# Status and Prospects of Next-generation Sequencing Technologies in Crop Plants

1

Tilak R. Sharma*, Basavantraya N. Devanna, Kanti Kiran, Pankaj K. Singh, Kirti Arora, Priyanka Jain, Ila M. Tiwari, Himanshu Dubey, Banita K. Saklani, Mandeep Kumari, Jyoti Singh, Rajdeep Jaswal, Ritu Kapoor, Deepak V. Pawar, Shruti Sinha, Deepak S. Bisht, Amolkumar U. Solanke and Tapan K. Mondal

ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi, India.

*Correspondence: trsharma1965@gmail.com and trsharma@nrcpb.org

https://doi.org/10.21775/9781910190654.02

## Abstract

The history of DNA sequencing dates back to 1970s. During this period, the two first-generation nucleotide sequencing techniques were developed. Subsequently, Sanger's dideoxy method of sequencing gained popularity over Maxam and Gilbert's chemical method of sequencing. However, in the last decade, we have observed revolutionary changes in DNA sequencing technologies leading to the emergence of next-generation sequencing (NGS) techniques. NGS technologies have enhanced the throughput and speed of sequencing combined with bringing down the overall cost of the process over a time. The major applications of NGS technologies being genome sequencing and resequencing, transcriptomics, metagenomics in relation to plant–microbe interactions, exon and genome capturing, development of molecular markers and evolutionary studies. In this review, we present a broader picture of evolution of NGS tools, its various applications in crop plants, and future prospects of the technology for crop improvement.

## Introduction

The overall growth, development and behavioural characteristics of every living creature are largely determined by its genetic constitution. Subsequent to the famous double-helix model of DNA, proposed by Watson and Crick (1953), scientists began to find the ways and means to determine the nucleotide sequence of DNA. The first significant breakthrough in this area was achieved in late 1970s when two groups working independently reported two different approaches for DNA sequencing (Maxam and Gilbert, 1977; Sanger *et al.*, 1977). Though Maxam and Gilbert's approach for DNA sequencing was preferred initially, it was Sanger's sequencing technology which subsequently got popularized among the scientific community. The classical genome sequencing projects such as the Human Genome Project (HGP), the *Arabidopsis* Genome Initiative and the International Rice Genome Sequencing Project were successfully completed using Sanger's sequencing approach. Subsequently, many plant genomes were sequenced using this sequencing technology. Though Sanger's dideoxy sequencing method is considered as gold standard with respect to genome sequencing, there

are many shortcomings in this approach. The important shortcomings of Sanger's sequencing method are that it is time-consuming, low throughput and high cost, and needs more labour, *in vivo* cloning of DNA fragments, etc. Therefore, scientists and bioengineers tried to develop new sequencing techniques also known as second-generation (2GS) or next-generation sequencing (NGS) technologies. The success in this direction was reported in 2005, when the first NGS system was developed by 454 and commercialized by Roche as the GS20 (Margulies *et al.*, 2005). Subsequently, many NGS/2GS systems have been reported, which include Solexa GA2 (now Illumina), Applied Biosystem's SOLiD and Ion Torrent www.appliedbiosystems.com; www.illumina.com; www.iontorrent.com).

The developments in the field of NGS technologies have led to a revolution in the field of genetics and genomics. In spite of their inherent limitation of producing shorter reads with higher error rates than Sanger sequencing, NGS technologies are still gaining popularity, largely due to their ability to produce massive quantities of data at a relatively low cost and in a short time period. Further developments in NGS technologies have consistently led to increased read lengths and this is an active area of research in the future of NGS technologies. The present review compiles the different NGS technologies and their applications in the field of plant biology with special emphasis on agricultural plants (Fig. 1.1). In this review we focused primarily on the crop genomes giving weightage to genome sequencing techniques implied and their challenges, applications of NGS in plant biology with emphasis on agricultural crops and conclude with the future strategies and perspective in the area of plant genomics.

## Evolutions of DNA sequencing technology

### First generation sequencing technologies

The double helix structure of DNA proposed by Watson and Crick (1953) was the milestone

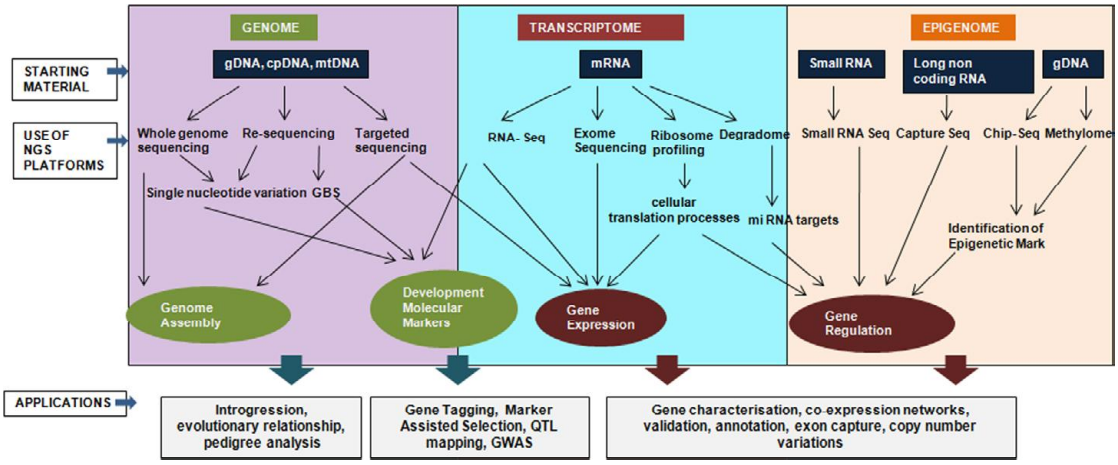


**Figure 1.1** NGS platforms as a tool for the plant genomic research. The phenotype of an organism is a manifestation of controlled expression of the underlying gene. The gene(s) expression is largely regulated at three stages: (a) genomic, (b) transcriptomic and (c) epigenetic. For unravelling the complex cellular machinery and its regulatory network the NGS platforms, with suitable modifications, can be utilized at any of these three stages (each stage is represented by different colour boxes). Genomic DNA, chloroplast DNA and mitochondrial DNA are the primary sources of genetic information which can be divulged by NGS assisted whole genome sequencing, high throughput re-sequencing and targeted sequencing. For example, high sequencing coverage obtained by NGS platforms is of great utility in determining the single nucleotide variations (SNVs) in genome which are of immense importance in establishing genotype phenotype relationships. In addition, NGS tools like MethylC-Seq are a 'gold standard' technique for studying genome wide methylation pattern. Similarly, sequencing of mRNA, small RNAs, long non coding RNAs, degradome and exome helps in understating the expression pattern of genes, their spatial and temporal regulations, co-expression networks and association with trait.

in molecular biology research. Subsequently, interest was shifted to decipher the nucleotide sequence of DNA molecules. The first major success in this field was reported simultaneously in 1977 by two independent groups of researchers when two chemistries for DNA sequencing were published (Maxam and Gilbert, 1977; Sanger *et al.*, 1977). Broadly referred to as first-generation sequencing technologies, these two methods utilized different chemistries for DNA sequencing. Due to its complex procedure and low resolution Maxam and Gilbert's method did not gain wide acceptance. Sanger's original dideoxynucleotide (ddNTPs) chain-termination sequencing method was comparatively less cumbersome and relatively accurate but required radiolabelled ddNTPs, and the chemistry involved four separate base specific reactions and autoradiography in order to detect the sequence of DNA molecule. Sanger's method was later on modified and improved. An account of improvement and evolution of Sanger's method can be found in previous reviews (França *et al.*, 2002; Ansorge, 2009; Mardis, 2013).

## Second-generation sequencing technologies

Second-generation sequencing technologies were capitalized by Illumina Genome Analyzer, Roche/454 FLX, Life technologies SOLiD and Ion Torrent PGM. Second-generation sequencers omitted the need of *in vivo* cloning, and here DNA can directly be fragmented to produce sequencing libraries of appropriate sizes *in vitro* by adapter ligated amplification using a PCR-based system. PCR-based amplification is required to produce millions of copies of the original DNA fragment, required to produce signal intensity sufficient to detect the incorporated bases during sequencing steps. Second-generation sequencing technologies are divided in two categories on the basis of reaction chemistry they use: (a) sequencing by synthesis used by Illumina, Roche/454 and Ion Torrent and (b) sequencing by ligation used by ABI's SOLiD sequencers.

## Third-generation sequencing technology: single molecule sequencing

Third-generation sequencing (3GS) technology differs from second-generation sequencing (2GS) technologies as it does not require the amplification step, leading to elimination of inclusion errors incurred by polymerase during library preparation. Second-generation sequencers use a cyclic wash-and-scan method, and as the number of washing and imaging cycles increases, addition of nucleotides becomes more asynchronous which increases sequencing errors and signal to noise ratio. This particularly limits the read length produced by second-generation sequencers. The advantage of some single molecule sequencing platforms is the ability to detect epigenetic modifications in the genome, which are largely diluted due to amplification step involved in 2GS techniques (Munera *et al.*, 2012). Also, the short reads generated by the 2GS technologies imposes problems in accurate *de novo* genome assembly, repetitive elements and large structural variation analyses (Fan *et al.*, 2010; Delcher *et al.*, 2010). The longer reads generated by third-generation sequencer helps in *de novo* genome sequencing and assembly, improving old assemblies, allowing more accurate analysis of structural variation within the genome and more contiguous reconstruction of genomes which were limited in second-generation sequencers (Carneiro *et al.*, 2013). At present, the major challenge with the single molecule sequencers is inherent higher error rate due to limited ability of detectors to identify and interpret very low levels of signal generated from individual molecules. The 3GS techniques are Pacific Biosciences Single Molecule Real Time sequencers (SMRT) and Helicos Genetic analysis system True single molecule sequencing (tSMS). Pacific Biosciences SMRT was the first 3GS technology to hit the market in 2010. With an improved chemistry (C4), the average read length of PacBio RS II system is over 10 kb and can generate 1 Gb of data per run. In sequel platform PacBio has increased the ZMWs density of loading of samples due to which the data output has increased drastically up to 5–10 Gb per run. Since this technology uses single DNA polymerase per ZMW, total length of a read is dependent on lifespan of the polymerase. Single molecule sequencers suffers from higher raw data error rate, which accounts for up to 15% of incorporated bases, however, this error rate can be minimized by multiple sequencing of same template. The tSMS from Helicos Genetic analysis system was the first commercially available true single molecule sequencing system based on

sequencing by synthesis chemistry. This platform can produce average read length of 30 bases. This platform suffers from higher deletion rate ranging from 1 to 5%, which can be reduced by sequencing from both the ends of a fragment (Buzby *et al.*, 2008).

### Fourth-generation sequencing technologies: nanopore sequencing

Fourth-generation sequencing (4GS) technologies omit the use of labelled nucleotides and does not rely on an optical system to detect incorporated nucleotides and also there is no need for the synchronous reagent addition and wash steps. This technology uses the electronic or chemical properties of each nucleotide to determine the sequence of DNA as it is threaded through a nanopore. Current nanopore sequencing technologies use the ionic current blockage method. Currently two nanopore-based systems, biological and solid-state, are being developed and refined. Oxford Nanopore technology has released MinION for test users which will be launched for commercial application. It is small, USB powered, easy to carry equipment. It can produce average read length of 5.4 kb and go up to 10 kb (www.nature.com/news/data-from-pocket-sized-genome-sequencer-unveiled-1.14724).

## NGS and plant genomics

Plants being the ultimate source of food and metabolic energy for nearly all animals, cannot manufacture their own food, but also provide sustenance, shelter, clothing, medicines, fuels, and the raw materials from which innumerable other products are made. The plant kingdom is filled with amazingly incomparable diversity and significance (Schatz *et al.*, 2010) pertaining to which sequencing of plant genomes is of great importance. Out of the hundreds of thousands of plant species around the world, only few of them have been sequenced (Michael and Van Buren, 2015; Table 1.1).

### Genomics of plants

Advanced high-throughput genome sequencing techniques have proved to be a boon in providing practical solutions to the challenges in the field of genomics, especially crops. The plant kingdom is majorly divided into spore bearing vascular plants and seed bearing vascular plants. The former is further classified into algae, mosses/liverworts and ferns, and the latter into non flowering (gymnosperms) and flowering (all angiosperms) plants. Since the first published genome sequence of *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000), a large number of plant genomes belonging to different phylum have been sequenced and published using both first-generation and NGS technologies (Turktas *et al.*, 2015).

### Sequenced plant genomes: links from vascular plants to angiosperms

The genome sequences of *Chlamydomonas reinhardtii* (120 Mb) and *Volvox carteri* (138 Mb) are amongst the first multi- and unicellular algae species, respectively (Merchant *et al.*, 2007; Prochnik *et al.*, 2010). Amongst the mosses, the genome of *Physcomitrella* with an assembled genome size of 500 Mb (Rensing *et al.*, 2008) and *Selaginella moellendorffii*, the first non-seed plant genome reported with an assembled genome size of 212.6 Mb (Banks *et al.*, 2011), are breakthrough model organisms in order to study the evolution of vascular land plants and their divergence. Further for a better understanding of the evolution of plants, the genome sequence information from other taxa, especially charophytes, ferns and gymnosperms, could serve as key references. In this context the three gymnosperm (conifer) genome sequences published recently, viz. the genome of Norway spruce (*Picea abies*) with an assembled size of 19.6 Gb (Nystedt *et al.*, 2013) is the first available gymnosperm sequence followed by the loblolly pine (*Pinus taeda*) genome (Zimin *et al.*, 2014; Neale *et al.*, 2014) and white spruce (*Picea glauca*) genome (Birol *et al.*, 2013) with an assembled genome size of 22 Gb and 20.8 Gb, respectively. These genomes helped to understand the divergence of angiosperms and gymnosperms (350 Myr ago) (Jiao *et al.*, 2011).

The development of water-conducting xylem cells and the reproductive development are the two major differences between angiosperms and gymnosperms. Nystedt *et al.* (2013) compared seven genomes, two from the basal plants and five from angiosperms and identified 1021 *P. abies*-specific gene families. *P. abies* homologs present in the angiosperms revealed that gymnosperms lack orthologues of flowering-responsible phosphatidylethanolamine-binding protein (PEBP) protein. Sequencing of angiosperm

**Table 1.1** Details of plant genomes sequenced and published during 2015 to 2016

| Number | Scientific name | Common name | Genome size (Mb) | Type | Reference |
|---|---|---|---|---|---|
| 1 | *Vigna angularis* | Adzuki bean | 538 | Dicot | Kang *et al*., 2015 |
| 2 | *Thlaspi arvense* | Field pennycress | 539 | Dicot | Dorn *et al*., 2015 |
| 3 | *Primula veris* | Cowslip | 480 | Dicot | Nowak *et al*., 2015 |
| 4 | *Hordeum vulgare* | Tibetan hulless barley | 4480 | Dicot | Zeng *et al*., 2015 |
| 5 | *Vaccinium corymbosum* | American blue berry | 500 | Dicot | Gupta *et al*., 2015 |
| 6 | *Arabis alpina* | Alpine rockcress | 375 | Dicot | Willing *et al*., 2015 |
| 7 | *Ipomea trifida* | Wild sweet potato | 520 | Dicot | Hirakawa *et al*., 2015 |
| 8 | *Gossypium hirsutum* | Upland cotton | 2340 | Dicot | Li *et al*., 2015 |
| 9 | *Boea hygrometrica* | | 1690 | Dicot | Xiao *et al*., 2015 |
| 10 | *Solanum commersonii* | Wild potato | 830 | Dicot | Aversano *et al*., 2015 |
| 11 | *Catharanthus roseus* | Madagascar periwinkle | 738 | Dicot | Kellner *et al*., 2015 |
| 12 | *Ocimum sanctum* | Holy basil | 386 | Dicot | Rastogi *et al*., 2015 |
| 13 | *Moringa oleifera* | Drumstick tree | 315 | Dicot | Tian *et al*., 2015 |
| 14 | *Zizania latifolia* | Jiaobei | 590 | Dicot | Guo *et al*., 2015 |
| 15 | *Cymbomonas tramitiformis* | | 850 | Algae | Burns *et al*., 2015 |
| 16 | *Gossypium barbadense* | Sea island cotton | 2470 | Dicot | Liu *et al*., 2015 |
| 17 | *Lolium perenne* | Perennial ryegrass | 2000 | Monocot | Byrne *et al*., 2015 |
| 18 | *Chlorella pyrenoidosa* | | 57 | Monocot | Fan *et al*., 2015 |
| 19 | *Anana comosus* | Pineapple | 526 | Monocot | Ming *et al*., 2015 |
| 20 | *Oropetium thomaeum* | | 245 | Monocot | Van Buren *et al*., 2015 |
| 21 | *Lemna minor* | Common duckweed | 481 | Monocot | Van Hoeck *et al*., 2015 |
| 22 | *Trifolium pratense* | Red clover | 420 | Dicot | De Vega *et al*., 2015 |
| 23 | *Salvia miltiorrhiza* | Chinese red sage | 641 | Dicot | Zhang *et al*., 2015 |
| 24 | *Parachlorella kessleri* | | 63 | Algae | Ota *et al*., 2016 |
| 25 | *Zostera marina* | Common eelgrass | 238 | Monocot | Olsen *et al*., 2016 |
| 26 | *Dendrobium catenatum* | Chained dendrobium | 1110 | Monocot | Zhang *et al*., 2016 |
| 27 | *Arachis duranensis* | Wild peanut A | 1250 | Dicot | Bertioli *et al*., 2016 |
| 28 | *Arachis ipaensis* | Wild peanut A | 1560 | Dicot | Bertioli *et al*., 2016 |
| 29 | *Cynara cardunculus* | Globe artichoke | 1084 | Dicot | Scaglione *et al*., 2016 |
| 30 | *Rosa roxburghii* | Chestnut rose | 481 | Dicot | Lu *et al*., 2016 |
| 31 | *Zoysia japonica* | | 390 | Monocot | Tanaka *et al*., 2016 |
| 32 | *Zoysia matrella* | Manila grass | 380 | Monocot | Tanaka *et al*., 2016 |
| 33 | *Zoysia pacifica* | | 370 | Monocot | Tanaka *et al*., 2016 |
| 34 | *Fagopyrum esculentum* | Common buckwheat | 1300 | Dicot | Yasvi *et al*., 2016 |
| 35 | *Gonium pectorale* | | 149 | Algae | Hanschen *et al*., 2016 |
| 36 | *Rubus accidentalis* | Black raspberry | 293 | Dicot | Van Buren *et al*., 2016 |
| 37 | *Petunia oxillaris* | White moon petunia | 1400 | Dicot | Bombarley *et al*., 2016 |
| 38 | *Pogostemon cablin* | Patchouli | 1570 | Dicot | He *et al*., 2016 |
| 39 | *Lepidium meyenii* | Maca | 751 | Dicot | Zhang *et al*., 2016 |
| 40 | *Daucus carota* | Carrot | 473 | Dicot | Iorizzo *et al*., 2016 |
| 41 | *Juglans regia* | Common walnut | 606 | Dicot | Martinez-Garcia *et al*., 2016 |
| 42 | *Drosera capensis* | Cape sundew | 293 | Dicot | Butts *et al*., 2016 |

**Table 1.1** Continued

| Number | Scientific name | Common name | Genome size (Mb) | Type | Reference |
|---|---|---|---|---|---|
| 43 | *Zostera muelleri* | Dwarf grass wrack | 390 | Dicot | Lee *et al*., 2016 |
| 44 | *Chenopodium quinoa* | Quinoa | 1500 | Dicot | Yasvi *et al*., 2016 |
| 45 | *Artocarpus camansi* | Breadnut | 669 | Dicot | Gardner *et al*., 2016 |
| 46 | *Citrus paradisi* × *Poncirus trifoliata* | Swingle citrumela | 380 | Dicot | Zhang *et al*., 2016 |
| 47 | *Musa itinerans* | Yunnan banana | 615 | Monocot | Wu *et al*., 2016 |
| 48 | *Cicer reticulatum* | Wild chickpea | 817 | Dicot | Gupta *et al*., 2016 |
| 49 | *Trifolium subterraneum* | Subterranean clover | 540 | Dicot | Hirakawa *et al*., 2016 |
| 50 | *Quercus lobata* | Valley oak | 725 | Dicot | Sork *et al*., 2016 |
| 51 | *Brassica nigra* | Black mustard | 591 | Dicot | Yang *et al*., 2016 |
| 52 | *Brassica juncea* | Chinese mustard | 922 | Dicot | Yang *et al*., 2016 |
| 53 | *Rhazya stricta* | Harmal | 274 | Dicot | Sabir *et al*., 2016 |

genomes thus additionally served and immensely helped in understanding their divergence from rest of the plant genomes. With over 100 plant genome sequence data already published (Michael and Van Buren, 2015; https://genomevolution.org/wiki/index.php/Sequenced_plant_genomes), it is possible to study their complex life cycle, evolutionary history and genome structural organization. The crops and model plant genomes such as *Arabidopsis, Brachypodium distachyon, Physcomitrella patens* and *Setaria italica, Oryza sativa, Populus trichocarpa, Zea mays, Glycine max, Solanum lycopersicum* and *Pinus taeda* are extremely vital not just as crops but as functional models to enable genome-wide studies of various key genes/gene families, pathways and important traits (Arabidopsis Genome Initiative, 2000; Vogel *et al*., 2010; Rensing *et al*., 2008; Bennetzen *et al*., 2012; Zhang *et al*., 2012a; Goff *et al*., 2002; Yu *et al*., 2002; Tuskan *et al*., 2006; Schnable *et al*., 2009; Schmutz *et al*., 2010; Tomato Genome Consortium, 2012; Zimin *et al*., 2014). Nevertheless, non-functional models and non-crop plants serve as sources to explore the evolution of flowering plants and an in-depth understanding of plant genome architecture. The genome sequences of non-functional model plants like *Utricularia gibba* (bladderwort; 77 Mb), *Genlisea aurea* (corkscrew; 43.4 Mb), an aquatic non-grass monocot *Spirodela polyrhiza* (greater duckweed; 158 Mb), *Selaginella moellendorffii* (212.6 Mb), and *Amborella trichopoda* (870 Mb) are available (Banks *et al*., 2011; Albert *et al*., 2013; Ibarra-Laclette *et al*., 2013; Leushkin *et al*., 2013; Wang *et al*., 2014). These can help to identify the whole evolutionary link between basal vascular plants and the most complicated and diverse angiosperms. Most importantly, it can help in identifying plant-specific gene families responsible for the specialized characteristic of each family and genus of the plant kingdom.

