Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants

Chibuikem I. N. Unamba[1,2], Akshay Nag[1] and Ram K. Sharma[1]*

[1] Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Palampur, India, [2] Department of Plant Science and Biotechnology, Imo State University, Owerri, Nigeria

Non-model plants i.e., the species which have one or all of the characters such as long life cycle, difficulty to grow in the laboratory or poor fecundity, have been schemed out of sequencing projects earlier, due to high running cost of Sanger sequencing. Consequently, the information about their genomics and key biological processes are inadequate. However, the advent of fast and cost effective next generation sequencing (NGS) platforms in the recent past has enabled the unearthing of certain characteristic gene structures unique to these species. It has also aided in gaining insight about mechanisms underlying processes of gene expression and secondary metabolism as well as facilitated development of genomic resources for diversity characterization, evolutionary analysis and marker assisted breeding even without prior availability of genomic sequence information. In this review we explore how different Next Gen Sequencing platforms, as well as recent advances in NGS based high throughput genotyping technologies are rewarding efforts on *de-novo* whole genome/transcriptome sequencing, development of genome wide sequence based markers resources for improvement of non-model crops that are less costly than phenotyping.

Keywords: non-model, genomics, next generation sequencing, whole genome, transcriptome

GENOMICS IN THE VIEWPOINT OF NON-MODEL PLANT SYSTEMS

Plant genomics, which entails the application of recombinant DNA technologies, sequencing methods, and Bioinformatics tools for assembling and assigning the function and structure of plant genomes, is a key to understanding their genome via determining the order of DNA sequences which sequentially enable exploring the evolution of plant genome structure and inferring molecular phylogeny. It also helps in fathoming the interaction of genes in controlling the organism growth, development and adaptation to their environment. Most of the information we have about mechanisms underlying plant biological processes come from investigations on "model plants," commonly referred to as "plants" extensively studied at the whole genome level to elucidate various complex biological phenomena. High-throughput sequencing technologies are, however changing the approaches toward projects geared at genome sequencing, giving us a deeper understanding of plant biology by the generation of biologically important data sets from different plant species other than the model plants.

Prior to the development of next generation sequencing (NGS) in 2005 (Morozova and Marra, 2008; Schuster, 2008), nucleic acid sequencing for genomic studies was based on the Sanger method. This technique was successfully used to complete the human genome and the first sequenced plant genome, *Arabidopsis thaliana (The Arabidopsis Genome Initiative)* published in 2000. The authors, after outlining its many advantages for genome analysis, which include small size, homozygous nature, large number of offspring due to short gestation period and relatively small nuclear genomes, reported the plant as an important model system for identifying pathways genes and determining their functions. Some other plants sequenced using this first generation method and reported by Schatz et al. (2012) as models include *Oryza sativa* (rice) in 2002, *Carica papaya* (papaya) in 2008 and *Zea mays* (maize) in 2009. Further genome sequencing of other plants was therefore based on the idea that a single species among related plant species that share some similarities is chosen as a model, studied as a representative and information gathered can be applied to related organisms as required but Tagu et al. (2014) noted that model organisms are often not archetypal and do not replicate the biology of their close relatives or even the wide diversity of living mechanisms. Hirsch and Buell (2013) stated that the characteristics of the ideal plant genome are hinged on technological limitations of genome-sequencing and assembly methods, computation, and the desire for a whole genome sequence for downstream biological interpretations. Despite the successful use of Sanger technology in sequencing the model crops, its throughput and high cost posed some constraints to sequencing millions of plant species, especially those with large and complex genomes and this prompted high demand for new and improved sequencing technologies. In addition, several non-model plants are indispensable assets for food, feed, or energy resource with certain characteristics unique to them and thus intricate to study them by the use of a model plant (Carpentier et al., 2008); hence, genomics in these species was not known and posed some challenges until the recent progress made by the emergence of alternative sequencing platforms with increased throughput and lower sequencing cost collectively termed as NGS technologies.

This article is therefore an appraisal of the impact made by NGS technologies on the "genomics of non-model plants." It also tried to make out the future prospect of using these technologies in this group of plants.

GLIMPSE OF NEXT GENERATION SEQUENCING TECHNOLOGIES (NGS)

NGS incorporates technologies which at low cost and in short time produces millions of short DNA sequence read mostly in the range of 25 and 700 bp in length. According to Metzker (2010), they include a number of methods grouped broadly as template preparation, sequencing and imaging, and data analysis in which protocol distinguishes one technology from another and by the amount of the data produced from each platform. NGS has turned out to be a realistic method for maximizing sequencing in

a large number of non-model plants while reducing time and cost when compared to the traditional Sanger method. The Sanger method makes use of the 2′,3′-dideoxy and arabinonucleoside analogs of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase (Sanger et al., 1977) while in the NGS techniques, the DNA sequencing libraries are first clonally amplified *in vitro*, circumventing the time consuming and laborious cloning of the DNA library into bacteria unlike the Sanger method (Anderson and Schrijver, 2010). In addition, DNA templates are randomly read along the entire genome in a massively parallel sequencing by splitting the entire genome into small pieces followed by adapter ligation to the fragmented DNA (Zhang et al., 2011). Different technologies comprise NGS and while some of these technologies seem to have slight common features, they share key characteristics (Supplementary Table 1). The most commonly used platforms for high-throughput, useful genomic research, especially in non-model plant species include, Illumina/Solexa, 454/Roche, ABI/SOLiD, and Helicos. Results obtained from such research point to the fact that NGS techniques should not be restricted to the genomes of model organisms only as non-model plants have provided useful resources for genomic studies. Though they have their shortcomings, they are better off than the traditional Sanger method as shown in Supplementary Table 1.

NGS ENABLED GENOMIC RESEARCH IN NON-MODEL PLANTS

Whole Genome Sequencing

High coverage and quality reference genome sequences which give insight into the relatively complete information of genes, the regulatory elements that control their function, genome composition and an outline for understanding genomic variations (Feuillet et al., 2011) are the basics of "omics" investigations in a targeted species (Wei et al., 2013). The low cost of NGS is making it achievable for non-model plants, but as highlighted by Hirsch and Buell (2013), four major factors hinder the obtaining of quality genome assembly from non-model species: the extent of genome duplication (segmental, tandem, and whole-genome); the heterozygosity; the ploidy level; and repetitive sequence composition which have until now thwarted full genome sequencing and assembly of these plants. However, different methods are being applied to obtain a good quality sequence data as most sequencing projects of non-model plants are *de novo,* therefore, sequencing and assembly require high coverage and quality sequence data.

Various strategies are being employed to overcome the high level of heterozygosity and repetitive sequences that hinder the sequencing and assembly of plants using NGS technologies. Sequencing several independent libraries with different insertion sizes in different platforms and combining their data for assembly (Peng et al., 2014) wherein all data put together achieved high coverage of the genome and consequently enhanced the quality of the *de novo* assembly. Combined sequence data from paired end and mate pair libraries also produce assemblies with longer

contigs and fewer, larger scaffolds for maximizing coverage across the genome, thus many biological questions in these non-model plants can be answered. The large genome size of these plants is contributed by highly repetitive sequences that are similar or identical to sequences in the genome, are so abundant in occurrence such that even sequencing to higher depths by short-read technologies does not guarantee assembly quality. According to Hirsch and Buell (2013), their overrepresentation in the read pool of short-read sequences when joined with the inherent error rate in current NGS technologies confounds genome assembly. However, a hybrid approach that combines WGS sequencing data from different short reads platforms with high-density genetic and physical maps was utilized by Kane et al. (2011); Yang et al. (2013); Chen et al. (2013) wherein the maps can serve as scaffolds for the linear assembly of WGS sequences. Heterozygosity hampers contig assembly when a whole-genome shotgun strategy is used for sequencing. The negative effect of ploidy level and heterozygosity to the assembly of short-read sequence can be cushioned using homozygous genotypes derived from successive generations of self-fertilization (Shulaev et al., 2011; Wang et al., 2012a; Polashock et al., 2014). Wu et al. (2013) employed a novel combination of BAC-by-BAC (bacterial artificial chromosome) libraries with Illumina sequencing technology and Liu et al. (2014) used BAC libraries successfully, to overcome the major issues of high heterozygosity and high repeat content. This showed that a complex plant genome sequence can be assembled and characterized using NGS without a physical reference.

