Available online at www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.6441

ISSN: 2320 – 7051

Int. J. Pure App. Biosci. **6 (3):** 616-622 (2018)

Review Article

OPEN ACCESS

Application of Next Generation Sequencing System in Groundnut Improvement

Preeti Pachauri*, R. S. Shikarvar, Sushma Tiwari, Chitralekha Shyam and Asha Kushwah

Department of Genetics and Plant Breeding, College of Agriculture, Rajmata Vijayaraje Scindia Krishi Vishwa Vidyalaya, Gwalior (Madhya Pradesh)-474 002

*Corresponding Author E-mail: preetishainy786@gmail.com

Received: 27.04.2018 | Revised: 30.05.2018 | Accepted: 8.06.2018

ABSTRACT

Groundnut (Arachis hypogaea L.), is an important oilseed crop worldwide with an allotetraploid genome (AABB, 2n = 4x = 40). Tetraploid nature of cultivated peanut, narrow genetic base of primary gene pool and cultivation of limited genotypes in the process of domestication has resulted in deteriorating genetic resources and low variability for numerous traits. Peanut genome Sequencing has now become a unique choice due to arrival of Next-Generation Sequencing technology platforms. Next-generation sequencing (NGS) is a type of DNA sequencing technology that uses parallel sequencing of multiple small fragments of DNA to determine sequence. NGS technologies give faster sequence data generation and informatics tools to manage and analyze NGS data in a relatively very small time. There is tremendous progress in peanut genome sequencing efforts and its application for improvement i.e., high-quality RAD sequence reads to give out a reference sequence for tetraploid peanuts and diploid peanuts, respectively. Candidate disease resistance genes find by help of sequence information, molecular marker development, tetraploid transcript assembly's guidance and to find genetic exchange between cultivated peanut's sub genomes. Next-Generation Sequencing technology help in development of new chloroplast genomic resources for the genus Arachis based on whole chloroplast genomes from seven species that were sequenced. NGS and ddRAD library combination discovered various SNPs in cultivated peanut. The currently available DNA sequencing techniques such as next generation sequencing technique play key role in groundnut improvement for large number of genotype or to find out new genome region by applying high throughput genotyping or SNP discovery.

Key words: *Next-generation sequencing (NGS), Genotyping, Groundnut, ddRAD*

INTRODUCTION

Groundnut (*Arachis hypogaea* L.) is one of the major economically important legumes and is widely grown in the southern United States, and countries like China, India, and in South and Central America and Africa. Peanut is a good and major source of vegetable oil and protein for human nutrition globally. Peanut offers many health benefits, as well as it is one of the primary food allergens

Cite this article: Pachauri, P., Shikarvar, R.S., Tiwari, S., Shyam, C., and Kushwah, A., Application of Next Generation Sequencing System in Groundnut Improvement, *Int. J. Pure App. Biosci.* **6(3):** 616-622 (2018). doi: http://dx.doi.org/10.18782/2320-7051.6441

Within Asia—India, China, Indonesia, Myanmar, and Vietnam, groundnut oil is utilized for cooking purpose, whereas in Thailand it is consumed as food. Peanut kernels can be consumed directly as ready-to-eat products or indirectly as confectionary. The lower quality oil is utilized in making soaps, detergents, cosmetics, paints, candles, and lubricants. Peanut contain protein (8–15%), lipids (1–3%), minerals (9–17%), and carbohydrate (38–45%) at levels higher than cereal fodder[25]. Peanut ranking fifth among oilseed crops in the world after oil palm, soybean, rapeseed, and sunflower [12].

