Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

J P P Journal of Pharmacognosy and Phytochemistry

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2019; 8(3): 4546-4548
Received: 04-03-2019
Accepted: 08-04-2019

Suresh Yadav
Division of Genetics, Indian Agricultural Research Institute, New Delhi, India

Bhagchand OLA
Department of Genetics and Plant Breeding, Agriculture University, Jodhpur, Rajasthan, India

Poonam Yadav
Department of Livestock Production and Management, SKN College of Agriculture (SKNAU), Jobner, Jaipur, Rajasthan, India

Poonam Kumari Yadav
Department of Soil science and Agricultural Chemistry, SKN College of Agriculture (SKNAU), Jobner, Jaipur, Rajasthan, India

Jagdish Yadav
Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi, India

Kuldeep Sharma
Department of Entomology, Rajasthan college of Agriculture (MPUAT), Udaipur, Rajasthan, India

Correspondence
Suresh Yadav
Division of Genetics, Indian Agricultural Research Institute, New Delhi, India

Genome sequencing technologies: Applications in crop improvement

Suresh Yadav, Bhagchand OLA, Poonam Yadav, Poonam Kumari Yadav, Jagdish Yadav and Kuldeep Sharma

Abstract
Whole genome sequencing of plants with large genome sizes were considered to be a major challenge about a decade ago. Rapid developments in genome sequencing technologies, has now become faster, cheaper and technically less demanding. These sequencing technologies are undergoing never ending revolution with commercialization of next generation technologies capable of sequencing thousands of millions of nucleotide bases in each run. Using these sequencing technologies, it is possible to sequence or resequence entire plant genomes or sample entire transcriptomes in greater depth than ever before. Rather than sequencing individual genomes, now scientists envision the sequencing of hundreds of related genomes to sample genetic diversity within and between germplasm pools. The increasing availability of DNA sequence information in large number of crop plants enable the discovery of genes and molecular markers associated with diverse important agronomic traits creating new opportunities for crop improvement. Such huge and accurate DNA sequence information impacts many of the current uses of molecular tools in plant evolution, phylogenetics, fingerprinting, linkage mapping and marker assisted selection and breeding. Therefore, this review provides an overview of various genome sequencing technologies that are currently available and in future arrivals along with their applications in *de nova* sequencing and resequencing of crop species.

Keywords: sequencing technologies, crop improvement

Introduction
These advances in sequencing technologies are not only used for de novo sequencing and resequencing of various crop species but also in the area of plant breeding, mutation mapping, transcript profiling study of small RNAs and protein-DNA interaction. These technologies have also been applied for developing single nucleotide polymorphism (SNP) based markers in N number of plant species irrespective of reference genome availability or not. Efforts are underway employing NGS in metagenomic studies, epigenetic modification and gene mining.

Nascent sequencing of crop species: A major landmark in the history of genomics and molecular biology was the sequencing Arabidopsis thaliana, the first plant genome to be sequenced. It is now used as a model plant in molecular biology for understanding several plant traits. Arabidopsis offered several merits over other, being smaller in size (125Mbp), short generation time (annual), and high efficiency of transformation. Scientists from Japan and USA joined hands for sequencing this model plant, in 1996 (Bevan, 1997) [1] and within four years completed and published its genome sequence (The Arabidopsis Genome Initiative, 2000). Next in the que was the draft genome sequencing of indica and japonica rice, an important cereal crop as well as a model monocotyledon that was published in 2002 (Goff *et al.,* 2002; Yu *et al.,* 2002) [2, 3]. With the successful completion of whole genome sequencing of Arabidopsis and rice, many more genome initiatives were formed to sequence other important food crops of the world. *De novo* sequencing provides a golden path for discovery of novel genes and templet for SNP discovery for many plant species for which reference genome of closely related species are unavailable. For large and complex genome, reduced complexity of sequencing approach provides ample sequence depth for SNP discovery without the need of complete sequence. EST is one of the several methods to reduce the complexity of sequencing templet by reducing the low information content repetitive sequences. EST sequencing is a routine method for gene discovery and EST data is valuable tool for mining of plant SNPs (Batley *et al.,* 2007) [4].