Genome duplication is thought to be a factor in the evolution and diversification of plants. Whole genome duplication (WGD) creates gene duplicates in plants, some which might not be essential to cell functioning while some may evolve novel genes via non-functionalization, neofunctionalization, or subfunctionalization. WGD thus contributes to evolution by enabling the evolution of new gene functions, advancing genome rearrangement and perhaps driving speciation. Whole genome sequencing (WGS) and analysis methods by comparing the sequences of individual members of a family is helping to map out the individual gene duplications involved in the evolution of a family from a single progenitor gene that existed in an ancestral genome as seen in Albert et al. (2013) where genomic changes that accompanied the origin of angiosperms was identified. They showed an ancient genome duplication that predated angiosperm diversification indicating that the ancestral angiosperm was a polyploid with a large assemblage of both novel and ancient genes that survived to play key roles in angiosperm biology.

The complete genome sequence of a species nevertheless does not imply that all accessions of the species has the same nucleotide sequence but rather contains almost same set of genes with changes in their nucleotide sequence arising maybe from substitutions, insertions, deletions, and structural variations. The low cost of NGS has made sequencing of related genomes to estimate the genetic diversity within and between germplasm pools possible, and identification and tracking of genetic variation are now so efficient and precise that thousands of variants can be tracked within large populations (Varshney et al., 2009). In sequencing the genomic DNA and RNA of

Cannabis sativa (Purple Kush) using hybrid approaches of Illumina and 454 pyrosequencing, Van Bakel et al. (2011) reported a draft haploid genome sequence of the cultivar which, when compared with the genome of another cultivar *C. sativa* (Finola), showed more expression of cannabinoid pathway genes and the exclusive presence of the functional THCA synthase (THCAS) in the genome and transcriptome of Purple Kush. Deciphering domestication of plants requires identification of the important traits that have been altered during domestication. NGS have made the discovery of the genes that have been selected during domestication feasible. Investigation of the primary gene pool and of more distantly related wild relatives has potential to identify genes and alleles that can be used to improve the performance of major crop species (Tang et al., 2010). Mace et al. (2013) used WGS to give an account of a strong racial structure and complex domestication events in 44 accessions of Sorghum and showed that the modern cultivated sorghum is derived from a limited sample of racial variation, with the result pointing to the positive utilization of NGS in the understanding of genetic diversity at the genomic sequence level.

To date, a number of non-model crops have been successfully sequenced using the NGS technology (**Table 1**) charting a new course for future genomic and genetic research and crop improvement in these plants, and even turning some of the so called non-model plants into genetic models for studying certain biological processes.

GENE IDENTIFICATION AND EXPRESSION ANALYSIS

The field of molecular and evolutionary biology are being revolutionized by the accessibility to genome-scale information which has helped to answer biological questions like how the identical genetic makeup of cells can give rise to different cell types, with each playing a different role in the working of a multicellular organism that until recently were implausible. Earlier technique used for detection and quantification of specific RNA levels is the Northern blotting (Northern hybridization) developed by James Alwine and George Stark. In this technique, electrophoretically separated bands of RNA are transferred from an agarose gel to a paper strip. Specific RNA bands can be detected by hybridization with ^{32}P-labeled DNA probes followed by autoradiography. This procedure allows the detection of specific RNA bands with high sensitivity and low background (Alwine et al., 1977). But as noted by Streit et al. (2008), northern blotting has some disadvantages among which are risk of mRNA degradation during electrophoresis, which compromises the quality and quantification of expression; health and environmental implication of high doses of radioactivity and formaldehyde; low sensitivity of northern blotting in comparison with that of RT-PCR; detection with multiple probes is difficult; use of ethidium bromide, DEPC and UV light needs special training and attention. The RNase protection assay, an alternative, is a highly sensitive technique developed to detect and measure the abundance of specific mRNAs in samples of total cellular RNA (Ma et al., 1996). Another method

TABLE 1 | Non-model plants sequenced using next generation sequencing technology.

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
1	*Solanum commersonii*	Commerson's nightshade	840 Mbp	I	105x	838 Mb	98	44.29 Kb	6.5 Kb	37,662	The draft genome sequence of *S. commersonii* substantially increases our understanding of the domesticated germplasm, facilitating translation of acquired knowledge into advances in crop stability in light of global climate and environmental changes.	Aversano et al., 2015
2	*Gossypium hirsutum* TM-1	Cultivated cotton	2.25–2.43 Gb	I	181x	2.1 Gb	96.70	107 kb	20kb	76,943	The complex genome of *Gossypium hirsutum* has been elucidated in this study, which was proven difficult to sequence owing to its complex allotetraploid (AtDt) genome.	Li et al., 2014
3	*Conyza canadensis*	Horseweed	335 Mb	R, I and PB	~350x	420 Mbp	92.30	33.5Kbp	20.8 kbp	44,592	Reportedly the first published draft genome of an agricultural weed which is a useful genomic resource for understanding weediness and the evolution of herbicide resistance as well as development of control strategies.	Peng et al., 2014
4	*Vaccinium macrocarpon*	American cranberry	470 Mbp	I	20x	420 Mbp	93	4.2 kbp		36,364	The study demonstrated the use of an inbred genotype derived from five generations of selfing to reduce heterozygosity and identified candidate genes which will be useful for further studies on biochemical pathways and cellular processes as well as development of molecular markers for breeding.	Polashock et al., 2014
5	*Ziziphus jujuba*	Jujube	444 Mb	I	429.25x	437.65 Mb	98.60	301.04 kbp	33.95kbp	32,808	The study provides insights into jujube-specific biology and valuable genomic resources for the improvement of Rhamnaceae plants and other fruit trees.	Liu et al., 2014
6	*Solanum melongena*	Eggplant	1.1 Gb	I and R		833 Mb	74	64kbp		85,446	The study gave an insight into the eggplant genome structure and will be a milestone for understanding unexplored species of the Solanaceae.	Hirakawa et al., 2014
7	*Humulus lupulus*	Hops	2.57 Gb	I	164x	2.05 Gb	80	37 kbp		41,228	The study which utilized two cultivars suggested the significance of historical human selection process for enhancing aroma and bitterness biosyntheses in hop cultivars, and as well serve as crucial information for breeding varieties with high quality and yield.	Natsume et al., 2014

(Continued)

TABLE 1 | Continued

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
8	*Camelina sativa*	False flax	785	I and R	123x	641.45Mb	82	30.09Mb	33.41 Kb	89,418	The study provides first chromosome-scale high-quality reference genome sequence for *C. sativa*, representing a whole-genome triplication event relative to the crucifer model *Arabidopsis thaliana*.	Kagale et al., 2014
9	*Oryza glaberrima*	African Rice		R and S		316 Mb		217 kb		33,164	This study provides evolutionary history of domestication and selection in African rice and supports the hypotheses that, it was domesticated in a single region, as opposed to domestication events across Africa.	Wang et al., 2014a
10	*Oryza sativa* AA genom	Wild rice		I							The WGS among the closely related wild rice species in different continents gave insight into plant gene and genome evolution. The study identified genomic variations, including segmental duplication and diversifying natural selection. It also indicated specific genes responsible for the adaptations.	Zhang et al., 2014
	-*Oryza nivara*		395 Mb		~73x	375 Mb	94.90	511.54 kbp	19.02 bp	41,490		
	-*Oryza glaberrima*		370Mb		~56x	344 Mb	93.20	722.13 kbp	25.248bp	41,476		
	-*Oryza barthii*		376Mb		~51x	335 Mb	89.10	237.57 kbp	16.126bp	41,605		
	-*Oryza glumaepatula*		366Mb		~86x	344 Mb	91.40	129.69 kbp	17.474bp	39,106		
	-*Oryza meridionalis*		388Mb		~60x	340 Mb	87.8	117.67 kbp	14.633bp	42,283		
11	*Pinus taeda*	Loblolly pine	20.15 Gb	I	98%			66.9 kbp		50,172	In this study, the large genome of the Loblolly pine ($\approx$ 20–40Gb, $2n = 24$) has been annotated for the first time by whole-genome shotgun assembly which comprises 20.1 Gb of sequence.	Wegrzyn et al., 2014
12	*Spirodela polyrhiza*	Greater duckweed	158 Mbp	R and S	5x			3.76Mb		19,623	In this study, it has been observed that, *Spirodela* has a genome with no signs of recent retrotranspositions but signatures of two ancient whole-genome duplications, possibly 95 million years ago (mya), older than those in Arabidopsis and rice. Its genome has only 19,623 predicted protein-coding genes, which is 28% less than the dicotyledonous *Arabidopsis thaliana* and 50% less than monocotyledonous rice.	Wang et al., 2014b

(Continued)