Peanut farmers face high production cost is one more challenge, this can be exemplified by chemical control of foliar diseases. Comparative and Functional Genomics to aflatoxin contamination of peanut is another concern in Food safety which is important issue threatening human health. Transgenic peanuts with resistance to herbicide, fungus, virus, and insects; tolerance to drought and salinity and better grain quality are under testing at different levels[21]. Conventional breeding procedures together with tools for phenotyping were mostly used in groundnut improvement programs [25]. Crop genetic improvement and production enhanced by novel tools and resources of Genomic research[1]. Wild *Arachis* spp. is potental sources of novel genes for the genetic improvement of cultivated peanut[8]. Peanut is a self-pollinating allotetraploid (AABB, 2n = 4x = 40) with a large genome size (2800 Mb/1C) and a narrow genetic base, that's why leading to very low DNA polymorphism, development of molecular markers and genomic resources in peanut has always been a difficult task[29]. Due to of very fast growth of science, next generation sequencing technologies are currently the newest topic in the field of human, animal's and crop genomics researches[3]. Next-generation sequencing (NGS) is a type of DNA sequencing technology that uses parallel sequencing of multiple small fragments of DNA to determine sequence[27]. NGS can be used to sequence each nucleotide in an individual's DNA, or restricted to smaller portions of the genome such as the exome or a preselected subset of genes[20]. NGS as a clinical tool for genetic diagnosis is suitable in individuals for whom sequencing of a single gene is unlikely to provide a diagnosis especially in diseases with genetic heterogeneity[16]. This review will present the basic principles of the currently available DNA sequencing techniques such as next generation sequencing technique and how it play key role in groundnut improvement or to find out new genome by genome sequencing.

WHOLE GENOME SEQUENCING:

Whole genome sequencing proceeded by a three major technological revolutions: first *generation sequencing* (whole genome shotgun sequencing), *next generation sequencing* (NGS high throughput sequencing) and the *third generation of sequencing* (single molecule long read sequencing) [18].

First generation DNA sequencing (whole genome shotgun sequencing)

Sanger Chain Termination Method is also called a first generation DNA sequencing strategy. The chain termination method of DNA sequencing is important for fairly short strands of 100 to 1000 base pairs. Longer sequences are subdivided into smaller fragments that can be sequenced individually, and then they are re-assembled to give the overall sequence. There are two principal methods are: primer walking (or "chromosome walking") which progresses by the entire strand piece by piece, and shotgun sequencing, which is a faster but more complex process that uses random fragments. In shotgun sequencing, DNA is broken up randomly into many small segments, using the chain termination methods which are sequenced to obtain *reads*. Multiple overlapping reads for the target DNA are obtained by performing numerous rounds of this fragmentation and sequencing. Overlapping ends of different reads can be use by computer programs and assemble them into a continuous sequence.

***Next Generation Sequencing* (NGS high throughput sequencing)**

Next-generation sequencing (NGS) is a high-throughput methodology that enables rapid sequencing of the base pairs in DNA or RNA samples. On the basis of different methods which can be used to immobilize DNA on a solid substrate, there are three technologies which were commercialized; (a) high throughput pyrosequencing on beads, (b) sequencing by ligation on beads and **(c)** sequencing by synthesis on a glass substrate.

(a) High throughput pyrosequencing on beads. Firstly, DNA molecule is sheared with enzymatic based digestion or sonication and ligated with oligonucleotide adapters. Each ligated fragment is then attached to a 28-µm bead, PCR amplified in an oil-water emulsion and pyrosequenced[28]. The main drawback of this technique is reading through homopolymeric sequences, where on time n nucleotides are read as n-1 nucleotides, making this technology prone to high error rates. *Myxococcus xanthus*, a soil inhabitant, was the first bacterium to be sequenced using this technology[33]. Later in a survey of microbial populations from different environments this method was used *viz*. underground mine water, marine, fresh water, fish, corals terrestrial animals and mosquitoes[5].

(b) Sequencing by ligation on beads. Beads with attached PCR amplicons were immobilized on a solid planar substrate and hybridized with a universal PCR primer complementary to the adaptor. Each sequencing cycle proceeds by ligation of a fluorescently labelled DNA octamer to the universal primer revealing the nucleotide positional identity. Subsequent chemical cleavage leaves a pentamer on the DNA template and this process reveals the DNA sequence. Since this platform utilizes a two-base coding system, miscalls are more readily identified resulting in 99.94% accuracy.