## Role of third-generation sequencing platforms in decoding plant genomes

We have seen a generation of large volumes of sequencing information through NGS and assembly platforms. Although NGS systems producing up to 100 Gb of data per run have advanced genome information (Imelfort and Edwards, 2009; Mardis, 2008; Schatz *et al*., 2010; The 1000 Genomes Project Consortium, 2010), the diversity and variations of different genomes and short read lengths from NGS make it difficult to achieve a complete published genome information. Therefore, taking DNA sequencing to a further level and dramatically reducing costs, several companies have hit the market with the third-generation sequencing (3GS) technologies. The remarkable quality of genome sequences produced by 3GS and mapping technologies are hundreds to thousands of times more contiguous and enable improved analysis of nearly every aspect of the genome. The genomes are more complete and show accurate representation of genes, regulatory regions and other important genomic elements, as well as better resolution of the overall chromosome organization (Burton *et*

*al.*, 2013; Roberts *et al.*, 2013; Ross *et al.*, 2013; Cao *et al.*, 2014; Pendleton *et al.*, 2015; Lee *et al.*, 2016).

## Next-generation sequencing and plant transcriptome analysis

The rapid developments in NGS technologies have revolutionized the field of plant transcriptomics, specifically in those plants where no genome sequence information is available. The NGS-based RNA sequencing technologies are generally called mRNA-seq tools. These technologies provide a novel method for identifying, mapping, and quantifying transcriptomes under different developmental stages and stress conditions. Deep RNA sequencing is a powerful tool for comparing gene expression analysis and discovering the full length 5′ and 3′ untranslated regions, novel splice junctions and transcripts, alternative transcription start sites, and rare transcripts (Cloonan *et al.*, 2008; Mortazavi *et al.*, 2008; Wilhelm *et al.*, 2008; Zhang *et al.*, 2010; Fig. 1.2). RNA-seq data with both technical and biological replicates show good level of reproducibility (Cloonan *et al.*, 2008). The transcribed RNA from genomic DNA, besides coding for proteins, also has potential non-coding RNAs. Owing to their significance in regulation of gene expression, these non-coding RNA have attracted great attention in recent past.

## Applications of transcriptome profiling and long non-coding RNA (lncRNA) in plant

### Cereals

A large number of studies are carried out to generate transcriptomes from cereal crops. RNA-seq was used to understand the transcriptional regulatory network of *Opaque2* (*O2*). *O2* transcription factor in maize plays an important role as a central regulatory molecule during maize endosperm development by regulating multiple regulatory pathways (Li *et al.*, 2015). The RNA-seq analysis found 1605 differentially expressed genes and 383 differentially expressed long, non-coding RNAs between wild-type and *O2* endosperms, 15 days after pollination. Transcriptome sequencing in rice was used to study differentially expressed genes

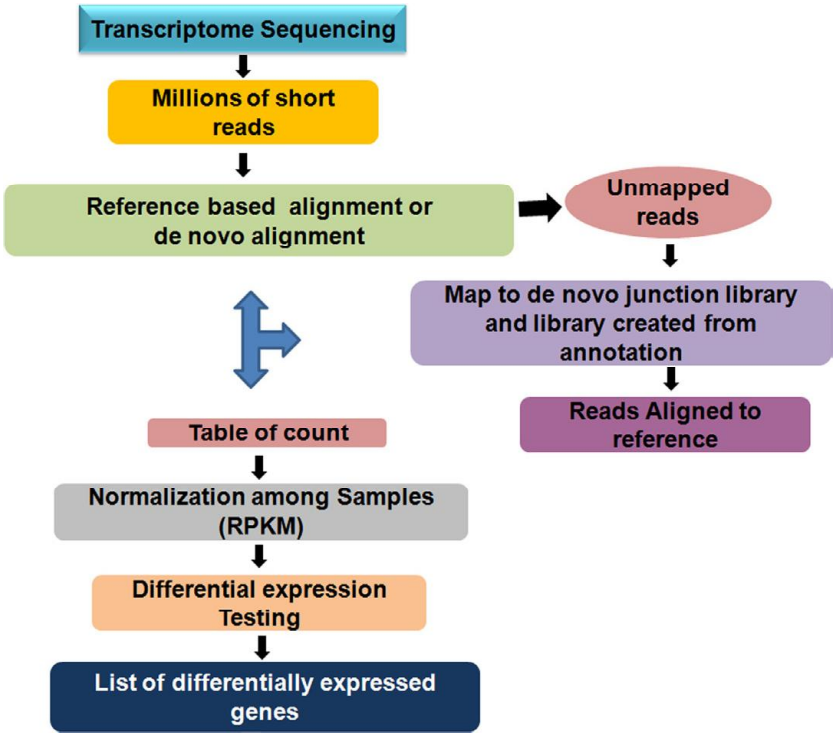


**Figure 1.2** Pipeline for transcriptome sequencing using NGS techniques.

(DEGs) upon infection with *Magnaporthe oryzae* and to probe the molecular response of rice to *Ustilaginoidea virens* infection (Kawahara *et al.*, 2012; Bagnaresi *et al.*, 2012; Chao *et al.*, 2014). Similarly, genome-wide transcription profiling of primed (Selenium and salicylic acid priming) and non-primed seedlings of rice was reported (Hussain *et al.*, 2016). Transcriptome of bread wheat provided a methodology for homeolog-specific sequence assembly (Schreiber *et al.*, 2012) and deep transcriptome sequencing of a wheat genome was used to construct a fine transcriptome map of the chromosome 3B (Pingault *et al.*, 2015). A NGS tool was also used in barley, finger millet and sorghum to study transcriptomes under different conditions (Bedada *et al.*, 2014; Rahman *et al.*, 2014; Tombuloglu *et al.*, 2015; Abdel-Ghany *et al.*, 2016).

Though many studies were conducted to profile the mRNA transcriptome of cereals using NGS technology, very few studies are available wherein NGS has been used for profiling of lncRNA. These studies include genome-wide analysis of non-coding parts of transcriptomes uncovering lncRNA in maize as well as rice and discovered long intergenic non-coding RNAs (lincRNA) which contain SNP associated with agriculturally important traits in maize and two lincRNAs in rice (Kim *et al.*, 2012; Wang *et al.*, 2015a).

### Pulses and oilseeds

Many studies of NGS applications in pulse crops are being reported in last few years. RNA-seq is successively used in chickpea, lentil, *Medicago truncatula* and *Vigna unguiculata* to identify developmental stage-specific transcriptomes, drought and salinity responsive transcriptomes, development of large scale genomics resources in lentil such as unigenes and 2393 EST-SSRs, to identify osmotic stress related lncRNAs and to generate a gene expression atlas of different plant tissues (Singh and Jain, 2014; Wang *et al.*, 2015b; Garg *et al.*, 2016; Yao *et al.*, 2016)

NGS technology has been used in oil-producing plants to study the transcriptome dynamics under different conditions. In soybean, RNA-seq atlas was built using NGS (Severin *et al.*, 2010). Transcriptome profiling of peanut, *Brassica napus* and sunflower was performed to identify various biochemical pathways. Further novel lncRNAs of *Brassica napus* in response to *Sclerotenia sclerotiorum* infection were identified. NGS technology was also used to understand the molecular basis of a higher rate of recombination in cultivated genotypes during meiosis as compared to their wild ancestors in sunflower ( Joshi *et al.*, 2016; Florez-Zapata *et al.*, 2016).

### Horticulture and ornamental crops

Li *et al.* (2012) identified 92 DEGs that were associated with the metabolism and anthocyanin synthesis during fruit ripening. Rowland *et al.* (2012) analysed transcriptomes of different blueberry tissues to identify genes that were associated with cold acclimation and fruit development. Various other transcriptome studies involving fruits, flowers and vegetables crops are given Table 1.2.

### Commercial crops

Cotton is an important fibre crops in India that has high commercial value. In the study aimed at analysing the lncRNAs in cotton, Wang *et al.* (2015c) characterized lncRNAs in *Gossypium* species. Recently, Lu *et al.* (2016) identified 10,820 lncRNAs that were associated with different abiotic stresses. Similarly, Zou *et al.* (2016) also identified a total of 5996 lncRNAs that were reported rapid and dynamic changes during early fibre and rapid elongation stages. NGS techniques were used for discovery of genes, regulatory sequences and non-coding RNA for the improvement of sugarcane as a biofuel crop. In a recent study, six genotypes of sugarcane were studied and a total of 72,269 unigenes were identified. Out of which more than 28,788 of these unigenes showed significant similarity to sorghum proteins indicating that both species share a higher degree of genetic lineage. In another study, transcriptome analysis of sugarcane during smut pathogen infection (*Sporisorium scitamineum*) identified 65,852 unigenes and most of the DEGs were related to metabolic pathways in resistant and susceptible genotypes (Que *et al.*, 2014).

Recently, lncRNAs have been found to play an important role in gene regulation by acting as endogenous target mimics. The size of lncRNAs is usually more than 200 nucleotides in length (Chen, 2009; Rinn and Chang, 2012). In plants, identification of long non-coding RNA is more recent and not as comprehensive compared to other eukaryotes (Ulitsky and Bartel, 2013; Zhang *et al.*, 2014). Non-coding RNA regulates the expression

**Table 1.2** List of mRNA and lncRNA studies and NGS platforms used for sequencing in different crops

| Number | Crop | Platform | Validation method | Condition | Reference |
|---|---|---|---|---|---|
| 1 | *Glycine max* | Illumina | – | Fourteen diverse tissues | Severin *et al*., 2010 |
| 2 | *Cucumis sativus* | Roche-454 | – | Flower buds | Guo *et al*., 2010 |
| 3 | *Lens culinaris* | Roche 454 | – | Developmental stages | Kaur *et al*., 2011 |
| 4 | *Daucus carota* | Illumina | – | Root and leaf tissues | Iorizzo *et al*., 2011 |
| 5 | *Oryza sativa* | Illumina | qRT-PCR | Biotic stress | Kawahara *et al*., 2012 |
| 6 | *Oryza sativa* | Illumina | qRT-PCR | Biotic stress | Bagnaresi *et al*., 2012 |
| 7 | *Triticum aestivum* | Illumina | – | Root and shoot tissue from seedlings | Schreiber *et al*., 2012 |
| 8 | *Arachis hypogaea* | Illumina | qRT-PCR | Immature seeds | Zhang *et al*., 2012 |
| 9 | *Cyanococcus* | Illumina | qRT-PCR | Fruit skin and pulp | Li *et al*., 2012 |
| 10 | *Cyanococcus* | Roche 454 | qRT-PCR | Cold acclimation | Rowland *et al*., 2012 |
| 11 | *Carnation* | GS FLX 454 pyrosequencing | – | Flower bud, flowers, leaves and stem (ethylene treated and control) | Tanase *et al*., 2012 |
| 12 | *Allium sativum* | Illumina | – | Vegetative buds | Sun *et al*., 2012 |
| 13 | *Saccharum* | Illumina | – | Leaves | Vicentini *et al*., 2012 |
| 14 | *Chrysanthemum* | Illumina | – | Stems and leaves | Wang *et al*., 2013 |
| 15 | *Chrysanthemum* | Illumina | – | Dehydration stress | Xu *et al*., 2013 |
| 16 | *Allium cepa* | Roche-454 | – | Vernalized bulbs, leaves, unopened umbels, bulbs, and roots | Duangjit *et al*., 2013 |
| 17 | *Oryza sativa* | Illumina | qRT-PCR | Biotic stress | Chao *et al*., 2014 |
| 18 | *Oryza sativa* | Illumina | qRT-PCR | Anthers before flowering, pistils before flowering, spikelets and shoots | Zhang *et al*., 2014 |
| 19 | *Hordeum vulgare* | Roche 454 | – | Drought stress | Bedada *et al*., 2014 |
| 20 | *Eleusine coracana* | Ion torrent | qRT-PCR | Salinity stress | Rahman *et al*., 2014 |
| 21 | *Cicer arietinum* | Illumina | – | Vegetative and reproductive tissues | Singh and Jain, 2014 |
| 22 | *Mangifera indica* | Illumina | qRT-PCR | Hot water treatment | Luria *et al*., 2014 |
| 23 | *Oryza sativa* | Illumina | qRT-PCR | Developmental stages | Wang *et al*., 2015b |
| 24 | *Zea mays* | Illumina | qRT-PCR and polyclonal antibody | Endosperm development | Li *et al*., 2015 |
| 25 | *Oryza sativa* and *Zea mays* | Illumina | qRT-PCR | Flower buds, flowers, flag leaves, roots, (before and after flowering stage), shoot and root tissues | Wang *et al*., 2015a |
| 26 | *Oryza sativa* and *Zea mays* | Illumina | qRT-PCR | Flower bud, milk grain and mature seed | Wang *et al*., 2015a |
| 27 | *Triticum aestivum* | Illumina | – | Root, leaf, stem, spike, and grain at three developmental stages | Pingault *et al*., 2015 |
| 28 | *Hordeum vulgare* | Illumina | qRT-PCR | Boron treatment | Tombuloglu *et al*., 2015 |
| 29 | *Medicago truncatula* | Illumina | qRT-PCR | Osmotic and salt stress | Wang *et al*., 2015 |

**Table 1.2** Continued

| Number | Crop | Platform | Validation method | Condition | Reference |
|---|---|---|---|---|---|
| 30 | *Mangifera indica* | Illumina | qRT-PCR | Fruit ripening | Dautt-Castro *et al*., 2015 |
| 31 | *Gossypium* | Illumina | qRT-PCR | Cotton fibre | Wang *et al*., 2015c |
| 32 | *Oryza sativa* | Illumina | qRT-PCR | Selinium and salicylic acid priming | Hussain *et al*., 2016 |
| 33 | *Sorghum bicolor* | SMRT(single-molecule real-time) cells (Pacific Biosciences) | qRT-PCR, cloning and sequencing | Seedlings | Abdel-Ghany *et al*., 2016 |
| 34 | *Sorghum bicolor* | Pacific Biosciences SMRT long-read isoform sequencing | RT-PCR | 8 day old seedlings | Abdel-Ghany *et al*., 2016 |
| 35 | *Helianthus annulus* | Illumina | qRT-PCR | Disc florets at R2 development stage | Florez-Zapata *et al*., 2016 |
| 36 | *Arabidopsis thaliana* | Illumina | qRT-PCR, RACE, western blot | Biotic stress | Gao *et al*., 2016 |
| 37 | *Actinidia deliciosa* | Illumina | qRT-PCR | Fruit samples from five 5-year-old plants | Tang *et al*., 2016 |
| 38 | *Raphanus sativus* | Illumina | qRT-PCR | Taproot sample at developmental stages | Yu *et al*., 2016 |
| 39 | *Cicer arietinum* | Illumina | qRT-PCR | Salinity and drought stress | Garg *et al*., 2016 |
| 40 | *Mangifera indica* | Roche 454 and Illumina | qRT-PCR | Fruit ripening | Srivastava *et al*., 2016 |
| 41 | *Actinidia chinensis* | Illumina | RT-PCR and qRT-PCR | Fruit samples from 5-year-old plant and 20, 120 and 127 days after pollination (DAP) | Tang *et al*., 2016 |
| 42 | *Rosa* | Illumina Hiseq | qRT-PCR | Cold stress | Zhang *et al*., 2016 |
| 43 | *Allium cepa* | Illumina | qRT-PCR | Floral inflorescences | Liu *et al*., 2016a |
| 44 | *Raphanus sativus* | Illumina | qRT-PCR | Bolting and flowering stage root, stem, leaf, flower, floral buds | Nie *et al*., 2016 |
| 45 | *Raphanus sativus* | Illumina | qRT-PCR | Salt stress | Sun *et al*., 2016 |
| 46 | *Brassica napus* | Illumina | qRT-PCR | Biotic stress | Joshi *et al*., 2016 |
| 47 | *Helianthus* | Illumina | – | Non-meiosis versus meiosis-specific cell | Flórez-Zapata *et al*., 2016 |
| 48 | *Gossypium* | Illumina | qRT-PCR | Drought stress | Lu *et al*., 2016 |
| 49 | *Gossypium* | Illumina | qRT-PCR | Cotton fibres and leaves | Zou *et al*., 2016 |

level of target genes via various molecular mechanisms (Quan *et al.,* 2015; Zhu and Wang, 2012). lncRNAs act as a regulator in important biological processes by enhancing target site accessibility to RNA polymerases, formation of RNA-dsDNA triplex, inhibition of RNA polymerase activities as well as regulation of transcription factors (Lipshitz *et al.,* 1987; Nguyen *et al.,* 2001; Willingham *et al.,* 2005; Martianov *et al.,* 2007; Hirota *et al.,* 2008; Mariner *et al.,* 2008). They also have a role in post-transcriptional modulations of mRNA. lncRNAs are found to have role in regulating complex gene regulatory networks involved in plant development and stress management (Crespi *et al.,* 1994;

Ding *et al.*, 2012; Heo and Sung, 2011; Zhang *et al.*, 2014).

## Identification of small non-coding RNA by NGS technology

Small RNAs (sRNAs) are another important class of RNA molecules of 21–24 bases long and include microRNAs (miRNAs), short interfering RNAs (siRNAs), transacting siRNAs (ta-siRNAs) and cis/trans-natural antisense small-interfering RNAs (nat-siRNAs). These small RNAs are non-coding, essential entities that regulate gene expression in epigenetic processes almost in every domain of life (Ruiz-Ferrer and Voinnet, 2009; Zhang *et al.*, 2011). The sequence alterations in miRNA or other small non coding RNAs can be mapped by genomic sequencing or RNA-seq, while the expression levels can be determined by RNA-seq or deep sequencing or expression microarrays (Singh *et al.*, 2012). The alliance of gene expression data with a small RNA data set will help to understand how different biological processes coordinate together in a cellular context (Jain, 2012). In recent times, a large number of studies involving NGS technology have been reported primarily to identify novel as well as conserve miRNA, along with their targets under different stresses or developmentally regulated conditions (Table 1.3) sRNAome analysis of *Arabidopsis thaliana* revealed while expression of miR156, miR399, miR778, miR827, and miR2111 was induced, yet, expression of miR169, miR395, and miR398 was repressed under phosphate (Pi) deficiency indicating that these miRNAs may be involved in Pi uptake in *Arabidopsis* (Hsieh *et al.*, 2009). In *Brachypodium distachyon*, a model crop, three conserved and 25 novel miRNAs showed significant change of expression in response to cold stress and differential expression profiling of 94 conserved miRNAs from 28 families during vegetative and reproductive tissues (Zhang *et al.*, 2009; Wei *et al.*, 2009). A review describing the study of various aspects of miRNA using deep sequencing tools was documented for further information (Yang and Li, 2012).