TABLE 1 | Continued

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
13	*Capsicum annuum*	Pepper	3.48 Gb	I and S	186.6x	3.06 Gb	90				WGS of *Capsicum annuum* integrated with data from resequencing of two cultivated peppers and *de novo* sequencing of a wild species afforded an evolutionary view into the genome expansion, origin of pungency, distinct ripening process and disease resistance of *Capsicum annuum* and gave an insight to the capsaicinoid pathway.	Qin et al., 2014; Kim et al., 2014
	cv. CM334							2.47 Mb	30.0kb	34,903		
	cv. Zunla-1							1.23 Mb	55.4 kb	35,336		
14	*Amborella trichopoda*	Amborella	870 Mb	R and I	~30x	706 Mb	81	4.9 Mbp		26,846	Study showed an ancient genome duplication preceding angiosperm diversification providing basis for understanding major genomic events in angiosperm evolution including polyploid origin of angiosperms and hexaploidization event in eudicots.	Albert et al., 2013
15	*Lupinus angustifolius*	Lupin	1.153 Gb	I	26.9x	598 Mbp	51.90	12.5 kbp	5.8 kbp	57,807	The study demonstrated the cost effectiveness of NGS in generating genomic resources for genomic and genetic studies in lupin and other non-model plants by combination of medium-depth genome sequencing and a high-density genetic linkage map.	Yang et al., 2013
16	*Oryza brachyantha*	Wild rice	~297 Mb	I	~104x	262 Mb	96	1.0 Mbp	20.4 kbp	32,038	The high-quality reference genome sequence of *Oryza brachyantha* provides an important resource for functional and evolutionary studies in the genus *Oryza*.	Chen et al., 2013
17	*Pyrus bretschneideri*	Pear	527 Mb	I	194x	512 Mb	97.10	540.8 kbp	35.7 kbp	42,812	The WGS of the plant in addition to providing an invaluable new resource for biological research of *Pyrus* gave insights into mechanisms underlying important biological processes, including stone cell formation, sugar accumulation, and aroma formation and release.	Wu et al., 2013
18	*Nelumbo nucifera*	Sacred lotus	929 Mb	I/R	101/5.1x	804 Mb	86.50	3.4 Mb	38.8 kbp	26,685	WGS reported a lineage-specific duplication in *Nelumbo nucifera* and lack of triplication event, characteristic of other eudicots thus making it a model for reconstructing the pan-eudicot genome and comparing eudicots and monocots.	Ming et al., 2013

(Continued)

TABLE 1 | Continued

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
19	*Genlisea aurea*	Corkscrew plant	63.6 Mb	–			43.4 Mb	68		5.78 kbp	17,755	Identified low number of genes despite being a carnivorous plant but introns and intergenic regions are unusually short and observed that reduction of genome size in the *G. aurea* lineage was due to both gene loss and non-coding sequences shrinking, but not to intron loss.	Leushkin et al., 2013
20	*Betula nana*	Dwarf birch	448 Mb	–	66x				18.6 kbp	5 kbp		The work presented a preliminary study of allele sharing among species, demonstrating the utility of the data for introgression studies and for the identification of species-specific alleles.	Wang et al., 2013
21	*Actinidia chinensis*	kiwifruit	758 Mb	–	140x	616.1 Mb	81.30	646.8 kbp	58.8 kbp	39,040	The study revealed WGD events undergone by the plant, detected heterozygous sites revealing high level of heterozygosity of the plant while providing a valuable resource for biological discovery, crop improvement and comparative genomic analysis.	Huang et al., 2013
22	*Hevea brasiliensis*	Rubber	~2.15 Gb	I, R and S	~43x	~1.1 Gb		3 kbp		68,955	The WGS in addition to key genes associated with rubber biosynthesis, rubberwood formation, disease resistance, and allergenicity identified a higher percentage of repetitive sequences which posed a challenge to the whole genome assembling.	Rahman et al., 2013
23	*Citrullus lanatus*	Water melon	~425 Mb	–	108.6x	353.5 Mb	83.20	2.38 Mb	26.38 kb	23,440	The WGS identified genomic regions that were preferentially selected as well as many disease-resistance genes lost during domestication in addition to providing insights into aspects of phloem-based vascular signaling in common between watermelon and cucumber and identification of genes crucial to valuable fruit-quality traits, including sugar accumulation and citrulline metabolism.	Guo et al., 2013
24	*Triticum urartu*	Einkorn wheat	4.94 Gb	–	~91x	4.66 Gb	94	63.69 kbp	3.42 kbp	34,879	The genome assembly provides a diploid reference for analysis of polyploid wheat genomes and is a valuable resource for the genetic improvement of wheat.	Ling et al., 2013

(Continued)

TABLE 1 | Continued

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
25	*Beta vulgaris*	Sugar beet	714–758 Mbp	I, R, and S	93x	569.0 Mb		2.01 Mb	1.7 Mb	27,421	Phylogenetic analyses in this study provided evidence for the separation of Caryophyllales before the split of asterids and rosids, and revealed lineage-specific gene family expansions and losses.	Dohm et al., 2014
26	*Gossypium raimondii*	Cotton		I	103.6x	775.2 Mb	88.10	2284 kbp	44.9 kb	40,976	The study observed evidence of the hexaploidization event shared by the eudicots as well as of a cotton-specific whole-genome duplication ~13–20 MYA.	Wang et al., 2012a
27	*Linum usitatissimum*	Flax	373 Mb	I	94x	318 Mb	85	694 kbp	20.1 kbp	43,384	The results from the demonstrated that *de novo* assembly of whole-genome shotgun short-sequence reads is an efficient means of obtaining nearly complete genome sequence information for some plant species.	Wang et al., 2012b
28	*Prunus mume*	Chinese Plum/Mei	280 Mb	I	101x	237 Mb	84.60	577.8 Kbp	31.8 Kbp	1154	The *P. mume* genome sequence contributes to the understanding of Rosaceae evolution and provided essential data for improvement of fruit trees.	Zhang et al., 2012
29	*Cyanophora paradoxa*	Glaucophyte	70 Mb	I, R, and S		70.2 Mbp			2.7 Kbp		In this study, analyses of the draft genome and transcriptome data from the basally diverging alga *Cyanophora paradoxa* was done and evidence for a single origin of the primary plastid in the eukaryote supergroup Plantae was established.	Price et al., 2012
30	*Bathycoccus prasinos* BBAN7	Green algae	15 Mb	S	22x	15.1 Mb				7847	The minimal genomes of the Mamiellophyceae provide a baseline for evolutionary and functional analyses of metabolic processes in green plants.	Moreau et al., 2012
31	*Chondrus crispus*	Red algae	105 Mb	S	14x	104.8		240 kb	64 kb	9606	Genome sequence of economically important red sea weed has been reported in this study.	Collén et al., 2013
32	*Cannabis sativa*	Cannabis	~820 Mb	I and R	110x	787 Mb	96	16.2 kbp		30,074	The study is an aid to the development of therapeutic marijuana strains with tailored cannabinoid profiles and provides a basis for the breeding of hemp with improved agronomic characteristics.	Van Bakel et al., 2011

(Continued)

TABLE 1 | Continued

S/N	PL	CN	EGS	SP	SC	AGS	AGC (%)	SN50	CN50	NG	GSO	REFERENCES
33	*Fragaria vesca*	Woodland strawberry	~240Mb	R, I, and S	39x	209.8 Mb	95%	1.3 Mb		34,809	WGS of fourth-generation inbred line of the *F. vesca* demonstrated that NGS can solely be used for assembling and characterization of a contiguous plant genome sequence while reporting the lack of large genome duplications seen in other rosids in the plant's sequence.	Shulaev et al., 2011
34	*Phoenix dactylifera*	Date palm	~658Mb	I		381 Mb	~60	30.48 kbp		28,890	The study identified a region of the genome linked to gender and found evidence that date palm employs an XY system of gender inheritance.	Al-Dous et al., 2011

PL, Plant Name; CN, Common Name; EGS, Estimated Genome size; SP, Sequencing Platform; SC, Sequencing Coverage; AGS, Assembled Genome Size; AGC, Assembly Genome Coverage; SN50, Scaffold N50; CN50, Contig N50; NG, Number of Genes; GSO, Genome Sequence Outcome; I, Illumina/Solexa; R, 454/Roche; S, SOLiD/ABI; H, Helicos; PB, Pac Bio System.

of gene expression analysis, hybridization of antisense RNA corresponding to a known complementary target sequence prevents target digestion by single strand–specific RNase activity. This process results in the degradation of all remaining single-stranded RNAs (i.e., those not hybridized to the probe sequence), enabling the accurate quantitation of specific target sequences (VanGuilder et al., 2008). However, the complex procedures as well as relatively large amounts of RNA involved pose some restrictions in the use of these methods. The development of real-time qPCR has increased the throughput of gene expression while reducing the required quantity of RNA. It has become a routine approach for measuring the expression of genes of interest, validating microarray experiments, and monitoring biomarkers (VanGuilder et al., 2008). Real-time PCR amplifies a specific target sequence in a sample, then monitors the amplification progress using fluorescent technology (Valasek and Repa, 2005). Despite the fact that real-time PCR technology is an invaluable tool for many scientists in gene expression analysis, its one major shortcoming is the prerequisite for prior sequence data of the specific target gene of interest, hence q-PCR can only be used for targeting of known genes.