(c) Sequencing by synthesis on a glass solid phase surface. The Illumina Genome Analyser (SOLEXA) was explaining in 2006 and 2008[10, 32]. Both forward and reverse primers, with complementarity to the adaptor, are attached to a glass surface by a flexible linker. Attached adaptor flanked DNA fragments on glass surface are hybridized on to the forward and reverse primers. Bridge PCR then amplifies the DNA fragment using formamide based denaturation and *Bst* DNA polymerase, resulting in clonal amplicons "*cluster*" formation. Amplicons formed from a single DNA fragment will cluster in a single physical location on the array. Following cluster generation, the sequencing primer hybridizes to the universal sequence flanking the region of interest. After that Sequencing proceeds in cycles with four nucleotides and a modified DNA polymerase. Chemically cleavable fluorescent reporters labelled on nucleotides are with a group at the 3′-OH end thereby allowing only a single base incorporation in each cycle. Each cycle extends a single base followed by the chemical cleavage of the fluorescent reporter that will identify the incorporated nucleotide.

Third generation sequencing (single molecule long read sequencing-SMRT)

Third generation sequencing technology was designed to deal with the limitation of NGS technologies. In this technology single DNA molecules are directly sequenced there by reducing low error rates by avoiding amplification associated bias, intensity averaging, phasing or synchronization problems. First commercially released long read methodology was single-molecule-real-time (SMRT) technology[9]. The main advantage of the SMRT sequencing is the read length obtained. The original C1 generation sequencer formed a read length of about 1,500 bp. Currently, C4 chemistry protocols provided for an average read length of 10-kbp. The typical throughput of a PacBio RS II

system is 0.5–1 billion bases per SMRT cell. However, the platform has considerably higher error rates (approximately 11- to 15%).

Role of Next Generation Sequencing In Groundnut:

This method provided great opportunities for low-cost and fast DNA sequencing. NGS overcome the limitations of conventional DNA sequencing methods and has found usage in a wide range of molecular biology applications[31]. Utilization of these NGS technologies for de novo sequencing, genome re-sequencing and whole genome and transcriptome analysis. New generation of sequencers, like the Single-Molecule Real-Time (SMRT) Sequencer, HeliScope Single Molecule Sequencer, and the Personal Genome Machine (PGM) are becoming available that is responsible to generating longer sequence reads in a shorter time and having a lower cost for sequencing[19]. Genetic mapping of induced or natural genomic variation remains a powerful approach to understand the function of genes in so many biological processes. Single nucleotide polymorphism (SNP) discovery and genotyping are essential to genetic mapping. Isolation of combined RAD marker with massively parallel, high-throughput Illumina sequencing helps to make a genotyping platform that can quickly and cost-effectively discover new SNP markers and simultaneously many individuals genotype [23]. SNPs are the most powerful type of genetic marker and high density nature of it makes them ideal for studying inheritance of genomic regions[2]. On former side, reduced representation RAD (restriction-site associated DNA tag) libraries of 11 peanut genotypes were sequenced on Illumina HiSeq 2000 platform. A total of 102 million reads, approximately 10 Gb of sequence data, were collected. High-quality RAD sequence reads of genotype "Tiffrunner" is a runner market-type peanut with a high level of resistance to TSWV (tomato spotted wilt virus), and moderate resistance to early (Cercospora arachidicola) and late leaf spot (Cercosporidium personatum) and accession *A. ipaensis* were clustered to make a non-redundant set of representative sequences and to serve as a reference sequence for tetraploid peanuts and diploid peanuts, respectively[30].

NGS technology used for peanut transcriptome analysis has been employed to develop a large number of ESTs and unigenes from various tissue and peanut seeds developmental stages. Large numbers of ESTs generated through high throughput sequencing technologies will greatly allows peanut functional genomics studies. Using next-generation sequencing technologies, new chloroplast genomic resources have been developed for the genus *Arachis* based on whole chloroplast genomes from seven species that were sequenced. Hence, using a next-generation sequencing platform, chloroplast genomes sequencing of seven *Arachis* species give a valuable chloroplast genome information, such as SNPs, microsatellites, indels, and highly variable regions for this genus, by comparing of chloroplast genomes to one another[8]. However, the chloroplast genome is non-recombining and uniparentally inherited, creation of a valuable information source for phylogenetics improvement, species identification, and resolution[6].

