene expression (DGE) profiling, five genes encoding for DOF protein, a blue light receptor, a lectin, a syntaxin-like protein, and a protein with unknown function were found to be specifically associated with photoperiodic tuberization [25]. Hamilton et al. [26] published transcriptome sequencing of three potato cultivars and identified a total of 55,340 SNPs using the Maq SNP filter. In 2013, a whole-genome transcript analysis of the pollen mRNA of *Solanumtuberosum*, *S. demissum*, and their reciprocal $F_1$ hybrids was performed using Illumina GAII platform [27]. A total of 12.6 billion bases were obtained and were assembled into 13,020 transcripts. They identified the transcriptional differences between these samples and also identified nuclear genes that contributed to the differences observed in reciprocal crosses. Very recently, a comparative transcriptome analysis of white and purple potato was reported using Illumina HiSeq 2000 platform [28]. *De novo* assembly of the reads was performed for each cultivar using Trinity version r20131110 (http://trinityrnaseq.source-forge.net/). A total of 209 million paired-end reads were assembled into 60,930 transcripts. They identified candidate genes encoding transcription factors involved in anthocyanin biosynthesis. In a very interesting study, Aulakh et al. [29] reported global gene expression comparisons between wild-type (Bintje) and an activation-tagged mutant *underperformer* (*up*) using RNA-seq and identified approximately 1600 genes that were differentially expressed between them, thereby suggesting the modification of various biological pathways in the mutant variety.

### 3.1.2. Tomato

Tomato is an important vegetable crop that supplies vitamins and nutrients and consumed in different forms around the world. Whole transcriptome sequencing of six tomato accessions

*Solanum pimpinellifolium* was performed by sequencing by synthesis method of Illumina GAII [30]. This resulted in the generation of 17 Gb of sequence data with 291,915,037 high-quality reads and represented an average of 32.5 Mb of transcriptomic sequence per accession. By using these data, a large number of SNPs were identified to analyze genetic variation in cultivated and wild populations. A leaf transcriptome sequence data of tomato cv. Hongtaiyang 903 were generated using Illumina RNA-seq, which resulted in 50,616 transcripts [31]. Eighty-four percent of these transcripts were functionally annotated in the NCBI nr database and 94.5% in the tomato reference genome [24]. Of these, 14,371 transcripts were found to be involved in 310 pathways. An expression analysis revealed that 2787 transcripts showed significant expression after exogenous ABA treatment. These transcripts were related to ABA signaling pathway, various transcription factors, heat shock proteins, and pathogen resistance. The RNA-seq of one cultivated (*Solanum lycopersicum* M82) and five wild species with two red-fruited (*Solanumpimpinellifolium* and *Solanum galapagense*) and three green-fruited (*Solanum habrochaites*, *Solanum chmielewski*, and *Solanum pennellii*) varieties of tomato was performed to study the changes in gene expression and diversity in DNA sequence of these six species [32]. From this analysis, they identified several distinguishable polymorphic positions between cultivated and wild genotypes. Further, to examine the effect of the fungal symbiosis of tomato root on tomato fruit metabolism, Zouari et al. [33] performed an RNA-Seq of *S. lycopersicum* cv. Moneymaker using Illumina GA and studied transcriptome profiling during fruit maturation. A total of 712 differentially expressed genes in fruits from mycorrhizal and control plants were identified. The majority of the regulated genes were involved in various functions such as photosynthesis, stress response, transport, amino acid synthesis, and carbohydrate metabolism. Further, it was found that AM fungi can serve as a replacement of exogenous fertilizer for the growth of tomato plant with nutrient rich fruits. In addition, to examine the hormonal response in tomato roots, Gupta et al. [34] published a transcriptome atlas of tomato root using Illumina RNA-Seq method. By mapping the 165 million reads onto the tomato reference genome (*S. lycopersicum*), they identified differential expression pattern after various hormonal treatments. To look into regulatory and metabolic pathways specific to fruit tissues, Matas et al. [35] reported a transcriptome study coupled with laser capture microdissection. Five fruit pericarp tissues were sequenced by the pyrosequencing method of GSFLX platform (Roche) and identified 20,976 high-quality expressed unigenes, which included genes that showed expression specific to particular cell type and tissue. Very recently, Mou et al. [36] performed a global analysis of transcriptome of cherry tomato (*Lycopersicon esculentum* var. *cerasiforme* "XinTaiyang") fruit after exogenous treatment of ABA and nordihydroguaiaretic acid (an inhibitor of ABA biosynthesis) to study their effect on fruit ripening process. Of the total 25,728 genes, 10,388 were found to be differentially expressed. The data also revealed the upregulation and downregulation of pigment-related genes after exogenous ABA and NDGA treatment, respectively. Moreover, they also suggested the transcriptional abundance of candidate genes involved in photosynthesis during inhibition of endogenous ABA, which highlighted the significance of ABA in the regulation of ripening process in tomato fruit. Further, to utilize the large amount of transcriptome data for tomato for studying gene expression analysis, Bostan and Chiusano [37] recently presented a web-based platform, i.e., NexGenEx-Tom, that contain collection of high quality transcriptome data of several tissue at

various stages of the development of different tomato genotypes and serve as a useful approach for analysis of gene expression profiling and comparisons in various tissues/ genotypes.

### 3.1.3. Pepper (Capsicum)

The capsicum is a diploid, $2x = 2n = 12$, and self-pollinating plant. Capsicum is closely related to other members of the Solanaceae family, such as potato, tomato, and tobacco, that originated in the New World. The genus contains 39 species of which only six species are cultivated, such as *C. annuum*, *C. baccatum*, *C. frutescence*, *C. chinense*, *C. pubescens*, and *C. assamicum* [38, 39]. These Capsicum species are grouped as pungent (hot/spicy) and nonpungent (sweet) pepper based on the presence and absence of capsaicinoid compounds, respectively, and therefore used as a major ingredient in various cuisines around the world. The fruit contains beneficial metabolites such as carotenoids (provitamin A), vitamins C and E, flavonoids, and capsaicinoids. It is also used as a coloring agent in food and also have several medicinal properties and thus used in making of traditional medicine. Moreover, several studies have suggested an effective role of capsaicinoids in inhibiting the growth of cancer [40–42], the painkiller in arthritis, reducing appetite, and weight management [43–45]. For chili pepper, a large number of varieties are available that are well adapted in diverse climate conditions around the world [46]. Many studies were targeted toward various aspects, including the development of genetic and genomic resources for crop improvement [39]. A *Capsicum* transcriptome database (DB, http://www.bioingenios.ira.cinvestav.mx:81/Joomla/) was developed by the sequencing of *C. annuum* transcriptome from different tissues [47]. They have obtained 1,324,516 raw reads from which 32,314 high-quality contigs, and 51,118 singletons were assembled. Functional annotation of the 75% of the contigs was done resulting in 7481 novel sequences. Further, using 454 GS-FLX pyrosequencing platform, the transcriptome analysis of red pepper (*C. annuum* L. TF68) was carried out [48]. They obtained approximately 30.63 Mb of EST data with 9818 contigs and 23,712 singletons. In another study, Nicolai et al. [49] performed transcriptome analysis using Roche 454 pyrosequencing, and this consists of 23,748 contigs and 60,370 singletons. Using the data, they identified a total of 11,849 SNPs and 853 SSRs. However, in a separate study, Ashrafi et al. [50] used three chili genotypes, namely, Maor, Early Jalapeno, and Criollo de Morelos-334 (CM334) for transcriptome sequencing. From the first assembly, they identified a total of 4236 SNPs and 2489 SSRs, while the second transcriptome assembly based on Illumina GAII resulted in 22,000 high-quality putative SNPs and 10,398 SSRs. Recently, the Pepper GeneChip array from Affymetrix in *Capsicum* for polymorphism detection and expression analysis was reported [51]. Further, the hybridization of genomic DNA from 40 diverse *C. annuum* lines and few lines from other cultivated species such as *C. frutescens*, *C. chinense*, and *C. pubescens* resulted in generation of 33,401 single-position marker (SPP) from 13,323 unigenes. Liu et al. [52] constructed *de novo* transcriptome assembly in *C. frutescens* and obtained 54,045 high-quality unigenes in which a total of 4072 SSRs were identified, including three candidate genes i.e., dihydroxyacid dehydratase (DHAD), Thr deaminase (TD), and prephenate aminotransferase (PAT) involved in the capsaicinoid biosynthesis pathway. Additionally, a total of 9150 putative SNPs in 3349 contigs were identified between *C. frutescens* and *C. annuum*. In another study, a high-throughput tran-

scriptome profiling in two *C. annuum* varieties resulted in 279,221 and 316,357 sequenced reads with a total of 120.44 and 142.54 Mb of sequence data. A total of 9701 and 12,741 potential SNPs were identified [53].

### 3.1.4. Eggplant

Eggplant or brinjal (*S. melongena* L.), an autogamous diploid ($2n = 2x = 24$), is the third most important vegetable crop from the genus *Solanum* after potato (*S. tuberosum*) and tomato (*S. lycopersicum*). The eggplant is widely grown in Asia, the Middle and Near East, Southern Europe, and Africa [54]. The eggplant fruit serves as an excellent source of antioxidants like anthocyanin and phenolics [55, 56] and the tolerance to abiotic and biotic stresses [57]. Therefore, several genetic studies have been carried out from the last two decades targeting various fruit traits such as size/shape and color. Moreover, the different origin of eggplant from other Solanaceae spp. makes it an important crop for comparative and evolutionary studies. In this regard, various aspects have been focused by researchers such as the development of genetic resources like molecular markers and genetic map that have been utilized for comparative analysis with other spp. of the Solanaceae family. The eggplant belongs to the *Leptostemonum* clade, which is far lagged behind the potato and tomato (potato clade) in terms of the development of genomic resources as only a total of 226,664 nucleotide sequences were available in NCBI database, of which majority (98,086) were obtained from ESTs generated by Fukuoka et al. [58]. These 98,086 ESTs were assembled into 16,245 unigenes that covered only a limited portion of eggplant transcriptome. Later, transcriptome sequencing was carried out using Illumina sequencing and reads were assembled into contigs using Trinity program [59]. Of these, 80% (27,393) of unigenes showed matches with the sequences available in NCBI nr database. A total of 29,717 genes were functionally annotated. A comparison of eggplant with 11 plant proteomes resulted in 276 high-confidence single-copy orthologous groups and revealed that eggplant and its wild *Leptostemonum* clade relative "turkey berry" split ~6.66 million years ago in the late Miocene and the *Leptostemonum* split ~15.75 Mya from the potato clade in the middle Miocene.

## 3.2. Whole genome and transcriptome sequencing of Solanaceae spp.

Whole genome sequencing of a species reveals the structural organization of genome, including a number of protein-coding and non-protein-coding genes and repetitive elements and serves as the basis for finding genome-wide analysis of genetic variation, QTL mapping, diversity analysis, association mapping of agronomically important traits for crop improvement, and comparative study of genome evolution between different species.