The transcriptome is the set of all RNA molecules (mRNA, rRNA, tRNA, and other non-coding RNA) transcribed by an organism. Wang et al. (2009) had posited that the fundamental principle for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues, and also for understanding development and disease is gaining insight into the transcriptome. Microarray is a technique widely employed for analyzing the transcriptome for patterns of gene expression. It has the ability to measure the expression levels of thousands of genes in a single experiment, but lacks the capacity to detect novel transcripts and sensitivity to expression levels of genes. NGS have rapidly advanced next-generation RNA sequencing (RNA-seq) for rapid generation of large expression datasets for gene discovery and expression analysis in non-model species (Marioni et al., 2008; Li et al., 2012). As stated by De Wit et al. (2012), RNA-seq focuses on sequencing only mRNA from the genes that are expressed in the tissue or transcriptome wherein a considerable proportion of adaptively interesting variations are located. It shows a record of how many mRNAs from a particular exon are in the sample and includes variations in the sequences that elucidate functional polymorphisms. Unlike the microarray techniques, RNA-seq can assemble reads *de novo* without mapping to reference genomic sequence, a feature that makes it an invaluable asset for identification of novel genes in non-model plants. Zhou et al. (2012) demonstrated the use of *de novo* assembly in *Ammopiptanthus*, a genus with evergreen broadleaf habit in the desert and arid regions of the Mid-Asia, playing a critical role in conserving the desert ecosystems, which is critical in controlling desertification. To understand the genetic mechanisms underlying deep, flourishing root system for water absorption to adapt these plants to harsh conditions, *de novo* transcriptome sequencing of *A. mongolicus* was carried out using 454 pyrosequencing to discover putative genes associated with drought tolerance. The potential drought stress related transcripts identified in the study provided a

foundation for further investigation into the drought adaptation in *Ammopiptanthus*. Transcriptome sequencing has, however caused a significant upshot in the expressed sequence tags (ESTs) collections, including the non-model plant species (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

MicroRNAs (miRNA; 21–24 nucleotide) are a class of non-coding endogenous small RNAs that are transcribed from a gene, but the transcript is never translated into a protein (Phelps-Durr, 2010) therefore are involved in regulating gene expression in different organisms including non-model plants. Since the discovery of the first miRNAs, *Lin-4*, Lee et al. (1993), there has been an increased interest in understanding post transcriptional gene expression regulation during development. According to Axtell and Bartel (2005), miRNAs affect the morphology of flowering plants by the post transcriptional regulation of genes involved in critical developmental events. They, however postulated that an understanding of the spatial and temporal dynamics of miRNA activity is fundamental to elucidate the functions of miRNAs. Achard et al. (2004) described the role of microRNA (miR159) in the regulation of short-day photoperiod flowering time and of anther development. Other plant developmental processes involving miRNAs include leaf morphogenesis and polarity (Floyd and Bowman, 2004), floral development and timing defects (Aukerman and Sakai, 2003) among others. Zhang et al. (2006) identified four existing approaches for identifying miRNAs which include genetic screening, direct cloning after isolation of small RNAs, computational strategy, and ESTs analysis but observed that these approaches have different advantages and shortcomings and postulated that combining these methods, more miRNAs will be quickly discovered. As reported by Lakhotia et al. (2014), a large number of miRNAs are evolutionary conserved among diverse species, while several miRNAs, that are considered to be recently evolved show species-specificity and often express at lower levels relative to conserved miRNAs and as a result of their low expression levels, most of the species-specific miRNAs remained unidentified in many plant species. With improved methods of NGS technologies in investigating the transcriptome, enormous progress, especially with regard to regulatory pathways have been made in identifying and understanding non-coding RNAs such as miRNAs. RNA sequencing using high-throughput NGS platforms has the advantage of high accuracy in distinguishing miRNAs that are very similar in sequence and can detect novel miRNAs. Gao et al. (2015) identified 50 novel miRNAs, representing 19 families from three sRNA libraries of tobacco in addition to 165 miRNAs representing 55 conserved families using Solexa sequencer. Similarly, using high-throughput sequencing of small RNAs and analysis of transcriptome data, Zhu et al. (2013), identified 132 putative conserved miRNAs belonging to 31 known miRNA families and 10 novel miRNAs in *Caragana intermedia*. They in addition, predicted 38 potential targets for the conserved and novel miRNAs and validated four of them by 5′ RACE. These including identifications of miRNA in various non-model crops, Lakhotia et al. (2014) show the value of high throughput sequencing approach to miRNA discovery, especially novel miRNAs in non-model crops without a reference genome.

NGS IN AID OF MOLECULAR MARKER DEVELOPMENT AND BREEDING

Molecular markers are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next (Semagn et al., 2006). With the need to amplify the agricultural output to meet up with the challenge of producing enough food for the rising world population, advances in genomic technologies have provided new tools for discovering and tagging novel alleles and genes. These tools can enhance the efficiency of breeding programs through their use in marker-assisted selection (MAS), linkage mapping or quantitative trait locus (QTL) mapping, Phylogenetics, positional cloning, genetic diversity assessment, genotypic profiling etc. According to Kumpatla et al. (2012), the ability to deduce the underlying molecular mechanisms of a trait, understand the gene regulatory mechanisms, determine gene expression differences and variations in expressed gene sequences, and other structural variations such as copy number variations (CNV) and presence-absence variations (PAV) is to a large extent dependent on the availability of reference genome/transcriptome sequence.

Identification of polymorphic sequences, basic to a trait of interest enables the development of functional markers. The advent of NGS has enabled the exploration of thousands of markers across the entire genome using several approaches, enabling comprehensive genome-wide association studies, even in populations with little or any previous genetic information as in non-model plants (Sakiyama et al., 2014). SNP markers are the most abundant in a genome and appropriate for analysis on a wide range of genomic scales. SNPs are markers, which untangle polymorphism between individuals or populations due to change of a single nucleotide. Illumina transcriptome sequencing data was used to discover 2987 high-quality putative SNP in Turkish Olive Genotypes (Kaya et al., 2013). These were successfully used to access genetic diversity among 96 olive genotypes. A whole-genome resequencing of two cabbage inbred lines using Illumina (Lee et al., 2015) identified 674,521 SNPs. From these, 167 dCAPS markers were developed for genetic map construction which identified novel QTLs for black rot resistance. Similarly, a high-throughput and specific-locus amplified fragment sequencing (SLAF-seq) approach was also used by Wei et al. (2014a) to construct a high-density SNP map for cucumber. It contained 1800 high quality SNPs, spanning 890.79 cM with an average marker interval of 0.50 cM and further detected fruit-related QTLs. Also, genotyping-by-sequencing (GBS) approach via NGS identified 21,471 SNPs in oil palm (Pootakham et al., 2015). It enabled the construction of linkage map containing 1085 markers distributed over 17 linkage groups and identified quantitative trait loci (QTL) affecting trunk height and bunch weight.

Simple sequence repeat (SSR) markers which have the advantage of high abundance, random distribution within the genome, high polymorphism information content and co-dominant inheritance have been developed at large scale and lower costs via NGS. In *Myrica rubra* with an estimated genome size of 323 Mb, highly heterozygous but with little

duplication, Jiao et al. (2012) identified 28,602 SSRs from a WGS sequencing using Illumina. Polymorphic markers among these also successfully transferred to other Myrica species. Likewise, in Sesame genome, 23,438 putative SSRs were identified by whole-genome *de novo* sequencing and successfully used to screen accession across 12 countries (Wei et al., 2014b). *De novo* genic SSRs have been developed at large scale and used in a number of non-model, including but not limited to *Caragana korshinskii* Kom (Long et al., 2015), *Hevea brasiliensis* (Salgado et al., 2014), *Prosopis alba* (Torales et al., 2013).