### Agricultural crops

Rice being a major cereal crop, various transcriptome studies to identify non-coding RNAs have been reported. NGS was used to categorize different classes of small RNA regulatory elements from mature rice grain and seedlings (Heisel *et al.*, 2008; Guo *et al.*, 2012; Campo *et al.*, 2013; Li *et al.*, 2014). These studies together signified the role of miRNAs in rice blast disease and stripe virus resistance. Besides biotic factors, a study involving abiotic stresses has also been reported and elucidated 294 known and 539 novel heat-responsive miRNAs during heat stress (Li *et al.*, 2015b).

In maize, the epigenome was critically scrutinized by NGS technologies for its relationships to mRNA and small RNA transcriptomes (Wang *et al.*, 2009; Liu *et al.*, 2014; Zhou *et al.*, 2016) to understand the molecular aspects underlying maize ear development and to elucidate the rice *black-streaked dwarf* virus-responsive pathway in maize.

Wheat being hexaploid, its genome is considered to be genetically complex. In an earlier NGS-based study on wheat, 58 miRNAs comprising 43 miRNA families were identified, of which 20 families were conserved and 23 were found to be novel (Yao *et al.*, 2007). Further, in another study, 170 conserved miRNAs representing 25 families and 23 novel miRNAs were also identified (Wei *et al.*, 2009). Further, in wheat, a total of 2076 small RNAs were identified (Yao *et al.*, 2010). Similarly a set of wheat miRNAs which play roles in regulating wheat response to powdery mildew pathogen *Erysiphe graminis* f. sp. *tritici* and heat stress were also identified using high-throughput sequencing (Xin *et al.*, 2010). Tang *et al.* (2012) analysed the male sterility system of thermosensitive genic male sterile (TGMS) lines of wheat and concluded that miR167 and tasiRNA-ARF play a role in the developmental response to cold stress and the regulatory pathways of sRNA that were linked with male sterility. A brief summary of work related to wheat is given in Table 1.3.

Though initial efforts were concentrated to study the small RNAs in cereals, of late efforts are also being made to understand the role of these RNA molecules in pulse, oilseeds and other crops of economic importance. In *Medicago truncatula*, an important model legume, a small RNA library was generated and conserved and novel small RNAs (miR1507, miR2118, miR2119 and miR2199) were identified. They also identified three novel transcripts encoding TIR-NBS-LRR disease resistance (Jagadeeswaran *et al.*, 2009). The miRNA (miR408) could induce drought tolerance in chickpea (Jain *et al.*, 2014; Srivastava *et al.*, 2015;

**Table 1.3** Small-RNA studies and NGS platforms used for sequencing in different crops

| Number | Crop | Platform used | Validation method | Experimental conditions | Reference |
|---|---|---|---|---|---|
| 1 | *Triticum aestivum* L. line 3338 w | 454 Life Sciences™ technology | RNA gel blot analysis and RT-PCR | Wheat and monocot-specific miRNAs | Yao *et al*., 2007 |
| 2 | *Oryza sativa* spp. *japonica* cv. Nipponbare | 454 Life Sciences platform | Small RNA blotting and 5′RACE | Mature rice grain and seedlings | Heisel *et al*., 2008 |
| 3 | *Solanum lycopersicum* | 454 Life Sciences™ technology | Northern blotting and 5′RACE | Fruit ripening | Moxon *et al*., 2008 |
| 4 | *Arabidopsis thaliana* ecotype Columbia | Solexa sequencing technology (Illumina) | RLM-5'RACE | Phosphate deficiency | Hsieh *et al*., 2009 |
| 5 | *Medicago truncatula* Gaertner cv. Jemalong | 454 Life Sciences™ technology | Small RNA blotting and 5′RACE | Legume specific | Jagadeeswaran *et al*., 2009 |
| 6 | *Zea mays* inbred line B73 | Solexa Sequencing | Computational validation | Tissue-specific miRNA distribution | Wang *et al*., 2009 |
| 7 | *Brachypodium distachyon* BD21-3 cv. Chinese Spring | Solexa Sequencing | Northern blotting, RT-PCR and 5′RACE | vegetative and reproductive growth stage | Wei *et al*., 2009 |
| 8 | *Brachypodium distachyon* | Illumina-Solexa 1 G Genetic Analyzer | RNA gel blot analysis | Cold stress | Zhang *et al*., 2009 |
| 9 | *Triticum aestivum* L line JD8-Pm30 | Solexa sequencing | RNA gel blot analysis and qRT-PCR | Powdery mildew infection and heat stress | Xin *et al*., 2010 |
| 10 | *Triticum aestivum* L | 454 Life Sciences™ technology | Northern blotting and RT-PCR | heat, cold, salt and dehydration stress | Yao *et al*., 2010 |
| 11 | *Arachis hypogaea* L. cultivar Huayu19 | Solexa 1G Genome Analyzer | Stem–loop RT-PCR and qRT-PCR | Tissue- and/ or growth stage specific expression | Chi *et al*., 2011 |
| 12 | *Glycine max* inbred line of ‘HJ-1’ | Solexa sequencing | Northern blot analysis and qRT-PCR | Drought, salinity, and alkalinity stress | Li *et al*., 2011 |
| 13 | *Carthamus tinctorius L.* | Illumina Solexa Genome Analyzer | Computational validation | Tissue specific (Seed, leaf and petal) | Li *et al*., 2011 |
| 14 | *Glycine max* cv. Heinong44 | Solexa sequencing | RT-PCR and RLM-5′RACE | miRNA biogenesis | Song *et al*., 2011 |
| 15 | *Brassica rapa* ssp. *chinensis* | Illumina Genome Analyzer | Northern blot analysis, qRT-PCR, and cRT-PCR | Heat responsive | Wang *et al*., 2011 |
| 16 | *Brassica rapa* ssp. *chinensis* | Illumina Genome Analyzer | Northern Blotting, RT-PCR and 5′RACE | Heat stress | Yu *et al*., 2012 |
| 17 | *Oryza sativa* L. *japonica*. cv. Nipponbare | Illumina Solexa sequencing | sRNAs gel blot analysis and qRT-PCR | Virus infected | Guo *et al*., 2012 |
| 18 | *Triticum aestivum* TGMS line BS366 | Illumina Genome Analyzer | RLM-5′RACE and In situ hybridisation | Cold stress | Tang *et al*., 2012 |
| 19 | *Brassica napus* line, Westar | Solexa sequencing (Illumina) | qRT-PCR and RLM-5'RACE | miRNAome | Xu *et al*., 2012 |
| 20 | *Oryza sativa* L. cv. Nipponbare, *Oryza glaberrima* and wild rice species | 454 Life Sciences Technology and Microarray | Northern Blotting, qRT-PCR | *Magnaporthe oryzae* responsive | Campo *et al*., 2013 |

**Table 1.3** Continued

| Number | Crop | Platform used | Validation method | Experimental conditions | Reference |
|---|---|---|---|---|---|
| 21 | *Carthamus tinctorius L.* | Illumina HiSeq™ 2000 | Northern blot hybridization, stem–loop RT-PCR | Developing seeds with high linoleic and oleic acid content | Cao *et al*., 2013 |
| 22 | *Saccharum* spp. cultivars RB867515 | Solexa platform | Stem–loop RT-PCR and qRT-PCR | Drought responsive | Gentile *et al*., 2013 |
| 23 | *Brassica napus* L. cv.DH12075 | SOLiD v3 sequencing | qRT-PCR | Seed maturation | Huang *et al*., 2013 |
| 24 | *Triticum aestivum* L. | Pyrosequencing | Small RNA blot analysis | siRNA-mediated silencing of TAS3 transcripts | Li *et al*., 2013 |
| 25 | *Gossypium hirsutum* | Illumina Genome Analyzer | qRT-PCR and RLM-5'RACE | Somatic embryogenesis | Yang *et al*., 2013 |
| 26 | *Solanum melongena* L. | Illumina/Solexa | qRT-PCR | *Verticillium dahliae* infection | Yang *et al*., 2013 |
| 27 | *Solanum tuberosum* group Andigena (line ADG573) | Illumina GAIIX sequencer | Computational validation | miRNAome | Zhang *et al*., 2013 |
| 28 | Miniature Tomato cv. Micro-Tom | Illumina Solexa system | qRT-PCR | Cucumber mosaic virus responsive | Feng *et al*., 2014 |
| 29 | *Cicer arietinum* L. genotype ICC4958 | Illumina Genome Analyser | qRT-PCR | miRNAome profiling | Jain *et al*., 2014 |
| 30 | *Solanum tuberosum* cv. Kufri Chandramukhi | Illumina GAIIx | qRT-PCR and RLM-RACE | Different developmental stages | Lakhotia *et al*., 2014 |
| 31 | *Oryza sativa* | Illumina sequencing | sRNAs gel blot analysis, overexpressing transgenic and qRT-PCR | *Magnaporthe oryzae* infected (0, 12, 24, 48, 72 hpi) | Li *et al*., 2014 |
| 32 | *Zea mays* inbred lineB73 | Solexa Sequencing | qRT-PCR | Maize ear development | Liu *et al*., 2014 |
| 33 | *Coffea canephora* | Illumina HiSeq 2000 | Computational validation | miRNAome | Loss-Morrais *et al*., 2014 |
| 34 | *Brassica napus* L. | Solexa sequencing (Illumina) | qRT-PCR and RLM-5′RACE | *Verticillium longisporum* infection | Shen *et al*., 2014 |
| 35 | Broccoli (*Brassica oleracea* var. *italica*) | Illumina HiSeq | qRT-PCR | Salinity stress | Tian *et al*., 2014 |
| 36 | *Poncirus trifoliata* (L.) Raf. | Illumina Genome Analyser | qRT-PCR | Cold responsive | Zhang *et al*., 2014 |
| 37 | *Solanum tuberosum* tetraploid cultivar 'Zihuabei'. | Solexa sequencing technology | qRT-PCR | Drought stress | Zhang *et al*., 2014 |
| 38 | *Solanum linnaeanum*, brinjal | Illumina GAIIx | qRT-PCR | Salt stress | Zhuang *et al*., 2014 |
| 39 | *Vigna unguiculata* | Illumina Genome Analyzer | Northern blot hybridization | Drought stress | Barrera-Figueroa *et al*., 2011 |
| 40 | *Salicornia europaea* | Illumina sequencing | qRT-PCR and RLM-5′RACE | Salt stress | Feng *et al*., 2015 |

**Table 1.3** Continued

| Number | Crop | Platform used | Validation method | Experimental conditions | Reference |
|---|---|---|---|---|---|
| 41 | *Brassica napus* L. cv.DH12075 | SOLiD v3 sequencing system | qRT-PCR | Seed maturation | Hayzadeh *et al*., 2015 |
| 42 | *Rosa multiflora* Thunb. | Solexa-Illumina platform | RT-PCR | Pathogen stress (Virus and Viroids) | He *et al*., 2015 |
| 43 | *Cajanus cajan* L. | Whole genome shotgun sequencing | Computational validation | Plant growth and development | Kompelli *et al*., 2015 |
| 44 | *Oryza sativa* cultivar Nagina 22 | Ion Proton Sequencer | qRT-PCR | Heat stress | Li *et al*., 2015 |
| 45 | *Cymbidium ensifolium* 'Tiegusu' | Solexa technology | RT-qPCR | Floral development | Li *et al*., 2015 |
| 46 | Turnip cultivar 'Chang Huang Man Jing' | Illumina HiSeqTM 2000 | qRT-PCR | Tuberous root development | Li *et al*., 2015 |
| 47 | *Raphanus sativus* L. advanced inbred line 'NAU-YH' | Solexa sequencing (Illumina) | RT-qPCR | Chromium stress | Liu *et al*., 2015 |
| 48 | *Solanum pimpinellifolium* L3708 | HiSeq 2000 Sequencing System | qRT-PCR | *Phytophthora infestans* (Pathogen resistance) | Luan *et al*., 2015 |
| 49 | *Triticum aestivum* L. cv. Hanxuan10 and Zhengyin1 w | Genome Analyzer IIx System | Northern blotting, RT-PCR and qPCR | Dehydration stress | Ma *et al*., 2015 |
| 50 | *Raphanus sativus* L. inbred line 'NAU-LU127' | Illumina HiSeq™ 2000 | RT-qPCR | Bolting and flowering time related | Nie *et al*., 2015 |
| 51 | *Camelina sativa* | Illumina HiSeq 2000 | RT-PCR | Different developmental stages | Poudel *et al*., 2015 |
| 52 | *Brassica napus* cultivars Tapidor and Ningyou7 | Illumina HiSeq 2000 | qRT-PCR and RLM-RACE | Double haploid lines | Shen *et al*., 2015 |
| 53 | *Cicer arietinum* L. genotype IC4958 | Illumina Genome Analyser | Small RNA gel blot and 5'RACE | Phasi-siRNAs discovery | Srivastava *et al*., 2015 |
| 54 | *Raphanus sativus* L. advanced inbred line 'NAU-YH' | Illumina HiSeq™ 2000 | RT-qPCR | Salinity stress | Sun *et al*., 2015 |
| 55 | *Brassica rapa* ssp. *pekinensis* | Solexa platform | qRT-PCR and Microscopy | Pollen abortion and Bud development | Wei *et al*., 2015 |
| 56 | *Gossypium hirsutum* L. cultivar TM-1 | Illumina HiSeq high-throughput sequencing platform | Stem–loop RT-PCR and qRT-PCR | Drought- and salinity-responsive | Xie *et al*., 2015 |
| 57 | *Solanum lycopersicum* L. cv. 'Ailsa Craig' | Illumina Genome Analyzer I | qRT-PCR | Tomato pedicel abscission | Xu *et al*., 2015 |
| 58 | *Arachis hypogaea* L. Luhua 14 and *A. glabrata* | Illumina Genome Analyzer | qRT-PCR | Pathogen resistance *Ralstonia solanacearum*, bacterial wilt | Zhao *et al*., 2015 |
| 59 | *Camellia sinensis* (L.) O. Kuntze cv. lLongjing 43' | Illumina HiSeq 2000 platform | qRT-PCR | Chilling and freezing | Zheng *et al*., 2015 |
| 60 | *Ipomoea batatas* L.,cv. Ningzishu 1 | Solexa sequencing technology | Computational validation | miRNAome | Bian *et al*., 2016 |

**Table 1.3** Continued

| Number | Crop | Platform used | Validation method | Experimental conditions | Reference |
|---|---|---|---|---|---|
| 61 | *Brassica napus* | Illumina HiSeq2000 | Computational validation | Synthetic variety | Fu *et al*., 2016 |
| 62 | *Camellia sinensis* | Illumina HiSeq2500 | qRT-PCR | Drought stress | Liu *et al*., 2016b |
| 63 | *Moringa oleifera* | Illumina HiSeq | qRT-PCR and Western blotting | Pharmacological potential properties | Pirrò *et al*., 2016 |
| 64 | *Solanum tuberosum* Zhuangshu No. 3 | Illumina HiSeq2000 | qRT-PCR | Secondary metabolism | Qiao *et al*., 2016 |
| 65 | *Raphanus sativus* L. advanced inbred line 'NAU-YH' | Illumina HiSeq system | RT-qPCR | Taproot thickness | Yu *et al*., 2016 |
| 66 | 'Summer Black' grapevine (hybrids of *V. vinifera* and *V. Labrusca* | Solexa sequencing (Illumina) | qRT-PCR and RLM-RACE | In response to exogenous ethylene | Zhao *et al*., 2016 |
| 67 | *Zea mays* inbredlineB73 | Illumina HiSeq | qRT-PCR | Black-streaked dwarf virus, maize rough dwarf disease (MRDD) | Zhou *et al*., 2016 |
| 68 | *Solanum lycopersicum* 'Rui Xin' | Illumina HiSeq2500 | Computational validation | Chilling injury | Zuo *et al*., 2016 |

Hajyzadeh *et al*., 2015). Similarly novel miRNAs and their targets involved in stress response were also identified in pigeonpea and chickpea (Kompelli *et al*, 2015; Barrera-Figueroa *et al*., 2011).

Oilseed crops, largely grown for human consumption, are important for Indian agriculture. One of the important oilseed crops, soybean, is grown all over the world. To understand the regulatory network of miRNAs and their functions during seed development as well as to the miRNAome profiling associated with abiotic stress responses in soybean, NGS technology was used to identify the differentially expressed miRNAs (Song *et al*., 2011; Li *et al*., 2011b). Besides soybean, members of Brassicaceae family, specifically *Brassica napus*, is a major source of vegetable oil. There are many reports in *Brassica napus*, which used NGS techniques to decipher various molecular mechanisms during different stages of growth, during their interaction with biotic and abiotic stresses, oleic acid content, etc. (Table 1.3). Commercial or cash crops are another important category belonging to agricultural crops. Some of the major crops of India on which miRNA studies have been conducted are cotton (*Gossypium hirsutum*), tea (*Camellia sinensis* L.), *Coffea canephora* and sugarcane. In cotton, using NGS tools, 25 novel miRNAs that were associated with somatic embryogenesis were identified (Yang *et al*., 2013). Similarly, miRNAs in response to drought and salinity stress were also studied in cotton, and miRNAs which played role in drought and fibre development were identified (Xie *et al*., 2015). A brief summary of works is given in Table 1.3.

### Horticulture crops

Besides agricultural crops, horticultural crops are also a significant component of farming systems and sources for many dietary supplements. Various studies based on NGS technologies were conducted to identify and characterize miRNAs in these crops (Table 1.3). For instance, miRNAs involved in tomato fruit during ripening, *Phytophthora* infections and cucumber mosaic resistance, and chilling injury were identified (Moxon *et al*., 2008; Luan *et al*., 2015; Zou *et al*., 2016). Similarly, *Verticillium dahlia* responsive miRNAs in *Solanum linnaeanum* (Yang *et al*., 2013), and those involved in drought stress in *Solanum tuberosum* (Lakhotia *et al*., 2014) were also identified. Overall, small non-coding

RNAs is a broad class of regulatory RNAs, and behaves as protein counterparts involved in regulating post transcriptional gene silencing and translational repression. Deep sequencing employs efficient, economical, massive parallel sequencing, generating millions of small RNA sequence reads from a given sample. sRNAome by deep sequencing quantifies absolute abundance and allows for the discovery of novel microRNAs that have escaped previous cloning and standard sequencing efforts (Creighton *et al.*, 2009). Some miRNAs can significantly affect plant traits, including virus resistance, nematode resistance, drought and salinity tolerance, heavy metal detoxification, biomass yield, grain yield, fruit development and flower development. Therefore, miRNAs are considered as a newly identified gene resource for the genetic improvement of crop plants (Zheng and Qu, 2015). Therefore, these all studies which analysed and identified miRNAs have the potential to enhance food security by helping breed crop cultivars with improved agronomic traits (Zhou and Luo, 2013). Studies have indicated that small non-coding RNA loci, like protein-coding genes, could be targets of domestication selection and play an important role in crop domestication, and improvement, abiotic stress, plant–pathogen interactions and breeding useful traits (Ruiz-Ferrer and Voinnet, 2009; Wang *et al.*, 2010).

## NGS and metagenomics of plant–microbe interaction

Microbes are the most abundant living organisms on earth. In spite of their abundance, only small fraction of them have been explored and very few of them can be grown in laboratory conditions. Overall, large chunks of the microbial population are out of the reach of scientists largely due to an inability to culture them under laboratory conditions. Quite intriguingly, it has been observed that microorganisms growing on media and those directly obtained from natural samples showed very large differences in their number, displaying so called the great plate count anomaly (Staley *et al.*, 1985). Therefore, lack of standard culture media and growth conditions has largely led to the restricted scientific exploration of these microbial populations. In the context of foregoing observations, it becomes necessary to study the microbial populations in their natural conditions so as to capture the microbial community or metagenome (Woese *et al.*, 1987). Metagenome is the use of advanced techniques to analyse microbial communities taken directly from the natural sources without growing them under laboratory conditions (Chen *et al.*, 2005).

