International Journal of Agriculture, Environment and Biotechnology
Citation: IJAEB: 9(6): 947-950, December 2016
DOI: 10.5958/2230-732X.2016.00121.2

BIOTECHNOLOGY

Next Generation Sequencing Platforms and its Applications in Genomics

Bommesh J.C.[1]*, Kattula Nagaraju[1], Sunilkumar M.K.[2], Manjunatha Gowda D.C.[3], Mallik M.[4] and Ravi Y.[5]

[1]Division of Vegetable Crops, ICAR- IIHR, Bengaluru-560089, India
[2]Department of Vegetable Science, University of Horticultural Sciences, Bagalkot – 587104, India
[3]Directorate of Onion and Garlic Research, Rajgurunagar-410505, India
[4]Division of Genetics ICAR- IARI, New Delhi - 110012, India
[5]Department of Plantation, Spice, , Medicinal and Aromatic Crops , University of Horticultural Sciences, Bagalkot – 587104, India

*Corresponding author: bommesh.jc@icar.gov.in

Paper No. 524 Received: 13 July 2016 Accepted: 13 December 2016

Abstract

DNA sequencing technology is undergoing a revolution with the commercialization of next generation technologies. Over the past eight to ten years massively parallel DNA sequencing platforms have become widely available with reducing the cost of DNA sequencing. Next generation platform (NGS) includes Helicos Heliscope™, Pacific Biosciences SMRT, Ion Torrent, Oxford Nanopore, *etc*. These platforms have the potential to dramatically accelerate biological research, by enabling the comprehensive analysis of genomes, transcriptomes and interactomes to become inexpensive, routine and widespread. Variant discovery by re-sequencing targeted regions of interest or whole genomes, de novo assemblies of prokaryotic and eukaryotic genomes. Cataloguing the transcriptomes of cells, genome-wide profiling of epigenetic marks and chromatin structure is using other seq-based methods and species classification and gene discovery by metagenomics studies.

Highlights

- NGS provids unprecedented opportunities for high-throughput functional genomic research
- NGS technologies have been applied in a variety of contexts, *viz.* whole genome sequencing, targeted resequencing, discovery of transcription factor binding sites, and noncoding RNA expression profiling.

Keywords: Next generation sequencing (NGS), Template libraries, *de novo*

The sequencing of the reference human genome was the capstone for many years of hard work spent developing high-throughput, high-capacity production DNA sequencing and associated sequence finishing pipelines. Over the past eight years, there has been a fundamental shift away from the application of automated Sanger sequencing for genome analysis to NGS. Next generation sequencing (NGS) technologies constitute various strategies that rely on a combination of template preparation, sequencing and imaging, and genome alignment and assembly methods (Metzker, 2010). The arrival of NGS technologies in the marketplace has changed the way we think about scientific approaches in basic, applied and clinical research (Schadt, E.E. *et al.,* 2013).

The broadest application of NGS may be the re-sequencing of human genomes to enhance our understanding of how genetic differences affect health and disease (Mardis E.R. 2012 and Frese, K.S. *et al.,* 2013). The variety of NGS features makes it likely that multiple platforms will coexist in the marketplace, with some having clear advantages for particular applications over others.

Overview of next generation sequencing platforms

In the 1970s, Sanger and colleagues and Maxam and Gilbert developed methods to sequence DNA by chain termination and fragmentation techniques, respectively due its own limitations, the second generation sequencing platforms are evolved are discussed below.

Helicos Heliscope Gemone Sequencer

The Helicos sequencer based on work by Quake's group also relies on cyclic interrogation of a dense array of sequencing features. However, a unique aspect of this platform is that no clonal amplification is required. Instead, a highly sensitive fluorescence detection system is used to directly interrogate single DNA molecules *via* sequencing by synthesis. Template libraries, prepared by random fragmentation and poly-A tailing (that is, no PCR amplification), are captured by hybridization to surface-tethered poly-T oligomers to yield a disordered array of primed single-molecule sequencing templates (Table 1).

At each cycle, DNA polymerase and a single species of fluorescently labeled nucleotide are added, resulting in template-dependent extension of the surface-immobilized primer-template duplexes. After acquisition of images tiling the full array, chemical cleavageand release of the fluorescent label permits the subsequent cycle of extension and imaging. Several hundred cycles of single-base extension yield average read lengths of 25 bp or greater (Harris *et al.*, 2012).

