Meridian Vol. 5 (1), 1- 5, 2016. ISSN 2278-750x

RECENT ADVANCES IN NEXT GENERATION SEQUENCING (NGS) AND ITS IMPACT IN PLANT RESEARCH

Aiswariya K. S. & Geethu Elizabath Thomas*

With it's unprecedented through put, scalability and speed Next Generation Sequencing (NGS) enables researchers to study biological systems at a level never before. Recent advancement in NGS includes Nanopore DNA sequencing, Tunneling currents DNA sequencing, sequencing by hybridization, microscopy based techniques, micro fluidic Sanger sequencing, RNAP sequencing, in vitro virus high throughput sequencing etc. Next generation sequencing applies to genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA protein interactions (chip sequencing) and epigenome characterization. NGS is applied in plant research for genotyping, crop improvement and for rapid marker development in molecular plant breeding. It is used for plant transcriptomics including gene discovery, transcript quantification and marker discovery for non model plants as well as transcript annotation and quantification, small RNA discovery and transcription analysis for model plants. NGS enables the collection of genome-wide information on genetic variation, single nucleotide polymorphism (SNP) markers and identification of genes of adaptive importance which will help considerably in investigating the mechanisms that are important in a conservation genetic context such as inbreeding depression and local adaptation.

Keywords: RNA, single nucleotide polymorphism (SNP), genotyping, epigenome.

With fast development and wide applications of next-generation sequencing (NGS) technologies, genomic sequence information is within reach to aid the achievement of goals to decode life mysteries, make better crops, detect pathogens, and improve life qualities. NGS technologies create a vast amount of data, presenting many problems to computational biologists, bioinformaticians, and end-users endeavouring to assemble and analyze NGS data in novel ways. NGS methods have increased capabilities far beyond that of traditional Sanger sequencing (Sanger et al., 1977), allowing millions of bases to be sequenced in one round at a fraction of the cost. As the costs and capabilities of these technologies continue to improve, whole new fields of study are being opened, allowing us to analyze a variety of data sets and approach questions never possible before.

Next generation sequencing methods

The high demand for low-cost sequencing has driven the development of high-throughput sequencing (or next-generation sequencing) technologies that parallelize the sequencing process, producing thousands or millions of sequences concurrently. High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods. The latest sequencing techniques are introduced here. Nano pore sequencing technique involves single-molecule detection and analytical capabilities that are achieved by electrophoretically driving molecules in solution through a nano-scale pore. This method is based on the readout of electrical signals occurring at nucleotides passing by alpha-hemolysin pores covalently bound with cyclodextrin (Dekker, 2007; Brantonet al., 2008; Stoddartet al., 2009).Tunnelling currents DNA sequencing uses measurements of the electrical tunnelling currents across single-strand DNA as it moves through a

[1]Department of Botany, St. Thomas' College, Thrissur- 01
*geethuelizabath@gmail.com; Mob: 9447797920

Paper submitted October 2015. Revised and Accepted January 2016.

channel. Depending on its electronic structure each base affects the tunnelling current differently, allowing differentiation between different bases (Di Ventra, 2013). Sequencing by hybridization is a non-enzymatic method that uses DNA microarray. A single pool of DNA whose sequence is to be determined is fluorescently labelled and hybridized to an array containing known sequences. Strong hybridization signals from a given spot on the array identify its sequence in the DNA being sequenced (Hanna et al., 2000; Qin et al., 2012). Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), has specifically been investigated as an alternative method to gel electrophoresis for visualizing DNA fragments. With this method, DNA fragments generated by chain-termination sequencing reactions are compared by mass rather than by size. The mass of each nucleotide is different from the others and this difference is detectable by mass spectrometry. Single-nucleotide mutations in a fragment can be more easily detected with MS than by gel electrophoresis alone (Monforte & Becker, 1997; Edwards et al., 2005; Hall et al., 2005). The entire thermocycling amplification of DNA fragments as well as their separation by electrophoresis is done on a single glass wafer (approximately 10 cm in diameter) thus reducing the reagent usage as well as cost in Microfluidics Sanger sequencing (Kanet al., 2004; Chen et al., 2010). Identification of DNA base pairs within intact DNA molecules by enzymatically incorporating modified bases, direct visualization and identification of individually labelled bases using electron microscopy was demonstrated in 2012 (Bell et al.) and has be acclaimed a next generation sequencing technique. In vitro virus high-throughput sequencing has been developed to analyze full sets of protein interactions using a combination of 454 pyrosequencing and an in vitro virus mRNA display method. The combined method was titled IVV-HiTSeq and can be performed under cell-free conditions (Fujimori et al., 2012).