### Approaches used in metagenomics

Broadly two approaches are used to study metagenomics, i.e. sequence-based metagenomics and function-based metagenomics. The sequence-based study depends on identifying the complete genetic sequence of microorganisms present in the sample. In most of the studies targeting microbial communities, 16S rRNA gene sequences were used to identify the species. This method is useful to study the evolutionary conservation and phylogenetic analysis of the microbes present in the samples (Claesson *et al.*, 2010; Creer *et al.*, 2016). On the contrary, function-based metagenomics explores the products of the microbial community to study that specific community (Coughlan *et al.*, 2015). Subsequently it was sequencing techniques, particularly the introduction of NGS techniques, that revolutionized the plant–microbial metagenomic studies. This sequencing technology opened a new way for discovering and analysing these organisms at genome level, which is a culture independent technology that utilizes a combination of various research methods, specifically NGS and bioinformatics tools. Plants and microbes are in continuous interaction with each other in the environment. The association of microbes with the plant can be of endophytic (inside the plant) or epiphytic (attached to the plant). Plant–microbial associations could be positive interactions, neutral associations or negative interactions (pathogens). Almost every part of a plant is inhabited by a microbial community, yet rhizosphere and phyllosphere are the major sites for plant microbe interactions for most of the microbes. Microbiota of these two regions (rhizosphere and phyllosphere) are so closely associated with plants that even they are called as second genome of the plant (Berendsen *et al.*, 2012). Several studies have been carried out to analyse the microbiome of various plants sampled from the phyllosphere, rhizosphere, and various parts, and were found beneficial to the plant (Lugtenberg *et al.*, 2009; Porras-Alfaro *et al.*, 2011; Vorholt *et al.*, 2012). The details of the

metagenomics studies aimed at plant–microbial interactions and plant microbiota are given in Tables 1.4 and 1.5. Recently, the 454 NGS technique was used to study the diversity, community structure, and dynamics of endophytic bacteria in different plant species and it found four classes of *Proteobacteria* with *Alphaproteobacteria* as the dominant class and revealed that host plant species had greater influences on type of bacterial communities (Ding and Melcher, 2016).

Besides their application in metagenomics, NGS techniques also have great applications in the study of metatranscriptomics. Metatranscriptomics includes analysis of gene transcripts directly isolated from the entire community of organisms. Metatranscriptome study is also referred to as environmental transcriptomes, microbial community gene expression profiles, microbial community RNAs and whole community transcripts. The metatranscriptome field has opened a door to study various aspects around environmental community, i.e. active community members and metabolic pathways (Urich *et al.*, 2008). The metatranscriptomics approach has been used to identify various genes actively involved in *Eichhornia crassipes* and *Fusarium verticillioidies* association, in strawberry plants for defining the fungal communities associated with different organs of plants (Luo *et al.*, 2015; Abdelfattah *et al.*, 2016). Recently, a study combining NGS and metagenomic analysis was conducted to generate large number of cDNAs using model system tomato pepino mosaic virus. Subsequently, the same approach was used in the study of globe amaranth (*Gomphrena globose*) infected with an unknown pathogen. Therefore, this method hastens the process of development of routine assays for new viral pathogens (Adams *et al.*, 2009).

NGS technique was also used for sequencing whole viral genomes to undertake plant metagenomic studies to discover new viruses. Adams *et al.* (2013) identified the presence of *Maize chlorotic mottle virus* and *Sugarcane mosaic virus*, causal agents

**Table 1.4** Plant–microbial interaction studies using NGS-based metagenomics

| Number | Host | Interacting partner | Environment | Conclusions | Platform | Reference |
|---|---|---|---|---|---|---|
| 1 | Wheat | *Azospirillum brasilense* | Rhizosphere | Up-regulated genes related to nutrient uptake, cell cycle, and nitrogen assimilation that enhance productivity and growth | SOLiD | Camilios-Nato *et al*., 2014 |
| 2 | Wheat | Microbiome | Rhizosphere | Fertilizers with high nutrient availability and long-term storage leads to less microbe interaction with crop | 454 Pyrosequencing | Ai *et al*., 2015 |
| 3 | *Eichhornia crassipes* | *Fusarium verticillioidies* | Rhizosphere | Mutualistic action of plant and fungi was efficient for bioremediation | Illumina Hi Seq 2500 | Luo *et al*., 2015 |
| 4 | Potato | *Burkholderia phytofirmans* PsJN | Rhizosphere | Stress signal perceived by plant are also affects plant endophyte association | Illumina Hi Seq 2000 | Tezerji *et al*., 2015 |
| 5 | Soybean | Mycovirus | Phyllosphere | Novel mycoviruses were identified | Illumina Hi Seq 2500 | Marzona *et al*., 2016 |
| *6* | Strawberry | Fungi | Phyllosphere | Diverse fungal organisms inhabits on plants and *Botrytis* and *Cladosporium* were dominant. | 454 GS FLX+System | Abdelfattah *et al*., 2016 |
| 7 | Rice | *S. epidermidis* | Rhizosphere | *S. epidermidis* of plant and animal origin are diversified at genome level | Illumina MiSeq | Chaudhry *et al*., 2016 |
| 8 | Non-cultivated plants | Endophytic bacteria | Phyllosphere | Proteobacteria found to be highest phylum (85.42%). Acidobacteria (0.59%) lowest on leaf of all five plants | 454 Pyrosequencing | Ding *et al*., 2016 |

**Table 1.5** Studies of plant associated microbiota using NGS techniques

| Number | Host | Environment | Conclusions | Platform | References |
|---|---|---|---|---|---|
| 1 | Soybean | Phyllosphere | Composition of microbiota and proteomes remain consistent in different plants species | Roche 454 | Delmotte *et al*., 2009 |
| 2 | Rice | Phyllosphere and Rhizosphere | Phyllosphere and rhizosphere region analysis to identify bacteria and archea in association with rice | Roche 454 | Knief *et al*., 2012 |
| 3 | Tamarisk, soybean, *Arabidopsis thaliana* | Phyllosphere | Phyllosphere contains different groups of phototrophic organisms that are phylogenetically diverse | Roche 454 | Atamna-Ismaeel *et al*., 2012 |
| 4 | Barley | Rhizosphere | Mineral phosphate solubilization genes were identified | Roche 454 | Chhabra *et al*., 2013 |
| 5 | Lotus | Rhizosphere | No major microbial communities changes with respect to phytic acid utilization. Phytic acid utilization genes were identified | Roche 454 | Unno and Shinano, 2013 |
| 6 | Tomato | Phyllosphere | Different organs of plants have distinct microbial communities | Roche 454 GS Titanium FLX | Ottesen *et al*., 2013 |
| 7 | Soybean | Rhizosphere | Plants prefer a specific microbial community beneficial to its growth and function | Roche 454 GS-FLX | Mendes *et al*., 2014 |
| 8 | *Aloe vera* | Rhizosphere | *Firmicutes*, *Actinobacteria* and *Bacteriodetes*, four prominent phyla were identified as beneficial for bioactive compound production | Illumina MiSeq | Akinsanya *et al*., 2015 |
| 9 | *Genlisea species* | Phyllosphere | Complex food interaction between plant *Genlisea* and its entrapped microbiome | Illumina Hi Seq 2000 | Cao *et al*., 2015 |

of lethal necrosis in Kenyan maize using NGS (Wangai *et al.,* 2016). A detailed review describing the various roles of NGS in viral diagnostics has already been published by Boonham *et al.* (2014).

NGS technology has also been used successfully to determine the begomoviral genome and their associated satellites from begomovirus-affected tomato and okra plants by using metagenomics (Idris *et al.,* 2014). In another study, NGS and metagenomics approaches together were used to study native bacterial microflora diversity in different anatomical organs of *Solanum lycopersicum* (Ottesen *et al.,* 2013). In tomato, the root microbiome was studied in order to understand the endophytes and their association with root-knot nematodes (*Meloidogyne* spp.) (Tian *et al.,* 2015). Similarly in rice, a metagenomic rice endophyte DNA library was constructed and further 16S rRNA gene sequence information was studied (Sessitsch *et al.,* 2012). Similarly, metagenomic analysis of olive-knot to decipher the role of different bacterial species in the disease establishment has been reported (Passos da Silva *et al.,* 2014). Even the role of the enzymatic repertoire (glycoside hydrolases) of a microbial community in degradation of crude cotton biomass was reported using NGS-based metagenomics approach (Zhang *et al.,* 2016).

## NGS applications in exome and captured sequencing

Selection of genomic regions of interest and enrichment of these regions is called captured sequencing. This technology is a revolutionary process for the selective enrichment of targeted genomic regions from the complex genomic DNA. Targeted-capture is used to enrich the sequences of interest before carrying out NGS, such as repetitive sequences (Syring *et al.,* 2016), exome (Neves *et al.,* 2013), and gene space regions (Zhou and Holliday, 2012). This technology permits the isolation of target loci from the background of the entire genome. In comparison with other sequencing technologies, captured sequencing is inexpensive, quick and simple. The scale of capture can range from several targeted loci to over a million target regions (Agilent 2011; Microarray 2011; NimbleGen 2011), making it adaptable for both small-scale and large scale projects. This technology holds promise for plant genomes because they are large, complex

and contain a large amount of repetitive elements. In the present review, we are focusing on captured sequencing and exome sequencing. Exome are comprised of those sequences in DNA that code for a protein. It may also include the functional non-protein-coding sequences such as micro-RNAs, long intergenic non-coding RNA, etc. as well as specific candidate regions. The technique to select and enrich the exomes followed by their sequencing using NGS technology is called exome sequencing (Warr *et al.*, 2015). Initially, PCR was widely used to capture the sequence (Mamanova *et al.*, 2010), followed by circularization of capture sequences using a suitable circularization technique such as molecular inversion probes (Nilsson *et al*, 1994) and spacer multiplex amplification reaction (Krishnakumar *et al.*, 2008), and hybridisation capture methods such as array-based hybrid selection (Albert *et al.*, 2007; Okou *et al.*, 2007) and solution hybridization capture (Gnirke *et al.*, 2009).

## Captured and exome sequencing studies across different crops

Application of this approach has gained momentum in plant systems recently. Targeted sequencing has been done in several crop and tree species using the captured method of sequencing (Table 1.6). Captured sequencing has been used to identify DNA polymorphisms in many polyploidy crop species. It is a powerful tool for studying evolution, population genetics and phylogeographic studies (Carstens *et al.*, 2013; Smith *et al.*, 2014; McCormack *et al.*, 2015). Various targeted studies conducted using NGS techniques have been detailed in Table 1.6.

Exome sequencing is most widely used targeted sequencing method that aims only an informative subset of the genome which varied between 1% and 2% of the genome. Saintenac *et al.* (2011) captured 3.5 Mb of the exonic sequence with coverage of 3.5–7.0% of the exome and studied 3497 genes in durum wheat accessions. In the subsequent year, a novel exome capturing protocol for wheat based on a NimbleGen array was developed by Winfield *et al.* (2012). This protocol was used to sequence exomes of the eight wheat accessions by capturing a 56.5 Mb genomic region to identify SNPs for the genotypic classification of a segregating locus in polyploid wheat (Allen *et al.*, 2013). Other studies in rice and wheat include screening of novel mutations of rice and durum wheat, targeted capture of 107 Mb of non-redundant regions in 62 lines of wheat, exome capture for rapid cloning of *R*-genes in hexaploid bread wheat and exome sequencing in the genomes of rice somaclonal variant (salt-tolerant and drought-tolerant) and parental cultivar (Henry *et al.*, 2014; Udomchalothorn *et al.*, 2014; Jordan *et al.*, 2015; King *et al.*, 2015; Steuernagel *et al.*, 2016).

In maize, exome capture was performed to understand endosperm filling and maturation and to create a population of 1788 lines (Jia *et al.*, 2016). Exome capture kit has also been developed for barley (*Hordeum vulgare* L.) to selectively enrich 61.6 Mb of protein coding sequence (Mascher *et al.*, 2013) which was used for sequencing of exome X-ray mutagenized mutants and wild type genotypes to identify a candidate gene HvMND belonging to the CYP78A family that may affect

**Table 1.6** Plant genome capturing studies performed using NGS platforms

| Number | Plant | Genome coverage | Method | Reference |
|---|---|---|---|---|
| 1 | *Triticum aestivum* | 56.5 Mb | Nimblegen Array technology | Winfield *et al*., 2012 |
| 2 | *Populus trichocarpa* (black cottonwood) | 20.76 Mb | Agilent technologies | Zhou *et al*., 2012 |
| 3 | *Saccharum officinarum* | 5.8 Mb | SureSelect Target Enrichment System | Bundock *et al*., 2012 |
| 4 | *Saccharum hybrid* | 5.8 Mb | SureSelect Target Enrichment System | Bundock *et al*., 2012 |
| 5 | *Pinus taeda* L. | 21.7 Gbp | Probe based and hybridization capture | Neves *et al*., 2013 |
| 6 | *Fragaria vesca* | 100× | Mycroarray, Mybait and Illumina | Tennessen *et al*., 2013 |
| 7 | *Brassica napus* | 5.8 Mb | SureSelectXT | Schiessl *et al*., 2014 |
| 8 | *Pinus albicaulis* (whitebark pine) | 27 Gb | Hybridization-based target capture | Syring *et al*., 2016 |

many agricultural traits (Mascher *et al.*, 2014). Kono *et al.* (2016) discovered the distribution of hundreds of SNPs in cultivated and wild accessions of barley and soybean using exome sequencing. *Hordium bulbosum* is the wild relative of cultivated barley that has superior pathogen resistance and stress tolerance which can be crossed to cultivated barley genotype. There was lack of suitable molecular tools to characterize the genetic introgressions from *H. bulbosum* in order to select the beneficial variants and exclude the variants that were not important from breeding point of view. Recently many exome capture studies were reported in this crop for the development of genic markers and genome introgression studies (Wendler *et al.*, 2014, 2015). Russell *et al.* (2016) studied the environmental adaptation in the georeferenced landraces and wild accessions by exome capture.

Besides, exome capturing in switchgrass *(Panicum virgatum*), a potential biofuel feedstock crop, allowed assessment of the genome variation in its two primary ecotypes and identification of variation in *CONSTANS (CO)* and *EARLY HEADING DATE 1 (EHD1)* genes (Evans *et al.*, 2014, 2015). Exome sequencing has also been reported in Eucalypts, black spruce (*Picea mariana*) and black cottonwood (*Populus trichocarpa*) (Zhou and Holliday, 2012; Dasgupta *et al.*, 2015; Pavy *et al.*, 2016) primarily to study the genetic variation.

## Next-genomics sequencing and molecular markers

The discovery and application of molecular markers for the detection and exploitation of DNA polymorphism is one of the most significant achievements in the area of molecular genetics and plant breeding. Applicability of these markers depends on various factors, viz. its physical properties and genomic location, the cost involved, ease of use, and degree of throughput required ( Jonah *et al.*, 2011). Molecular markers are generally categorized into two groups, macro-molecules (proteins and deoxyribonucleic acid) and biochemical constituents (secondary metabolites in plants). Secondary metabolites are restricted to the plants and its applicability is not as wide as DNA makers ( Joshi *et al.*, 2011). Among the macro-molecules, availability of protein markers is very limited and its analysis is difficult and more tedious than DNA markers. Among all the molecular markers, DNA markers are the most commonly used markers in the field of molecular genetics and plant breeding for various purposes. Single nucleotide polymorphisms (SNPs) have been recognized as potential markers of choice for genome-wide studies due to even distribution throughout the genome (like simple sequence repeats, SSRs), having the advantage over SSRs of being easily typed in large numbers (in high-throughput manner), and signifying variation in both coding and noncoding regions of the genome (Altshuler *et al.*, 2000; Brumfield *et al.*, 2003; Slate *et al.*, 2009). With the advancement of genome sequencing technologies, molecular markers with a known genomic location are becoming more useful and applicable. The existence of various kinds of molecular markers, and differences in their principles, methodological adaptability, and application's suitability need cautious consideration in opting for one or more of such methods for crop improvement programmes. However, these markers are generally based on electrophoretic resolution of DNA fragments, which limits capturing of genetic differences and also this method cannot resolve genetic polymorphisms with less than 5 bp differences (Semagn *et al.*, 2006). Genotyping of considerably large plant populations may take longer duration depending on how to do, what kind of maker system to adopt and how much throughput the adopted system could generate. NGS technology fulfils all the demands of the coming age plant breeding experiments. It is an efficient technology to develop low cost, high-throughput molecular markers for genotyping of such a large plant population in a short period. Using the NGS technologies, several molecular markers were developed to decipher the complex sequences at thousand loci in the genome of all the individuals of a large plant population sample. These NGS technologies include reduced-representation libraries (RRLs; Gore *et al.*, 2009; Hyten *et al.*, 2010), complexity reduction of polymorphic sequences (CRoPS; Mammadov *et al.*, 2010), low coverage multiplexed shotgun genotyping (MSG; Andolfatto *et al.*, 2011), restriction-site associated DNA sequencing (RAD-seq; Pfender *et al.*, 2011), genotyping by sequencing (GBS; Elshire *et al.*, 2011), high-density array (HDR) genotyping (Gunderson, 2009), and sequence-based polymorphic (SBP) marker technology (Sahu *et al.*, 2012).

All of these methods comprise the following basic steps: the digestion of multiple samples of genomic DNA extracted from individuals or set of populations with one or more restriction enzymes; a selection or reduction of the resulting restriction fragments; and NGS of the final set of selected fragments, which should be less than 1 kb in size (avoiding the read-length limits of most of current NGS platforms, except PacBio); bioinformatic analysis to study the association between traits and called variants; and ultimately infer the biological importance from the analysed dataset. Variations in the resulting sequenced fragments can be used as molecular markers in crop breeding programmes (Davey *et al*., 2011). Though NGS-based markers have enhanced crop breeding programmes, still there are challenges in high-throughput marker generation. These are posed in the form of a way to design the experiment, how many individuals to be screened, which NGS platform and method will be well-suited to minimize the per-sample sequencing cost, etc.

## Major applications of high-throughput markers in crop plants

A wide range of applications and methodologies of genetic markers has been reported in various crop plants (Semagn *et al*., 2006). Molecular plant breeding aims to improve crop variety in terms of its quantity and quality by applying the latest inventions made in the fields of genetics and genomics. Our understanding about the relationship between genotype and phenotype has been continuously increasing with the help of advanced genomics tools. Some of these applications include (i) surveying allelic diversity in breeding material or natural populations to select the desired genotypes; (ii) marker-assisted selection (MAS) strategies for variety development and germplasm improvement; and (iii) gene pyramiding for gathering multiple agronomically desirable genes within the same cultivar (Jain *et al*., 2002; Gupta and Varshney, 2004). Of these several applications, one of the major applications of markers are identification of DNA sequences associated with desired traits in crop breeding. This type of application has been described in several crop breeding programmes and greatly benefited the plant breeders in the easy selection of genotypes where phenotypic expressions become difficult to detect individually with utmost breeding value. These expressions may be hindered by many factors such as tissue and age of plant, environmental conditions, expression observing methods and time-frame of expression. These constraints can be easily avoided by the application of molecular markers in selection of genotypes with particular trait(s). Some major applications of high-throughput marker technologies are given with separate subheadings.

### Resequencing, genotyping and diversity analysis

High-throughput genotyping derived from NGS is one of the major applications of molecular markers. Approximately seven million plant accessions, including wild relatives, landraces and human-made advanced varieties/cultivars have been preserved in several, around 1750 national and international, gene banks worldwide (FAO, 2010) and it is a well-known fact that the whole world is dependent on these plants for food, fibre and fuel. Therefore, it is very important for biologists to characterize these accessions and make them available for further crop improvement programmes. In this order, McCouch *et al*. (2012) presented a vision for the potential of genotyping at large-scales such as gene bank collections. In this vision, authors also outlined the constraints in genotyping work at the gene bank level and suggested that applications of NGS may solve many problems related to genetic characterization efforts in gene banks. These major challenges include the need to correctly identify accessions and eliminate duplicate accessions from gene bank collections. Such characterization work has begun with rice genomics, whereby 3000 rice genotypes have been characterized, including identification of SNPs and other structural variations of the genome (3K RGP, 2014). However, almost every county has legal provision to protect the country's genetic resources by not moving or transferring any such materials without approval of the competent institutions or genetic resource management governing body of the country.

### Linkage and association mapping

With the use of NGS and genotyping technologies, it is possible to develop high-throughput molecular markers as well as assume genotyping at large scale in both major and minor genes that can be exercised

for generating high-resolution genetic and physical maps. Meanwhile, these approaches can also be exploited to identify genetic variation in germplasm collections of cultivars, landraces and wild species. This genetic variation can be introgressed in elite cultivar or genotype of interest through linkage and association mapping approaches. Moreover, the major genes or superior alleles of the genes or QTLs for the desirable traits can also be identified and introgressed or pyramided in elite cultivars or genotypes of interest using advanced plant breeding approaches such as marker-assisted back crossing (MABC), marker-assisted recurrent selection (MARS), advanced-backcross QTL (AB-QTL), multi-parent advanced generation intercross (MAGIC), or genome-wide selection (GWS) (Varshney and Dubey, 2009).