Resequencing of well characterized species: Species for which genome or expressed sequence tag (EST) data are available, resequencing proves a boon to scientist involving in molecular breeding. SNP discovery by whole genome and targeted genome sequencing is the very first

application of resequencing. Sequencing of parental genotypes through NGS technologies aligned to reference genome to identify the variation between genotype even at SNP level, which further can be exploited to develop SNP based molecular markers. For instance, sequencing of two Arabidopsis divergent accessions (Bur-0 and Tsu-1) (Using Solexa technology) and aligning with reference accession (Col-1) resulted in 823,325 unique SNP and 79,961 unique 1-3 bp indel polymorphism (Vera *et al.,* 2008) [5]. In resequencing of 6 elite Maize inbred lines including productive commercial hybrid parents from China uncovered more than 1,000,000 SNPs, and 30,000 indel polymorphism (Lai *et al.,* 2010) [7]. With the acquaintance of B73 genome sequence features, gene fraction of maize genome was targeted for resequencing in founder inbred lines of the Nested Associated Mapping (NAM) population (McMullen *et al.,* 2009) [8]. Two datasets comprising 3.3 million SNPs were used to produce first Haplotype (HapMap) which was utilized to study distribution of recombination and diversity along the maize chromosomes. This HapMap and comparative genome hybridization (CGH) experiments enabled identification of >100 low diversity regions possibly associated with domestication and geographic distribution of maize. Very recently whole genome resequencing of *Phytophthora infestans* (HP1031 strain of A2 mating type) was completed covering 10x of its genome size (240Mb) along with three phylotypes of *Ralstonia solanacearum* (RS 50 [phylotype I], RS2 and RS56 [phylotype II], and RS75 [phylotype IV]) covering 50x of the 5.8 Mb genome to decipher the genome wide SNP variations.

Molecular breeding: Determining DNA sequence variation within genome is highly informative for crop genetics and breeding. Genetic variation can be assayed using a variety of molecular markers. NGS proves cheap and efficient methods for identification of SNPs and SSRs (simple sequence repeats) marker development (Robinson *et al.,* 2004; Jewell *et al.,* 2006; Duran *et al.,* 2009) [9, 10]. Even partial genome sequencing would also facilitate maker assisted breeding programs for an efficient introduction of desired traits (Xu *et al.,* 2011) [12]. Through marker assisted selection (MAS) one can select desired lines from large scale population. Once marker has been linked to trait of interest which is more economical using NGS technologies as compared to conventional methods, MAS can be used to modulate the breeding program for crop improvement. Nowadays, plant breeding is dependent on molecular markers for rapid and cost-effective analysis of germplasm and trait mapping. Molecular markers enhance understanding of genetic association that can modify breeding strategy. When a desired trait is under genetic control and phenotypic trials are unsuccessful and unreliable, MAS allows breeder not only early selection of trait but also to carry forward the desired allele to a large number of populations. The availability of sequence data for identifying genes through various sequencing project has led to development of genic or functional markers (FMs) from transcribed region of genome which can be well utilized for extracting putative function. Moreover, GMMs (genic molecular markers) or FMs have been developed using the transcript sequence data available in public domain. Applications of these GMM are accelerating because their discovery is inexpensive and putative functions can often be extracted by homology searches. NGS methods for developing molecular markers in crop breeding can be effectively used in two circumstances; in major crop species for which genome, or transcriptome sequence data already available, and in less characterized species with no or limited genome resources (Vera *et al.,* 2008) [5]. SNP involves finding differences between two sequences even at single nucleotide level. Traditionally SNPs were determined by PCR amplification of genes/genomic region of interest. Individuals are selected from questioned population followed by either direct sequencing of amplicons or cloning, which is further more expensive and time consuming as far as identification of large number of SNPs for application like genetic mapping and association studies are considered. Large amount of data generated by NGS technologies is a treasure for mining SNPs that can be subsequently used in developing molecular markers (Imelfort *et al.,* 2009) [6]. Till date numerous examples of SNP discoveries in various crop species.