These developed markers are also used for association mapping studies in non-model plants. Association mapping (linkage disequilibrium mapping) identifies QTLs that accounts for phenotypic variation among individuals or species. It helps in the dissection of complex genetic traits and enhances crop breeding for traits as disease resistance, salinity and drought tolerance. In an association mapping analyses, accounting for population structure study by Gupta et al. (2014), eight out 50 SSR markers representing the nine chromosomes of foxtail millet used in testing population structure in 184 accessions were shown to have significant association with nine agronomic traits. Also, association analysis using 20 SSR markers to detect the marker loci linked to morphological traits and physiological traits in a wild *Populus simonii* population Wei et al. (2014c), revealed that three SSR markers were identified for seven traits, one was associated with five morphological traits while two of the markers were associated with one morphological trait and one physiological trait, respectively. These studies infer that the identified markers are suitable for MAS breeding, target gene detection or QTL.

Genome sequencing have aided in deciphering the influence of transposable elements in the function and evolution of genes and genomes. Most of these repetitive sequences are found in different regions across the genome and have been implicated in genome diversity and phenotypic variation. In view of these, molecular markers are being developed from these elements and used for diversity characterization and construction of genetic linkage maps. In foxtail millet, genome-wide analysis, Yadav et al. (2014) identified 30,706 TEs, which led to the development of 20,278 TE-based markers from namely Retrotransposon-Based Insertion Polymorphisms (4801), Inter-Retrotransposon Amplified Polymorphisms (3239), Repeat Junction Markers (4451), Repeat Junction-Junction Markers (329), Insertion-Site-Based Polymorphisms (7401) and Retrotransposon-Microsatellite Amplified Polymorphisms (57). Of these, 30 out of 134 Repeat Junction Markers screened in 96 accessions of *Setaria italica* and three wild *Setaria* accessions showed polymorphism. This demonstrates that transposable elements can serve as genomic resources for genotyping. Insertions and Deletions (Indels), are other genomic resources distributed across the genome that can also be used as molecular markers for Phylogenetics. 2687 InDel-based markers were developed from Illumina sequence data from three genotypes of *Phaseolus vulgaris* L (Moghaddam et al., 2014). These markers were successfully used to construct a phylogenetic tree and a genetic map, deducing that InDel markers are reliable, simple, and accurate. Introns are non-coding RNA transcripts that are spliced out before the translation of the RNA molecule into a protein. Markers developed from introns have high evolutionary rate, possibly because they are flanked by exons which consign conserved primers that may function across a wide range of species. Intron Length Polymorphic (ILP) markers are thus designed via exon-primed intron-crossing PCR (EPIC-PCR) by designing primers in exons flanking the target intron. NGS sequence data from a potato cultivar was used to design ILP markers (Ahmadvand et al., 2014). These markers were used to test diversity in other potato genotypes and cross transferability was investigated in other *Solanum* species. The results demonstrated ILPs as genomic resources in diverse molecular analyses, including cross-species studies. Similarly, Muthamilarasan et al. (2014) developed 5123 ILP markers, of which 4049 were physically mapped onto nine chromosomes of foxtail millet. They further showed the applicability of the markers in germplasm characterization, transferability, Phylogenetics and comparative mapping studies in millets and bioenergy grass species.

UNDERSTANDING BIOSYNTHETIC PATHWAYS OF SPECIALIZED PLANT METABOLITES IN NON-MODEL PLANTS

Plants manufacture a huge and diverse group of organic compounds called secondary metabolites. These compounds appear to have no direct role in growth and other physiological processes in plants, but are implicated in their adaptation to their environment such as control of seed germination, symbiosis regulation, defense against herbivores and pathogens, and chemical inhibition of competing plant species. Contrasting the primary metabolites (sugars, amino acids, acyl lipids, and nucleotides) which are found in all plants, these secondary metabolites only pertain to a plant species or group of related plant species. They were initially thought to be waste products of metabolism until research showed that these secondary metabolites are useful in pharmaceuticals, flavors, industrial materials, and chemicals consequently increasing interest for their use. Most of these compounds occur in non-model plants for which genomic sequence information is not yet available (Xiao et al., 2013). The genus *Panax*, for instance, consists of at least nine species (Leung and Wong, 2010), most commonly referred to as ginsengs which are known from research to have anticancerous, antidiabetic, immunomodulatory, anti-inflammatory, and antiallergic, effects among other medicinal uses. The mode of action of ginseng was however not known until ginsenosides were isolated in 1963 (Shibata et al., 1963, 1965). Christensen (2008) reported that ginsenosides are found nearly exclusively in *Panax* species (ginseng) with more than 150 naturally occurring ginsenosides being isolated from the roots, leaves/stems, fruits, and/or flower heads of ginseng. Since then, research effort on evaluating the function and elucidating the molecular mechanism of each ginsenoside has been on the increase. Researchers have generated genomic information about ginsengs, identifying several candidate genes encoding enzymes responsible for the biosynthesis of the secondary metabolites

ginsenoside using different NGS platforms (Sun et al., 2010; Luo et al., 2011; Li et al., 2013; Jayakodi et al., 2014).

Access to some of these secondary metabolic compounds was often poor because of a lack of understanding of how these metabolites are synthesized (Oksman-Caldentey and Inzé, 2004), partly owing to the fact that the enzymes and biochemical pathways in their synthesis were either unknown or having complexities that make identification of the enzymes that catalyze the numerous metabolic cycles difficult. In some of these plants, a number of regulatory enzymes are involved in the biosynthesis process. Many of the genes in plant genomes code enzymes for secondary metabolism and transcriptomics data mining however have proven to be an efficient way to discover genes or gene families encoding enzymes involved in various metabolic pathways (Xiao et al.,

2013). *Podophyllum* species are sources of podophyllotoxin, an aryltetralin lignan used for semi-synthesis of various powerful and extensively employed cancer-treating drugs, its biosynthetic pathway, however, remains largely unknown. NGS/Bioinformatics and metabolomics analysis of *Podophyllum hexandrum* and *P. peltatum* plant tissues gave two putative genes in podophyllotoxin biosynthesis (Marques et al., 2013). Further studies using integrated omics technologies (including advanced mass spectrometry/metabolomics, transcriptome sequencing/gene assemblies, and Bioinformatics) in the two *Podophyllum* plants (Marques et al., 2014) enabled discovery of the aporphine alkaloid pathway in Podophyllum species, result which suggest evolutionary linkages between both lignan and alkaloid biosynthetic pathways. The authors reported that RNA-seq transcriptome sequencing and Bioinformatics
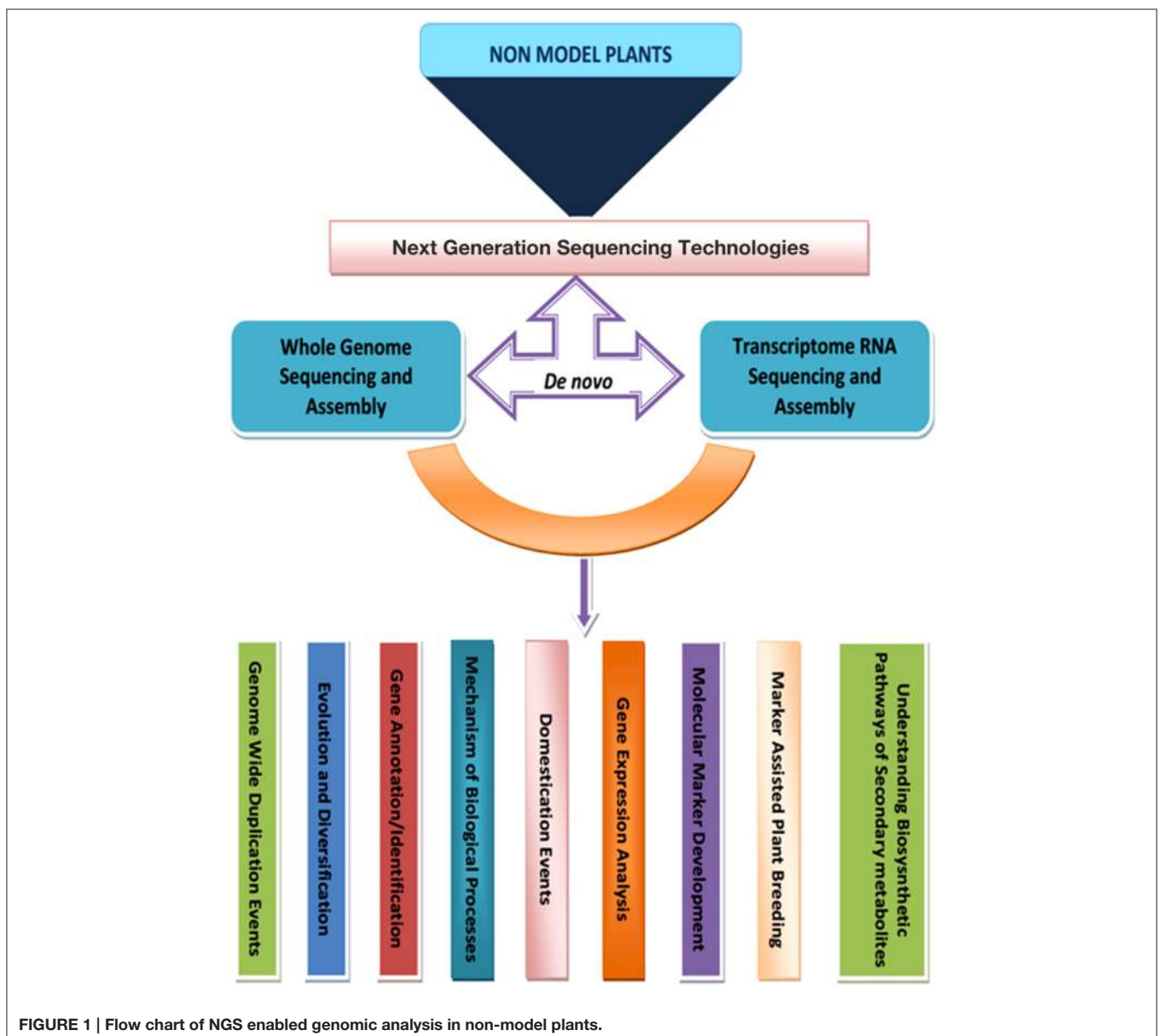