### Phylogenetic and evolutionary studies

Before the development of NGS technologies, large-scale genome-wide studies were restricted to a few model organisms whose genomes were sequenced. During that time whole genome sequencing was very time-consuming and a laborious project and it was done with collaboration of many countries. Now, the scenario is totally different and even a single laboratory may afford the whole sequencing work that is only possible by the advancement in sequencing technologies. Discovery of molecular markers is directly related to the NGS techniques and the role of markers in the phylogenetic and evolutionary studies is well known. For good quality of phylogeny as well as evolutionary studies, ecologists and evolutionary biologists need data from large numbers of individuals. These studies along with the power of NGS technology have been reported in many non-model organisms for determining the gene flow, population divergence, diversity level at intra- or inter-population, phylogeography, domestication process, and phylogenetic and evolutionary analyses (Grover *et al.*, 2012).

## Conclusion and future prospects of next-generation sequencing

Next-generation sequencing technologies have revolutionized the field of plant genomics. As of today, the large chunk of this success is largely attributed to the second-generation sequencing tools. The areas which were significantly influenced by NGS technologies include genome sequencing, captured sequencing, exome sequencing, metagenomics, and plant transcriptomics; which includes mRNAs and non-coding RNAs and molecular markers and plant breeding. The continuous improvements in the field of sequencing technologies and associated decrease in sequencing costs has opened the door to small laboratories to take up plant genome sequencing projects. In spite of these developments, still there are considerable challenges in sequencing more complex genomes such as cross-pollinated crops and crop with higher polyploids. Nevertheless, researchers have taken interest in sequencing particular genomic regions and exomes, called as captured sequencing. This approach has considerably advanced our knowledge of those crops for which there is no sequence information available or which are only partially sequenced. NGS also has a significant role in metagenomics and enhanced the rate of analysis, especially in those circumstances where it would have been difficult to analyse using traditional tools and approaches. NGS-based transcriptome analysis has contributed significantly in transcriptome profiling of those plant species where no sequence information is available and also improve their ability to immensely contribute for novel genes to plant biology. NGS is playing a vital role in development of high throughput molecular markers as well.

However, the benefits of developments in the field of NGS tools can only be harvested by integrating the different genomics and genetics technologies and also advancements in biometrics and bioinformatics tools and techniques. This is going to be a challenging task in coming years. Finally, though captured sequencing has gained importance in past few years, largely due to sequencing costs and complexity involved in it, the future holds bright for whole genome sequencing, as the cost of sequencing is expected to decrease further and also the developments in the third-generation sequencing techniques have the potential to decrease the complexity by easily assembling of more complex genomes.

## Acknowledgements

TRS is thankful to the Indian council of Agricultural Research for funding under NPTC project and Department of Science and Technology Government of India for JC Bose National Fellowship.

## References


3K RGP (2014). The 3,000 rice genomes project. Giga Science *3*, 7.

Abdelfattah, A., Wisniewski, M., Nicosia, M.G.L.D., Cacciola, S.O., and Schena, L. (2016). Metagenomic analysis of fungal diversity on strawberry plants and the effect of management practices on the fungal community structure of aerial organs. PLOS ONE, *11*, e0160470.

Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., and Reddy, A.S. (2016). A survey of the sorghum transcriptome using single-molecule long reads. Nat. Commun. *7*, 11706. http://dx.doi.org/10.1038/ncomms11706

Adams, I.P., Miano, D.W., Kinyua, Z.M., Wangai, A., Kimani, E., Phiri, N., Reeder, R., Harju, V., Glover, R., Hany, U., *et al.* (2013). Use of next-generation sequencing for the identification and characterization of *Maize chlorotic mottle virus* and *Sugarcane mosaic virus* causing maize lethal necrosis in Kenya. Plant Pathol., 62(4), 741–749.

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., and Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol. Plant Pathol. *10*, 537–545. http://dx.doi.org/10.1111/j.1364-3703.2009.00545.x

Ai, C., Liang, G., Sun, J., Wang, X., He, P., Zhou, W., and He, X. (2015). Reduced dependence of rhizosphere microbiome on plant-derived carbon in 32-year long-term inorganic and organic fertilized soils. Soil Biol. Biochem. *80*, 70–78.

Akinsanya, M.A., Goh, J.K., Lim, S.P., and Ting, A.S. (2015). Metagenomics study of endophytic bacteria in Aloe vera using next-generation technology. Genom. Data *6*, 159–163. http://dx.doi.org/10.1016/j.gdata.2015.09.004

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., *et al.* (2007). Direct selection of human genomic loci by microarray hybridization. Nat. Methods *4*, 903–905.

Albert, V.A., Barbazuk, W.B., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., Soltis, D.E., *et al.* (2013). The *Amborella* genome and the evolution of flowering plants. Science *342*,1241089.

Allen, A.M., Barker, G.L., Wilkinson, P., Burridge, A., Winfield, M., Coghill, J., Uauy, C., Griffiths, S., Jack, P., Berry, S., *et al.* (2013). Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotechnol. J. *11*, 279–295.

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., and Lander, E.S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature *407*, 513–516. http://dx.doi.org/10.1038/35035083

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. *21*, 610–617. http://dx.doi.org/10.1101/gr.115402.110

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. New Biotechnol. *25*, 195–203. http://dx.doi.org/10.1016/j.nbt.2008.12.009

*Arabidopsis* Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796–815

Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F., Spudich, J.L., von Mering, C., Vorholt, J.A., Iluz, D., *et al.* (2012b). Microbial rhodopsins on leaf surfaces of terrestrial plants. Environ. Microbiol. *14*, 140–146. http://dx.doi.org/10.1111/j.1462-2920.2011.02554.x

Bagnaresi, P., Biselli, C., Orrù, L., Urso, S., Crispino, L., Abbruscato, P., Piffanelli, P., Lupotto, E., Cattivelli, L., and Valè, G. (2012). Comparative transcriptome profiling of the early response to *Magnaporthe oryzae* in durable resistant vs susceptible rice (*Oryza sativa* L.) genotypes. PLOS ONE *7*, e51609. http://dx.doi.org/10.1371/journal.pone.0051609

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J., L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, Naoki., Aoyama, T., Ambrose, B.A., Ashton, N.W., *et al.* (2011). The compact *Selaginella* genome identifies changes in gene content associated with the evolution of vascular plants. Science *332*, 960–963.

Barrera-Figueroa, B.E., Gao, L., Diop, N.N., Wu, Z., Ehlers, J.D., Roberts, P.A., Close, T.J., Zhu, J.K., and Liu, R. (2011). Identification and comparative analysis of drought-associated microRNAs in two cowpea genotypes. BMC Plant Biol. *11*, 127. http://dx.doi.org/10.1186/1471-2229-11-127

Bedada, G., Westerbergh, A., Muller, T., Galkin, E., Bdolach, E., Moshelion, M., Fridman, E., and Schmid, K.J. (2014). Transcriptome sequencing of two wild barley (*Hordeum spontaneum* L.) ecotypes differentially adapted to drought stress reveals ecotype-specific transcripts. BMC genomics *15*, 995. http://dx.doi.org/10.1186/1471-2164-15-995.

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., *et al.* (2012). Reference genome sequence of the model plant Setaria. Nat. Biotechnol. *30*, 555–561. http://dx.doi.org/10.1038/nbt.2196

Berendsen, R.L., Pieterse, C.M., and Bakker, P.A. (2012). The rhizosphere microbiome and plant health. Trends Plant Sci. *17*, 478–486. http://dx.doi.org/10.1016/j.tplants.2012.04.001

Bertioli, D.J., Cannon, S.B., Froenicke, L., Huang, G., Farmer, A.D., Cannon, E.K., Liu, X., Gao, D., Clevenger, J., Dash, S., *et al.* (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat. Genet. *48*, 438–446. http://dx.doi.org/10.1038/ng.3517

Bian, X., Ma, P., Jia, Z., Guo, X., and Xie, Y. (2016). Identification of miRNAs in sweet potato by Solexa sequencing. Russ. J. Plant Physiol. *63*, 283–292.

Birol, I., Raymond, A., Jackman, S.D., Pleasance, S., Coope, R., Taylor, G.A., Yuen, M.M., Keeling, C.I., Brand, D., Vandervalk, B.P., *et al.* (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics *29*, 1492–1497. http://dx.doi.org/10.1093/bioinformatics/btt178

Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C.S., Bliek, M., Boersma, M.R., Borghi, L., Bruggmann, R., *et al.* (2016). Insight into the evolution

of the Solanaceae from the parental genomes of *Petunia hybrida*. Nat. Plants *2*, 16074. http://dx.doi.org/10.1038/nplants.2016.74

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J., and Mumford, R. (2014). Methods in virus diagnostics: from ELISA to next generation sequencing. Virus Res *186*, 20–31. http://dx.doi.org/10.1016/j.virusres.2013.12.007

Brumfield, R.T., Beerli, P., Nickerson, D.A., and Edwards, S.V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. *18*, 249–256. http://dx.doi.org/10.1016/S0169-5347(03)00018-1

Bundock, P.C., Casu, R.E., and Henry, R.J. (2012). Enrichment of genomic DNA for polymorphism detection in a non-model highly polyploid crop plant. Plant Biotechnol. J. *10*, 657–667. http://dx.doi.org/10.1111/j.1467-7652.2012.00707.x

Burns, J.A., Paasch, A., Narechania, A., and Kim, E. (2015). Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. Genome Biol. Evol. *7*, 3047–3061. http://dx.doi.org/10.1093/gbe/evv144

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125. http://dx.doi.org/10.1038/nbt.2727

Butts, C.T., Bierma, J.C., and Martin, R.W. (2016). Novel proteases from the genome of the carnivorous plant *Drosera capensis*: Structural prediction and comparative analysis. Proteins *84*, 1517–33.

Byrne, S.L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., Mayer, K., Campbell, J.D., Czaban, A., Hentrup, S., *et al.* (2015). A synteny-based draft genome sequence of the forage grass *Lolium perenne*. Plant J. *84*, 816–826. http://dx.doi.org/10.1111/tpj.13037

Camilios-Neto, D., Bonato, P., Wassem, R., Tadra-Sfeir, M.Z., Brusamarello-Santos, L.C., Valdameri, G., Donatti, L., Faoro, H., Weiss, V.A., Chubatsu, L.S., *et al.* (2014). Dual RNA-seq transcriptional analysis of wheat roots colonized by *Azospirillumbrasilense* reveals up-regulation of nutrient acquisition and cell cycle genes. BMC genomics, *15*, 378.

Campo, S., Peris-Peris, C., Siré, C., Moreno, A.B., Donaire, L., Zytnicki, M., Notredame, C., Llave, C., and San Segundo, B. (2013). Identification of a novel microRNA (miRNA) from rice that targets an alternatively spliced transcript of the *Nramp6* (Natural resistance-associated macrophage protein 6) gene involved in pathogen resistance. New Phytol. *199*, 212–227.

Cao, H.X., Schmutzer, T., Scholz, U., Pecinka, A., Schubert, I., and Vu, G.T. (2015). Metatranscriptome analysis reveals host-microbiome interactions in traps of carnivorous Genlisea species. Front. Microbiol. *6*, 526. http://dx.doi.org/10.3389/fmicb.2015.00526

Cao, H., Hastie, A.R., Cao, D., Lam, E.T., Sun, Y., Huang, H., Liu, X., Lin, L., Andrews, W., Chan, S., *et al.* (2014). Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. GigaScience *3*, 34. http://dx.doi.org/10.1186/2047-217X-3-34

Cao, J.Y., Xu, Y.P., Zhao, L., Li, S.S., and Cai, X.Z. (2016). Tight regulation of the interaction between *Brassica napus* and *Sclerotinia sclerotiorum* at the microRNA level. Plant Mol. Biol. *92*, 39–55. http://dx.doi.org/10.1007/s11103-016-0494-3

Cao, S., Zhu, Q.H., Shen, W., Jiao, X., Zhao, X., Wang, M.B., Liu, L., Singh, S.P., and Liu, Q. (2013). Comparative profiling of miRNA expression in developing seeds of high linoleic and high oleic safflower (*Carthamus tinctorius* L.) plants. Front. Plant Sci. *4*, 489. http://dx.doi.org/10.3389/fpls.2013.00489

Carstens, B.C., Brennan, R.S., Chua, V., Duffie, C.V., Harvey, M.G., Koch, R.A., McMahan, C.D., Nelson, B.J., Newman, C.E., Satler, J.D., *et al.* (2013). Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. Mol. Ecol. *22*, 4014–4028. http://dx.doi.org/10.1111/mec.12347

Chao, J., Jin, J., Wang, D., Han, R., Zhu, R., Zhu, Y., and Li, S. (2014). Cytological and transcriptional dynamics analysis of host plant revealed stage-specific biological processes related to compatible rice-*Ustilaginoidea virens* interaction. PLOS ONE, *9*(3), p.e91391. http://dx.doi.org/10.1371/journal.pone.0091391.

Chaudhry, V., and Patil, P.B. (2016). Genomic investigation reveals evolution and lifestyle adaptation of endophytic Staphylococcus epidermidis. Scientific reports, *6*.

Chen, K., and Pachter, L. (2005). Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput. Biol. *1*, 106–112. http://dx.doi.org/10.1371/journal.pcbi.0010024

Chen, X. (2009). Small RNAs and their roles in plant development. Annu. Rev. Cell Dev. Biol. *25*, 21–44. http://dx.doi.org/10.1146/annurev.cellbio.042308.113417

Chhabra, S., Brazil, D., Morrissey, J., Burke, J.I., O Gara, F., N Dowling, D. (2013). Characterization of mineral phosphate solubilization traits from a barley rhizosphere soil functional metagenome. Microbiologyopen *2*, 717–724. http://dx.doi.org/10.1002/mbo3.110

Chi, X., Yang, Q., Chen, X., Wang, J., Pan, L., Chen, M., Yang, Z., He, Y., Liang, X., and Yu, S. (2011). Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. PLoS One *6*, e27530. http://dx.doi.org/10.1371/journal.pone.0027530

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O'Toole, P.W. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic acids research gkq873.

Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., *et al.* (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat. Methods *5*, 613–619. http://dx.doi.org/10.1038/nmeth.1223

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635–641. http://dx.doi.org/10.1038/nature11119

Coughlan, L.M., Cotter, P.D., Hill, C., and Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional

metagenomics in the food and pharmaceutical industries. Frontiers in microbiology 6.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C., and Bik, H.M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution. http://dx.doi.org/10.1111/2041-210X.12574

Creighton, C.J., Reid, J.G., and Gunaratne, P.H. (2009). Expression profiling of microRNAs by deep sequencing. Briefings Bioinf. *10*, 490–497. http://dx.doi.org/10.1093/bib/bbp019

Crespi, M.D., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., and Kondorosi, A. (1994). *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J. *13*, 5099–5112.

Dasgupta, M.G., Dharanishanthi, V., Agarwal, I., and Krutovsky, K.V. (2015). Development of genetic markers in *Eucalyptus* species by target enrichment and exome sequencing. PLOS ONE *10*, e0116528. http://dx.doi.org/10.1371/journal.pone.0116528

Dautt-Castro, M., Ochoa-Leyva, A., Contreras-Vergara, C.A., Pacheco-Sanchez, M.A., Casas-Flores, S., Sanchez-Flores, A., Kuhn, D.N., and Islas-Osuna, M.A. (2015). Mango (*Mangifera indica* L.) cv. Kent fruit mesocarp *de novo* transcriptome assembly identifies gene families important for ripening. Frontiers in plant science, 6, 62. http://dx.doi.org/10.3389/fpls.2015.00062.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. *12*, 499–510. http://dx.doi.org/10.1038/nrg3012

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C., and Vorholt, J.A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc. Natl. Acad. Sci. U.S.A. *106*, 16428–16433. http://dx.doi.org/10.1073/pnas.0905240106

De Vega, J.J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J.L., Ergon, Å., Rognli, O.A., Jones, C., Swain, M., Geurts, *et al.* (2015). Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. Sci. Rep. *5*, 17394. http://dx.doi.org/10.1038/srep17394

Ding, T., and Melcher, U. (2016). Influences of Plant species, season and location on leaf endophytic bacterial communities of non-cultivated plants. PLOS ONE *11*, e0150895. http://dx.doi.org/10.1371/journal.pone.0150895

Dorn, K.M., Fankhauser, J.D., Wyse, D.L., and Marks, M.D. (2015). A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. DNA Res. *22*, 121–131. http://dx.doi.org/10.1093/dnares/dsu045

Duangjit, J., Bohanec, B., Chan, A.P., Town, C.D., and Havey, M.J. (2013). Transcriptome sequencing to produce SNP-based genetic maps of onion. Theor. Appl. Genet. *126*, 2093–2101. http://dx.doi.org/10.1007/s00122-013-2121-x

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLOS ONE *6*, e19379. http://dx.doi.org/10.1371/journal.pone.0019379

Evans, J., Crisovan, E., Barry, K., Daum, C., Jenkins, J., Kunde-Ramamoorthy, G., Nandety, A., Ngan, C.Y., Vaillancourt, B., Wei, C.L., *et al.* (2015). Diversity and population structure of northern switchgrass as revealed through exome capture sequencing. Plant J. *84*, 800–815. http://dx.doi.org/10.1111/tpj.13041

Evans, J., Kim, J., Childs, K.L., Vaillancourt, B.,Crisovan, E., Nandety, A., Gerhardt, D.J., Richmond, T.A., Jeddeloh, J.A., Kaeppler, S.M., *et al.* (2014). Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass *Panicum virgatum*. Plant J. *79*, 993–1008.

Fan, J., Ning, K., Zeng, X., Luo, Y., Wang, D., Hu, J., Li, J., Xu, H., Huang, J., Wan, M., *et al.* (2015). Genomic foundation of starch-to-lipid switch in Oleaginous *Chlorella* spp. Plant. Physiol. *169*, 2444–2461. http://dx.doi.org/10.1104/pp.15.01174

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., and Jabado, O.J., *et al.* (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nature biotechnology *30*, 1232–1239.

FAO. (2010). The second report on the state of the world's plant genetic resources for food and agriculture. Food and Agriculture Organization, Rome, Italy.

Feng, J., Liu, S., Wang, M., Lang, Q., and Jin, C. (2014). Identification of microRNAs and their targets in tomato infected with *Cucumber mosaic virus* based on deep sequencing. Planta *240*, 1335–1352. http://dx.doi.org/10.1007/s00425-014-2158-3

Feng, J., Wang, J., Fan, P., Jia, W., Nie, L., Jiang, P., Chen, X., Lv, S., Wan, L., Chang, S., *et al.* (2015). High-throughput deep sequencing reveals that microRNAs play important roles in salt tolerance of euhalophyte *Salicornia europaea*. BMC Plant Biol. *15*, 63. http://dx.doi.org/10.1186/s12870-015-0451-3

Flórez-Zapata, M.V.N., Reyes-Valdés, M.H., and Martínez, O. (2016). Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates. BMC Genomics *17*, 490 http://dx.doi.org/10.1186/s12864-016-2776-1

França, L.T., Carrilho, E., and Kist, T.B. (2002). A review of DNA sequencing techniques. Q. Rev. Biophys. *35*, 169–200.

Fu, Y., Xiao, M., Yu, H., Mason, A.S., Yin, J., Li, J., Zhang, D., and Fu, D. (2016). Small RNA changes in synthetic *Brassica napus*. Planta *244*, 607–622. http://dx.doi.org/10.1007/s00425-016-2529-z

Gao, R., Liu, P., Irwanto, N., Loh, R., and Wong, S.M. (2016). Upregulation of LINC-AP2 is negatively correlated with AP2 gene expression with Turnip crinkle virus infection in *Arabidopsis thaliana*. Plant Cell Rep. *35*, 2257–2267. http://dx.doi.org/10.1007/s00299-016-2032-9

Gardner, E.M., Johnson, M.G., Ragone, D., Wickett, N.J., and Zerega, N.J. (2016). Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Appl. Plant Sci. *4*, apps.1600017.

Garg, R., Shankar, R., Thakkar, B., Kudapa, H., Krishnamurthy, L., Mantri, N., Varshney, R.K., Bhatia,

S., and Jain, M. (2016). Transcriptome analyses reveal genotype-and developmental stage-specific molecular responses to drought and salinity stresses in chickpea. Scientific Reports, *6*. http://dx.doi.org/10.1038/srep19228.