### 3.2.1. Potato genome

The draft sequence of 844 Mb genome of a homozygous double-monoploid genotype named DM (DM1-3 516R44) was sequenced using three methods, namely, Sanger method, Roche/454 Pyrosequencing, and Illumina sequencing-by-synthesis method and assembled using the SOAPdenovo assembly algorithm (PGSC; The Potato Genome Sequencing Consortium, 2011) [6]. A heterozygous diploid line, i.e., RH (RH89-039-16) was also sequenced using shotgun

sequencing of BACs and WGS, and its reads were mapped to the reference assembly of DM genome (http://potatogenome.net). About 86% of the genome was anchored and assembled into pseudomolecules. A total of 39,031 protein-coding genes were obtained; of them, 90% were located on 12 pseudomolecules. To overcome the problem of heterozygosity and inbreeding depression, which is the major drawback in potato improvement using traditional breeding practices, the researchers selected a homozygous, double-monoploid form, referred as DM for sequencing and integrated with sequence data of heterozygous diploid line RH. The potato genome was the first among the asterid species to be sequenced, and a total of 2642 high-confidence asterid-specific and 3372 potato lineage-specific genes were identified and also found the collinearity with 97.5% identity between DM and RH genome. Furthermore, they identified 3.67 million SNPs and 275 gene-specific presence/absence variations and concluded that the homozygous alleles were the reason for the reduced level of vigor in DM line. They also studied the evolution of tuber development, which revealed that about 15,235 genes were found to be expressed in developing tubers.

### 3.2.2. Tomato genome

In the year 2012, the Tomato Genome Consortium (TGC, 2012) [5] reported the draft genome sequence of inbred cultivar of tomato "Heinz 1706" using a combination of NGS technologies (454/Roche GS FLX, Illumina Genome Analyser, and SOLiD sequencing). They predicted the genome size of 900 Mb, which were assembled in 91 scaffolds aligned to 12 chromosomes. The data revealed only 0.6% nucleotide divergence (in two tomato genotypes) compared to 8% divergence with potato. The alignment of tomato–potato orthologous regions confirmed nine large inversions during evolution. They predicted about 34,727 (in tomato) and 35,004 (in potato) protein-coding genes. The analysis suggested that the genome triplications could have added new gene family members such as RIN (ripening-inhibitor), CNR (colorless nonripening), ACS (associated with ethylene biosynthesis), PHYB1/PHYB2 for red light photoreceptors, and PSY1/PSY2 (phytoene synthase) for lycopene biosynthesis that mediate important fruit-specific functions such as fleshiness and color. Further, the study reported the presence of noncoding RNAs (ncRNA) with the identification of 96 miRNA genes in tomato and 120 miRNA genes in potato genome. In another study, Aflitos et al. [60] performed the resequencing of 84 tomato accessions and explored the genetic variability present among those cultivated tomato and its wild progenitor. They identified more than 10 million SNPs in wild species, signifying the dramatic genetic erosion of tomato. Furthermore, through comparative sequence alignment, group-, species-, and accession-specific polymorphism was observed, which may be linked to agronomically important fruit traits. Such information may be easily used by recent high-throughput genotyping methods for the detection of genetic variability across extensive populations. The genomic information provided by these projects could be used for comparative genetic and genomic studies and in-depth sequence analysis in Solanaceae.

### 3.2.3. Pepper genome

The recent advancement in the sequencing and development of NGS technologies has accelerated the genetics and genomics studies of capsicum. Recently, a draft genome sequence

of a diploid hot pepper, i.e., "*C. annum* cv 'CM334'" (Criollo de Morelos 334), was published [7]. The variety CM334 has been utilized in breeding practices as it showed resistance against *Phytophthora capsici*, pepper mottle virus, and root-knot nematodes. The authors reported sequencing a total of 650.2 Gb pepper genome, which is approximately equal to 186.6× genome coverage of 3.48 Gb estimated *C. annum* genome by utilizing Illumina platform. Filtered reads were assembled into 37,989 scaffolds using SOAPdenovo and SPACE (total 3.06 Gb). Anchoring of those contigs on high-density genetic map could assembled 86% of the (2.63 Gb; 1357 scaffolds) scaffolds onto 12 pseudomolecules of capsicum genome. A total of 34,903 protein-coding genes were identified using the PGA annotation pipeline. The comparative analysis showed a high level of conservation with its closest relative, i.e., tomato, as 17,397 orthologous gene sets were identified, and their expression studies revealed that 8.8% of them showed expression in leaf tissue and 46.4% were found to be expressed in pericarp tissue. As the pepper genome is four times larger than tomato, the genome size increment seen is mainly due to the presence of a large number of repetitive elements such as LTR retrotransposons. Of the reported retrotransposons, the Gypsy family was found to present 12-fold more than the Copia family when compared to another genome such as tomato, maize, and barley. Moreover, the expression analysis of different capsaicinoid pathway genes showed that all genes were expressed at 16 DPA, 25 DPA, and mature green stages of pepper fruit, but their orthologous genes hardly showed any expression in tomato and potato fruits. This study confirms the specificity of capsaicinoid pathway in the development of pungent flavor in pepper fruit.

To provide a better understanding of evolution and domestication of capsicum, Qin et al. [61] reported two reference genome sequences of cultivated Zunla-1 (*C. annuum* L.) and wild Chiltepin (*C. annuum* var. glabriusculum) pepper. They estimated the genome size of 3.26 Gb and 3.07 Gb, respectively. The reads were assembled in scaffolds comprising 3.48 and 3.35 Gb, respectively. They found different transposable elements (TEs) that covered ~2.7 Gb (81%) of the genome and estimated that the pepper genome expanded ~0.3 Mya. Approximately 79% of 3.48 Gb scaffolds contained 34,476 protein-coding genes that were anchored to chromosomes by a high-density genetic map. Using an in-house-generated program, they identified 6527 long noncoding RNAs (lncRNAs), which comprised 5976 intergenic and 222 intron-overlapping lncRNAs. In addition, the sequencing of small RNAs from five different tissues allowed the identification of 5581 phased siRNAs. Based on plant micro-RNAs (miRNAs) miRBase database, a total of 176 miRNAs were discovered of which 35 were found to be specific to pepper. They also predicted 1104 target genes that have putative functions such as dihydrolipoamide dehydrogenase (*Capana12g000245*) and α-CT (*Capana09g001602*) genes from capsaicinoid biosynthetic pathway, suggesting the regulation of capsaicinoid biosynthesis by miRNAs. Further, they identified 31% constitutively expressed genes and also 3670 genes that were showing tissue-specific expression. The annotation of these genes resulted in the identification of candidate genes for various traits. By a comparison of cultivated and wild pepper genomes with data of 20 resequencing accessions, they identified genes for domestication, which revealed molecular footprints of artificial selection. Moreover, they identified 51 gene families involved in capsaicinoid biosynthesis, and based on the phylogenetic analysis, they concluded that independent pepper-specific duplications in 13 gene families had occurred compared with tomato, potato, and *Arabidopsis*.

*3.2.4. Eggplant genome*

To elucidate the genome structure and complexity, a draft genome sequence of eggplant has recently been published in which the whole genome shotgun sequencing of eggplant variety named as "Nakate-Shinkuro" was carried out using HiSeq 2000 sequencer (Illumina) [62]. The high-quality reads were assembled using SOAPdenovo v1.05 into 1,321,157 scaffolds and presented a draft genome assembly "SME_r2.5.1" that spanned approximately 74% (833.1 Mb) of the total 1127 Mb of the eggplant genome. Also, transcriptome sequencing of "AE-P03" and "LS1934" was carried out using Roche/454 FLX sequencer (Roche Diagnostics, Basel, Switzerland). By merging the two data sets, a hybrid assembly was produced using PCAP.rep that constituted 81,273 hybrid scaffolds of a total of 836.8 Mb in size. They predicted about 42,035 protein-coding genes in SME_r2.5.1 by Augustus 2.7. A total of 16,573 genes were located on superscaffolds and showed an orthologous relationship with tomato.

## 3.3. Sequence-based molecular marker discovery and genetic mapping

Sequence-based molecular markers have been used in many comparative and functional genomics studies because of their preferable features like genome-wide distribution, chromosome-specific location, co-dominant inheritance, and reproducibility. The high-throughput NGS technologies produce a huge amount of data, which is highly suitable for the identification of a large number of sequence variations in genome or transcriptome. For SNP identification, various SNP calling programs such as SOAPsnp [63], MAQ [64], Atlas-SNP2 [65], SAMtools [66], and GATK [67, 68] have been used commonly [69].

In tomato, Sim et al. [70] developed the first large-scale SNP genotyping array using 8784 SNPs based on NGS-derived transcriptome sequences of six different genotypes [71]. They constructed three high-density linkage maps using interspecific $F_2$ populations (with various accessions of *S. lycopersicum* and *S. pennellii*). The physical positions of about 7666 SNPs were identified relative to the draft tomato genome sequence and found that the genetic and the physical distances were persistent. Such maps help to provide details of genetic order and recombination, also to improve gene assemblies and to dissect the complex traits. In another study, the genome-wide SNP genotyping was carried out with 7617 SNPs in 40 tomato lines and identified 6474 polymorphic SNPs [72]. Further, the effect of SNPs on protein function was studied, which revealed that the function of about 200 genes was altered by the substitutions phenomenon.

In eggplant, Barchi et al. [73] mapped QTLs associated with anthocyanin pigmentation using inter- and intraspecific linkage maps. They used a combination of the restriction site-associated DNA (RAD) strategy with high throughput sequencing (Illumina) to generate SNPs. A total of 415 of the 431 markers were assembled into twelve major and one minor linkage group, covering 1390 cM distance.

Very recently, in pepper, Devran et al. [74] developed molecular markers tightly linked to *potyvirus resistance* 4 (Pvr4) by sequencing the parental lines and progenies using Illumina Hi-Seq2500 in combination with bulked segregant analysis (BSA) approach. By comparative analysis, they identified the syntenic regions between resistant and susceptible progenies, and

more than 5000 single-nucleotide variants (SNVs) were identified that were converted into CAPS markers and used to map *Pvr4* locus using $F_2$ mapping populations. In a separate study, intron-targeting (IT) markers were developed from the NGS (5500xl SOLiD)-derived transcripts in tetraploid potato cv. White lady [75]. These markers were tested on various potato genotypes and in other *Solanum* species. A detailed list of reports of NGS-based molecular marker is given in Table 1.