Evolutionary genetic studies: In recent years NGS technologies have been used to study whole population rather than just individuals. The study known as population genetics and Handelsman in 1998 coined the term Metagenomics for the same (Handelsman, 2004) [14]. In today's world metagenomics studies are expanding due to the decreasing cost of sequencing. It has the power of exploring the varying microbial population, community structure and composition with respect to diverse environmental condition like soil (Leininger *et al.,* 2006), deep sea (Sogin *et al.,* 2006) [15] and deep mines (Edwards *et al.,* 2006) [16]. Metagenomics is the field of research that allows the study of genomes recovered from various environmental samples bypassing the need for isolation and laboratory cultivation of individual species (Shakira *et al.,* 2009) [17]. Large scale shotgun sequencing approaches allow the discovery of many novel genes found in the environments independent of cultivation efforts.

Mini organ genome sequencing: Mitochondria and chloroplast genome represent a rich source of molecular markers for a range of applications including population genetics, systematics and ecology (Jex *et al.,* 2008) [18]. Both these organelles are result of endosymbiosis, mitochondria are the result of group of aerobic bacteria (the α-proteobacteria), while chloroplasts are plant cell organelle of cynobacterial origin. Understanding their structure provides basis of investigating intercellular physiology and biochemistry. They have a major role in performing essential metabolic and biosynthetic functions of global significance including photosynthesis and amino acid biosynthesis (Kleffmann *et al.,* 2004) [19]. Sequencing of nuclear as well as organelle genome from phylogenetically diverse species will help to know how these genomes have evolved and importance of their encoded genes. Till date a total of 2075 organelle genomes have been sequenced covering 1386 organisms. Of these, 1898 are complete mitochondrial and 122 are complete chloroplast genome sequences (EnterZ Genomes). Sequencing of such huge number of organelle genomes was possible only after the arrival of NGS technologies. The first land plant whose mitochondrial DNA sequenced was that of the liverwort (*Marchantia polymorpha*) having genome size of 186 Kb. Very recently complete chloroplast and mitochondrial genomes of *Boea hygrometrica* was sequenced to have detailed insight into the evolution of plant organellar genomes. Sequencing of these organellar genomes revealed that the smaller chloroplast genome (150kb) contains more coding genes (147 genes covering 72% genome size) and large mitochondrial (510.5 kb) genome contains less genes (65 genes covering 12% of genome). The study also revealed the horizontal gene transfers between the organelles may have begun early in the land plants lineage. Male sterility genes which are significant for

developing hybrid crops are present in mitochondrial genome and therefore sequence analysis of its genome helps improving hybrid crop production (Varshney *et al,* 2009) [13]. Sequencing of chloroplast genome was first accomplished in *Nicotina tabaccum* using Sanger technology.

Functional genomics in model crops: Sequenced legume genome and many other crop species provide rich opportunities for translational biology. The genome sequence of three legume crops, soybean (*Glycine max*) barrel medic (*Medicago truncatula*) and birdsfoot trefoil (*Lotus japonicas*) provide wealthy opportunities for translational biology. Soybean, a valuable protein rich and edible oil source crop while other two are forage crop. Used barrel medic to map based clone of RCT1 (for resistance to *C. trifolii*) gene that confers the resistance to multiple species of anthracnose (*Colletotrichum trifolii*). TERMINAL FLORAL 1 (TLF1) gene a floral regulatory gene, identified in Arabidopsis was used to find gene responsible for determinacy trait in common bean (*Phaseolus vulagris*) (Kwak *et al.,* 2008) [20]. Medicago and lotus have been extensively utilized in studies of nodulation, mycorrhization and plant symbiont signaling while studies are going on to understand the phenylpropanoid and isoflavinoid pathways and secondary metabolites various defense responses, abiotic stress tolerance with the help of barrel medic system.