Gentile, A., Ferreira, T.H., Mattos, R.S., Dias, L.I., Hoshino, A.A., Carneiro, M.S., Souza, G.M., Calsa, T., Nogueira, R.M., Endres, L., *et al*. (2013). Effects of drought on the microtranscriptome of field-grown sugarcane plants. Planta *237*, 783–798. http://dx.doi.org/10.1007/s00425-012-1795-7

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., *et al*. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. *27*, 182–189. http://dx.doi.org/10.1038/nbt.1523

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al*. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science *296*, 92–100. http://dx.doi.org/10.1126/science.1068275

Gore, M.A., Chia, J.M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., *et al*. (2009). A first-generation haplotype map of maize. Science *326*, 1115–1117. http://dx.doi.org/10.1126/science.1177837

Grover, C.E., Salmon, A., and Wendel, J.F. (2012). Targeted sequence capture as a powerful tool for evolutionary analysis. Am. J. Bot. *99*, 312–319. http://dx.doi.org/10.3732/ajb.1100323

Gunderson, K.L. (2009). Whole-genome genotyping on bead arrays. Methods Mol. Biol. *529*, 197–213. http://dx.doi.org/10.1007/978-1-59745-538-1_13

Guo, L., Qiu, J., Han, Z., Ye, Z., Chen, C., Liu, C., Xin, X., Ye, C.Y., Wang, Y.Y., Xie, H., *et al.* (2015). A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. Plant J. *83*, 600–609. http://dx.doi.org/10.1111/tpj.12912

Guo, W., Wu, G., Yan, F., Lu, Y., Zheng, H., Lin, L., Chen, H., and Chen, J. (2012). Identification of novel *Oryza sativa* miRNAs in deep sequencing-based small RNA libraries of rice infected with *Rice stripe virus*. PLOS ONE *7*, e46443. http://dx.doi.org/10.1371/journal.pone.0046443

Gupta, P.K., and Varshney, R.K. (2004). Cereal genomics: an overview. In P.K. Gupta, and R.K. Varshney, eds. (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 639–643.

Gupta, V., Estrada, A.D., Blakley, I., Reid, R., Patel, K., Meyer, M.D., Andersen, S.U., Brown, A.F., Lila, M.A., and Loraine, A.E. (2015). RNASeq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. GigaScience *4*, 5.

Hajyzadeh, M., Turktas, M., Khawar, K.M., and Unver, T. (2015). miR408 overexpression causes increased drought tolerance in chickpea. Gene *555*, 186–193. http://dx.doi.org/10.1016/j.gene.2014.11.002

Hanschen, E.R., Marriage, T.N., Ferris, P.J., Hamaji, T., Toyoda, A., Fujiyama, A., Neme, R., Noguchi, H., Minakuchi, Y., Suzuki, M., *et al.* (2016). The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. Nat. Commun. *7*, 11370. http://dx.doi.org/10.1038/ncomms11370

Harris, T.D., Buzby, P.R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., Dimeo, J., Efcavitch, J.W., *et al*. (2008). Single-molecule DNA sequencing of a viral genome. Science *320*, 106–109. http://dx.doi.org/10.1126/science.1150427

He, Y., Xiao, H., Deng, C., Xiong, L., Nie, H., and Peng, C. (2016). Survey of the genome of Pogostemon cablin provides insights into its evolutionary history and sesquiterpenoid biosynthesis. Sci. Rep. *6*, 26405. http://dx.doi.org/10.1038/srep26405

He, Y., Yang, Z., Hong, N., Wang, G., Ning, G., and Xu, W. (2015). Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. Mol. Plant Pathol. *16*, 449–458. http://dx.doi.org/10.1111/mpp.12202

Heisel, S.E., Zhang, Y., Allen, E., Guo, L., Reynolds, T.L., Yang, X., Kovalic, D., and Roberts, J.K. (2008). Characterization of unique small RNA populations from rice grain. PLOS ONE 3(8), e2871.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., *et al.* (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell *26*, 1382–1397.

Heo, J.B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science *331*, 76–79. http://dx.doi.org/10.1126/science.1197349

Hirakawa, H., Okada, Y., Tabuchi, H., Shirasawa, K., Watanabe, A., Tsuruoka, H., Minami, C., Nakayama, S., Sasamoto, S., Kohara, M., *et al*. (2015). Survey of genome sequences in a wild sweet potato, *Ipomoea trifida* (H. B. K.) G. Don. DNA Res. 171–9. http://dx.doi.org/10.1093/dnares/dsv002

Hirakawa, H., Kaur, P., Shirasawa, K., Nichols, P., Nagano, S., Appels, R., Erskine, W., and Isobe, S.N. (2016). Draft genome sequence of subterranean clover, a reference for genus *Trifolium*. Sci. Rep. *6*, 30358. http://dx.doi.org/10.1038/srep30358

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature *456*, 130–134. http://dx.doi.org/10.1038/nature07348

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. Plant Physiol. *151*, 2120–2132. http://dx.doi.org/10.1104/pp.109.147280

Huang, D., Koh, C., Feurtado, J.A., Tsang, E.W., and Cutler, A.J. (2013). MicroRNAs and their putative targets in *Brassica napus* seed maturation. BMC Genomics *14*, 140. http://dx.doi.org/10.1186/1471-2164-14-140

Hussain, S., Yin, H., Peng, S., Khan, F.A., Khan, F., Sameeullah, M., Hussain, H.A., Huang, J., Cui, K., and Nie, L. (2016). Comparative transcriptional profiling of primed and non-primed rice seedlings under

submergence Stress. Front. Plant Sci. 7, 1125. http://dx.doi.org/10.3389/fpls.2016.01125.

Hyten, D.L., Cannon, S.B., Song, Q., Weeks, N., Fickus, E.W., Shoemaker, R.C., Specht, J.E., Farmer, A.D., May, G.D., and Cregan, P.B. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics *11*, 38. http://dx.doi.org/10.1186/1471-2164-11-38

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.H., Lan, T., Welch, A.J., Juárez, M.J., Simpson, J., *et al.* (2013). Architecture and evolution of a minute plant genome. Nature *498*, 94–98. http://dx.doi.org/10.1038/nature12132

Idris, A., Al-Saleh, M., Piatek, M.J., Al-Shahwan, I., Ali, S., and Brown, J.K. (2014). Viral metagenomics: analysis of begomoviruses by illumina high-throughput sequencing. Viruses, *6*(3), 1219–1236.

Imelfort, M., and Edwards, D. (2009). *De novo* sequencing of plant genomes using second-generation technologies. Briefings Bioinf. *10*, 609–618. http://dx.doi.org/10.1093/bib/bbp039

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., Cavagnaro, P., *et al.* (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. Nat. Genet. *48*, 657–666. http://dx.doi.org/10.1038/ng.3565

Iorizzo, M., Senalik, D.A., Grzebelus, D., Bowman, M., Cavagnaro, P.F., Matvienko, M., Ashrafi, H., Van Deynze, A., and Simon, P.W. (2011). De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. BMC Genomics *12*, 389. http://dx.doi.org/10.1186/1471-2164-12-389

Jagadeeswaran, G., Zheng, Y., Li, Y.F., Shukla, L.I., Matts, J., Hoyt, P., Macmil, S.L., Wiley, G.B., Roe, B.A., Zhang, W., *et al.* (2009). Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families. New Phytol. *184*, 85–98. http://dx.doi.org/10.1111/j.1469-8137.2009.02915.x

Jain, M. (2012). Next-generation sequencing technologies for gene expression profiling in plants. Brief. Funct. Genomics. *11*, 63–70. http://dx.doi.org/10.1093/bfgp/elr038

Jain, M., Chevala, V.V., and Garg, R. (2014). Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. J. Exp. Bot. *65*, 5945–5958. http://dx.doi.org/10.1093/jxb/eru333

Jain, S.M., Brar, D.S., and Ahloowalia, B.S. (2002). Molecular techniques in crop improvement. (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp 433–444.

Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmöckel, S.M., Li, B., Borm, T.J., Ohyanagi, H., Mineta, K., Michell, C.T., Saber, N., *et al.* (2016). The genome of *Chenopodium quinoa*. Nature *542*, 307–312.

Jia, S., Li, A., Morton, K., Avoles-Kianian, P., Kianian, S.F., Zhang, C., and Holding, D. (2016). A population of deletion mutants and an integrated mapping and exome-seq pipeline for gene discovery in maize. G3 *6*, 2385–2395. http://dx.doi.org/10.1534/g3.116.030528

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., *et al.* (2011). Ancestral polyploidy in seed plants and angiosperms. Nature. *473*, 97–100.

Jonah, P.M., Bello, L.L., Lucky, O., Midau, A., and Moruppa, S.M. (2011). Review: the importance of molecular markers in plant breeding programmes. GJSFR 11, 5–12.

Jordan, K.W., Wang, S., Lun, Y., Gardiner, L.J., MacLachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K.L., Sharpe, A.G., *et al.* (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol. *16*, 48. http://dx.doi.org/10.1186/s13059-015-0606-4

Joshi, R.K., Megha, S., Basu, U., Rahman, M.H., and Kav, N.V. (2016). Genome wide identification and functional prediction of long non-coding RNAs responsive to *Sclerotinia sclerotiorum* infection in *Brassica napus*. PLOS ONE *11*. http://dx.doi.org/10.1371/journal.pone.0158784.

Joshi, S.P., Prabhakar, K., Ranjekar, P.K., and Gupta, V.S. (1999). Molecular markers in plant genome analysis. http://www.ias.ac.in/currsci/jul25/articles15.htm. pp 1–19.

Kang, Y.J., Satyawan, D., Shim, S., Lee, T., Lee, J., Hwang, W.J., Kim, S.K., Lestari, P., Laosatit, K., Kim, K.H., *et al.* (2015). Draft genome sequence of adzuki bean, *Vigna angularis*. Sci. Rep. *5*, 8069. http://dx.doi.org/10.1038/srep08069

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., and Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLOS ONE *7*, e49423. http://dx.doi.org/10.1371/journal.pone.0049423

Kehoe, M.A., Coutts, B.A., Buirchell, B.J., and Jones, R.A. (2014). Plant virology and next generation sequencing: experiences with a Potyvirus. PLOS ONE *9*, e104580. http://dx.doi.org/10.1371/journal.pone.0104580

Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., *et al.* (2015). Genome-guided investigation of plant natural product biosynthesis. Plant J. *82*, 680–692. http://dx.doi.org/10.1111/tpj.12827

Kim, E.D., and Sung, S. (2012). Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends Plant Sci. *17*, 16–21. http://dx.doi.org/10.1016/j.tplants.2011.10.008

King, R., Bird, N., Ramirez-Gonzalez, R., Coghill, J.A., Patil, A., Hassani-Pak, K., Uauy, C., and Phillips, A.L. (2015). Mutation scanning in wheat by exon capture and next-generation sequencing. PLOS ONE *10*, e0137549. http://dx.doi.org/10.1371/journal.pone.0137549

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., and Vorholt, J.A. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. ISME J. *6*, 1378–1390. http://dx.doi.org/10.1038/ismej.2011.192

Kompelli, S.K., Kompelli, V.S.P., Enjala, C., and Suravajhala, P. (2015.) Genome-wide identification of miRNAs in pigeonpea (*Cajanus cajan* L.). Aust. J. Crop Sci *9*, 215–222.

Kono, T.J., Fu, F., Mohammadi, M., Hoffman, P.J., Liu, C., Stupar, R.M., Smith, K.P., Tiffin, P., Fay, J.C., and Morrell, P.L. (2016). The role of deleterious substitutions in crop

genomes. Mol. Biol. Evol. *33*, 2307–2317. http://dx.doi.org/10.1093/molbev/msw102

Krishnakumar, S., Zheng, J., Wilhelmy, J., Faham, M., Mindrinos, M., and Davis, R. (2008). A comprehensive assay for targeted multiplex amplification of human DNA sequences. Proc. Natl. Acad. Sci. U.S.A. *105*, 9296–9301. http://dx.doi.org/10.1073/pnas.0803240105

Lakhotia, N., Joshi, G., Bhardwaj, A.R., Katiyar-Agarwal, S., Agarwal, M., Jagannath, A., Goel, S., and Kumar, A. (2014). Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (*Solanum tuberosum* L.) by high-throughput sequencing. BMC Plant Biol. *14*, 6. http://dx.doi.org/10.1186/1471-2229-14-6

Lee, H., Golicz, A.A., Bayer, P.E., Jiao, Y., Tang, H., Paterson, A.H., Sablok, G., Krishnaraj, R.R., Chan, C.K., Batley, J., *et al.* (2016). The genome of a southern hemisphere seagrass species (*Zostera muelleri*). Plant Physiol. *172*, 272–83.

Lee, H., Gurtowski, j., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., McCombie, W.R., and Schatz, M. (2016). Third-generation sequencing and the future of genomics. http://dx.doi.org/10.1101/048603

Leushkin, E.V., Sutormin, R.A., Nabieva, E.R., Penin, A.A., Kondrashov, A.S., and Logacheva, M.D. (2013). The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short noncoding sequences. BMC Genomics *14*, 476.17.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. Science *299*, 682–686. http://dx.doi.org/10.1126/science.1079700

Li, C., Qiao, Z., Qi, W., Wang, Q., Yuan, Y., Yang, X., Tang, Y., Mei, B., Lv, Y., Zhao, H., *et al.* (2015). Genome-wide characterization of cis-acting DNA targets reveals the transcriptional regulatory framework of *opaque2* in maize. Plant. Cell *27*, 532–545. http://dx.doi.org/10.1105/tpc.114.134858

Li, F., Fan, G., Wang, K., Sun, F., Yuan, Y., Song, G., Li, Q., Ma, Z., Lu, C., Zou, C., *et al.* (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. Nat. Genet. *46*, 567–572. http://dx.doi.org/10.1038/ng.2987

Li, H., Dong, Y., Sun, Y., Zhu, E., Yang, J., Liu, X., Xue, P., Xiao, Y., Yang, S., Wu, J., *et al.* (2011a). Investigation of the microRNAs in safflower seed, leaf, and petal by high-throughput sequencing. Planta *233*, 611–619. http://dx.doi.org/10.1007/s00425-010-1327-2

Li, H., Dong, Y., Yin, H., Wang, N., Yang, J., Liu, X., Wang, Y., Wu, J., and Li, X. (2011b). Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing. BMC Plant Biol. *11*, 170. http://dx.doi.org/10.1186/1471-2229-11-170

Li, J., Ding, Q., Wang, F., Zhang, Y., Li, H., and Gao, J. (2015a). Integrative analysis of mRNA and miRNA expression profiles of the tuberous root development at seedling stages in turnips. PLOS ONE *10*, e0137983. http://dx.doi.org/10.1371/journal.pone.0137983

Li, J., Wu, L.Q., Zheng, W.Y., Wang, R.F., and Yang, L.X. (2015b). Genome-wide identification of microRNAs responsive to high temperature in rice (*Oryza sativa*) by high-throughput deep sequencing. J. Agron. Crop Sci. *201*, 379–388.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., *et al.* (2010). The sequence and *de novo* assembly of the giant panda genome. Nature *463*, 311–317. http://dx.doi.org/10.1038/nature08696

Li, X., Jin, F., Jin, L., Jackson, A., Ma, X., Shu, X., Wu, D., and Jin, G. (2015). Characterization and comparative profiling of the small RNA transcriptomes in two phases of flowering in *Cymbidium ensifolium*. BMC Genomics *16*, 622. http://dx.doi.org/10.1186/s12864-015-1764-1

Li, X., Sun, H., Pei, J., Dong, Y., Wang, F., Chen, H., Sun, Y., Wang, N., Li, H., and Li, Y. (2012). *De novo* sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. Gene *511*, 54–61. http://dx.doi.org/10.1016/j.gene.2012.09.021

Li, Y., Lu, Y.G., Shi, Y., Wu, L., Xu, Y.J., Huang, F., Guo, X.Y., Zhang, Y., Fan, J., Zhao, J.Q., *et al.* (2014). Multiple rice microRNAs are involved in immunity against the blast fungus *Magnaporthe oryzae*. Plant Physiol. *164*, 1077–1092. http://dx.doi.org/10.1104/pp.113.230052

Li, Y.F., Zheng, Y., Jagadeeswaran, G., and Sunkar, R. (2013). Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. Plant Sci. *203*, 17–24.

Lipshitz, H.D., Peattie, D.A., and Hogness, D.S. (1987). Novel transcripts from the Ultrabithorax domain of the bithorax complex. Genes. Dev. *1*, 307–322.

Liu, H., Qin, C., Chen, Z., Zuo, T., Yang, X., Zhou, H., Xu, M., Cao, S., Shen, Y., Lin, H., *et al.* (2014). Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. BMC Genomics *15*, 25. http://dx.doi.org/10.1186/1471-2164-15-25

Liu, Q., Lan, Y., Wen, C., Zhao, H., Wang, J., and Wang, Y. (2016a). Transcriptome sequencing analyses between the cytoplasmic male sterile line and its maintainer line in Welsh onion (*Allium fistulosum* L.). Int. J. Mol. Sci. 17(7). http://dx.doi.org/10.3390/ijms17071058

Liu, S.C., Xu, Y.X., Ma, J.Q., Wang, W.W., Chen, W., Huang, D.J., Fang, J., Li, X.J., and Chen, L. (2016b). Small RNA and degradome profiling reveals important roles for microRNAs and their targets in tea plant response to drought stress. Physiol. Plant http://dx.doi.org/10.1111/ppl.12477.

Liu, W., Xu, L., Wang, Y., Shen, H., Zhu, X., Zhang, K., Chen, Y., Yu, R., Limera, C., and Liu, L. (2015). Transcriptome-wide analysis of chromium-stress responsive microRNAs to explore miRNA-mediated regulatory networks in radish (*Raphanus sativus* L.). Sci. Rep. *11*, 5. http://dx.doi.org/10.1038/srep14024.

Liu, X., Zhao, B., Zheng, H.J., Hu, Y., Lu, G., Yang, C.Q., Chen, J.D., Chen, J.J., Chen, D.Y., Zhang, L., *et al.* (2015). *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci. Rep. *5*, 14139. http://dx.doi.org/10.1038/srep14139

Loss-Morais, G., Ferreira, D.C., Margis, R., Alves-Ferreira, M., and Corrêa, R.L. (2014). Identification of novel and conserved microRNAs in *Coffea canephora* and *Coffea arabica*. Genet. Mol. Biol. *37*, 671–682. http://dx.doi.org/10.1590/S1415-47572014005000020

Lu, M., An, H., and Li, L. (2016). Genome survey sequencing for the characterization of the genetic background of *Rosa roxburghii* tratt and leaf ascorbate metabolism genes. PLOS ONE *11*, e0147530. http://dx.doi.org/10.1371/journal.pone.0147530

Lu, X., Chen, X., Mu, M., Wang, J., Wang, X., Wang, D., Yin, Z., Fan, W., Wang, S., Guo, L., *et al.* (2016). Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.). PLOS ONE *11*, e0156723. http://dx.doi.org/10.1371/journal.pone.0156723

Luan, Y., Cui, J., Zhai, J., Li, J., Han, L., and Meng, J. (2015). High-throughput sequencing reveals differential expression of miRNAs in tomato inoculated with *Phytophthora infestans*. Planta *241*, 1405–1416.

Lugtenberg, B., and Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. Annu. Rev. Microbiol. *63*, 541–556. http://dx.doi.org/10.1146/annurev.micro.62.081307.162918

Luo, B., Gu, W., Zhong, J., Wang, Y., and Zhang, G. (2015). Revealing crosstalk of plant and fungi in the symbiotic roots of sewage-cleaning *Eichhornia crassipes* using direct *de novo* metatranscriptomic analysis. Scientific Reports, *5*. http://dx.doi.org/10.1038/srep15407.

Luria, N., Sela, N., Yaari, M., Feygenberg, O., Kobiler, I., Lers, A., and Prusky, D. (2014). De-novo assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment. BMC Genomics *15*, 957. http://dx.doi.org/10.1186/1471-2164-15-957

Ma, X., Xin, Z., Wang, Z., Yang, Q., Guo, S., Guo, X., Cao, L., and Lin, T. (2015). Identification and comparative analysis of differentially expressed miRNAs in leaves of two wheat (*Triticum aestivum* L.) genotypes during dehydration stress. BMC Plant Biol. *15*, 1. http://dx.doi.org/10.1186/s12870-015-0413-9.

Mahuku, G., Wangai, A., Sadessa, K., Teklewold, A., Wegary, D., Adams, I., Smith, J., Bottomley, E., Bryce, S., Braidwood, L., *et al.* (2015). First report of *Maize chlorotic mottle virus* and Maize lethal necrosis on maize in Ethiopia. *Plant Disease*, *99*(12), 1870.

Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., and Turner, D.J. (2010). Target-enrichment strategies for next-generation sequencing. Nat. Methods *7*, 111–118. http://dx.doi.org/10.1038/nmeth.1419

Mammadov, J.A., Chen, W., Ren, R., Pai, R., Marchione, W., Yalçin, F., Witsenboer, H., Greene, T.W., Thompson, S.A., and Kumpatla, S.P. (2010). Development of highly polymorphic SNP markers from the complexity reduced portion of maize [*Zea mays* L.] genome for use in marker-assisted breeding. Theor. Appl. Genet. *121*, 577–588. http://dx.doi.org/10.1007/s00122-010-1331-8

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends Genet. *24*, 133–141. http://dx.doi.org/10.1016/j.tig.2007.12.007

Mardis, E.R. (2013). Next-generation sequencing platforms. Annu. Rev. Anal. Chem. *6*, 287–303. http://dx.doi.org/10.1146/annurev-anchem-062012-092628

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol. Cell *29*, 499–509. http://dx.doi.org/10.1016/j.molcel.2007.12.013

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.

Martínez-García, P.J., Crepeau, M.W., Puiu, D., Gonzalez-Ibeas, D., Whalen, J., Stevens, K.A., Paul, R., Butterfield, T.S., Britton, M.T., Reagan, R.,L., *et al.* (2016). The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis of non-structural polyphenols. Plant J. *87*, 507–32.

Marzano, S.Y.L., andDomier, L.L. (2016). Reprint of 'Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllospherephytobiomes'. Virus Res. *219*, 11–21.

Mascher, M., Jost, M., Kuon, J.E., Himmelbach, A., Aßfalg, A., Beier, S., Scholz, U., Graner, A., and Stein, N. (2014). Mapping-by-sequencing accelerates forward genetics in barley. Genome Biol. *15*, R78. http://dx.doi.org/10.1186/gb-2014-15-6-r78

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., *et al.* (2013). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. Plant J. *76*, 494–505. http://dx.doi.org/10.1111/tpj.12294

Maxam, A.M., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. *74*, 560–564.

McCormack, J.E., Tsai, W.L., and Faircloth, B.C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour. *16*, 1189–1203. http://dx.doi.org/10.1111/1755-0998.12466

McCouch, S.R., McNally, K.L., Wang, W., and Sackville Hamilton, R. (2012). Genomics of gene banks: A case study in rice. Am. J. Bot. *99*, 407–423. http://dx.doi.org/10.3732/ajb.1100385

Mendes, L.W., Kuramae, E.E., Navarrete, A.A., van Veen, J.A., and Tsai, S.M. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. ISME J. *8*, 1577–1587. http://dx.doi.org/10.1038/ismej.2014.17

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., *et al.* (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science *318*, 245–250.

Michael, T.P., and VanBuren, R. (2015). Progress, challenges and the future of crop genomes. Curr. Opin. Plant Biol. *24*, 71–81. http://dx.doi.org/10.1016/j.pbi.2015.02.002

Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., Lyons, E., Wang, M.L., Chen, J., Biggers, E., *et al.* (2015). The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. *47*, 1435–1442. http://dx.doi.org/10.1038/ng.3435

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian

transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628. http://dx.doi.org/10.1038/nmeth.1226

Moxon, S., Jing, R., Szittya, G., Schwach, F., Rusholme Pilcher, R.L., Moulton, V., and Dalmay, T. (2008). Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. *18*, 1602–1609. http://dx.doi.org/10.1101/gr.080127.108

Neale, D.B., Wegrzyn, J.L., Stevens, K.A., Zimin, A.V., Puiu, D., Crepeau, M.W., Cardeno, C., Koriabine, M., Holtz-Morris, A.E., Liechty, J.D., *et al.* (2014). Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biol. *15*, R59. http://dx.doi.org/10.1186/gb-2014-15-3-r59

Neves, L.G., Davis, J.M., Barbazuk, W.B., and Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. Plant. J. *75*, 146–156. http://dx.doi.org/10.1111/tpj.12193

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322–325. http://dx.doi.org/10.1038/35104581

Nie, S., Li, C., Xu, L., Wang, Y., Huang, D., Muleke, E.M., Sun, X., Xie, Y., and Liu, L. (2016). *De novo* transcriptome analysis in radish (*Raphanus sativus* L.) and identification of critical genes involved in bolting and flowering. BMC Genomics *17*, 389. http://dx.doi.org/10.1186/s12864-016-2633-2

Nie, S., Xu, L., Wang, Y., Huang, D., Muleke, E.M., Sun, X., Wang, R., Xie, Y., Gong, Y., and Liu, L. (2015). Identification of bolting-related microRNAs and their targets reveals complex miRNA-mediated flowering-time regulatory networks in radish (*Raphanus sativus* L.). Sci. Rep. 5, http://dx.doi.org/10.1038/srep14034.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P., and Landegren, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. Science *265*, 2085–2088.

Nowak, M.D., Russo, G., Schlapbach, R., Huu, C.N., Lenhard, M., and Conti, E. (2015). The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. Genome Biol. *16*, 12. http://dx.doi.org/10.1186/s13059-014-0567-z

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., *et al.* (2013). The Norway spruce genome sequence and conifer genome evolution. Nature *497*, 579–584. http://dx.doi.org/10.1038/nature12211

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., and Zwick, M.E. (2007). Microarray-based genomic selection for high-throughput resequencing. Nat. Methods *4*, 907–909.

Olsen, J.L., Rouzé, P., Verhelst, B., Lin, Y.C., Bayer, T., Collen, J., Dattolo, E., De Paoli, E., Dittami, S., Maumus, *et al.* (2016). The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature *530*, 331–335. http://dx.doi.org/10.1038/nature16548

Ota, S., Oshima, K., Yamazaki, T., Kim, S., Yu, Z., Yoshihara, M., Takeda, K., Takeshita, T., Hirata, A., Bišová, K., *et al.* (2016). Highly efficient lipid production in the green alga *Parachlorella kessleri*: draft genome and transcriptome endorsed by whole-cell 3D ultrastructure. Biotechnol. Biofuels *9*, 13.

Ottesen, A.R., González Peña, A., White, J.R., Pettengill, J.B., Li, C., Allard, S., Rideout, S., Allard, M., Hill, T., Evans, P., *et al.* (2013b). Baseline survey of the anatomical microbial ecology of an important food plant: *Solanum lycopersicum* (tomato). BMC Microbiol. *13*, 114. http://dx.doi.org/10.1186/1471-2180-13-114

Pankin, A., Campoli, C., Dong, X., Kilian, B., Sharma, R., Himmelbach, A., Saini, R., Davis, S.J., Stein, N., Schneeberger, K. *et al.* (2014). Mapping-by-sequencing identifies HvPHYTOCHROME C as a candidate gene for the early maturity 5 locus modulating the circadian clock and photoperiodic flowering in barley. Genetics *198*, 383–396.

Passos da Silva, D., Castañeda-Ojeda, M.P., Moretti, C., Buonaurio, R., Ramos, C., and Venturi, V. (2014). Bacterial multispecies studies and microbiome analysis of a plant disease. Microbiology *160*, 556–566. http://dx.doi.org/10.1099/mic.0.074468-0

Pavy, N., Gagnon, F., Deschênes, A., Boyle, B., Beaulieu, J., and Bousquet, J. (2016). Development of highly reliable in silico SNP resource and genotyping assay from exome capture and sequencing: an example from black spruce (*Picea mariana*). Mol. Ecol. Resour. *16*, 588–598. http://dx.doi.org/10.1111/1755-0998.12468

Pendleton, M., Sebra, R., Pang, A.W., Ummat, A., Franzen, O., Rausch, T., Stütz, A.M., Stedman, W., Anantharaman, T., Hastie, A., *et al.* (2015). Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat. Methods *12*, 780–786. http://dx.doi.org/10.1038/nmeth.3454

Pfender, W.F., Saha, M.C., Johnson, E.A., and Slabaugh, M.B. (2011). Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. Theor. Appl. Genet. *122*, 1467–1480. http://dx.doi.org/10.1007/s00122-011-1546-3

Pingault, L., Choulet, F., Alberti, A., Glover, N., Wincker, P., Feuillet, C., and Paux, E. (2015). Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. Genome Biol. *16*, 29. http://dx.doi.org/10.1186/s13059-015-0601-9

Pirrò, S., Zanella, L., Kenzo, M., Montesano, C., Minutolo, A., Potestà, M., Sobze, M.S., Canini, A., Cirilli, M., Muleo, R., *et al.* (2016). MicroRNA from *Moringa oleifera*: identification by high throughput sequencing and their potential contribution to plant medicinal value. PLOS ONE *11*, e0149495. http://dx.doi.org/10.1371/journal.pone.0149495

Porras-Alfaro, A., and Bayman, P. (2011). Hidden fungi, emergent properties: endophytes and microbiomes. Annu. Rev. Phytopathol. *49*, 291–315. http://dx.doi.org/10.1146/annurev-phyto-080508-081831

Poudel, S., Aryal, N., and Lu, C. (2015). Identification of microRNAs and transcript targets in *Camelina sativa* by deep sequencing and computational methods. PLOS ONE *10*, e0121542. http://dx.doi.org/10.1371/journal.pone.0121542

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., *et al.* (2010). Genomic analysis of organismal complexity in the multicellular green alga

*Volvox carteri*. Science *329*, 223–226. http://dx.doi.org/10.1126/science.1188800

Qiao, Y., Zhang, J., Zhang, J., Wang, Z., Ran, A., Guo, H., Wang, D., and Zhang, J. (2016). Integrated RNA-seq and sRNA-seq analysis reveals miRNA effects on secondary metabolism in *Solanum tuberosum* L. Mol. Genet. Genomics 1–16, http://dx.doi.org/10.1007/s00438-016-1253-5

Quan, M., Chen, J., and Zhang, D. (2015). Exploring the secrets of long noncoding RNAs. Int. J. Mol. Sci. *16*, 5467–5496. http://dx.doi.org/10.3390/ijms16035467

Que, Y., Su, Y., Guo, J., Wu, Q., and Xu, L., 2014. A global view of transcriptome dynamics during *Sporisorium scitamineum* challenge in sugarcane by RNA-Seq. PLOS ONE, 9(8), p.e106476, http://dx.doi.org/10.1371/journal.pone.0106476.

Rahman, H., Jagadeeshselvam, N., Valarmathi, R., Sachin, B., Sasikala, R., Senthil, N., Sudhakar, D., Robin, S., and Muthurajan, R. (2014). Transcriptome analysis of salinity responsiveness in contrasting genotypes of finger millet (*Eleusine coracana* L.) through RNA-sequencing. Plant molecular biology *85*, 485–503.

Rastogi, S., Kalra, A., Gupta, V., Khan, F., Lal, R.K., Tripathi, A.K., Parameswaran, S., Gopalakrishnan, C., Ramaswamy, G., and Shasany, A.K. (2015). Unravelling the genome of Holy basil: an 'incomparable' 'elixir of life' of traditional Indian medicine. BMC Genomics *16*, 413. http://dx.doi.org/10.1186/s12864-015-1640-z

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., *et al.* (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science *319*, 64–69.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166. http://dx.doi.org/10.1146/annurev-biochem-051410-092902

Roberts, R.J., Carneiro, M.O., and Schatz, M.C. (2013). The advantages of SMRT sequencing. Genome Biol. *14*, 405. http://dx.doi.org/10.1186/gb-2013-14-6-405

Ross, M.G., Russ, C., Costello, M., Hollinger, A., Lennon, N.J., Hegarty, R., Nusbaum, C., and Jaffe, D.B. (2013). Characterizing and measuring bias in sequence data. Genome Biol. *14*, R51. http://dx.doi.org/10.1186/gb-2013-14-5-r51

Rowland, L.J., Alkharouf, N., Darwish, O., Ogden, E.L., Polashock, J.J., Bassil, N.V., and Main, D. (2012). Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. BMC plant biology, *12*, 1. http://dx.doi.org/10.1186/1471-2229-12-46.

Ruiz-Ferrer, V., and Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. Annu. Rev. Plant Biol. *60*, 485–510. http://dx.doi.org/10.1146/annurev.arplant.043008.092111

Russell, J., Mascher, M., Dawson, I.K., Kyriakidis, S., Calixto, C., Freund, F., Bayer, M., Milne, I., Marshall-Griffiths, T., Heinen, S., *et al.* (2016). Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. Nature Genet. *48*, 1024–1030. http://dx.doi.org/10.1038/ng.3612.

Sahu, B.B., Sumit, R., Srivastava, S.K., and Bhattacharyya, M.K. (2012). Sequence based polymorphic (SBP) marker technology for targeted genomic regions: its application in generating a molecular map of the *Arabidopsis thaliana* genome. BMC Genomics *13*, 20. http://dx.doi.org/10.1186/1471-2164-13-20

Saintenac, C., Jiang, D., and Akhunov, E.D. (2011). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol. *12*, R88. http://dx.doi.org/10.1186/gb-2011-12-9-r88

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. *74*, 5463–5467.

Scaglione, D., Reyes-Chin-Wo, S., Acquadro, A., Froenicke, L., Portis, E., Beitel, C., Tirone, M., Mauro, R., Lo Monaco, A., Mauromicale, G., *et al.* (2016). The genome sequence of the outbreeding globe artichoke constructed *de novo* incorporating a phase-aware low-pass sequencing strategy of F1 progeny. Sci. Rep. *6*, 19427.

Schatz, M.C., Delcher, A.L., and Salzberg, S.L. (2010). Assembly of large genomes using second-generation sequencing. Genome Res. *20*, 1165–1173. http://dx.doi.org/10.1101/gr.101360.109

Schiessl, S., Samans, B., Hüttel, B., Reinhard, R., and Snowdon, R.J. (2014). Capturing sequence variation among flowering-time regulatory gene homologs in the allopolyploid crop species Brassica napus. Front. Plant Sci. *5*, 404. http://dx.doi.org/10.3389/fpls.2014.00404

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., *et al.* (2010). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178–183. http://dx.doi.org/10.1038/nature08670

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., *et al.* (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112–1115. http://dx.doi.org/10.1126/science.1178534

Schreiber, A.W., Hayden, M.J., Forrest, K.L., Kong, S.L., Langridge, P., and Baumann, U. (2012). Transcriptome-scale homoeolog-specific transcript assemblies of bread wheat. BMC Genomics *13*, 492. http://dx.doi.org/10.1186/1471-2164-13-492

Semagn, K., Bjornstad, A., and Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. Afr. J. Biotech. *5*, 2540–2568.

Sessitsch, A., Hardoim,P., Döring,J., Weilharter, A., Krause, A. *et al.* (2012). Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. Mol. Plant Microbe Interact., http://dx.doi.org/10.1094/MPMI-08-11-0204

Severin, A.J., Woody, J.L., Bolon, Y.T., Joseph, B., Diers, B.W., Farmer, A.D., Muehlbauer, G.J., Nelson, R.T., Grant, D., Specht, J.E., *et al.* (2010). RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol. *10*, 160. http://dx.doi.org/10.1186/1471-2229-10-160

Sheibani-Tezerji, R., Rattei, T., Sessitsch, A., Trognitz, F., and Mitter, B. (2015). Transcriptome profiling of the endophyte *Burkholderia phytofirmans* PsJN indicates sensing of the plant environment and drought stress.

mBio *6*, e00621–15. http://dx.doi.org/10.1128/mBio.00621-15

Shen, D., Suhrkamp, I., Wang, Y., Liu, S., Menkhaus, J., Verreet, J.A., Fan, L., and Cai, D. (2014). Identification and characterization of microRNAs in oilseed rape (*Brassica napus*) responsive to infection with the pathogenic fungus *Verticillium longisporum* using Brassica AA (*Brassica rapa*) and CC (*Brassica oleracea*) as reference genomes. New Phytol. *204*, 577–594.

Shen, E., Zou, J., Hubertus Behrens, F., Chen, L., Ye, C., Dai, S., Li, R., Ni, M., Jiang, X., Qiu, J., *et al.* (2015). Identification, evolution, and expression partitioning of miRNAs in allopolyploid *Brassica napus*. J. Exp. Bot. *66*, 7241–7253. http://dx.doi.org/10.1093/jxb/erv420

Singh, D., Singh, P.K., Chaudhary, S., Mehla, K., and Kumar, S. (2012). Exome sequencing and advances in crop improvement. Adv. Genet. *79*, 87–121. http://dx.doi.org/10.1016/B978-0-12-394395-8.00003-7

Singh, V.K., and Jain, M. (2014). Transcriptome profiling for discovery of genes involved in shoot apical meristem and flower development. Genom. Data *2*, 135–138. http://dx.doi.org/10.1016/j.gdata.2014.06.004

Slate, J., Gratten, J., Beraldi, D., Stapley, J., Hale, M., and Pemberton, J.M. (2009). Gene mapping in the wild with SNPs: guidelines and future directions. Genetica *136*, 97–107. http://dx.doi.org/10.1007/s10709-008-9317-z

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., and Brumfield, R.T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic biology *63*, 83–95.

Song, Q.X., Liu, Y.F., Hu, X.Y., Zhang, W.K., Ma, B., Chen, S.Y., and Zhang, J.S. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol. *11*, 5. http://dx.doi.org/10.1186/1471-2229-11-5

Srivastava, S., Singh, R.K., Pathak, G., Goel, R., Asif, M.H., Sane, A.P., and Sane, V.A. (2016). Comparative transcriptome analysis of unripe and mid-ripe fruit of *Mangifera indica* (var. 'Dashehari') unravels ripening associated genes. Scientific Reports, *6*, http://dx.doi.org/10.1038/srep32557.

Srivastava, S., Zheng, Y., Kudapa, H., Jagadeeswaran, G., Hivrale, V., Varshney, R.K., and Sunkar, R. (2015). High throughput sequencing of small RNA component of leaves and inflorescence revealed conserved and novel miRNAs as well as phasiRNA loci in chickpea. Plant. Sci. *235*, 46–57. http://dx.doi.org/10.1016/j.plantsci.2015.03.002

Staley, J.T., and Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. *39*, 321–346. http://dx.doi.org/10.1146/annurev.mi.39.100185.001541

Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D., *et al.* (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. Nat. Biotechnol. *34*, 652–655. http://dx.doi.org/10.1038/nbt.3543

Sun, X., Xu, L., Wang, Y., Luo, X., Zhu, X., Kinuthia, K.B., Nie, S., Feng, H., Li, C., and Liu, L. (2016). Transcriptome-based gene expression profiling identifies differentially expressed genes critical for salt stress response in radish (*Raphanus sativus* L.). Plant cell reports, 35(2), 329–346.

Sun, X., Xu, L., Wang, Y., Yu, R., Zhu, X., Luo, X., Gong, Y., Wang, R., Limera, C., Zhang, K. *et al.* (2015). Identification of novel and salt-responsive miRNAs to explore miRNA-mediated regulatory network of salt stress response in radish (*Raphanus sativus* L.). BMC genomics 16(1), 1. http://dx.doi.org/10.1186/s12864-015-1416-5

Syring, J.V., Tennessen, J.A., Jennings, T.N., Wegrzyn, J., Scelfo-Dalbey, C., and Cronn, R. (2016). Targeted capture sequencing in whitebark pine reveals range-wide demographic and adaptive patterns despite challenges of a large, repetitive genome. Front. Plant Sci. *7*, 484.

Tanaka, H., Hirakawa, H., Kosugi, S., Nakayama, S., Ono, A., Watanabe, A., Hashiguchi, M., Gondo, T., Ishigaki, G., Muguerza, M., *et al.* (2016). Sequencing and comparative analyses of the genomes of zoysiagrasses. DNA Res. *23*, 171–180. http://dx.doi.org/10.1093/dnares/dsw006

Tanase, K., Nishitani, C., Hirakawa, H., Isobe, S., Tabata, S., Ohmiya, A., and Onozaki, T. (2012). Transcriptome analysis of carnation (*Dianthus caryophyllus* L.) based on next-generation sequencing technology. BMC Genomics *13*, 292. http://dx.doi.org/10.1186/1471-2164-13-292

Tang, W., Zheng, Y., Dong, J., Yu, J., Yue, J., Liu, F., Guo, X., Huang, S., Wisniewski, M., Sun, J., *et al.* (2016). Comprehensive transcriptome profiling reveals long noncoding RNA expression and alternative splicing regulation during fruit development and ripening in kiwifruit (*Actinidia chinensis*). Front. Plant Sci. *7*, 335, http://dx.doi.org/10.3389/fpls.2016.00335.

Tang, Z., Zhang, L., Xu, C., Yuan, S., Zhang, F., Zheng, Y., and Zhao, C. (2012). Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. Plant Physiol. *159*, 721–738. http://dx.doi.org/10.1104/pp.112.196048

Tennessen, J.A., Govindarajulu, R., Liston, A., and Ashman, T.L. (2013). Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). G3 3, 1341–1351.

Tian, B.Y., Cao, Y., and Zhang, K.Q. (2015). Metagenomic insights into communities, functions of endophytes, and their associates with infection by root-knot nematode, *Meloidogyne incognita*, in tomato roots. Scientific Reports, 5.