| S. No. | Type of study | Population/species | Number of SSRs | Number of SNPs/ InDels | NGS platform | Reference |
|---|---|---|---|---|---|---|
| | **Capsicum** | | | | | |
| 1 | Transcriptome profiling | TF68 (*Capsicum annuum*) | 751 | 1536 SNPs 101 InDels | 454 GS-FLX | [39] |
| 2 | Transcriptome profiling | Yolo Wonder and Criollo de Morelos 334 (both *C. annuum*) | 853 | 11,849 | 454 GS-FLX and Illumina | [40] |
| 3 | Transcriptome profiling | Bukang (*C. annuum*) First assembly | 2,489 | 4,236 | Illumina | [41] |
| | | Second assembly | 10,398 | 22,000 | Illumina | |
| 4 | Transcriptome profiling | Xiaomila (*Capsicum frutescens*) | 4,072 | 9,150 | Illumina | [43] |
| 5 | Transcriptome profiling | Mandarin (*C. annuum*) | – | 1025 | 454 GS-FLX | [44] |
| | | Blackcluster (*C. annuum*) | – | 1059 | | |
| 6 | Whole genome re-sequencing | BA3 (*C. annuum*) | – | 154,519 InDels | Illumina | [76] |
| | | BA07 (*C. annuum*) | – | 149,755 InDels | | |
| 7 | Genome sequencing with BSA | SR231 and Criollo de Morelos334 (*C. annuum* L.) | – | 5,000 SNV | Illumina HiSeq 2500 | [74] |
| | **Tomato** | | | | | |
| 1 | Whole genome re-sequencing | Ailsa Craig, Furikoma, M82, Tomato Chuukanbonhon Nou 11, Ponderosa and Regina (All are inbred lines of *Solanum lycopersicum*) | – | 1536 SNPs were selected for genotyping of which 1293 successfully genotyped and 1248 found polymorphic | Re-sequencing with ABI SOLiD and Genotyping by Illumina GoldenGate Assay | [77] |
| 2 | Whole transcriptome sequencing | 8 accessions of (*S. lycopersicum*) and 1 of (*Solanum pimpinellifolium*) | – | 62,576 non redundant putative SNPs | Illumina | [30] |

| S. No. | Type of study | Population/species | Number of SSRs | Number of SNPs/ InDels | NGS platform | Reference |
|---|---|---|---|---|---|---|
| | **Capsicum** | | | | | |
| 3 | Whole genome re-sequencing and transcriptome re-sequencing | Several accessions of *S. lycopersicum* and *S. pimpinellifolium* | – | 4,812,432 non-redundant SNPs | Illumina and 454 GS-FLX | [78] |
| 4 | Whole genome sequencing | *S. pimpinellifolium* | – | 4,680,647 | Illumina and 454 GS-FLX | [78] |
| 5 | Whole genome re-sequencing | 'Micro-Tom' and 'Heinz 1706' of *S. lycopersicum* | – | 1,231,191 | Illumina and 454 | [79] |
| 6 | Genome sequencing and transcriptome sequencing | *S. lycopersicum* accessions | – | 6,000 (identified) 5528 (validated) | Illumina and 454 | [80] |
| | **Eggplant** | | | | | |
| 1 | Genome sequencing | accessions of *Solanum melongena* and *Solanum aethiopicum* | 2,000 putative SSRs | 10,089 SNPs 874 (InDels) | Illumina | [81] |
| | **Potato** | | | | | |
| 1 | Transcriptome sequencing | *Solanum tuberosum* | – | 575,340 SNPs | Illumina | [26] |
| 2 | Genome sequencing | *S. tuberosum* | – | 111,212 SNPs 13,094 InDels | Illumina | [82] |

Note: SNP—single-nucleotide polymorphism, SNV—single-nucleotide variant, SSR—simple sequence repeat, InDels—insertion/deletion.

**Table 1.** List of transcriptome and whole genome sequencing using NGS technologies for development of genomic resources in Solanaceae crop plants

### 3.4. Epigenomics during the age of next-generation sequencing technologies

Molecular breeding has a crucial role in the improvement of crops. Although conventional breeding program brought a substantial increment of food production, however, with rapid population growth worldwide, crop improvement should be accelerated so that climate resilient, biotic stress-resistant, high-nutritional, and high-productivity cultivars could be developed. The advent of NGS made it possible to study phenotypic variations caused by genetic and epigenetic modification to facilitates crop improvement. The term epigenotype was first introduced by Conrad H. Waddington to demonstrate the sum of interrelated developmental pathways that enable one genome to give rise to multiple epigenomes and consequently to multiple cell types that make up the whole organism. Nowadays, the term epigenetics is commonly referred to all kinds of heritable changes that are not caused by

changes in the alteration of DNA sequences but are triggered by chemical modifications on the DNA (cytosine methylation) or on histone modifications (e.g., acetylation, methylation) bringing about modulation of chromatin structure and function [83]. In recent years, small RNAs have been emerged as key players in controlling epigenetic changes throughout the plant genome.

### 3.4.1. DNA methylation

DNA methylation refers to the covalent addition of methyl group to the cytosine base at position 5 by the action of DNA methyl transferases. In mammals, cytosine methylation occurs mostly at CG sites and rarely at non-CG sites, while in plants, cytosine methylation can occur in both CG and non-CG contexts. Non-CG methylation involves both symmetrical and asymmetrical sites, CHG and CHH, respectively (H = A, T, or C). Much of our knowledge with respect to DNA methylation is based on the studies performed on model plant *Arabidopsis thaliana*. DNA methylation in plants is being catalyzed principally by three different enzymes. The maintenance of symmetrical CG methylation during DNA replication is carried out by *Methyltransferase1* (*MET1*) (homolog of animal DNA methyltransferase DNMT1), while CHG methylation is catalyzed by the plant-specific *chromomethylase 3* (CMT3) and asymmetric CHH methylation is mediated by *domains rearranged methyltransferase 2* (*DRM2*) (similar to the mammalian *DNMT3* family) activity, which works through RNA-directed DNA methylation (RdDM) pathway [83, 84].

The first ever single-base resolution methylomes of tomato fruits were established, which revealed that fruit epigenome is not static, and the changes occur continuously during different stages of fruit development. The whole genome bisulfite sequencing was employed to study four different stages of fruit development. This study identified 52,095 differentially methylated regions of the 90% of the genome covered in this analysis in wild-type tomato fruits. Comparative analysis of fruits from two nonripening mutants of tomato viz ripening-inhibitor (rin) and Colorless nonripening (Cnr) demonstrated the changes in the methylation patterns in the wild type and the mutants [85]. The *Cnr* mutation in tomato restricts normal ripening process in tomato resulted in a colorless fruits develop a colorless pericarp [86]. Silencing of the *SlCMT3* gene in tomato resulted in the increased expression of *LeSPL-CNR* that encodes for SBP-box transcription factor, which was located in the *Cnr* locus that ultimately triggers *Cnr* fruits to ripen normally. These studies revealed that the induced ripening of *Cnr* fruits is associated with a reduction of methylation at CHG sites of the *LeSPL-CNR* promoter, while a decrease of DNA methylation in differentially methylated regions associated with the *LeMADS-RIN* binding sites [87, 88].

### 3.4.2. Histone modifications

The interaction between DNA and proteins has a crucial role in the regulation of gene expression. Chromatin immunoprecipitation (ChIP) can be employed to study such interactions. These interactions can be explored using a technique called ChIP, microarray platforms (ChIP-on-chip or ChIP-chip) [89, 90]. More recently, NGS-based techniques are being used for studying histone modifications where ChIP-Seq combines ChIP with massively parallel direct

sequencing. ChIP-enriched DNA is sequenced directly, using the Solexa/Illumina platform, and the readings were mapped to the reference genome. Histone modification phenomenon includes methylation, acetylation, phosphorylation, ubiquitination, sumoylation, and ADP-ribosylation. These modifications bring changes directly and cause structural changes to the chromatin or indirectly through the mediator proteins. All histone modifications are reversible and provide versatile ways for regulating gene expression during plant development and their responses to environmental stimuli. The study found that the reversible acetylation and deacetylation of specific *Lys* residues on core histone N-terminal tails catalyzed by histone acetyltransferases (HDA) and histone deacetylases (HDAC), respectively [91, 92]. The action of both enzymes regulates biological processes like transcriptional regulation. It was found that generally, hyperacetylated histones are associated with gene activation, whereas hypoacetylated histones were involved in gene inactivation. ChIP-seq was employed to identify the targets of *ASR1* starting out with the purification of *ASR1*, by using the high-quality anti-*ASR1* antibody. ChIP-seq data generated through this helped in identifying the genes encoding aquaporins and those associated with the cell wall; these genes were associated with drought stress response [93]. There are several studies reported where ChIP-seq along with ChIP-chip methods were used to search genomes for locations associated with binding of several transcription factors (TFs) such as *RIN* and *fruitful* homologs (FUL1/FUL2) [85, 94, 95]. The investigation of genome-wide targets for the main regulators of fruit ripening viz. *RIN*, *FUL1*, and *FUL2* by combining RNA-Seq with ChIP-chip assay identified a total of 292, 860, and 878 target ripening-associated genes in tomato [85, 95]. Therefore, a combination of ChIP-seq and RNA-Seq with ChIP-chip are imperative tools nowadays and can be employed for better understanding of transcriptional networks underlying tomato development.

## 3.5. Noncoding RNA (ncRNAs) in crop improvement

Recent advances in next-generation genome and transcriptome sequencing with thorough bioinformatics and computational analysis laid to the discovery of numerous RNA types. The ncRNAs are one of the great examples of such techniques. The ncRNAs has emerged as a key product of eukaryotic transcriptionary machinery with a critical role in the regulatory mechanism. The ncRNAs are being classified as housekeeping ncRNAs and regulatory ncRNAs [96]. The rRNAs, tRNAs, small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs) are under the "housekeeping" ncRNAs, whereas the "regulatory" ncRNAs are known as small ncRNAs (such as miRNAs and siRNAs) and long noncoding RNA (lncRNAs) [96, 97].

### 3.5.1. Role of long noncoding RNAs in Solanaceae

The lncRNAs are defined as a non-protein-coding functional RNAs of more than 200 bp in length with regulatory function and principally transcribed by RNA polymerase II. The identification of lncRNA in plants and especially in Solanaceae is still at infancy as compared with the human/animal genome. The application of high-throughput NGS technologies toward identification and the characterizations of lncRNAs are being reported. Recently, by analyzing around 200 *A. thaliana* transcriptome data sets, about 6480 lncRNAs were identified

in the intergenic regions of the genome [98]. Further, 439 lncRNAs were identified in maize [99], and in a more comprehensive way by integrating all available data sets for maize transcriptome, high confidence 1704 lncRNAs were identified [100]. However, a systemic study on lncRNAs in Solanaceae has not been done except some few reports. In pepper, a total of 5976 long intergenic ncRNAs (lincRNAs), 222 intronic overlapping lncRNAs, and 329 bidirectional overlapping lncRNAs were identified from RNA-seq data of unopened flower buds [44]. Recently, a genome-wide identification of lncRNAs in tomato was reported [101]. The study identified a total of about 3679 lncRNAs from wild-type AC tomato and mutant ripening fruit (*rin*). The analysis further reported that out of 3530 and 3679 lncRNAs identified in wild-type and *rin* mutant tomatoes, only 23 and 126 lncRNAs were transcribed specifically in wild-type and *rin* mutant tomatoes, respectively. Most of the lncRNAs are derived from intergenic regions. It was also found that 490 lncRNAs were upregulated in ripening mutant fruits, while 187 lncRNAs were downregulated, suggesting the involvement of lncRNAs in the regulation of fruit ripening. However, the function of lncRNAs has not been fully understood and studied. In a more conclusive study, the role of lncRNAs known as *COOLAIR* (cool-assisted intronic noncoding RNA) and *COLDAIR* (cold-assisted intronic noncoding RNA) during vernalization was investigated. These lncRNAs are involved in the epigenetic silencing of *FLC* gene that subsequently promotes flowering [102]. The identification and the characterization of novel lncRNAs have enormous potential to open new windows for crop improvement. Therefore, databases of lncRNAs named as PLncDB (plant long noncoding RNA database) [103] and PNRD (plant ncRNA Database) [104] have been developed which provide information about the functions and role of lncRNAs in plants.

### 3.5.2. Role of miRNAs in regulation of gene expression

MicroRNAs (miRNAs) are approximately 21 nucleotides long in length, and they are a class of noncoding RNAs that play an important role in regulating gene expression in plants [105–107]. Plant miRNAs mostly exert their effects by cleavage of target mRNA with full complementarity, and their target sites are mostly found in coding regions thus altering the gene expression [105–107]. Recent studies have shown that plant miRNAs also repress translation via a slicer-independent mechanism and, therefore, mediates the expression of the genes posttranscriptionally [108, 109].