Pacific Bioscience SMRT

The single-molecule real-time (SMRT) sequencing approach developed by Pacific Biosciences is the first TGS approach to directly observe a single molecule of DNA polymerase as it synthesizes a strand of DNA, directly leveraging the speed and processivity of this enzyme to address many of the shortcomings of SGS. Given that a single DNA polymerase molecule is of the order of 10 nm in diameter, two important obstacles needed to be overcome to enable direct observation of DNA synthesis as it occurs in real time are: (i) confining the enzyme to an observation volume that was small enough to achieve the signal-to-noise ratio needed to accurately call bases as they were incorporated into the template of interest; and (ii) labeling the nucleotides to be incorporated in the synthesis process such that the dye–nucleotide linker is cleaved after completion of the incorporation process so that an actural strand of DNA remains for continued synthesis and so that multiple dyes are not held in the confinement volume at a time (something that would destroy the signal-to-noise-ratio). The problem of observing a DNA polymerase working in real time, detecting the incorporation of a single nucleotide taken from a large pool of potential nucleotides during DNA synthesis, was solved using zero-mode waveguide (ZMW) technology (Metzker M.L. 2010).

Table 1: Comparison between Next generation sequencing platforms

Instrument	Run time	Millions of reads/ run	Bases/ read	Yield MB/ run
Nanopore minion	≤ 6 hrs	0.1	9,000	1,000
Ion Torrent – '314' chip	4 hrs.	0.1	400	40
Ion Torrent – Proton I	≤ 4 hrs.	70	≤ 200	10,000
Illumina MiSeq	26 hrs.	4	150+150	1,200
Illumina HiSeq 1000	8.5 days	≤1500	100+100	≤300,000
SOLiD – 5500 x 1^{e}	8 days	> 1,410	75+35	155,100

Source: Metzker M.L. 2010.

DNA sequencing with nanopores

Most nanopore sequencing technologies rely on transit of a DNA molecule or its component bases through a hole and detecting the bases by their effect on an electric current or optical signal. Because this type of technology uses single molecules of unmodified DNA, they have the potential to work quickly on extremely small amounts of input material. Both biological nanopores constructed from engineered proteins and entirely synthetic nanopores are under development. In particular, there is potential to use atomically thin sheets of grapheme as a matrix supporting nanopores and also carbon nanotubes (Mardis E.R. 2012).

NGS applications in genomics

The production of large numbers of low-cost reads makes the NGS platforms described above useful for many applications. These include variant discovery by re-sequencing targeted regions of interest or whole genomes, *de novo* assemblies of bacterial and lower eukaryotic genomes, cataloguing the transcriptomes of cells, tissues and organisms (RNA-seq), genome-wide profiling of epigenetic marks and chromatin structure using other seq-based methods (ChIP-seq, methyl-seq and DN ase-seq), and species classification and/ or gene discovery by metagenomics studies (Adcocok *et al.*, 2001).

With this many applications, which platform is best suited for a given biological experiment? For example, the Illumina/Solexa and Life/APG platforms are well suited for variant discovery by re-sequencing human genomes because gigantic volumes of high quality bases are produced per run (Backhed *et al.*, 2006). Furthermore, the Helicos Bio-Sciences platform is well suited for applications that demand quantitative information in RNA-seq or direct RNA sequencing, as it sequences RNA templates directly without the need to convert them into cDNAs provides an overview of NGS technologies, instrument performance and cost, pros and cons, and recommendations for biological applications (Bentley DR, 2006); however, the rapid pace of technological advances in the field could change this information in the near future. Readers are directed to several excellent reviews on RNA-seq, ChIP-seqand metagenomics.

Human genome studies aim to catalogue SNVs and SVs and their association to phenotypic differences, with the eventual goal of personalized genomics for medical purposes. In 2004, the International Human Genome Sequencing Consortium published the first, and still only, finished-grade human reference genome (currently National Center for Biotechnology Information (NCBI) build 36. Its cost was estimated at US$300 million (Lander *et al.*, 2001).

In October 2007, Venter and col-leagues described the genome sequence of J. Craig Venter using a whole-genome shotgun approach coupled with automated Sanger sequencing. When the Venter genome was compared with the reference genome, 3.2 million SNVs were identified (Margulies *et al.*, 2005). In addition, there were over 900,000 SVs, which altogether accounted for more variant bases than the SNVs Personal genomics is also being applied to the study of disease. For example, Mardis have reported the sequencing of two acute myeloid leukaemia cancer genomes using the Illumina/ Solexa platform, and both studies identified somatic mutations that may be associated with the disease (Mardis, 2012). Gibbs and colleagues have recently described the elucidation of both allelic variants in a family with a recessive form of Charcot-Marie-Tooth disease using the Life/APG platform (Gibbs, 2007).