FIGURE 1 | Flow chart of NGS enabled genomic analysis in non-model plants.

guided gene assemblies/analyses *in silico*, specifically suggested presence of transcripts homologous to genes encoding all known steps in aporphine alkaloid biosynthesis. Miettinen et al. (2014) had stated that the biotechnological production progress of the monoterpenoid indole alkaloids (MIAs), produced by *Catharanthus roseus* in extremely low levels and used as anticancer drugs, from other sources is hampered by the lack of knowledge of the enzymes responsible for their biosynthesis. They nevertheless reported the characterization of the last missing steps of the *C. roseus* secoiridoid pathway using an integrated transcriptomics and proteomics approach for gene discovery, followed by biochemical characterization of the isolated candidates and further reported the reconstitution of the entire MIA pathway up to strictosidine in the plant host *Nicotiana benthamiana*, by heterologous expression of the newly identified genes in combination with the previously known biosynthesis genes. This new technology of NGS has helped in explicating the progression of events that lead to the production of these secondary compounds of interest in non-model plants, accelerating gene discovery for secondary metabolite pathways without preexisting sequence knowledge of the genes studied.

Many secondary metabolites have a complex and unique structure and their production is often enhanced by both biotic and abiotic stress conditions (Dixon, 2001). Ryan et al. (2002) provided valuable insights into the biochemical response of plants to UV stress, which results in the production of a more protective flavonoid profile. Rezaeieh et al. (2012) however noted that biotic and abiotic stresses exert an outstanding influence on the biosynthesis of several secondary metabolites in medicinal plants. Often, it is difficult to predict the complex signaling pathway that are activated or deactivated in response to different abiotic stresses but the complex molecular regulatory system involved in stress tolerance and adaptation in plants can be easily deciphered with the help of different omics study (Chawla et al., 2011). In response to various abiotic stresses, plant continuously needs to adjust their transcriptome profile (Gupta et al., 2013) thus NGS based transcriptome shotgun sequencing (RNA-seq), which targets the genes that are expressed in a tissue at a particular time is invaluable. A comprehensive transcriptome analysis of a salinity tolerant *Phaseolus vulgaris* L. variety by Illumina sequencing showed genes related to salt tolerance in plant (Hiz et al., 2014). This and other studies using transcriptomic approaches in non-model plants (Xu et al., 2013) for drought stress in *chrysanthemum* have continued to generate functional genomics resource, giving an unfathomable understanding of the molecular mechanisms underlying plant's responses to stress conditions.

CONCLUSION AND FUTURE PROSPECTS

From the foregoing, it is evidently clear that the cost effective and timely sequencing provided by different NGS technology platforms has impacted positively in advancing the course of non-model plants which earlier had no place in genomics. The technology has enabled scientists to explore the plants to their own benefit and in understanding mechanisms underlying processes of gene expression and secondary metabolism in addition to creation of genomic resources for diversity analysis and marker assisted breeding (**Figure 1**) through *de novo* analysis which hitherto was impossible due to lack of reference genomes.

The decreasing cost of this technology is however an open door to the possibility of sequencing genomes of individuals of a particular species. This if utilized properly will immensely assist comparative genomics in acquiring vital information about the evolutionary history of non-model plant species by studying the order of their DNA sequences, which had relied on chromosome numbers and ploidy levels. Moreso, protein seq (proteomics) combined with the increasing number of WGS will aid functional genomics in protein identification and consequently perform functional prediction of hypothetical proteins/genes which usually form the largest category during functional (BLASTX) annotations in non-model plants as well as in metabolomics which involves large scale measurements of metabolites level as non-model plants are large repositories of secondary metabolites of economic interest. It will also enable Phenomics for development of large scale phenotypic data for understanding how interactions of genotypes with the environment translate into phenotypic variations in non-model plants. In addition, improvements in these technologies will also advance Bioinformatics in data handling processes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2015.01074

REFERENCES

Achard, P., Herr, A., Baulcombe, D. C., and Harberd, N. P. (2004). Modulation of floral development by a gibberellin-regulated microRNA. *Development* 131, 3357–3365. doi: 10.1242/dev.01206

Ahmadvand, R., Poczai, P., Hajianfar, R., Kolics, B., Gorji, A. M., Polgár, Z., et al. (2014). Next generation sequencing based development of intron-targeting markers in tetraploid potato and their transferability to other *Solanum* species. *Gene* 540, 117–121. doi: 10.1016/j.gene.2014.02.045

Albert, V. A., Barbazuk, W. B., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., et al. (2013). The Amborella genome and the evolution of flowering plants. *Science* 342:1241089. doi: 10.1126/science.1241089

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., et al. (2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* 29, 521–527. doi: 10.1038/nbt.1860

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and

hybridization with DNA probes. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5350–5354. doi: 10.1073/pnas.74.12.5350

Anderson, M. W., and Schrijver, I. (2010). Next generation DNA sequencing and the future of genomic medicine. *Genes* 1, 38–69. doi: 10.3390/genes1010038

Aukerman, M. J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* 15, 2730–2741. doi: 10.1105/tpc.016238

Aversano, R., Contaldi, F., Ercolano, M. R., Grosso, V., Iorizzo, M., Tatino, F., et al. (2015). The *Solanum* commersonii genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. *Plant Cell* 27, 954–968. doi: 10.1105/tpc.114.135954

Axtell, M. J., and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17, 1658–1673. doi: 10.1105/tpc.105.032185

Carpentier, S. C., Panis, B., Vertommen, A., Swennen, R., Sergeant, K., and Renaut, J., et al. (2008). Proteome analysis of non-model plants: a challenging but powerful approach. *Mass Spectrom. Rev.* 27, 354–377. doi: 10.1002/mas.20170

Chawla, K., Barah, P., Kuiper, M., and Bones, A. M. (2011). "Systems biology: a promising tool to study abiotic stress responses," in *Omics and Plant Abiotic Stress Tolerance,* eds N. Tuteja, S. S. Gill and R. Tuteja (Bentham Science Publishers), 163–172. doi: 10.2174/978160805092511101010163

Chen, J., Huang, Q., Gao, D., Wang, J., Lang, Y., and Liu, T., et al. (2013). Whole-genome sequencing of *Oryza* brachyantha reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4, 1595. doi: 10.1038/ncomms2596

Christensen, L. P. (2008). Ginsenosides: chemistry, biosynthesis, analysis, and potential health effects. *Adv. Food Nutr. Res.* 55, 1–99. doi: 10.1016/S1043-4526(08)00401-4