There are mainly two major approaches for identifying miRNAs in plants: (1) experimental and (2) bioinformatic approaches. An experimental approach includes forward genetics, direct cloning, and next-generation high-throughput sequencing. High-throughput sequencing technology showed significant progress in small RNA identification and has become commonly available and affordable tool nowadays. A large number of miRNAs have been identified by means of high-throughput sequencing and available in online database (http://www.mirbase.org, accessed June 21, 2014), which currently holds 35,828 mature miRNA products from 223 species. The majority of miRNAs identified so far have been obtained from only a few model plant species, such as *A. thaliana*, *Oryza sativa*, *Glycine max*, and *Medicago truncatula*. Despite the largest family in the plant kingdom, the annotated miRNAs are still

very limited in Solanaceae [110–113]. It is necessary to understand the function of miRNAs in Solanaceae. The study of the miRNAs in pepper has been reported based on identification using an *in silico* approach [114]. However, there is a need to employ high-throughput sequencing approaches on the pepper to discover miRNAs. Recently high-throughput sequencing technologies have been employed to identify miRNAs in pepper from ten different tissues such as leaf, stem, root, flower, and six developmental stages of fruits. Based on a bioinformatics pipeline, the researchers successfully identified 29 and 35 families of conserved and novel miRNAs, respectively. Moreover, their miRNA targets were also predicted com-putationally, many of which were experimentally validated using 5′ rapid amplification of cDNA ends (RACE) analysis. Among them, one of the confirmed novel targets of miR-396 was a domain-rearranged *methyltransferase*, the major *de novo* methylation enzyme responsible for RNA-directed DNA methylation in plants. These studies carried out using NGS technologies provide a basis for understanding the functional roles of miRNAs in pepper that can be explored for the crop improvement [115].

Kim et al. [114] identified miRNAs and their target genes by analyzing expressed sequence tag (EST) data from five different species of Solanaceae, wherein they revealed the presence of at least 11 miRNAs and 54 target genes in pepper (*C. annuum* L.) and 22 miRNAs with 221 target genes in potato (*S. tuberosum* L.). Apart from this, they identified a total of 12 miRNAs with 417 target genes in tomato, 46 miRNAs with 60 target genes in tobacco (*Nicotiana tabacum* L.), and 7 miRNAs with 28 target genes in *Nicotiana benthamiana*. Further, the identified miRNAs with their target genes were submitted to the SolmiRNA database, (http://gene-pool.kribb.re.kr/SolmiRNA). They showed the presence of both conserved and specific miRNAs, which may play crucial roles in the growth and development of Solanaceae plants. In addition, 12 miRNAs were randomly selected from a differentially expressed conserved miRNA family and subjected to qRT-PCR validation. Of these, the expression level of nta-miR167d was highly enriched in the leaf tissue, whereas the expression level of nta-miR319a and nta-miR160c were specifically found in stem and root tissues, respectively. The target prediction showed that most of the targets genes were those which codes for transcription factors involved in cellular and metabolic processes [116]. Similar study was performed where deep sequencing of leaf, stem, and root, and four early developmental stages of tubers were performed [117]. The study revealed a total of 89 conserved miRNAs belonging to 33 families and 147 novel miRNAs with 112 candidate potato-specific miRNAs. Digital expression profiling based on TPM (transcripts per million) and qRT-PCR analysis of conserved and potato-specific miRNAs revealed that some of the miRNAs showed tissue-specific expression (leaf, stem, and root), while a few demonstrated tuber-specific expressions. Further, targets were predicted for the identified conserved and potato-specific miRNAs. The predicted targets of four conserved miRNAs are as follows, *ARF16* (*auxin response factor* 16) for miR160, NAM (*no apical meristem*) for miR164, RAP1 (*relative to Apetala2* 1) for miR172, and HAM (*hairy meristem*) for miR171. Later they were experimentally validated using 5′ RLM-RACE (RNA ligase mediated rapid amplification of cDNA ends). The list of databases for miRNA identifi-cation is presented as Table 2.

| Database | Description | Link | Reference |
|---|---|---|---|
| miRBase | Database of published miRNA sequences and their annotation | http://www.mirbase.org/ | [118–122] |
| deepBase | A platform for annotating and discovering small and long ncRNAs (microRNAs, siRNAs, and piRNAs) from next generation sequencing data | http://deepbase.sysu.edu.cn/ | [123] |
| miRanda-microRNA.org | Database for predicted microRNA targets, target downregulation scores and experimentally observed expression patterns | http://www.microrna.org/microrna/home.do | [124] |
| DIANA-mirGen 2.0 | Database of miRNA genomic information and regulation | http://diana.cslab.ece.ntua.gr/mirgen/ | [125] |
| miRNAMap | miRNAMap Genomic maps of miRNA genes and their target genes in human, mouse, rat, and other metazoan genomes | http://mirnamap.mbc.nctu.edu.tw/ | [126, 127] |
| PMRD | Plant miRNA database with large information of plant microRNAs data, consisting of microRNA sequence and their target genes, secondary dimension structure, expression profiling, genome browser, etc. | http://bioinformatics.cau.edu.cn/PMRD/ | [128] |

**Table 2.** List of databases for miRNA identification

### 3.5.3. miRNAs in plant growth and development

To investigate the role of miRNAs in ovary and fruit development of tomatoes, transgenic plants were generated by overexpressing MIR167. The transgenic plants showed a reduction in leaf size and internode length as well as shortened petals, stamens, and styles. The RNA-Seq analysis identified many genes with altered expression patterns in tomato. Of these, *SpARF6* and *SpARF8* genes involved in flower maturation in *Arabidopsis* have been found to be significantly down regulated [129]. In a separate study, it was found that transgenic tomato plants harboring AtMIR156b (*A. thaliana* miRNA 156b family) precursor resulted in abnormal flower and fruit morphology; in addition, the fruits were characterized by the growth of extra carpels and ectopic structures [130]. Moreover, these transgenic lines also displayed increased the expression of genes, which are involved in maintenance of meristem and formation of new organs such as *LeT6/TKN2* (a KNOX-like class I gene) and *GOBLET* (a NAM/CUC-like gene). Overall, these observations suggest that the miR156 is involved in the maintenance of the meristematic activity of ovary tissues and participates in the normal fleshy fruit development.

Several miRNAs have been identified in the fruit tissue. However, no miRNA has been experimentally validated to be involved in fruit ripening. Recently, *SlymiR157* and *Sly-miR156* have been shown to regulate ripening and softening of tomato fruits. SlymiR157 governs the expression of key ripening gene *LeSPL-CNR* by miRNA-induced mRNA degradation and by translational repression. Furthermore, qRT-PCR profiling of key ripening-related genes reveals that the SlymiR157-target LeSPL-CNR may also affect the expression of *LeMADS-RIN*, *LeHB1*, *SlAP2a*, and *SlTAGL1* [131]. Table 3 contains the list of databases for miRNA target gene prediction.

| Database | Description | Link | References |
|---|---|---|---|
| starBase | Interaction Networks of lncRNAs, miRNAs, competing endogenous RNAs (ceRNAs), RNA-binding proteins (RBPs), and mRNAs from large-scale CLIP-Seq (HITS-CLIP, PAR-CLIP, iCLIP, and CLASH) data | http://starbase.sysu.edu.cn/ | [132, 133] |
| miRwalk 2.0 | Database with collection of predicted and experimentally verified miRNA–target interactions with various novel and unique feature | http://zmf.umm.uniheidelberg.de/apps/zmf/mirwalk2/index.html | [134] |
| targetScan | Database and Webserver for predicted miRNA targets in animals | http://www.targetscan.org/ | [135–137] |
| DIANA-TarBase v7.0 | DIANA-TarBase v7.0 provides for the first time hundreds of thousands of high quality manually curated experimentally validated miRNA–gene interactions | http://diana.imis.athenainnovation.gr/DianaTools/index.php?r=tarbase/index | [138, 139] |
| DIANA -microT v3.0 | Accurate microRNA target prediction database | http://diana.cslab.ece.ntua.gr/microT/ | [140, 141] |
| miRecords | Manually curated database of experimentally validated miRNA–target interactions | http://c1.accurascience.com/miRecords/prediction_query.php | [142] |
| picTar | PicTar: a computational method for identifying common targets of microRNAs | http://pictar.mdc-berlin.de/ | [143] |
| RNA22 | Web based browser to identity miRNA targets | https://cm.jefferson.edu/rna22/Interactive/ | [144] |
| micTarBase | miRTarBase has accumulated more than fifty thousand miRNA–target interactions (MTIs) | http://mirtarbase.mbc.nctu.edu.tw/ | [126, 127] |

| Database | Description | Link | References |
|---|---|---|---|
| RNALogo | Database with novel graphical representation of the patterns in an aligned RNA sequences with a consensus structure | http://rnalogo.mbc.nctu.edu.tw/ | [145] |
| miRGator | Database with microRNA diversity, expression profiles, and target relationships | http://mirgator.kobic.re.kr/ | [146–148] |
| miRNAMap | miRNAMap Genomic maps of miRNA genes and their target genes in human, mouse, rat, and other metazoan genomes | http://mirnamap.mbc.nctu.edu.tw/ | [112] |
| miRDB | Webserver for miRNA target prediction and functional annotation | http://mirdb.org/miRDB/ | [149] |
| RNA hybrid | This tool is primarily meant as a means for microRNA target prediction | http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/ | [150] |
| miRU, psRNAtarget | A Plant Small RNA Target Analysis Server | http://plantgrn.noble.org/ psRNATarget/ | [151] |
| miRNEST | miRNEST is an integrative collection of animal, plant and virus microRNA data | http://rhesus.amu.edu.pl/mirnest/ copy/browse.php | [152] |
| PMTED | Plant MicroRNA Target Expression Database | http://pmted.agrinome.org/ by_mirna.jsp | [153] |
| MIREX | A platform for comparative exploration of plant pri-miRNA expression data | http://www.comgen.pl/mirex2/ | [154] |
| TAPIR | Target prediction for plant microRNAs | http://bioinformatics.psb.ugent.be/ webtools/tapir/ | [155] |
| PASmiR | A database for miRNA molecular regulation in plant abiotic stress | http://pcsb.ahau.edu.cn:8080/ PASmiR/ | [156] |

**Table 3.** List of databases for miRNA target gene prediction

### 3.5.4. miRNAs in biotic stress

miRNAs have been identified in many plants with their diverse regulatory roles in biotic stresses. miRNA sequencing was used to investigate the miRNA expression difference between the tomatoes treated with and without *Phytophthora infestans*. Using high-throughput sequencing technologies, they could identify a total of 207 known miRNAs and 67 novel miRNAs. In addition to this, a total of 70 miRNAs were differentially regulated in the plants treated with *P. infestans*; of these, 50 were downregulated and 20 were upregulated. Also, a total of 73 target genes were identified for 28 differentially expressed miRNAs by using psRNATarget analysis [157].

The fungus *Fusarium oxysporum* f. sp. *lycopersici* causes vascular wilt disease in tomato. A comparative miRNA profiling of susceptible (Moneymaker) and resistant (Motelle) tomato cultivars were performed to explore the role of miRNAs in tomato defense against *F. oxysporum*. *SlmiR482f* and *SlmiR5300* were repressed during infection of Motelle with *F. oxysporum*. Four predicted mRNA targets, two each of slmiR482f and slmiR5300, displayed increased expression in resistant Motelle. This was further confirmed by co-expression analysis in *N. benthamiana*. Silencing of the targets in the resistant Motelle cultivar compromised the resistance to *F. oxysporum* and confirmed the role of these genes in fungal resistance [158].