Forage peanut (*Arachis pintoi*), locally famous as 'Mani Forrajero Perenne' is a Leguminosae plant native in central Brazil[13] is a diploid perennial herb (2n=20) of tetrafoliolate leaves having a heights between 20 to 40 cm and is well useful for clonal propagation by way of stolons[15, 26]. Forage peanut (*A. pintoi*) transcriptome from a bulk sample of leaves using Next-Generation sequencing technique, three dissimilar virus species from the *Potyviridae*, *Luteoviridae* and *Alphaflexiviridae* families were identified. PeMoV and TuYV has been earlier reported infecting peanut; in contrast, the *Alphaflexiviridae* member possibly corresponds to a new virus distantly related to the genus *Allexi virus* and which would be a first and new report of this family in *A. pintoi*. *A. pintoi* infected by at least three viruses in

Colombia then it is important to initiate a germplasm clean-up program in the seed and propagation material which can be further used[24]. Detail knowledge of RNA and DNA viruses as well as viriods in a plant sample at single-nucleotide resolution is made possible by one NGS run of total small RNAs, followed by data mining with homology-dependent and homology-independent computational algorithms. Application of NGS technologies to pathogen discovery are major challenges[35]. Plant genotyping can give benefit to plant breeding through selection of individuals resistant to biotic and abiotic stress causing substantial losses in agriculture. Next-generation sequencing (NGS) platforms play role to change the impact of sequencing on our knowledge of crop genomes and gene regulation.

CURRENT DEMAND OF NGS IN PLANT BREEDING:

Genome sequencing of some important crops is becoming a first step for ascertainment of the genome and evolution. Determination of the sequence helps in modification of specific genes using genome editing or identification of appropriate mutations and gets a new allelic form. Typical plant breeding programmes are mostly based on morphological traits, but due to rising of knowledge in genetic background of important agronomic traits, there has been a quick demand for genotype-based selection[22]. High-throughput sequencing is a key foctor in breeding having a large amount of genomic data helps in systematic characterization of phenotypes for a wide range of traits. In various plant species, GWAS (genome-wide association studies) has been widely used to overcome some of the limitation essential in bi-parental linkage mapping[17] and are applicable to a wider germplasm base. MAS (marker-assisted selection) use molecular marker which is located inside or nearby a locus with the known phenotypic effect, and this process is known as marker-assisted selection. In plant breeding use of this approach result that some traits are difficult to control by the standard phenotypic selection and some traits expression dependent on environmental conditions or developmental stages[36]. Another problem is difficulty in maintenance of recessive alleles during backcrossing or pyramiding multiple monogenic traits. This process is utilized in selection of small numbers of genes with the major phenotypic effect. MAS connected with genetic mapping, and this process comprises multiple consecutive steps from development of mapping populations, genetic mapping, and marker validation to MAS application. MAS were used in several important crops, such as wheat[14], apple[11] or peanut[1].

CONCLUSION

In the plant research area, next-generation sequencing (NGS) technologies is a good approach for assembly of crop reference genomes, transcriptome sequencing for the gene expression study, whole-genome molecular marker development, and markers identification in known-function genes. Few of them become useful in the breeding of various crops. NGS technologies give faster sequence data generation and informatics tools to manage and analyze NGS data in a relatively very small time. There is tremendous progress in peanut genome sequencing efforts and its application for improvement i.e., high-quality RAD sequence reads to give out a reference sequence for tetraploid peanuts and diploid peanuts, respectively. Further, it would be possible to combine genetic and physical maps in order to facilitate gene cloning and molecular breeding effectively. Peanut genomic resources continue to develop large EST data sets by NGS technologies and thousands of SNP and SSR markers will be developed.