Next generation sequencing platforms

Roche 454 System

Roche 454 sequencer uses pyrosequencing technology which relies on the detection of pyrophosphate released during nucleotide incorporation. The library DNAs with 454-specific adaptors are denatured into single strand and captured by amplification beads followed by emulsion PCR. Then on a picotiter plate, one of dNTP (dATP, dGTP, dCTP, dTTP) will complement to the bases of the template strand with the help of ATP sulfurylase, luciferase, luciferin, DNA polymerase, and adenosine 5′ phosphosulfate (APS) and release pyrophosphate (PPi) which equals the amount of incorporated nucleotide. The ATP transformed from PPi drives the luciferin into oxyluciferin and generates visible light. At the same time, the unmatched bases are degraded by apyrase. Then another dNTP is added into the reaction system and the pyrosequencing reaction is repeated (Berkaet al., 2010).The software for processing data is GS FLX Titanium system. One disadvantage is that it has relatively high error rate in terms of poly-bases longer than 6 bp. But its library construction can be automated, and the emulsion PCR can be semi automated which could reduce the manpower in a great extent.

AB SOLiD System (Sequencing by Oligo Ligation Detection)

SOLiD was purchased by Applied Biosystems in 2006. The sequencer adopts the technology of two-base sequencing based on ligation sequencing. On a SOLiD flowcell, the libraries can be sequenced by 8 base-probe ligation which contains ligation site (the first base), cleavage site (the fifth base), and 4 different fluorescent dyes (linked to the last base). The fluorescent signal will be recorded during the probes complementary to the template strand and vanished by the cleavage of probes' last 3 bases. The sequence of the fragment can be deduced after 5 round of sequencing using ladder primer sets.The short read length and resequencing only in applications is still its major shortcoming (Mardis, 2008). Application of SOLiD includes whole genome resequencing, targeted resequencing, transcriptome research (including gene expression profiling, small RNA analysis, and whole transcriptome analysis), and epigenome (like ChIP-Seq and methylation). The employed software is SOLiD software.

Illumina GA/HiSeq System

In 2006, Solexa released the Genome Analyzer (GA), and in 2007 the company was purchased by Illumina. The sequencer adopts the technology of sequencing by synthesis (SBS). The library with fixed adaptors is denatured to single strands and grafted to the flowcell, followed by bridge amplification to form clusters which contains clonal DNA fragments. Before sequencing, the library splices into single strands with the help of linearization enzyme, and then four kinds of nucleotides (ddATP, ddGTP, ddCTP, ddTTP) which contain different cleavable fluorescent dye and a removable blocking group would complement the template one base at a time, and the signal could be captured by a (charge-coupled device) CCD. HiSeq software is used for processing.

Of the three NGS systems described before, the Illumina HiSeq 2000 features the biggest output and lowest reagent cost, the SOLiD system has the highest accuracy, and the Roche 454 system has the longest read length.

Applications of NGS

Next generation sequencing applies to genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA protein interactions (chip sequencing) and epigenome characterization. Genome sequencing or de-novo sequencing is figuring out the order of DNA nucleotides, or bases, in a genome—the order of As, Cs, Gs, and Ts that make up an organism's DNA (Skovgaardet al., 2011). Whole genome re-sequencing aims to sequence the individual whose reference genome is already known. As reference genome sequences become increasingly available for many species, cataloguing sequence variations and understanding their biological consequences have become major research goals (Mills et al., 2011). With next generation sequencing technology, all transcriptional activity, for both coding and non-coding regions, in any organism can be characterized without prior annotation information. This allows the identification of regulatory RNAs, annotation of coding SNPs, determination of the relative abundance of transcripts and delivers unbiased transcriptome information (Fehnigeret al., 2010). Epigenomics refers to a more global analysis of epigenetic changes across the entire genome. The epigenetic changes include DNA methylation and histone modification, which regulate high-order DNA structure and gene expression (Guet al., 2011).ChIP-Seqcombines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify binding sites of DNA-associated proteins, and can be used to precisely map global binding sites for any protein of interest (Johnson et al., 2007).