Collén, J., Porcel, B., Carré, W., Ball, S. G., Chaparro, C., Tonon, T., et al. (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the *Archaeplastida. Proc. Natl. Acad. Sci. U.S.A.* 110, 5247–5252. doi: 10.1073/pnas.1221259110

De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., et al. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Mol. Ecol. Resour.* 12, 1058–1067. doi: 10.1111/1755-0998.12003

Dixon, R. A. (2001). Natural products and plant disease resistance. *Nature* 411, 843–847. doi: 10.1038/35081178

Dohm, J. C., Minoche, A. E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., et al. (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505, 546–549. doi: 10.1038/nature12817

Feuillet, C., Leach, J. E., Rogers, J., Schnable, P. S., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. *Trends Plant Sci.* 16, 77–88. doi: 10.1016/j.tplants.2010.10.005

Floyd, S. K., and Bowman, J. L. (2004). Gene regulation: ancient microRNA target sequences in plants. *Nature* 428, 485–486. doi: 10.1038/428485a

Gao, J., Yin, F., Liu, M., Luo, M., Qin, C., Yang, A., et al. (2015). Identification and characterisation of tobacco microRNA transcriptome using high-throughput sequencing. *Plant Biol.* 17, 591–598. doi: 10.1111/plb.12275

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470

Gupta, B., Sengupta, A., Saha, J., and Gupta, K. (2013). Plant abiotic stress:'Omics' approach. *Plant Biochem. Physiol.* 1:e108. doi: 10.4172/2329-9029.1000e108

Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S. K., and Prasad, M. (2014). Population structure and association mapping of yield contributing agronomic traits in foxtail millet. *Plant Cell Rep.* 33, 881–893. doi: 10.1007/s00299-014-1564-0

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., et al. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *Solanum* species indigenous to the old world. *DNA Res.* 21, 649–660. doi: 10.1093/dnares/dsu027

Hirsch, C. N., and Buell, C. R. (2013). Tapping the promise of genomics in species with complex, non-model genomes. *Annu. Rev. Plant Biol.* 64, 89–110. doi: 10.1146/annurev-arplant-050312-120237

Hiz, M. C., Canher, B., Niron, H., and Turet, M. (2014). Transcriptome analysis of salt tolerant common bean (*Phaseolus vulgaris* L.) under saline conditions. *PLoS ONE* 9:e92598. doi: 10.1371/journal.pone.0092598

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., et al. (2013). Draft genome of the kiwifruit *Actinidia chinensis. Nat. Commun.* 4:2640. doi: 10.1038/ncomms3640

Jayakodi, M., Lee, S. C., Park, H. S., Jang, W., Lee, Y. S., Choi, B. S., et al. (2014). Transcriptome profiling and comparative analysis of *Panax* ginseng adventitious roots. *J. Ginseng Res.* 38, 278–288. doi: 10.1016/j.jgr.2014.05.008

Jiao, Y., Jia, H. M., Li, X. W., Chai, M. L., Jia, H. J., Chen, Z., et al. (2012). Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 13:201. doi: 10.1186/1471-2164-13-201

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., et al. (2014). The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat. Commun.* 5:3706. doi: 10.1038/ncomms4706

Kane, N. C., Gill, N., King, M. G., Bowers, J. E., Berges, H., Gouzy, J., et al. (2011). Progress towards a reference genome for sunflower. *Botany* 89, 429–437. doi: 10.1139/b11-032

Kaya, H. B., Cetin, O., Kaya, H., Sahin, M., Sefer, F., Kahraman, A., et al. (2013). SNP discovery by Illumina-based transcriptome sequencing of the olive and the genetic characterization of Turkish olive genotypes revealed by AFLP, SSR and SNP markers. *PLoS ONE* 8:e73674. doi: 10.1371/journal.pone.0073674

Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877

Kumpatla, S. P., Abdurakhmonov, I. Y., Mammadov, J. A., and Buyyarapu, R. (2012). *Genomics-assisted Plant Breeding in the 21st Century: Technological Advances and Progress.* INTECH Open Access Publisher.

Lakhotia, N., Joshi, G., Bhardwaj, A. R., Katiyar-Agarwal, S., Agarwal, M., Jagannath, A., et al. (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. doi: 10.1186/1471-2229-14-6

Lee, J., Izzah, N. K., Jayakodi, M., Perumal, S., Joh, H. J., Lee, H. J., et al. (2015). Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. *BMC Plant Biol.* 15, 32. doi: 10.1186/s12870-015-0424-6

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854.

Leung, K. W., and Wong, A. S. (2010). Pharmacology of ginsenosides: a literature review. *Chin. Med.* 5:20. doi: 10.1186/1749-8546-5-20

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 non-coding sequences. *BMC Genomics* 14:476. doi: 10.1186/1471-2164-14-476

Li, C., Zhu, Y., Guo, X., Sun, C., Luo, H., Song, J., et al. (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax* ginseng CA Meyer. *BMC Genomics* 14:245. doi: 10.1186/1471-2164-14-245

Li, D., Deng, Z., Qin, B., Liu, X., and Men, Z. (2012). *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13:192. doi: 10.1186/1471-2164-13-192

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

Ling, H. Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu. Nature* 496, 87–90. doi: 10.1038/nature11997

Liu, M. J., Zhao, J., Cai, Q. L., Liu, G. C., Wang, J. R., Zhao, Z. H., et al. (2014). The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* 5:5315. doi: 10.1038/ncomms6315

Long, Y., Wang, Y., Wu, S., Wang, J., Tian, X., and Pei, X. (2015). *De novo* assembly of transcriptome sequencing in *Caragana korshinskii* kom. and characterization of EST-SSR markers. *PLoS ONE* 10:e0115805. doi: 10.1371/journal.pone.0115805

Luo, H., Sun, C., Sun, Y., Wu, Q., Li, Y., Song, J., et al. (2011). Analysis of the transcriptome of *Panax* notoginseng root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics* 12(Suppl. 5):S5. doi: 10.1186/1471-2164-12-S5-S5

Ma, Y. J., Dissen, G. A., Rage, F., and Ojeda, S. R. (1996). RNase protection assay. *Methods* 10, 273–278. doi: 10.1006/meth.1996.0102

Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop *Sorghum*. *Nat. Commun.* 4:2320. doi: 10.1038/ncomms3320

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517. doi: 10.1101/gr.079558.108

Marques, J. V., Dalisay, D. S., Yang, H., Lee, C., Davin, L. B., and Lewis, N. G. (2014). A multi-omics strategy resolves the elusive nature of alkaloids in *Podophyllum* species. *Mol. Biosyst.* 10, 2838–2849. doi: 10.1039/C4MB00403E

Marques, J. V., Kim, K. W., Lee, C., Costa, M. A., May, G. D., Crow, J. A., et al. (2013). Next generation sequencing in predicting gene function in podophyllotoxin biosynthesis. *J. Biol. Chem.* 288, 466–479. doi: 10.1074/jbc.M112.400689

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., et al. (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* 5:3606. doi: 10.1038/ncomms4606

Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L. T., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo* nucifera Gaertn.). *Genome Biol.* 14:R41. doi: 10.1186/gb-2013-14-5-r41

Moghaddam, S. M., Song, Q., Mamidi, S., Schmutz, J., Lee, R., Cregan, P., et al. (2014). Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. *Front. Plant Sci.* 5:185. doi: 10.3389/fpls.2014.00185

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., et al. (2012). Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* 13:R74. doi: 10.1186/gb-2012-13-8-r74

Morozova, O., and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92, 255–264. doi: 10.1016/j.ygeno.2008.07.001

Muthamilarasan, M., Suresh, B. V., Pandey, G., Kumari, K., Parida, S. K., and Prasad, M. (2014). Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. *DNA Res.* 21, 41–52. doi: 10.1093/dnares/dst039

Natsume, S., Takagi, H., Shiraishi, A., Murata, J., Toyonaga, H., Patzak, J., et al. (2014). The draft genome of hop (*Humulus lupulus*), an essence for brewing. *Plant Cell Physiol.* 56, 428–441. doi: 10.1093/pcp/pcu169

Oksman-Caldentey, K. M., and Inzé, D. (2004). Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. *Trends Plant Sci.* 9, 433–440. doi: 10.1016/j.tplants.2004.07.006

Peng, Y., Lai, Z., Lane, T., Nageswara-Rao, M., Okada, M., Jasieniuk, M., et al. (2014). *De novo* genome assembly of the economically important weed horseweed using integrated data from multiple sequencing platforms. *Plant Physiol.* 166, 1241–1254. doi: 10.1104/pp.114.247668

Phelps-Durr, T. L. (2010). MicroRNAs in *Arabidopsis*. *Nat. Educ.* 3, 51.