REFERENCES

1. Guo, B.Z., Chen, C.Y., Chu, Y., Holbrook, C.C., Akins, P.O. and H. T., Stalker Advances in genetics and genomics for sustainable peanut production, in Sustainable Agriculture and New Biotechnologies, *N. Benkeblia, Ed.,* pp. 341–367 (2012).
2. Berger, J., Suzuki, T., Senti, K.A., Stubbs, J., Schaffner, G., Genetic mapping with

SNP markers in Drosophila, *Nat Genet* **29:** 475–481J (2001).

3. Chandra, S.P., Rafal, S., Andrzej, T., Sequencing technologies and genome sequencing, *J Appl Genet*, **52:** 413-435 (2011).
4. Chu, Y., Wu, C.L., Holbrook, C.C., Tillman. B.L., Person. G., Ozias-Akins, P., Marker-assisted selection to pyramid Nematode resistance and the high oleic trait in peanut, *Plant Genome*, **4:** 110–117 (2011).
5. Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Functional metagenomic profiling of nine biomes, *Nature*, **452:** 629–632 (2008).
6. Dong, W., Liu, J., Yu, J., Wang, L. & Zhou, S., Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding, *PLOS ONE* **7** (2012).
7. Dongmei, Yin, Yun, Wang, Xingguo Zhang, Xingli Ma, Xiaoyan He & Jianhang Zhang, Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of *Arachis*," Scientific Reports **7:** Article number: 11649 (2017).
8. Dongmei, Yin, Yun, Wang, Xingguo Zhang, Xingli Ma, Xiaoyan He & Jianhang Zhang Development of chloroplast genome resources for peanut (*Arachis hypogaea* L.) and other species of *Arachis*, Scientific Reports **7** (2017).
9. Eid J., Fehr, A., Gray, J., Luong K., Lyle J., Otto G., Real-time DNA sequencing from single polymerase molecules, Exch. Organ. *Behav. Teach. J*. **323:** 133–138 (2009).
10. Fedurco, M., Romieu, A., Williams, S., Lawrence, I., Turcatti, G., BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies, *Nucleic Acids Res.* **34: e** 22. 10 (2006).
11. Flachowsky, H., Le Roux, P.M., Peil, A., Patocchi, A., Richter, K., Hanke, M.V., Application of a high-speed breeding technology to apple (*Malus* × *domestica*) based on transgenic early flowering plants and marker-assisted selection, *New Phytologist*, **192:** 364–377 (2011).
12. Food Agriculture Organization of United Nations Available online at: http://faostat.fao.org/DesktopDefault.aspx?PageID=339 (2013).
13. Krapovickas, A. and Gregory, W.C., Taxonomía del género *Arachis* (Leguminosae), *Bonplandia* **8(1):** 1-186 (1994).
14. Kumar, J., Mir, R.R., Kumar, N., Kumar, A., Mohan, A., Prabhu, K.V., Balyan, H.S., Gupta, P.K., Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat, *Plant Breeding*, **129:** 617–621 (2010).
15. Lavia, G.I., Ortiz, A.M., Robledo, G., Fernández, A. and Seijo, G., Origin of triploid *Arachis pintoi* (Leguminosae) by autopolyploidy evidenced by FISH and meiotic behaviour, *Annals of Botany* **108(1):** 103-111 (2011).
16. Lee, H., Deignan, J.L., Dorrani, N., Strom, S.P., Kantarci, S., Clinical exome sequencing for genetic identification of rare Mendelian disorders, JAMA **312:** 1880-1887 (2014).
17. Lehmensiek, A., Bovill, W., Wenzl, P., Langridge, P., Appels, R., Genetic mapping in the triticeae. In: Feuillet C., Muehlbauer G.J. (eds): Genetics and Genomics of the Triticeae, *Heidelberg, Springer:* 201–236 (2009).
18. Loman N. J., Pallen M. J., Twenty years of bacterial genome sequencing *Nat. Rev. Microbiol*. **13:** 1–9 (2015).
19. M. Thudi, Y. Li, S. A. Jackson, G. D. May, and Varshney, R. K., Current state-of-art of sequencing technologies for plant genomics research, Briefings in Functional Genomics, **11:** pp. 3–11 (2012).
20. Mardis, E.R., Next-generation DNA sequencing methods, *Annu Rev Genomics Hum Genet*, **9:** 387-402 (2008).
21. Murali, T., Variath, and Janila, P., Economic and Academic Importance of Peanut, Springer International Publishing AG 2017 (2017).