Tian, Y., Tian, Y., Luo, X., Zhou, T., Huang, Z., Liu, Y., Qiu, Y., Hou, B., Sun, D., Deng, H., *et al.* (2014). Identification and characterization of microRNAs related to salt stress in broccoli, using high-throughput sequencing and bioinformatics analysis. BMC Plant Biol. *14*, 226. http://dx.doi.org/10.1186/s12870-014-0226-2

Tian, Y., Zeng, Y., Zhang, J., Yang, C., Yan, L., Wang, X., Shi, C., Xie, J., Dai, T., Peng, L., *et al.* (2015). High quality reference genome of drumstick tree (*Moringa oleifera* Lam.), a potential perennial crop. Sci. China Life Sci. *58*, 627–638. http://dx.doi.org/10.1007/s11427-015-4872-x

Tombuloglu, G., Tombuloglu, H., Sakcali, M.S., and Unver, T. (2015). High-throughput transcriptome analysis of barley (*Hordeum vulgare*) exposed to excessive boron. Gene *557*, 71–81. http://dx.doi.org/10.1016/j.gene.2014.12.012

Turktas, M., Kurtoglu, K.Y., Dorado, G., Zhang, B., and Hernandez, P. (2015). Sequencing of plant genomes – a review. Turk J. Agric. For. *39*, 361–376. http://dx.doi.org/10.3906/tar-1409–93

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., *et al.* (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science *313*, 1596–1604.

Udomchalothorn, T., Plaimas, K., Comai, L., Buaboocha, T., and Chadchawan,S. (2014). Molecular karyotyping and exome analysis of salt-tolerant rice mutant from somaclonal variation. The Plant Genome 7, http://dx.doi.org/10.3835/plantgenome2014.04.0016.

Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46. http://dx.doi.org/10.1016/j.cell.2013.06.020

Unno, Y., and Shinano, T. (2013). Metagenomic analysis of the rhizosphere soil microbiome with respect to phytic acid utilization. Microbes. Environ. *28*, 120–127.

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., and Schuster, S.C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. PLOS ONE 3, e2527. http://dx.doi.org/10.1371/journal.pone.0002527

VanBuren, R., Bryant, D., Bushakra, J.M., Vining, K.J., Edger, P.P., Rowley, E.R., Priest, H.D., Michael, T.P., Lyons, E., Filichkin, S.A., *et al.* (2016). The genome of black raspberry (*Rubus occidentalis*). Plant J. *87*, 535–547. http://dx.doi.org/10.1111/tpj.13215

VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., *et al.* (2015). Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature *527*, 508–511. http://dx.doi.org/10.1038/nature15714

Van Hoeck, A., Horemans, N., Monsieurs, P., Cao, H.X., Vandenhove, H., and Blust, R. (2015). The first draft genome of the aquatic model plant Lemna minor opens the route for future stress physiology research and biotechnological applications, Biotechnol. Biofuels *8*, 1.

Varshney, R.K., and Dubey, A. (2009). Novel genomic tools and modern genetic and breeding approaches for crop improvement. J. Pant Biochem. Biot. *18*, 127–138.

Vicentini, R., Bem, D., Van Sluys, M.A., Nogueira, F.T.S., and Vincentz, M. (2012). Gene content analysis of sugarcane Public ESTs reveals thousands of missing coding-genes and an unexpected pool of grasses conserved ncRNAs. Tropical Plant Biol. *5*, 199–205, http://dx.doi.org/10.1007/s12042-012-9103-z

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, Miranda., Lail, K., *et al.* (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463,763–768.

Vorholt, J.A. (2012). Microbial life in the phyllosphere. Nat. Rev. Microbiol. *10*, 828–840. http://dx.doi.org/10.1038/nrmicro2910

Wang, H., Jiang, J., Chen, S., Qi, X., Peng, H., Li, P., Song, A., Guan, Z., Fang, W., Liao, Y., *et al.* (2013). Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. PLOS ONE, 8(4), 62293, http://dx.doi.org/10.1371/journal.pone.0062293.

Wang, H., Niu, Q.W., Wu, H.W., Liu, J., Ye, J., Yu, N., and Chua, N.H. (2015a). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. Plant. J. *84*, 404–416. http://dx.doi.org/10.1111/tpj.13018

Wang, L., Yu, X., Wang, H., Lu, Y.Z., de Ruiter, M., Prins, M., and He, Y.K. (2011). A novel class of heat-responsive small RNAs derived from the chloroplast genome of Chinese cabbage (*Brassica rapa*). BMC Genomics *12*, 289. http://dx.doi.org/10.1186/1471-2164-12-289

Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., Wang, P., Liu, N., Lindsey, K., and Zhang, X. (2015c). Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). New Phytol. *207*, 1181–1197. http://dx.doi.org/10.1111/nph.13429

Wang, T.Z., Liu, M., Zhao, M.G., Chen, R., and Zhang, W.H. (2015b). Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. BMC Plant Biol. *15*, 131. http://dx.doi.org/10.1186/s12870-015-0530-5

Wang, W., Haberer, G., Gundlach, H., Gläßer, C., Nussbaumer, T., Luo, M.C., Lomsadze, A., Borodovsky, M., Kerstetter, R.A., Shanklin, J., *et al.* (2014). The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat. Commun. *5*, 3311. http://dx.doi.org/10.1038/ncomms4311

Wang, X., Elling, A.A., Li, X., Li, N., Peng, Z., He, G., Sun, H., Qi, Y., Liu, X.S., and Deng, X.W. (2009). Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell *21*, 1053–1069. http://dx.doi.org/10.1105/tpc.109.065714

Wang, Y., Shen, D., Bo, S., Chen, H., Zheng, J., Zhu, Q.H., Cai, D., Helliwell, C., and Fan, L. (2010). Sequence variation and selection of small RNAs in domesticated rice. BMC Evol. Biol. *10*, 119. http://dx.doi.org/10.1186/1471-2148-10-119

Wangai, A.W., Redinbaugh, M.G., Kinyua, Z.M., Miano, D.W., Leley, P.K., Kasina, M., Mahuku, G., Scheets, K., and Jeffers, D. (2016). First report of *maize chlorotic mottle virus* and maize lethal necrosis in Kenya. Virus Research.

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., and Watson, M. (2015). Exome sequencing: current and future perspectives. G3 *5*, 1543–1550. http://dx.doi.org/10.1534/g3.115.018564

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Wei, B., Cai, T., Zhang, R., Li, A., Huo, N., Li, S., Gu, Y.Q., Vogel, J., Jia, J., Qi, Y., *et al.* (2009). Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.)

and *Brachypodium distachyon* (L.) Beauv. Funct. Integr. Genomics *9*, 499–511.

Wei, X., Zhang, X., Yao, Q., Yuan, Y., Li, X., Wei, F., Zhao, Y., Zhang, Q., Wang, Z., Jiang, W., *et al.* (2015). The miRNAs and their regulatory networks responsible for pollen abortion in Ogura-CMS Chinese cabbage revealed by high-throughput sequencing of miRNAs, degradomes, and transcriptomes. Front. Plant Sci. *6*, 894, http://dx.doi.org/10.3389/fpls.2015.00894.

Wendler, N., Mascher, M., Himmelbach, A., Johnston, P., Pickering, R., and Stein, N. (2015). Bulbosum to go: A toolbox to utilize *Hordeum vulgare/bulbosum* introgressions for breeding and beyond. Mol. Plant *8*, 1507–1519. http://dx.doi.org/10.1016/j.molp.2015.05.004

Wendler, N., Mascher, M., Nöh, C., Himmelbach, A., Scholz, U., Ruge-Wehling, B., and Stein, N. (2014). Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol. J. *12*, 1122–1131. http://dx.doi.org/10.1111/pbi.12219

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature *453*, 1239–1243.

Willing, E.M., Rawat, V., Mandáková, T., Maumus, F., James, G.V., Nordström, K.J., Becker, C., Warthmann, N., Chica, C., Szarzynska, B., *et al.* (2015). Genome expansion of Arabis alpina linked with retrotransposition and reduced symmetric DNA methylation. Nat. Plants *1*, 14023. http://dx.doi.org/10.1038/nplants.2014.23

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570–1573.

Winfield, M.O., Wilkinson, P.A., Allen, A.M., Barker, G.L., Coghill, J.A., Burridge, A., Hall, A., Brenchley, R.C., D'Amore, R., Hall, N., *et al.* (2012). Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol. J. *10*, 733–742. http://dx.doi.org/10.1111/j.1467-7652.2012.00713.x

Woese, C.R. (1987). Bacterial evolution. Microbiol. Rev. *51*, 221–271.

Wu, W., Yang, Y.L., He, W.M., Rouard, M., Li, W.M., Xu, M., Roux, N., and Ge, X.J. (2016). Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus. Sci. Rep. *6*, 31586. http://dx.doi.org/10.1038/srep31586

Xiao, L., Yang, G., Zhang, L., Yang, X., Zhao, S., Ji, Z., Zhou, Q., Hu, M., Wang, Y., Chen, M., *et al.* (2015). The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration. Proc. Natl. Acad. Sci. U.S.A. *112*, 5833–5837. http://dx.doi.org/10.1073/pnas.1505811112

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J., Chakrabarty, R., Desgagné-Penix, I., Haslam, T.M., Kim, Y.B., Liu, E., et al. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J. Biotechnol. *166*, 122–134.

Xie, F., Wang, Q., Sun, R., and Zhang, B. (2015). Deep sequencing reveals important roles of microRNAs in response to drought and salinity stress in cotton. J. Exp. Bot. *66*, 789–804. http://dx.doi.org/10.1093/jxb/eru437

Xin, M., Wang, Y., Yao, Y., Xie, C., Peng, H., Ni, Z., and Sun, Q. (2010). Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). BMC Plant Biol. *10*, 123. http://dx.doi.org/10.1186/1471-2229-10-123

Xu, M.Y., Dong, Y., Zhang, Q.X., Zhang, L., Luo, Y.Z., Sun, J., Fan, Y.L., and Wang, L. (2012). Identification of miRNAs and their targets from *Brassica napus* by high-throughput sequencing and degradome analysis. BMC Genomics *13*, 421. http://dx.doi.org/10.1186/1471-2164-13-421

Xu, T., Wang, Y., Liu, X., Lv, S., Feng, C., Qi, M., and Li, T. (2015). Small RNA and degradome sequencing reveals microRNAs and their targets involved in tomato pedicel abscission. Planta *242*, 963–984. http://dx.doi.org/10.1007/s00425-015-2318-0

Xu, X., Passey, T., Wei,F., Saville, R., and Harrison, R.J. (2015). Amplicon-based metagenomics identified candidate organisms in soils that caused yield decline in strawberry. Hortic. Res. 2:15022. http://dx.doi.org/10.1038/hortres.2015.22..

Xu, Y., Gao, S., Yang, Y., Huang, M., Cheng, L., Wei, Q., Fei, Z., Gao, J., and Hong, B. (2013). Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. BMC Genomics *14*, 662. http://dx.doi.org/10.1186/1471-2164-14-662

Yang, L., Jue, D., Li, W., Zhang, R., Chen, M., and Yang, Q. (2013). Identification of MiRNA from eggplant (*Solanum melongena* L.) by small RNA deep sequencing and their response to *Verticillium dahliae* infection. PLOS ONE *8*, e72840. http://dx.doi.org/10.1371/journal.pone.0072840

Yang, X., and Li, L. (2012). Analyzing the microRNA transcriptome in plants using deep sequencing data. Biology *1*, 297–310. http://dx.doi.org/10.3390/biology1020297

Yang, X., Wang, L., Yuan, D., Lindsey, K., and Zhang, X. (2013). Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis. J. Exp. Bot. *64*, 1521–1536. http://dx.doi.org/10.1093/jxb/ert013

Yao, S., Jiang, C., Huang, Z., Torres-Jerez, I., Chang, J., Zhang, H., Udvardi, M., Liu, R., and Verdier, J. (2016). The *Vigna unguiculata* Gene Expression Atlas (VuGEA) from *de novo* assembly and quantification of RNA-seq data provides insights into seed maturation mechanisms. Plant J. *88*, 318–327. http://dx.doi.org/10.1111/tpj.13279

Yao, Y., Guo, G., Ni, Z., Sunkar, R., Du, J., Zhu, J.K., and Sun, Q. (2007). Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). Genome Biol. *8*, R96.

Yao, Y., Ni, Z., Peng, H., Sun, F., Xin, M., Sunkar, R., Zhu, J.K., and Sun, Q. (2010). Non-coding small RNAs responsive to abiotic stress in wheat (*Triticum aestivum* L.). Funct. Integr. Genomics. *10*, 187–190. http://dx.doi.org/10.1007/s10142-010-0163-6

Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S.J., Aii, J., Sato, S., and Mori, M. (2016). Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. DNA Res. *23*, 215–224. http://dx.doi.org/10.1093/dnares/dsw012

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science *296*, 79–92. http://dx.doi.org/10.1126/science.1068037

Yu, R., Wang, J., Xu, L., Wang, Y., Wang, R., Zhu, X., Sun, X., Luo, X., Xie, Y., *et al.* (2016). Transcriptome profiling of taproot reveals complex regulatory networks during taproot thickening in radish (*Raphanus sativus* L.). Front. Plant Sci. *7*, 1210, http://dx.doi.org/10.3389/fpls.2016.01210.

Yu, X., Wang, H., Lu, Y., de Ruiter, M., Cariaso, M., Prins, M., van Tunen, A., and He, Y. (2012). Identification of conserved and novel microRNAs that are responsive to heat stress in *Brassica rapa*. J. Exp. Bot. *63*, 1025–1038. http://dx.doi.org/10.1093/jxb/err337

Zeng, X., Long, H., Wang, Z., Zhao, S., Tang, Y., Huang, Z., Wang, Y., Xu, Q., Mao, L., Deng, G., *et al.* (2015). The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. Proc. Natl. Acad. Sci. U.S.A. *112*, 1095–1100. http://dx.doi.org/10.1073/pnas.1423628112

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X., *et al.* (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res. *20*, 646–654. http://dx.doi.org/10.1101/gr.100677.109

Zhang, G., Liu, P., Zhang, L., Wei, W., Wang, X., Wei, D., and Wang, W. (2016). Bioprospecting metagenomics of a microbial community on cotton degradation: Mining for new glycoside hydrolases. J. Biotechnol. *234*, 35–42.

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., Wang, W., *et al.* (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat. Biotechnol. *30*, 549–554. http://dx.doi.org/10.1038/nbt.2195

Zhang, G., Tian, Y., Zhang, J., Shu, L., Yang, S., Wang, W., Sheng, J., Dong, Y., and Chen, W. (2015). Hybrid de novo genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). Gigascience *4*, 62. http://dx.doi.org/10.1186/s13742-015-0104-3

Zhang, G.Q., Xu, Q., Bian, C., Tsai, W.C., Yeh, C.M., Liu, K.W., Yoshida, K., Zhang, L.S., Chang, S.B., Chen, F., *et al.* (2015). The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Sci. Rep. *5*, 19029.

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., Zhang, Q., Liang, X., and Li, Y. (2012). *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics *13*, 90. http://dx.doi.org/10.1186/1471-2164-13-90

Zhang, J., Tian, Y., Yan, L., Zhang, G., Wang, X., Zeng, Y., Zhang, J., Ma, X., Tan, Y., Long, N., *et al.* (2016). Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high-altitude adaptation in the central Andes. Mol. Plant. *9*, 1066–1077. http://dx.doi.org/10.1016/j.molp.2016.04.016

Zhang, J., Xu, Y., Huan, Q., and Chong, K. (2009). Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response. BMC Genomics *10*, 449. http://dx.doi.org/10.1186/1471-2164-10-449

Zhang, N., Yang, J., Wang, Z., Wen, Y., Wang, J., He, W., Liu, B., Si, H., and Wang, D. (2014). Identification of novel and conserved microRNAs related to drought stress in potato by deep sequencing. PLOS ONE *9*, e95489. http://dx.doi.org/10.1371/journal.pone.0095489

Zhang, R., Marshall, D., Bryan, G.J., and Hornyik, C. (2013). Identification and characterization of miRNA transcriptome in potato by high-throughput sequencing. PLOS ONE *8*, e57233. http://dx.doi.org/10.1371/journal.pone.0057233

Zhang, X., Zou, Z., Zhang, J., Zhang, Y., Han, Q., Hu, T., Xu, X., Liu, H., Li, H., and Ye, Z. (2011). Over-expression of sly-miR156a in tomato results in multiple vegetative and reproductive trait alterations and partial phenocopy of the *sft* mutant. FEBS Lett. *585*, 435–439. http://dx.doi.org/10.1016/j.febslet.2010.12.036

Zhang, X.N., Li, X., and Liu, J.H. (2014). Identification of conserved and novel cold-responsive microRNAs in trifoliate orange (*Poncirus trifoliata* (L.) Raf.) using high-throughput sequencing. Plant Mol. Biol. Rep. *32*, 328–341.

Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S., and Chen, Y.Q. (2014). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biol. *15*, 512. http://dx.doi.org/10.1186/s13059-014-0512-1

Zhang, Y.C., Yu, Y., Wang, C.Y., Li, Z.Y., Liu, Q., Xu, J., Liao, J.Y., Wang, X.J., Qu, L.H., Chen, F., *et al.* (2013). Overexpression of microRNA OsmiR397 improves rice yield by increasing grain size and promoting panicle branching. Nat. Biotechnol. *31*, 848–852. http://dx.doi.org/10.1038/nbt.2646

Zhao, C., Xia, H., Cao, T., Yang, Y., Zhao, S., Hou, L., Zhang, L., Li, C., and Zhang, X. (2015). Small RNA and degradome deep sequencing reveals peanut microRNA roles in response to pathogen infection. Plant Mol. Biol. Rep. *33*, 1013–1029.

Zhao, F., Wang, C., Han, J., Zhu, X., Li, X., Wang, X., and Fang, J. (2016). Characterization of miRNAs responsive to exogenous ethylene in grapevine berries at whole genome level. Funct. Integr. Genomics 1–23. http://dx.doi.org/10.1007/s10142-016-0514-z

Zheng, C., Zhao, L., Wang, Y., Shen, J., Zhang, Y., Jia, S., Li, Y., and Ding, Z. (2015). Integrated RNA-Seq and sRNA-Seq analysis identifies chilling and freezing responsive key molecular players and pathways in tea plant (*Camellia sinensis*). PLOS ONE *10*, e0125031. http://dx.doi.org/10.1371/journal.pone.0125031

Zheng, L.L., and Qu, L.H. (2015). Application of microRNA gene resources in the improvement of agronomic traits in rice. Plant Biotechnol. J. *13*, 329–336. http://dx.doi.org/10.1111/pbi.12321

Zhou, L., and Holliday, J.A. (2012). Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene

space using sequence capture. BMC Genomics *13*, 703. http://dx.doi.org/10.1186/1471-2164-13-703

Zhou, M., and Luo, H. (2013). MicroRNA-mediated gene regulation: Potential applications for plant genetic engineering. Plant Mol. Biol. *83*, 59–75.

Zhou, Y., Xu, Z., Duan, C., Chen, Y., Meng, Q., Wu, J., Hao, Z., Wang, Z., Li, M., Yong, H., *et al.* (2016). Dual transcriptome analysis reveals insights into the response to *Rice black-streaked dwarf virus* in maize. J. Exp. Bot. *67*, 4593–4609. http://dx.doi.org/10.1093/jxb/erw244

Zhu, Q.H., and Wang, M.B. (2012). Molecular Functions of Long Non-Coding RNAs in Plants. Genes *3*, 176–190. http://dx.doi.org/10.3390/genes3010176

Zhuang, Y., Zhou, X.H., and Liu, J. (2014). Conserved miRNAs and their response to salt stress in wild eggplant *Solanum linnaeanum* roots. Int. J. Mol. Sci. *15*, 839–849. http://dx.doi.org/10.3390/ijms15010839

Zimin, A., Stevens, K.A., Crepeau, M.W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J.L., de Jong, P.J., *et al.* (2014). Sequencing and assembly of the 22-gb loblolly pine genome. Genetics *196*, 875–890. http://dx.doi.org/10.1534/genetics.113.159715

Zou, C., Wang, Q., Lu, C., Yang, W., Zhang, Y., Cheng, H., Feng, X., Prosper, M.A., and Song, G. (2016). Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). Sci. China. Life. Sci. *59*, 164–171. http://dx.doi.org/10.1007/s11427-016-5000-2

Zuo, J., Wang, Q., Han, C., Ju, Z., Cao, D., Zhu, B., Luo, Y., and Gao, L. (2016). SRNAome and degradome sequencing analysis reveals specific regulation of sRNA in response to chilling injury in tomato fruit. Physiol. Plant 6, http://dx.doi.org/10.1111/ppl.12509.