Agronomic treasure gene mining: With the inception of faster, simple and cheaper sequencing platforms and progress in plant breeding in terms of development of superior and high yielding varieties of agricultural crops, vast amount of sequence information has been available in public databases. It is important to use this genomic information for identification of novel and superior agronomically important genes to develop improved varieties. Structuring of rice genome by clone by clone shotgun approach produced massive amount of sequence data which was very helpful in discovering new functional genes controlling important agronomic traits. Discovery of new genes in rice by sequencing of EST resulted in out characterization of new rice small GTP binding protein coding (osrab5B) and Glucose-6-phosphate dehydrogenase genes. In cultivated barley (*Hordeum vulgare* L.) five SNP sites corresponding to substitution in protein sequence of β-amylase gene (Bmy 1) were genotyped by pyrosequencing and CAPS assay. As an outcome, six different haplotypes of the gene Bmy1 were discovered of which four were identified as previously described alleles Bmy1-Sd1, Bmy1-Sd2L, Bmy1-Sd2H and Bmy1-Sd3, while two were newly discovered.

References

1. Bevan M. Objective: the complete sequence of a plant genome. Plant Cell, 1997; 9:476-478
2. Goff SA, Ricke D, Lan TH, Presting GA. draft sequence of the rice (*Oryza sativa* L. spp. japonica). Science. 2002; 196:92-100.
3. Yu J, Hu S, Wang J, Wong GK. A draft sequence of the rice genome (*Oryza sativa* L. spp. indica). Science. 2002; 296:79-92.
4. Batley J, Edwards D. SNP applications in plants. Springer New York. 2007; 6:95-102.
5. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I *et al*. Rapid transcriptome characterization for nonmodel organism using 454 pyrosequencing, Molecular Ecology. 2008; 17:1636-47.
6. Imelfort M, Duran C, Batley J, Edwards D. Discovering genetic polymorphisms in next-generation sequencing data. J. Plant Biotech. 2009; 7:312-17.
7. Lai J, Li R, Xu X, Jin W, Xu M. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics*. 2010; 42:1027-30.
8. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Huihui Li, Qi Sun Q. Genetic properties of the maize nested association mapping population. Science. 2009; 325:737-40.
9. Robinson A, Love CG, Batley J, Barker G, Edwards D. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics. 2004; 20:1475-76.
10. Jewell E, Robinson A, Savage D, Erwin T, Love CG, Lim GAC *et al*. SR Primer and SSR taxonomy tree: Biome SSR discovery. Nucleic Acids Research. 2006; 34:656-59.
11. Duran C, Appleby N, Clark T, Wood D, Imelfort M, Batley J *et al*. Auto SNPdb: an annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Research. 2009; 37:951-953.
12. Xu P, Wu X, Luo J, Wang B, Liu Y, Ehlers JD *et al*. Partial sequencing of the bottle gourd genome reveals markers useful for phylogenetic analysis and breeding. BMC Genomics. 2011; 12:467-77.
13. Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotech. 2009; 27:522-30.
14. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiology and Molecular Biology Reviews. 2004; 68:669-85.
15. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM. Microbial diversity in the deep sea and the under-explored rare biosphere. Proceedings of the National Academy of Sciences USA. 2006; 103:12115-20.
16. Edwards RA, Rodriguez-Brito B, Wegley L, Haynes M, Breitbart M, Peterson DM. Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genomics. 2006; 7:57-70.
17. Shakira G, Azim A. Metagenomics and its application in rumen ecosystem: potential biotechnological prospects. Pakistan J. of Nutrition. 2009; 8:1309-15.
18. Jex AR, Hu M, Timothy D, Littlewood J, Waeschenbach A, Gasser RB. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics. 2008; 11:9-11.
19. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W *et al*. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Current Biology. 2004; 14:354-62.
20. Kwak M, Velasco D, Gepts P. Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). Journal of Heredity, 2008; 99:283-91.