### 3.5.5. miRNAs in abiotic stress

Abiotic stress (such as salt, drought, and heat) is becoming a major constraint to crop production due to the climate change. miRNAs have been found to play a significant role in tolerance to these stresses. For example, in tomato, transgenic lines were generated by the overexpression of miR169 family member: Sly-miR169c that displayed reduced stomatal opening, decreased transpiration rate, reduced water loss, and enhanced drought tolerance [159]. In eggplant, the high-throughput sequencing of salt tolerant species was performed and identified 98 conserved miRNAs from 37 families [160]. Some of them were found to be expressed under salt stress. These studies provide a better understanding about the regulation of gene expression under abiotic stresses for genetic improvement of crops.

## 4. High-throughput genotyping technologies

With the development of various NGS platforms, thousands to millions of SNPs have been identified from whole genome and transcriptome sequence data. Therefore, various high-throughput genotyping platforms were developed simultaneously for large-scale genotyping of SNPs in a large set of individuals. These platforms are the GoldenGate Genotyping Technology (GGGT; Illumina, San Diego, CA, USA) [161], BeadChip-based Infinium assay (Illumina) [162], SNPStream (Beckman Coulter, USA) [163], GeneChip (Affymetrix, USA) [164], and competitive allele-specific PCR, KASPar (KBioscience, UK) [165].

### 4.1. GoldenGate Genotyping Technology (GGGT)

The Illumina GGGT is a custom-based platform that covers construction of 96-1536 SNPs assay. The method is based on BeadArray technology, which includes immobilization of genomic DNA on avidin-coated particle. A further step is annealing of two allele-specific oligonucleotides and a locus-specific oligonucleotide for each SNP, later allele-specific primer extension for generating allele-specific products followed by PCR amplification with universal primers. It is a custom-based genotyping platform that allows screening of a vast number of samples (up to 3072 SNPs) using a single multiplexed assay. Shirasawa et al. [77] utilized 1536-plex SNP genotyping in tomato, of which 1293 were genotyped successfully. Moreover, 1248 SNPs showed clear polymorphism in 663 accessions. For eggplant, Barchi et al. [73] identified >10,000 potential SNPs. Of these, 384 highest quality SNPs were used to genotype 23 diverse eggplant

germplasm with respect to fruit shape and color, and observed polymorphic information content values ranged from 0.29 to 0.5 with a mean value of 0.43.

### 4.2. BeadChip-based Infinium assay (Illumina)

It includes whole genome amplification followed by hybridization to oligonucleotide probe attached to a bead, extension, and detection of fluorescence by iScan Reader. The assay considers up to four million SNPs in a single sample run, or even up to several hundred thousand multiple samples in the same array. The chemistry involves incubation of samples on bead chip where they anneal to locus-specific 50-mers covalently linked to beads followed by allele-specific single-base extension, fluorescent staining, signal amplification, scanning in a dual-color channel reader, and analysis. This technology is advantageous as one can use a premade array that is easily available commercially for selected species. Hamilton et al. [26] identified 69,011 high confidence SNPs from six potato cultivars and used for genotyping with the Infinium platform. A total of 96 of these SNPs were used to assess allelic diversity in 248 germplasms and found 82 informative SNPs for subsequent analyses. In 2012, Felcher et al. [166] reported "Infinium 8303 Potato Array" comprising of 8303 functional markers which includes 3018 from candidate genes of interest by utilizing the transcriptome data from Hamilton et al. [26]. These were used for the genotyping and development of linkage maps. In tomato, a large-scale SNP genotyping array using 8784 SNPs were obtained from transcriptome sequencing [30] and later used for construction of a high-density linkage map of tomato [70].

### 4.3. SNPStream (Beckman coulter)

This method involves a single-base extension assay and tag array technology. It starts with a multiplexed SNP-specific PCR followed by a primer extension reaction using tagged primers and fluorescent-labeled nucleotide terminators, i.e., ddNTPs. The products are captured on a tag array, which is then scanned to detect the hybridized extension primers and produce calls. It allows the processing of up to three million genotypes in 384 samples at a time. This genotyping system combines solid-phase primer extension assay and universal tags for SNP genotyping. The instrument allows processing of 4,600–3,000,000 genotypes per day [167].

### 4.4. GeneChip (Affymetrix, USA)

The GeneChip assays are based on allelic discrimination by the direct hybridization of genomic DNA to arrays containing locus- and allele-specific oligonucleotides (25 mers). Genomic DNA is digested with a restriction endonuclease and ligated to adaptors, which are then amplified by PCR using a single universal primer thereby creating a reduced representation of the genome [168]. These PCR amplicons are fragmented, end-labeled, and hybridized. The fluorescence signal is recorded by the GeneChip 3000 scanner (Affymetrix). The hybridization scanning is evaluated as positive and negative signals. Hill et al. [42] developed a GeneChip® array for analysis of polymorphism and expression in Capsicum. The array was designed from 30,815 unigenes, and hybridization was performed using genomic DNA of 40 diverse lines of *C. annum*. They detected 33,401 single-position polymorphisms within 13,323 unigenes. A total of 251 highly informative markers across these *C. annuum* lines were found. Also, a region of 8.7 cM was detected around Pun1 locus in nonpungent line that showed no polymorphism.

In tomato, an oligonucleotide array was developed with 22,821 probe sets, which correspond to 22,714 unigenes [169]. Genomic DNA isolated from three *S. lycopersicum* varieties, i.e., FL7600 (fresh-market), OH9242 (processing), and PI114490 (var. cerasiformae), were used to hybridize with that array. They identified 189 putative single feature polymorphisms, and a subset of these was utilized for validation which resulted in the identification of 279 SNPs and 27 InDels in 111 loci. Moreover, a subset of validated SNPs was used for analysis of genetic diversity in 92 tomato varieties and accessions.

### 4.5. KASPar (KBioscience, UK)

The KBioscience-competitive allele-specific PCR (KASPar) is a simple, cost-effective, and flexible way for determining both SNP and InDel in genotypes. It is a custom-based technology that covers 96-1536-well plate formats like Illumina's GGGT. It relies on the discrimination power of a novel form of competitive allele-specific PCR to determine the alleles at a specific locus. The improvement has been made by incorporating a 5′–3′ exonuclease cleaved *Taq* DNA polymerase (the engineered *Taq* increases its discrimination power) and a homogeneous fluorescence resonance energy transfer (FRET) detection system, which makes this technology more competent among the genotyping platforms. From the pepper transcriptome sequence data, Ashrafi et al. [41] identified a large number of SNPs. A subset of them was validated by KASPar assay and identified 78 polymorphic SNPs.

## 5. Genotyping By Sequencing (GBS)

This technology is comparatively new in which genomic DNAs from large mapping populations are sequenced followed by SNP identification. This allows a rapid way for dissecting QTLs for economically important traits in large mapping populations besides allowing genetic diversity and the phylogenetic study between large numbers of accessions/genotypes. This approach is based on reduced representation sequencing, which involves the digestion of genomic DNA with appropriate restriction enzyme to capture a targeted portion of the genome followed by adapter (DNA-barcoded) ligation, PCR amplification, and sequencing of multiplexed libraries [170, 171]. For sequencing, the Illumina's GAII and HiSeq and latest with the Torrent PGM and Proton (Life Technologies) are used. To analyze the large sequencing data, several automated pipelines are being developed, including TASSEL, UNEAK, and IGST. Besides *de novo* SNP discovery, it offers the greatest advantage for those crops in which the solid reference genome sequence is absent. GBS has emerged as a high-throughput, robust, and cost-effective tool for genome-wide association studies and genomics-assisted breeding in numbers of plant and animal species, in particular for those having a complex genome. The utility of GBS has been demonstrated very well for discovery and genotyping of large number of SNPs, genetic mapping, diversity analysis, and population structure [172]. Among Solanaceae family, in potato, a high-quality sequence data of 12.4 Gb was obtained from which 129,156 sequence variants have been identified and mapped to 2.1 Mb of the potato reference genome with average read depth of 636 per cultivar [173].

# 6. Genome-Wide Association Study (GWAS)

The advent of NGS technologies provides a large number of sequence variants (mainly SNPs) within a shorter period. These sequence variants can be utilized for QTL mapping, GWAS, and germplasm characterization. The establishment of an association between genotype and phenotype is a very challenging task. For crop improvement, it is necessary to determine the genetic basis of the agronomic trait. GWAS is a powerful technique for detecting natural variation and fine mapping of QTL underlying complex traits [174]. It requires a collection of individuals or a population of diverse genotypes and highly polymorphic markers that showed genome-wide distribution. This is a very robust method, in comparison to biparental cross-mapping, to map multiple traits simultaneously. In tomato, Shirasawa et al. [77] reported the whole genome resequencing of six tomato cultivars and detected 1.5 million SNPs by mapping the reads onto the reference genome (SL2.40). They utilized Illumina GoldenGate assay for genotyping of 1536 SNPs in 663 tomato accessions. There was no population structure observed when analyzing the genetic relationship using the STRUCTURE software. Further, they identified a total of nine SNP loci that were found to be associated with eight morpho‐logical traits. To overcome the low polymorphism in cultivated tomato (*S. lycopersicum*), they used genome admixture of the cultivated and its wild ancestor (*S. pimpinellifolium)* for association mapping in tomato [175].

# 7. Next-generation sequencing toward translational research

## 7.1. Fruit traits (size, shape, ripening, and development)

The transcriptome studies in Solanaceae crops such as potato revealed the identification of transcription factors associated with fruit development. A total of 632 lineage-specific genes were identified, of which 289 genes were asterid specific and 343 were potato specific [23]. They identified 290 genes, including *pectin esterase*, *lipoxygenase*, and *malate synthase*. *Leafy Cotyledon 1* (*LEC 1*) and transcriptional factor *B3* were found to be co-expressed in fruit tissues. These TFs are consistently found to be involved in plant embryo development.

In tomato, using NGS technologies, several SNPs successfully differentiating between cherry type and round/beef type tomatoes were identified [80]. The SNP data revealed that cherry tomatoes share more SNPs with *S. pimpinellifolium*, a wild relative of the tomato. This revealed a close phylogenetic relationship of cherry tomato with the wild type. Several SNPs belonged to the chromosomal region that harbors genes/QTLs related to fruit weight, size, shape, and color, indicating that the SNPs may be used to explore the other fruit traits. In a miRNA study, it was observed that the transgenic tomato plants harboring AtMIR156b precursor resulted in abnormal flower and fruit morphology [130], indicating that mir156b plays crucial role in ovary and normal fleshy fruit development.

## 7.2. Tuber

The transcriptome of tuber tissue showed the presence of several transcripts that are specific for tuber. Around 90 genes were co-expressed in tuber, including the genes involved in starch

biosynthesis pathway such as *glucose 6-phosphate/phosphate translocator* and storage proteins such as *patatin* [23]. The *APETALA* and *WRKY* transcription factors were specifically found to be expressed in tubers. Further, using DGE profiling, the photoperiodic tuberization-specific genes were identified and suggested that the potato tuberization may be controlled by the genes associated with flowering time in other plant species [25]. These data contribute toward the development of powerful resources that could be used in candidate gene mining for important agricultural traits.