NGS in plant research

The recent advances in genome sequencing, through the development of second generation sequencing technologies and beyond, provide opportunities to develop millions of novel markers, in non-model crop species, as well as identification of genes of agronomic importance. Mining of genes of agronomic importance provides the knowledge of the gene underlying a trait enables the transfer of the trait between cultivars and even species using genetic modification. Alternatively, the gene conferring the favourable trait may be incorporated into a cultivar by marker-assisted selection (MAS) breeding. NGS analysis of DNA/RNA of individuals is an emerging approach for SNP discovery in plant and animal species. NGS-based SNP discovery is very challenging in the species that do not have a reference genome because of poor alignment of short sequence reads of different individuals and genotypes generated by current NGS technologies (Azamet al., 2012). Zalapaet al. (2012)show the power of NGS for developing SSRs in plants through a review of their work in cranberry and 95 other studies that developed SSRs using Sanger, Illumina, and 454 technologies. Strickleret al. (2012) review methods to design RNA-seq projects and to analyse and interpret the data. They provide a set of ways in which transcriptome data can be used, such as characterizing differential expression or tissue-specific transcripts, and SNP identification that can be useful in designing markers for mapping and studying evolution (Ward et al., 2012; Yant, 2012; He et al., 2012).Grover et al. (2012)describe how targeted sequence capture, coupled with NGS, opens up genomic resources to nonmodel organisms, allowing us to address questions such as parentage, gene flow, population divergence, phylogeography, diversity, domestication and improvement, phylogeny, hybrid identification, introgression, and polyploid parentage.

Conclusion

As NGS technologies continue to improve, their scope and application will correspondingly expand within and across scientific disciplines. Plant biology has much to gain from increasing our technological capacity in genomics, with applications reaching from plant breeding to evolutionary studies. In terms of comparative genomics, the increasing number of fully sequenced plant genomes will enable greater understanding of genetic, genomic, developmental, and evolutionary processes that create the diversity of plant life on earth. The innovation and application of NGS technologies paints a bright future for plant biology and all areas of life science research.

REFERENCES

Azam,S., Thakur, V., Ruperao, P., Shah,T., Balaji, J., Amindala, B., andFarmer, D. 2012. Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results for the identification of SNPs in chickpea (Cicerarietinum; Fabaceae), a crop species without a reference genome. American Journal of Botany .99: 186–192.

Bell, D. C., Thomas, W. K., Murtagh, K. M., Dionne, C. A., Graham, A. C., Anderson, J. E., & Glover, W. R. 2012. DNA base identification by electron microscopy. Microscopy and Microanalysis, 18(05), 1049-1053.

Berka, J., Yi-Ju Chen, J. H.,Leamon, S.,Lefkowitz, K. L.,Lohman, V. B.,Makhijani, J. M., Rothberg, G. J.,Sarkis, M., Srinivasan, and Weiner,M.P. 2010. Bead emulsion nucleic acid amplification.U.S. Patent 7,842,457, issued November 30.

Branton, D., David, W.Deamer., Andre,Marziali., Hagan, Bayley., Steven, A .Benner., Thomas, Butler., Massimiliano,Di .Ventra., Slaven,Garaj., Andrew,Hibbs., Xiaohua, Huang., Stevan, B. Jovanovich., Predrag,S.Krstic., Stuart, Lindsay., Xinsheng, Sean .Ling., Carlos ,H .Mastrangelo., Amit ,Meller., John ,S. Oliver., Yuriy,V .Pershin., Michael ,Ramsey,J., Robert ,Riehn., Gautam, V.Soni., Vincent, Tabard-Cossa., Meni,Wanunu., Matthew ,Wiggin .andJeffery, A.Schloss. 2008. The potential and challenges of nanopore sequencing. Nature biotechnology, 26(10), 1146-1153.

Chen, Y. J., Roller, E. E., and Huang, X. 2010. DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device. Lab on a Chip,10(9), 1153-1159.

Dekker, C. 2007. Solid-state nanopores. Nature nanotechnology, 2(4), 209-215.

Di Ventra, M. 2013. Fast DNA sequencing by electrical means inches closer.Nanotechnology, 24(34), 342501-342501.