Several projects aimed at sequencing more individuals, including the cancer genome atlas and the 1000 Genomes Project, are also using the Illumina/ Solexa and Life/454 platforms to sequence whole genomes (Calin GA, 2007). Complete Genomics recently described the first genome sequence from a Caucasian male (PGP1) enrolled in the Personal Genome Projects. These projects should lead to a substantial increase in the number of personal genomes sequenced in the near future. The first group to apply NGS to whole human genomes was Roche/454 in collaboration with Gibbs and colleagues at the BCM-HGSC, who reported the diploid genome of James D. Watson.

When the Watson genome was compared with the reference genome, roughly 3.3 million SNVs were identified, but far fewer SVs were found than in the study by Venter and colleagues. This is important because undiscovered SVs could account for a substantial fraction of the total number of sequence variants, many of which could be potentially causative in disease (Schadt, E.E. *et al.*, 2013).

Within the past year, five additional human genomes have been described, one of which was sequenced on two different NGS platforms. As with the Watson genome, far fewer SVs were reported than in the study by Venter and colleagues (Venter JC *et al.*, 2004). In comparison with automated Sanger sequencing, NGS platforms have dramatically increased throughput and substantially lowered expenditure, with several groups reporting reagent costs of below US$100,000. However, there is variability among and within NGS platforms in terms of template size and construct, read-length, throughput, and base and genome coverage, and such variability makes it difficult to assess the quality (that is, the base accuracy, genome coverage and genome continuity) of genomes based on cost considerations (Mikkelsen TS, 2007 and Venter JC *et al*., 2004).

Conclusion

Compared to Sanger sequencing, advantages of the next-generation technologies mentioned thus far, including 454/Roche, Illumina/Solexa and ABI/SOLiD, alleviate the need for in vivo cloning by clonal amplification of spatially separated single molecules using either emulsion PCR (454/Roche and ABI/SOLiD) or bridge amplification on solid surface (Illumina/Solexa). In addition to providing a means for cloning-free amplification, these methods use single-molecule templates allowing for the detection of heterogeneity in a DNA, which is a significant advantage over Sanger sequencing. Next-generation sequencing technologies have found broad applicability in functional genomics research. Their applications in the field have included gene expression profiling, genome annotation, small ncRNA discovery and profiling, and detection of aberrant transcription, which are areas that have been previously dominated by microarrays.

Refrences

Adcock, G.J., Dennis, E.S., Easteal, S., Huttley, G.A. and Jermiin, L.S. 2001. Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. *Proc. Natl. Acad. Sci. USA* **98**: 537-42.

Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. and Gordon, J.I. 2005. Host-bacterial mutualism in the human intestine. *Science* **307**: 15-20.

Bentley, D.R. 2006. Whole-genome resequencing. *Curr. Opin. Genet. Dev.* **16**: 545-552.

Calin, G.A., Liu, C.G., Ferracin, M., Hyslop, T. and Spizzo, R. 2007. Ultra-conserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**: 215-229.

Edwards, R.A., Rodriguez-Brito, B., Wegley, L., Haynes, M. and Breitbart, M. 2006. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**: 57- 65.

Frese, K.S., Katus, H.A. and Meder, B. 2013. Next generation sequencing: From understanding biology to personalized medicine. *Biol.*, **2**: 378-398.

Gibbs, Bik E.M., Digiulio, D.B., Relman, D.A. and Brown, P.O. 2007. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**: 177.

Harris, Heidelberg, J.F., Halpern, A.L. and Rusch, D. 2012. Sequencing technologies - From understanding biology to personalized medicine. *Nature Rev. Genet.* **18**: 29-36.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C. and Zody, M.C. 2001. Initial sequencing and analysis of the human genome. *Nature.* **409**: 860-921.

Mardis, E.R. 2012. Next generation DNA sequencing methods. *Annu. Rev. Genom. Human Genet.,* **9**: 387-402.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S. and Bader, J.S. 2005. Genome sequencing in microfab-ricated high-density picolitrereactors. *Nature* **437**: 376-380.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H. and Chinwalla, A. 2001. A physical map of the human genome. *Nature* **409**: 934-941.

Metzker, M.L. 2010. Sequencing technologies - the next generation. *Nature Rev. Genet.* **11**: 31-46.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B. and Lieberman, E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature.* **448**: 553-60.

Schadt, E.E., Turner, S. and Kasarskis, A. 2013. A window into third-generation sequencing. *Human Mol. Genet.* **19**(2): 227-240.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L. and Rusch, D. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66-74.