### 7.3. Pungency

Pungency is a special and economically important quality trait only found in pepper fruits, and it has been studied extensively [7, 43]. NGS technology has a wide scope to explore this trait and provides insights into the capsaicinoid pathway revealing the genes/loci associated with pungency. The transcriptome profiling of *C. frutescens* revealed the identification of three structural genes, namely, dihydroxyacid dehydratase (DHAD), Thr deaminase (TD), and prephenate aminotransferase (PAT) involved in the capsaicinoid biosynthesis pathway [43]. They claimed the identification of several new candidate genes involved in the capsaicinoid pathway. The comparative transcriptomic study of pepper with potato and tomato showed that the different capsaicinoid pathway genes were expressed during placenta development at 16 DPA, 25DPA, and mature green stages of pepper fruits, but their orthologous genes hardly showed any expression in tomato and potato fruit [7]. The study confirmed the specificity of capsaicinoid pathway in the development of pungency in pepper fruit.

### 7.4. Disease resistance

Using NGS technology, single-nucleotide variants (SNVs) were identified in resistant and susceptible pepper population for potato virus Y and pepper mottle virus. The comparative genomic tools were used to align the SNVs with syntenic region/loci of tomato. Later, the SNVs were converted into PCR-based CAPS (cleaved amplified polymorphic site) marker to map *potyvirus resistance 4* (*pvr4*) locus. These molecular markers could be used in large-scale marker assisted selection (MAS) programs [74].

### 7.5. Hormone and stress

Global transcriptome profiling of exogenously applied ABA tomato seedling revealed the identification of a large number of genes related to various stress responses [31]. These included several transcription factors, heat shock proteins, and pathogen resistance. Apart from this, salicylic acid, jasmonic acid, and ethylene signaling pathways were upregulated by exogenous ABA. The study suggested the role of ABA in improving pathogen resistance and abiotic stress tolerance. Moreover, the tomato transgenic lines were developed with the overexpression of Sly-miR169c, a miR169 family member. The transgenic plants displayed reduced stomatal opening, decreased transpiration rate, reduced water loss, and enhanced drought tolerance [159].

## 8. Conclusion and future direction

As the sequencing technologies are advancing at a rapid rate, enormous genomic information is being generated for Solanaceae crop plants. The question at present is how to utilize this enormous NGS-generated information for Solanaceae translational research. The large-scale phenotyping and transcriptome and whole genome resequencing of diverse genotypes from each species and their correlation will help in the identification of genetic region and eventually of candidate genes in the genomes. The integration of classical genetics, QTL mapping, and whole genome and transcriptome sequencing would be helpful in accelerating the Solanaceae translational research. Consideration of noncoding RNAs and epigenetics mechanism while designing breeding strategies would expedite the manipulation of mechanisms underlying various developmental aspects of plant biology in Solanaceae. Furthermore, the use of NGS technology provides an opportunity to investigate and understand the structure and evolution of complex Solanaceae genomes.

## Author details

Sushil Satish Chhapekar, Rashmi Gaur, Ajay Kumar and Nirala Ramchiary[*]

*Address all correspondence to: nrudsc@gmail.com

School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

## References

[1] Schmidt MA, Barbazuk WB, Sandford M, et al. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. Plant Physiol. 2011;156:330–345. doi:10.1104/pp.111.173807.

[2] Suzuki T, Igarashi K, Dohra H, et al. A new omics data resource of *Pleurocybella porrigens* for gene discovery. PLoS One. 2013;8(7):e69681. doi:10.1371/journal.pone. 0069681.

[3] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74(12):5463–5467.

[4] Metzke ML. Sequencing technologies: the next generation. Nat Rev Genet. 2010;11:31–46. doi:10.1038/nrg2626.

[5] The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485(7400):635–641. doi:10.1038/nature11119.

[6]   Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475(7355):189–195. doi:10.1038/nature10158.

[7]   Kim S, Park M, Yeom SI, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet. 2014;46(3):270–278. doi: 10.1038/ng.2877.

[8]   Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437 (7057):376–80. doi:10.1038/nature03959.

[9]   Valouev A, Ichikawa J, Tonthat T, et al. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Res. 2008;18 (7):1051–1063. doi:10.1101/gr.076463.108

[10]  Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev. 2006;16(6):545–552. doi:10.1016/j.gde.2006.10.009.

[11]  Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348–52. doi:10.1038/nature10242.

[12]  Harris, TD, Buzby PR, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. Science. 2008;320:106–109. doi:10.1126/science.1150427.

[13]  Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–138. doi:10.1126/science.

[14]  Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7:461–465. doi:10.1038/nmeth.1459.

[15]  Kasianowicz, JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci U S A. 1996;93:13770–13773.

[16]  Braha, O. et al. Designed protein pores as components for biosensors. Chem Biol. 1997;4:497–505.

[17]  Branton D, Deamer DW, Marziali A, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol. 2008;26:1146–1153. doi:10.1038/nbt.1495.

[18]  Maitra RD, Kim J, Dunbar WB. Recent advances in nanopore sequencing. Electrophoresis. 2012;33(23):3418–3428. doi:10.1002/elps.201200272.

[19]  Flinn B, Rothwell C, Griffiths R, et al. Potato expressed sequence tag generation and analysis using standard and unique cDNA libraries. Plant Mol Biol. 2005;59:407–433. doi:10.1007/s11103-005-0185-y.

[20]  Ronning CM, Stegalkina SS, Ascenzi RA, et al. Comparative analyses of potato expressed sequence tag libraries. Plant Physiol. 2003;131:419–429. doi:http://dx.doi.org/10.1104/pp.013581.

[21] Rensink W, Hart A, Liu J, Ouyang S, Zismann V, Buell CR. Analyzing the potato abiotic stress transcriptome using expressed sequence tags. Genome. 2005;48:598–605.

[22] Li X, Griffiths R, Lague M, et al. EST sequencing and analysis from cold-stored and reconditioned potato tubers. Acta Hortic. 2007;745:491–493.

[23] Massa AN, Childs KL, Lin H, Bryan GJ, Giuliano G, Robin Buell C. The transcriptome of the reference potato genome *Solanum tuberosum* group Phureja clone DM1-3 516R44. PLoS One. 2011;6(10). doi:10.1371/journal.pone.0026801.

[24] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2015;4(1):1544–6115, doi:10.2202/1544-6115.1128.

[25] Shan J, Song W, Zhou J. Transcriptome analysis reveals novel genes potentially involved in photoperiodic tuberization in potato. Genomics. 2013;102(4):388–396. doi: 10.1016/j.ygeno.2013.07.001.

[26] Hamilton JP, Hansey CN, Whitty BR, et al. Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genomics. 2011;12(1):302. doi: 10.1186/1471-2164-12-302

[27] Sanetomo R, Hosaka K. Pollen transcriptome analysis of *Solanum tuberosum* (2*n*=4*x*=48), *S. demissum* (2n=6x=72), and their reciprocal F1 hybrids. Plant Cell Rep. 2013;32:623–636. doi:10.1007/s00299-013-1395-4.

[28] Liu Y, Lin-Wang K, Deng C, et al. Comparative transcriptome analysis of white and purple potato to identify genes involved in anthocyanin biosynthesis. PLoS One. 2015;10(6):e0129148. doi:10.1371/journal.pone.0129148.

[29] Aulakh SS, Veilleux RE, Dickerman AW, Tang GZ, Flinn BS. Characterization and RNA-seq analysis of underperformer, an activation-tagged potato mutant. Plant Mol Biol. 2014;84(6):635–658. DOI 10.1007/s11103-013-0159-4.

[30] Hamilton JP, Sim SC, Stoffel K, Van Deynze A, Buell CR, Francis DM. Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. Plant Genome. 2012;5(1):17–29. doi:10.3835/plantgenome2011.12.0033.

[31] Wang Y, Tao X, Tang XM, et al. Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. BMC Genomics. 2013.14:81. doi:10.1186/1471-2164-14-841.

[32] Koenig D, Jiménez-Gómez JM, Kimura S, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci USA. 2013;110:E2655–E2662. doi:10.1073/pnas.1309606110.

[33] Zouari I, Salvioli A, Chialva M, et al. From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. BMC Genomics. 2014;15:221. doi:10.1186/1471-2164-15-221.

[34] Gupta S, Shi X, Lindquist IE, Devitt N, Mudge J, Rashotte AM. Transcriptome profiling of cytokinin and auxin regulation in tomato root. J Exp Bot. 2013;64(2):695–704. doi:10.1093/jxb/ers365

[35] Matas AJ, Yeats TH, Buda GJ, et al. Tissue and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. Plant Cell. 2011;23:3893–3910. doi:http://dx.doi.org/10.1105/tpc.111.091173

[36] Mou W, Li D, Luo Z, Mao L, Ying T. Transcriptomic analysis reveals possible influences of ABA on secondary metabolism of pigments, flavonoids and antioxidants in tomato fruit during ripening. PLoS One. 2015;10(6):e0129598. doi:10.1371/journal.pone.0129598.

[37] Bostan H, Chiusano ML. NexGenEx-Tom: a gene expression platform to investigate the functionalities of the tomato genome. BMC Plant Biol. 2015;15:48. doi 10.1186/s12870-014-0412-2.

[38] Reifschneider FJB. Capsicum. Pimentas e pimentões no Brasil. Embrapa Comunicação para Transferência de Tecnologia, Brasilia, 2000;113pp.

[39] Ramchiary N, Kehei M, Brahma V, Kumaria S, Tandon P. Application of genetics and genomics towards Capsicum translational research. Plant Biotechnol Rep. 2014;8:101–123. doi:10.1007/s11816-013-0306-z.

[40] Surh YJ. Cancer chemoprevention with dietary phytochemicals. Nat Rev Cancer. 2003;3:768–780. doi:10.1038/nrc1189.

[41] Ito K, Nakazato T, Yamato K, et al. Induction of apoptosis in leukemic cells by homovanillic acid derivative, capsaicin, through oxidative stress: implication of phosphorylation of p53 at Ser-15 residue by reactive oxygen species. Cancer Res. 2004;64:1071–1078. doi:10.1158/0008-5472.CAN-03-1670.

[42] Mori A, Lehmann S, O'Kelly J, et al. Capsaicin, a component of red peppers, inhibits the growth of androgen-independent, p53 mutant prostate cancer cells. Cancer Res. 2006;66:3222–3229 doi:10.1158/0008-5472.CAN-05-0087.

[43] Lejeune MP, Kovacs EM, Westerterp-Plantenga MS. Effect of capsaicin on substrate oxidation and weight maintenance after modest body-weight loss in human subjects. Br J Nutr. 2003;90:651–659. doi:http://dx.doi.org/10.1079/BJN2003938.

[44] Westerterp-Plantenga MS, Smeets A, Lejeune MP. Sensory and gastrointestinal satiety effects of capsaicin on food intake. Int J Obesity. 2005;29:682–688. doi:10.1038/sj.ijo.0802862.

[45] Ludy MJ, Moore GE, Mattes RD. The effects of capsaicin and capsiate on energy balance. Critical review and meta-analyses of studies in humans. Chem Senses. 2012;37:103–121. doi:10.1093/chemse/bjr100.

[46] Lannes SD, Finger FL, Schuelter AR, Casali VWD. Growth and quality of Brazilian accessions of *Capsicum chinense* fruits. Sci Hortic. 2007;112:266–270. doi:10.1016/j.scienta.2006.12.029.