Edwards, J. R., Ruparel, H., and Ju, J. 2005. Mass-spectrometry DNA sequencing. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 573(1), 3-12.

Fehniger, T. A., Wylie,T., Germino,E., et al.2010. Next-generation sequencing identifies the natural killer cell microRNA transcriptome. Genome Research. vol. 20, no. 11, pp. 1590–1604.

Fujimori, S. ,Naoya, H. H.,Ohashi, K.,Masuoka, A.,Nishikimi, Y., Fukui, T.,Washio, T.,Oshikubo, T., Yamashita and Sato,E.M. 2012. Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. Scientific reports. 2 : 691

Gu, H. Z., Smith,D., Bock,C., Boyle,P., Gnirke,A., and Meissner,A.2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling.Nature Protocols. vol. 6, no. 4, pp. 468–481.

Hall, T. A., Budowle, B., Jiang, Y., Blyn, L., Eshoo, M., Sannes-Lowery, K. A.,Sampath, R.,Drader, J. J.,Hannis, J. C., Harrell, P., Samant, V., White, N.,Ecker, D. J. and Hofstadler, S. A. 2005. Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. Analytical biochemistry, 344(1), 53-69.

Hanna, George. J., Victoria. A. Johnson., Daniel. R. Kuritzkes., Douglas .D. Richman., Javier Martinez-Picado, Lorraine.Sutton., Darren Hazelwood,J andRichard,T.D. 2000. Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase.Journal of clinical microbiology 38, no. 7: 2715-2721.

He R., Kim,M.J., Nelson,W., Balbuena,T.S., Kim,R., Kramer,R., Crow,J.A.,etal. 2012. Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmitesaustralis (Poaceae), reveals genes involved in invasiveness and rhizome specificity. American Journal of Botany.99: 232–247.

Johnson, D. S.,Mortazavi,A., Myers,R.M., and Wold,B.2007. Genome-wide mapping of in vivo protein-DNA interactions.Science. vol. 316, no. 5830, pp. 1497–1502.

Kan, C. W., Fredlake, C. P., Doherty, E. A., and Barron, A. E. 2004. DNA sequencing and genotyping in miniaturized electrophoresis systems.Electrophoresis, 25(21-22), 3564-3588.

Liu, Lin, Li,Y., Li,S., Ni Hu, He,Y., Pong,R., Lin,D., Lu,L., and Law,M.2012. Comparison of next-generation sequencing systems.BioMed Research International.

Mardis.E. R. 2008.The impact of next-generation sequencing technology on genetics.Trends in Genetics.vol. 24, no. 3, pp. 133–141.

Mills, R. E., Walter,K., Stewart,C.,et al. 2011. Mapping copy number variation by population-scale genome sequencing. Nature. vol. 470, no. 7332, pp. 59–65.

Monforte, J. A.and Becker, C. H. 1997.High-throughput DNA analysis by time-of-flight mass spectrometry. NatureMedicine 3 (3): 360–362. doi:10.1038/nm0397-360.PMID 9055869

Qin, Y., Schneider, T. M., and Brenner, M. P. 2012.Sequencing by hybridization of long targets. PloSone.7(5), e35819.

Sanger, F., Nicklen,S., and Coulson,A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences,USA.74: 5463–5467.

Skovgaard, M.,Bak,andLøbner-Olesen,A.2011. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing.Genome Research. vol. 21, no. 8, pp. 1388–1393.

Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G.,and Bayley, H. 2009.Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proceedings of the National Academy of Sciences.106(19), 7702-7707.

Strickler S.R., Bombarely.A., and Mueller,L.A. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. American Journal of Botany 99: 257–266.

Ward ,J.A., Ponnala,L., and Weber,C.A. 2012. Strategies for transcriptome analysis in nonmodel plants. American Journal of Botany .99: 267–276.

Yant ,Y. 2012. Genome-wide mapping of transcription factor binding reveals developmental process integration and a fresh look at evolutionary dynamics.American Journal of Botany. 99: 277–290.

Zalapa ,J.E., Cuevas,H., Zhu,H., Steffan,S., Senalik,D., Zeldin,E., McCown,B.,etal.2012. Using next-generation sequencing approaches for the isolation of simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany. 99: 193–208.