Polashock, J., Zelzion, E., Fajardo, D., Zalapa, J., Georgi, L., Bhattacharya, D., et al. (2014). The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biol.* 14:165. doi: 10.1186/1471-2229-14-165

Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J. R., Sonthirod, C., Sangsrakru, D., et al. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* 105, 288–295. doi: 10.1016/j.ygeno.2015.02.002

Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P. M., et al. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335, 843–847. doi: 10.1126/science.1213561

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum

domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Rahman, A. Y. A., Usharraj, A. O., Misra, B. B., Thottathil, G. P., Jayasekaran, K., Feng, Y., et al. (2013). Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 14:75. doi: 10.1186/1471-2164-14-75

Rezaeieh, K. A. P., Gurbuz, B., and Uyanık, M. (2012). "Biotic and abiotic stresses mediated changes in secondary metabolites induction of medicinal plants," in *Tibbi ve Aromatik Bitkiler Sempozyumu* (Antalya), 13–15.

Ryan, K. G., Swinny, E. E., Markham, K. R., and Winefield, C. (2002). Flavonoid gene expression and UV photoprotection in transgenic and mutant Petunia leaves. *Phytochemistry* 59, 23–32. doi: 10.1016/S0031-9422(01)00404-6

Sakiyama, N. S., Ramos, H. C. C., Caixeta, E. T., and Pereira, M. G. (2014). Plant breeding with marker-assisted selection in Brazil. *Crop Breed. Appl. Biotechnol.* 14, 54–60. doi: 10.1590/S1984-70332014000100009

Salgado, L. R., Koop, D. M., Pinheiro, D. G., Rivallan, R., Le Guen, V., Nicolás, M. F., et al. (2014). *De novo* transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics* 15:236. doi: 10.1186/1471-2164-15-236

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. doi: 10.1073/pnas.74.12.5463

Schatz, M. C., Witkowski, J., and McCombie, W. R. (2012). Current challenges in *de novo* plant genome sequencing and assembly. *Genome Biol.* 13, 243. doi: 10.1186/gb-2012-13-4-243

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nat. Methods* 5, 16–18. doi: 10.1038/nmeth1156

Semagn, K., Bjørnstad, Å., and Ndjiondjop, M. N. (2006). An overview of molecular marker methods for plants. *Afr J. Biotech.* 5, 2540–2568. doi: 10.5897/AJB2006.000-5110

Shibata, S., Fujita, M., Itokawa, H., Tanaka, O., and Ishii, T. (1963). Studies on the constituents of Japanese and Chinese crude drugs. XI. Panaxadiol, a sapogenin of ginseng roots. *Chem. Pharm. Bull.* 11, 759–761.

Shibata, S., Tanaka, O., Soma, K., Iida, Y., Ando, T., and Nakamura, H. (1965). Studies on saponins and sapogenins of ginseng the structure of panaxatriol. *Tetrahedron Lett.* 6, 207–213. doi: 10.1016/S0040-4039(01)99595-4

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116. doi: 10.1038/ng.740

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., and Friess, H. (2008). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nat. Protoc.* 4, 37–43. doi: 10.1038/nprot.2008.216

Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., et al. (2010). *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* 11:262. doi: 10.1186/1471-2164-11-262

Tagu, D., Colbourne, J. K., and Nègre, N. (2014). Genomic data integration for ecological and evolutionary traits in non-model organisms. *BMC Genomics* 15:490. doi: 10.1186/1471-2164-15-490

Tang, H., Sezen, U., and Paterson, A. H. (2010). Domestication and plant genomes. *Curr. Opin. Plant Biol.* 13, 160–166. doi: 10.1016/j.pbi.2009.10.008

Torales, S. L., Rivarola, M., Pomponio, M. F., Gonzalez, S., Acuña, C. V., Fernández, P., et al. (2013). *De novo* assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species Prosopis alba. *BMC Genomics* 14:705. doi: 10.1186/1471-2164-14-705

Valasek, M. A., and Repa, J. J. (2005). The power of real-time PCR. *Adv. Physiol. Educ.* 29, 151–159. doi: 10.1152/advan.00019.2005

Van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.* 12:R102. doi: 10.1186/gb-2011-12-10-r102

VanGuilder, H. D., Vrana, K. E., and Freeman, W. M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* 44, 619. doi: 10.2144/000112776

Varshney, R. K., Nayak, S. N., May, G. D., and Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotech.* 27, 522–530. doi: 10.1016/j.tibtech.2009.05.006

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., et al. (2012a). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44, 1098–1103. doi: 10.1038/ng.2371

Wang, M., Yu, Y., Haberer, G., Marri, P. R., Fan, C., Goicoechea, J. L., et al. (2014a). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46, 982–988. doi: 10.1038/ng.3044

Wang, N., Thomson, M., Bodles, W. J., Crawford, R. M., Hunt, H. V., Featherstone, A. W., et al. (2013). Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol. Ecol.* 22, 3098–3111. doi: 10.1111/mec.12131

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

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., et al. (2012b). The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. *Plant J.* 72, 461–473. doi: 10.1111/j.1365-313X.2012.05093.x

Wegrzyn, J. L., Liechty, J. D., Stevens, K. A., Wu, L., Loopstra, C. A., Vasquez-Gross, H. A., et al. (2014). Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196, 891–909. doi: 10.1534/genetics.113.159996

Wei, L., Xiao, M., Hayward, A., and Fu, D. (2013). Applications and challenges of next-generation sequencing in *Brassica* species. *Planta* 238, 1005–1024. doi: 10.1007/s00425-013-1961-6

Wei, Q., Wang, Y., Qin, X., Zhang, Y., Zhang, Z., Wang, J., et al. (2014a). An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing. *BMC Genomics* 15:1158. doi: 10.1186/1471-2164-15-1158

Wei, X., Wang, L., Zhang, Y., Qi, X., Wang, X., Ding, X., et al. (2014b). Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. *Molecules* 19, 5150–5162. doi: 10.3390/molecules19045150

Wei, Z., Zhang, G., Du, Q., Zhang, J., Li, B., and Zhang, D. (2014c). Association mapping for morphological and physiological traits in *Populus simonii*. *BMC Genet.* 15:S3. doi: 10.1186/1471-2156-15-S1-S3

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., et al. (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* 23, 396–408. doi: 10.1101/gr.144311.112

Xiao, M., Zhang, Y., Chen, X., Lee, E. J., Barber, C. J., Chakrabarty, R., 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. doi: 10.1016/j.jbiotec.2013.04.004

Xu, Y., Gao, S., Yang, Y., Huang, M., Cheng, L., Wei, Q., et al. (2013). Transcriptome sequencing and whole genome expression profiling of *Chrysanthemum* under dehydration stress. *BMC Genomics* 14:662. doi: 10.1186/1471-2164-14-662

Yadav, C. B., Bonthala, V. S., Muthamilarasan, M., Pandey, G., Khan, Y., and Prasad, M. (2014). Genome-wide development of transposable elements-based markers in foxtail millet and construction of an integrated database. *DNA Res.* 22, 79–90. doi: 10.1093/dnares/dsu039

Yang, H., Tao, Y., Zheng, Z., Zhang, Q., Zhou, G., Sweetingham, M. W., et al. (2013). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS ONE* 8:e64799. doi: 10.1371/journal.pone.0064799

Zhang, B., Pan, X., Cobb, G. P., and Anderson, T. A. (2006). Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.* 289, 3–16. doi: 10.1016/j.ydbio.2005.10.036

Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J. Genet. Genomics* 38, 95–109. doi: 10.1016/j.jgg.2011.02.003

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., et al. (2012). The genome of *Prunus* mume. *Nat. Commun.* 3, 1318. doi: 10.1038/ncomms2290

Zhang, Q. J., Zhu, T., Xia, E. H., Shi, C., Liu, Y. L., Zhang, Y., et al. (2014). Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4954–E4962. doi: 10.1073/pnas.1418307111

Zhou, Y., Gao, F., Liu, R., Feng, J., and Li, H. (2012). *De novo* sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus* mongolicus. *BMC Genomics* 13:266. doi: 10.1186/1471-2164-13-266

Zhu, J., Li, W., Yang, W., Qi, L., and Han, S. (2013). Identification of microRNAs in *Caragana intermedia* by high-throughput sequencing and expression analysis of 12 microRNAs and their targets under salt stress. *Plant Cell Rep.* 32, 1339–1349. doi: 10.1007/s00299-013-1446-x