[47] Góngora-Castillo E, Fajardo-Jaime R, Fernández-Cortes A, et al. The capsicum transcriptome DB: a "hot" tool for genomic research. Bioinformation. 2012;8:043–047.

[48] Lu FH, Cho MC, Park YJ. Transcriptome profiling and molecular marker discovery in red pepper, *Capsicum annuum* L. TF68. Mol Biol Rep. 2012;39:3327–3335. doi:10.1007/s11033-011-1102-x.

[49] Nicolai M, Pisani C, Bouchet JP, Vuylsteke M, Palloix A. Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). Genet Mol Res. 2012;11: 2295–2300. doi:10.4238/(2012).

[50] Ashrafi H, Hill T, Stoffel K, et al. De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genomics. 2012;13:571. doi:10.1186/1471-2164-13-571.

[51] Hill TA, Ashrafi H, Reyes-Chin-Wo S, et al. Characterization of *Capsicum annuum* genetic diversity and population structure based on parallel polymorphism discovery with a 30K unigene pepper GeneChip. PLoS One. 2013;8: e56200. doi:10.1371/journal.pone.0056200.

[52] Liu S, Li W, Wu Y, Chen C, Lei J. *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. PLoS One. 2013;8:e48156. doi:10.1371/journal.pone.0048156.

[53] Ahn YK, Tripathi S, Kim JH, et al. Transcriptome analysis of *Capsicum annuum* varieties Mandarin and Blackcluster: assembly, annotation and molecular marker discovery. Gene. 2013;533:494–499. doi:10.1016/j.gene.2013.09.095.

[54] Daunay MC, Lester RN. The usefulness of taxonomy for Solanaceae breeders, with special reference to the genus *Solanum* and to *Solanum melongena* L. (eggplant). Capsicum Newslett. 1998;7:70–79

[55] Azuma K, Ohyama A, Ippoushi K, et al. Structures and antioxidant activity of anthocyanins in many accessions of eggplant and its related species. J Agric Food Chem. 2008;56:10154–10159. doi:10.1021/jf801322m.

[56] Stommel JR. Whitaker BD. Phenolic acid content and composition of eggplant fruit in a germplasm core subset. J Am Soc Hort Sci. 2003;128:704–710.

[57] Sakata Y, Monma S, Narikawa T, Komochi S. Evaluation of resistance to bacterial wilt and Verticillium wilt in eggplants (*Solanum melongena* L.) collected in Malaysia. J Jpn Soc Hort Sci. 1996;65:81–88. doi:http://doi.org/10.2503/jjshs.65.81.

[58] Fukuoka H, Yamaguchi H, Nunome T, Negoro S, Miyatake K, Ohyama A. Accumulation, functional annotation, and comparative analysis of expressed sequence tags in

eggplant (*Solanum melongena* L.), the third pole of the genus *Solanum* species after tomato and potato. Gene. 2010;450:76–84. doi:10.1016/j.gene.2009.10.006.

[59] Yang X, Cheng YF, Deng C, Ma Y, Wang ZW, Chen XH, Xue LB. Comparative transcriptome analysis of eggplant (*Solanum melongena* L) and turkey berry (*Solanum torvum* Sw): phylogenomics and disease resistance analysis. BMC Genomics. 2014;15:412. doi:10.1186/1471-2164-15-412.

[60] Aflitos S, Schijlen E, Jong H, et al. Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. Plant J. 2014;80(1):136–148. doi:10.1111/tpj.12616.

[61] Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci. 2014;111(14):5135–5140. doi:10.1073/pnas.1400975111.

[62] Hirakawa H, Shirasawa K, Miyatake K, et al. Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *Solanum* species indigenous to the Old World. DNA Res. 2014;21:649–60. doi:10.1093/dnares/dsu027.

[63] Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K. SNP detection for massively parallel whole-genome resequencing. Genome Res. 2009;19(6):1124–1132.

[64] Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18:1851–1858. doi:10.1101/gr.078212.108.

[65] Shen Y, Wan Z, Coarfa C, et al. A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome Res. 2010;20(2):273–280. doi:10.1101/gr.096388.109.

[66] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

[67] McKenna A, Hanna M, Banks E. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–1303. doi:10.1101/gr.107524.110.

[68] DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–498. doi:10.1038/ng.806.

[69] Pabinger S, Dander A, Fischer M, et al. A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform. 2013;15(2):256–78. doi:10.1093/bib/bbs086

[70] Sim S-C, Durstewitz G, Plieske J, et al. Development of a large snp genotyping array and generation of high-density genetic maps in tomato. PLoS One. 2012;7(7):e40563. doi:10.1371/journal.pone.0040563

[71] Hamilton JP, Sim S, Stoffel K, et al. Single nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis. Plant Genome. 2012;5:17–29. doi:10.3835/plantgenome2011.12.0033.

[72] Hirakawa H, Shirasawa K, Ohyama A, et al. Genome-wide SNP genotyping to infer the effects on gene functions in tomato. DNA Res. 2013;20(3):221–233. doi:10.1093/dnares/dst005.

[73] Barchi L, Lanteri S, Portis E, et al. A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation. PLoS One. 2012;7(8):e43740. doi:10.1371/journal.pone.0043740.

[74] Devran Z, Kahveci E, Ozkaynak E, Studholme DJ, Tor M. Development of molecular markers tightly linked to Pvr4 gene in pepper using next-generation sequencing. Mol Breed. 2015;35(4):101. doi:10.1007/s11032-015-0294-5.

[75] Ahmadvand R, Poczai P, Hajianfar R, et al. Next generation sequencing based development of intron-targeting markers in tetraploid potato and their transferability to other Solanum species. Gene. 2014;540(1):117–121. doi:10.1016/j.gene.2014.02.045.

[76] Li W, Cheng J, Wu Z, et al. An InDel-based linkage map of hot pepper (*Capsicum annuum*). Molecular Breeding. 2015;35(1), 1–10. doi:10.1007/s11032-015-0219-3.

[77] Shirasawa K, Fukuoka H, Matsunaga H, et al. Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. DNA Res. 2013;20(6), 593–603. doi:10.1093/dnares/dst033.

[78] Kim JE, Oh SK, Lee JH, Lee BM, Jo SH. Genome-wide SNP calling using next generation sequencing data in tomato. Mol Cells. 2014;37(1),36. doi:10.14348/molcells.2014.2241.

[79] Kobayashi M, Nagasaki H, Garcia V, et al. Genome-wide analysis of intraspecific DNA polymorphism in 'Micro-Tom', a model cultivar of tomato (*Solanum lycopersicum*). Plant Cell Physiol. 2014;55(2):445–454. doi:10.1093/pcp/pct181.

[80] Víquez-Zamora M, Vosman B, Van de Geest H, et al. Tomato breeding in the genomics era: insights from a SNP array. BMC Genomics. 2013;14:354. doi:10.1186/1471-2164-14-354.

[81] Barchi L, Lanteri S, Portis E, et al. Identification of SNP and SSR markers in eggplant using RAD tag sequencing. BMC Genomics. 2011;12:304. doi:10.1186/1471-2164-12-304.

[82] Uitdewilligen JG, Wolters AMA, Bjorn B, et al. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One. 2013;8(5), e62355. doi:10.1371/journal.pone.0062355.

[83] Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11:204–20. doi:10.1038/nrg2719.

[84] Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153:193–205. doi:10.1016/j.cell.2013.02.033.

[85] Zhong S, Fei Z, Chen YR, et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat Biotechnol 2013;31:154–159. doi:10.1038/nbt.2462.

[86] Thompson A, Tor M, Barry C, et al. Molecular and genetic characterization of a novel pleiotropic tomato-ripening mutant. Plant Physiol. 1999;120:383–90.

[87] Manning K, Tor M, Poole M, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nat Genet 2006;38:948–52. doi:10.1038/ng1841.

[88] Chen W, Kong J, Qin C, et al. Requirement of CHROMOMETHYLASE3 for somatic inheritance of the spontaneous tomato epimutation colourless non-ripening. Sci Rep. 2015b;5:9192. doi:10.1038/srep09192.

[89] Solomon MJ, Larsen PL, Varshavsky A. Mapping protein–DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell. 1988;53:937–47. doi:10.1016/S0092-8674(88)90469-2.

[90] Kaufmann K, Muino JM, Jauregui R, et al. Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. PLoS Biol. 2009;7:e1000090. doi:10.1371/journal.pbio.1000090.

[91] Pandey R, Muller A, Napoli CA, et al. Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res. 2002;30:5036–5055.

[92] Chen ZJ, Tian L. Roles of dynamic and reversible histone acetylation in plant development and polyploidy. Biochim Biophys Acta. 2007;1769:295–307. doi:10.1016/j.bbaexp.2007.04.007.

[93] Ricardi MM, González RM, Zhong S, et al. Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. BMC Plant Biol. 2014;14:29. doi:10.1186/1471-2229-14-29.

[94] Fujisawa M, Nakano T, Ito Y. Identification of potential target genes for the tomato fruit-ripening regulator RIN by chromatin immunoprecipitation. BMC Plant Biol. 2011;11:26. doi:10.1186/1471-2229-11-26.

[95]   Fujisawa M, Shima Y, Nakagawa H, et al. Transcriptional regulation of fruit ripening by tomato FRUITFULL homologs and associated MADS box proteins. Plant Cell. 2014;26(1):89–101. doi:10.1105/tpc.113.119453.

[96]   Kim ED, Sung S. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends Plant Sci. 2012;17:16–21. doi:10.1016/j.tplants.2011.10.008.

[97]   Zhu QH, Wang MB. Molecular functions of long non-coding RNAs in plants. Genes (Basel). 2012;3:176–190. doi:10.3390/genes3010176.

[98]   Liu J, Jung C, Xu J, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. Plant Cell. 2012;24(11):4333–4345. doi:10.1105/tpc. 112.102855.

[99]   Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. PLoS One. 2012;7(8):e43047. doi:10.1371/journal.pone.0043047.

[100]  Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Muehlbauer GJ. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol. 2014;15(2):R40. doi:10.1186/gb-2014-15-2-r40.

[101]  Zhu B, Yang Y, Li R, et al. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. J Exp Bot. 2015;66(15): 4483–4495. DOI:10.1093/jxb/erv203.

[102]  Song J, Angel A, Howard M, Dean C. Vernalization—a cold-induced epigenetic switch. J Cell Science. 2012;125(16):3723–3731. doi:10.1242/jcs.084764.

[103]  Jin J, Liu J, Wang H, Wong L, Chua NH. PLncDB: plant long noncoding RNA database. Bioinformatics. 2013;29(8):1068–1071. doi:10.1093/bioinformatics/btt107.

[104]  Yi X, Zhang Z, Ling Y, Xu W, Su Z. PNRD: a plant non-coding RNA database. Nucleic Acids Res. 2015;43(D1):D982–D989. doi:10.1093/nar/gku1162.

[105]  Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNA and their regulatory roles in plants. Annu Rev Plant Biol. 2006;57:19–53. doi:10.1146/annurev.arplant. 57.032905.105218.

[106]  Zhang B, Pan X, Cobb GP, Anderson TA. Plant microRNA: a small regulatory molecule with big impact. Dev Biol. 2006;289:3–16. doi:10.1016/j.ydbio.2005.10.036.

[107]  Voinnet O. Origin, biogenesis, and activity of plant microRNAs. Cell. 2009;136:669–687. doi:10.1016/j.cell.2009.01.046.