22. Myles, S., Chia, J., Hurwitz, B., Simon, C., Zhong, G.Y., Buckler, E., Ware, D. Rapid genomic characterization of the genus *Vitis*, *PLoS ONE*, **5:** e8219 (2010).
23. Nathan, A., Baird, Paul, D., Etter , Tressa S. Atwood, Mark, C., Currey, Anthony, L., Shiver, Zachary, A. Lewis, Eric U. Selker, William A. Cresko and Eric A., "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers," *Johnson Published*: October 13, 2008 (2008).
24. Pablo Andrés Gutiérrez Sánchez1, Helena Jaramillo Mesa1 and Mauricio Marín Montoya, Next generation sequence analysis of the forage peanut (Arachis pintoi)virome, *Rev. Fac. Nal. Agr. Medellín* **69(2):** (2016).
25. Janila, P., Nigam, S.N., Manish, K., Pandey, Nagesh, P. and Rajeev, K., Varshney, Groundnut improvement: use of genetic and genomic tools, Front Plant Sci. **4:** 23 (2013).
26. Rincon, C.A., Cuesta, P.A., Pérez, R., Lascano, C.E. and Ferguson, J., Mani forrajero perenne (*Arachis pintoi*). Una alternativa para ganaderosy agricultores," Boletin Tecnico No. 219. *ICA/CIAT, Palmira.* 23 p (1992).
27. Rizzo, J.M., Buck, M.J., Key principles and clinical applications of "next-generation" DNA sequencing, *Cancer Prev Res (Phila)*, **5:** 887-900 (2012).
28. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., Nyrén, P.L., Real-time DNA sequencing using detection of pyrophosphate release, *Anal. Biochem*, **242:** 84–89 (1996).
29. Stalker, H.T., Mozingo, L.G., Molecular markers of Arachis and marker-assisted selection, *Peanut Sci 2001*, **28:** 117–123 (2001).
30. Gupta, S.K., Baek, J., Garcia, N.C. and Varma, P.R., Genome-wide polymorphism detection in peanut using next-generation restriction-site-associated DNA (RAD) sequencing, *Molecular Breeding 2015*, **35:** 145 (2015).
31. Ari, S. and Arikan, M., Next-Generation Sequencing: Advantages, Disadvantages, and Future. Plant Omics, *Trends and Applications* pp 109-13 (2016).
32. Turcatti, G., Romieu, A., Fedurco, M., Tairi, A.P., A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis, *Nucleic Acids Res*. **36:** e25 (2008).
33. Vos, M., Velicer, G.J., Genetic population structure of the soil bacterium *Myxococcus xanthus*at the centimeter scale, *Appl. Environ. Microbiol*. **72:** 3615–3625 (2006).
34. Zhuang, W.J. H., Chen, P.K., Nancy, Isolation and characterization of important genes toward improvement peanut resistance to Aspergillus flavus, in Proceedings of the 5th International Conference of the Peanut Research Community on Advances in Arachis *through Genomics and Biotechnology*, p. 28 (2011).
35. Wu, Q., Ding, S., Zhang, Y. and Zhu, S., Identification of viruses and viroids by Next-Generation Sequencing and homology dependent and homology independent algorithms, *Annual Review of Phytopathology* **53:** 1-20 (2015).
36. Xu, Y., Developing marker-assisted selection strategies for breeding hybrid rice, *Plant Breeding Reviews*, **23:** 73–174 (2003).