[108]  Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, et al. Widespread translational inhibition by plant miRNAs and siRNAs. Science. 2008;320:1185–90. doi: 10.1126/science.1159151.

[109]  Lanet E, Delannoy E, Sormani R, et al. Biochemical evidence for translational repression by *Arabidopsis* microRNAs. Plant Cell. 2009;21:1762–8. doi:10.1105/tpc. 108.063412.

[110]  Itaya A, Bundschuh R, Archual AJ, et al. Small RNAs in tomato fruit and leaf development. Biochim Biophys Acta. 2008;1779:99–107. doi:10.1016/j.bbagrm.2007.09.003.

[111]  Moxon S, Jing R, Szittya G, et al. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. 2008;18:1602–9. doi:10.1101/gr.080127.108.

[112]  Zhang J, Zeng R, Chen J, Liu X, Liao Q. Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. Gene. 2008;423:1–7. doi:10.1016/j.gene. 2008.05.023.

[113]  Knapp S. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J Exp Bot. 2002;53:2001–2022. doi:10.1093/jxb/erf068.

[114]  Kim HJ, Baek KH, Lee BW, Choi D, Hur CG. In silico identification and characterization of microRNAs and their putative target genes in Solanaceae plants. Genome. 2011;54:91–98. doi:10.1139/G10-104.

[115]  Hwang D-G, Park JH, Lim JY, et al. The hot pepper (*Capsicum annuum*) microRNA transcriptome reveals novel and conserved targets: a foundation for understanding MicroRNA functional roles in hot pepper. PLoS One. 2013;8:e64238. doi:10.1371/journal.pone.0064238.

[116]  Gao J, Yin F, Liu M, et al. Identification and characterization of tobacco microRNA transcriptome using high-throughput sequencing. Plant Biol. 2014;17(3):591–598. doi: 10.1111/plb.12275.

[117]  Lakhotia N, Joshi G, Bhardwaj AR, et al. Identification and characterization of miR-NAome in root, stem, leaf and tuber developmental stages of potato (*Solanum tuberosum* L.) by high-throughput sequencing. BMC Plant Biol. 2014;14:6. doi: 10.1186/1471-2229-14-6.

[118]  Griffiths-Jones S. The microRNA registry. Nucleic Acids Res. 2004;32:D109–11. doi: 10.1093/nar/gkh023.

[119]  Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34:D140–4. doi:10.1093/nar/gkj112.

[120]  Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008;36:D154–8. doi:10.1093/nar/gkm952.

[121]  Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152–7. doi:10.1093/nar/gkq1027.

[122] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42:D68–73. doi:10.1093/nar/gkt1181.

[123] Yang JH, Shao P, Zhou H, Chen YQ, Qu LH. DeepBase: a database for deeply annotating and mining deep sequencing data. Nucleic Acids Res. 2009;38:D123–D130. doi: 10.1093/nar/gkp943.

[124] Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res. 2008;36:D149–53.doi:10.1093/nar/gkm995.

[125] Alexiou P, Vergoulis T, Gleditzsch M, et al. miRGen 2.0: a database of microRNA genomic information and regulation. Nucleic Acids Res. 2010;38:D137–41. doi: 10.1093/nar/gkp888.

[126] Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, Hsu PWC, Wong YH, Chen YH, Chen GH, Huang HD. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Res. 2008;36:D165–9. doi:10.1093/nar/gkm1012.

[127] Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update (2014): an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2014;42:D78–85. doi:10.1093/nar/gkt1266.

[128] Zhang Z, Yu J, Li D, et al. PMRD: plant microRNA database. Nucleic Acids Res. 2010;38:D806–13. DOI:10.1093/nar/gkp818.

[129] Liu N, Wu S, Van Houten J, et al. Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. J Exp Bot. 2014;65:2507–2520. doi:10.1093/jxb/eru141.

[130] Ferreira e Silva GF, Silva EM, Azevedo M da S, et al. microRNA156-targeted SPL/SBP box transcription factors regulate tomato ovary and fruit development. Plant J. 2014;78:604–18. doi:10.1111/tpj.12493.

[131] Chen W, Kong J, Lai T, et al. Tuning LeSPL-CNR expression by SlymiR157 affects tomato fruit ripening. Sci Rep. 2015;5:7852. doi:10.1038/srep07852.

[132] Yang JH, Li JH, Shao P, Zhou H, Chen YQ, Qu LH. starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. Nucleic Acids Res. 2011;39:D202–9. doi:10.1093/nar/gkq1056.

[133] Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. 2014;42:D92–7. doi:10.1093/nar/gkt1248.

[134] Dweep H, Sticht C, Pandey P, Gretz N. miRWalk-database: prediction of possible miRNA binding sites by 'walking' the genes of three genomes. J Biomedical Informatics. 2011;44:839–847. doi:10.1016/j.jbi.2011.05.002.

[135] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005;120:15–20. doi:10.1016/j.cell.2004.12.035.

[136] Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell. 2007;27:91–105. doi:10.1016/j.molcel.2007.06.017.

[137] Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19:92–105. doi:10.1101/gr. 082701.108.

[138] Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA. 2006;12:192–7. doi:10.1261/rna.2239606.

[139] Vlachos IS, Paraskevopoulou MD, Karagkouni D, et al. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids Res. 2015;43:D153–9. doi:10.1093/nar/gku1215.

[140] Maragkakis M, Alexiou P, Papadopoulos GL, et al. Accurate microRNA target prediction correlates with protein repression levels. BMC Bioinformatics. 2009;10:295. doi:10.1186/1471-2105-10-295.

[141] Maragkakis M, Reczko M, Simossis VA, et al. DIANA-microT web server: elucidating microRNA functions through target prediction. Nucleic Acids Res. 2009;37:W273–6. doi:10.1093/nar/gkp292.

[142] Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 2009;37:D105–10. doi:10.1093/nar/gkn851.

[143] Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. Nat Genet. 2005;37:495–500. doi:10.1038/ng1536.

[144] Loher P, Rigoutsos I. Interactive exploration of RNA22 microRNA target predictions. Bioinformatics. 2012;28:3322–3. doi:10.1093/bioinformatics/bts615.

[145] Chang TH, Horng JT, Huang HD. RNALogo: a new approach to display structural RNA alignment. Nucleic Acid Res. 2008;36:W91–6. doi:10.1093/nar/gkn258.

[146] Nam S, Kim B, Shin S, Lee S. miRGator: an integrated system for functional annotation of microRNAs. Nucleic Acids Res. 2008;36:D159–64. doi:10.1093/nar/gkm829.

[147] Cho S, Jun Y, Lee S, et al. miRGator v2.0: an integrated system for functional investigation of microRNAs. Nucleic Acids Res. 2011;39:D158–62. doi:10.1093/nar/gkq1094.

[148] Cho S, Jang I, Jun Y, et al. MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. Nucleic Acids Res. 2013;41:D252–7. doi:10.1093/nar/gks1168.

[149] Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015;43:D146–52. doi:10.1093/nar/gku1104.

[150] Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. RNA. 2004;10:1507–17. doi:10.1261/rna.5248604.

[151] Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 2011;39:W155–9. doi:10.1093/nar/gkr319.

[152] Szczesniak MW, Makalowska I. miRNEST 2.0: a database of plant and animal microRNAs. Nucleic Acid Res. 2014;42:D74–7. doi:10.1093/nar/gkt1156.

[153] Sun X, Dong B, Yin L, et al. PMTED: a plant microRNA target expression database. BMC Bioinformatics. 2013;14:174. doi:10.1186/1471-2105-14-174.

[154] Bielewicz D, Dolata J, Zielezinski A, et al. mirEX: a platform for comparative exploration of plant pri-miRNA expression data. Nucleic Acid Res. 2012;40:D191–7. doi: 10.1093/nar/gkr878.

[155] Bonnet E, He Y, Billiau K, Van de Peer Y. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. Bioinformatics. 2010;26:1566–1568. doi:10.1093/bioinformatics/btq233.

[156] Zhang S, Yue Y, Sheng L, et al. PASmiR: a literature-curated database for miRNA molecular regulation in plant response to abiotic stress. BMC Plant Biol. 2013;13:33. doi:10.1186/1471-2229-13-33.

[157] Luan Y, Cui J, Zhai J, Li J, Han L, Meng J. High-throughput sequencing reveals differential expression of miRNAs in tomato inoculated with *Phytophthora infestans*. Planta. 2015;1:1405–1416. doi:10.1007/s00425-015-2267-7.

[158] Ouyang S, Park G, Atamian HS, et al. MicroRNAs suppress NB domain genes in tomato that confer resistance to *Fusarium oxysporum*. PLoS Pathogens. 2014;10:e1004464. doi:10.1371/journal.ppat.1004464.

[159] Zhang X, Zou Z, Gong P, et al. Over-expression of microRNA169 confers enhanced drought tolerance to tomato. Biotechnology Letters. 2011;33:403–409. doi:10.1007/s10529-010-0436-0.

[160] Zhuang Y, Zhou XH, Liu J. Conserved miRNAs and their response to salt stress in wild eggplant *Solanum linnaeanum* roots. Int J Mol Sci. 2014;15:839–849. doi:10.3390/ijms15010839.

[161] Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. Cold Spring Harb. Symp Quant Biol. 2003;68:69–78

[162] Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques. 2002;32:S56

[163] Bell PA, Chaturvedi S, Gelfand CA, et al. SNPstream UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques. 2002;32:S70–77

[164] Matsuzaki H, Loi H, Dong S, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res. 2004;14:414–25. doi: 10.1101/gr.2014904

[165] Nijman IJ, Kuipers S, Verheul M, Guryev V, Cuppen E. A genome-wide SNP panel for mapping and association studies in the rat. BMC Genomics. 2008;9:95. doi: 10.1186/1471-2164-9-95

[166] Felcher KJ, Coombs JJ, Massa AN, et al. Integration of two diploid potato linkage maps with the potato genome sequence. PLoS One. 2012;7:e36347. doi:10.1371/journal.pone.0036347

[167] Gupta PK, Rustgi S, Mir RR. Array-based high-throughput DNA markers for crop improvement. Heredity. 2008;101:5–18. doi:10.1038/hdy.2008.35.

[168] Kennedy GC, Matsuzaki H, Dong S, et al. Large-scale genotyping of complex DNA. Nature Biotechnol. 2003;21:1233–1237. doi:10.1038/nbt869.

[169] Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM. Oligonucleotide array discovery of polymorphisms in cultivated tomato (*Solanum lycopersicum* L.) reveals patterns of SNP variation associated with breeding. BMC Genomics. 2009;10:10. doi: 10.1186/1471-2164-10-466.

[170] Elshire RJ, Glaubitz JC, Sun Q, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6:e19379. doi:10.1371/journal.pone.0019379.

[171] Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One. 2012;7:e32253. doi:10.1371/journal.pone.0032253.

[172] Mascher M, Wu S, Amand PS, Stein N, Poland J. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. PLoS One. 2013;8:e76925. doi:10.1371/journal.pone.0076925.

[173] Uitdewilligen JG, Wolters AM, D'hoop BB, et al. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One. 2013;8:e62355. doi:10.1371/journal.pone.0062355.

[174] Rafalski JA. Association genetics in crop improvement. Curr Opin Plant Biol. 2010;13:174–180. doi:10.1016/j.pbi.2009.12.004.

[175] Ranc N, Munos S, Xu J, et al. Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycope*rsicum var. cerasiforme. G3 (Bethesda). 2012;2:853–864. doi:10.1534/g3.112.002667.