Journal of Advances in Science and Technology
Vol. 12, Issue No. 24, November-2016, ISSN 2230-9659

An Analysis upon Various Technological Developments of High – Throughput DNA Sequencing: Challenges and Perspective

Shaiqua Imam[1]* Dr. Surendra Sarsiya[2]

[1]Research Scholar, SSSUTMS, Sehore

[2]UTD, SSSUTMS, Sehore

Abstract – Improvements in technology have rapidly changed the field of DNA sequencing. These improvements are boosted by bio-medical research. Plant science has benefited from this breakthrough, and a number of plant genomes are now available, new biological questions can be approached and new breeding strategies can be designed. The first part of this review aims to briefly describe the principles of the new sequencing methods, many of which are already used in plant laboratories. The second part summarizes the state of plant genome sequencing and illustrates the achievements in the last few years. Although already impressive, these results represent only the beginning of a new genomic era in plant science.

Finallywedescribe some of the exciting discoveries in the structure and evolution of plant genomes made possible by genome sequencing in terms of biodiversity, genome expression and epigenetic regulations. All of these findings have already influenced plant breeding and biodiversity protection. Finallywediscuss current trends, challenges and perspectives.

The human genome sequence has profoundly altered our understanding of biology, human diversity, and disease. The path from the first draft sequence to our nascent era of personal genomes and genomic medicine has been made possible only because of the extraordinary advancements in DNA sequencing technologies over the past 10 years. Here,we discuss commonly used high-throughput sequencing platforms, the growing array of sequencing assays developed around them, as well as the challenges facing current sequencing platforms and their clinical application.

Recent advances in DNA sequencing have revolutionized the field of genomics, making it possible for even single research groups to generate large amounts of sequence data very rapidly and at a substantially lower cost. These highthroughput sequencing technologies make deep transcriptome sequencing and transcript quantification, whole genome sequencing and resequencing available to many more researchers and projects. However, while the cost and time have been greatly reduced, the error profiles and limitations of the new platforms differ significantly from those of previous sequencing technologies.

The selection of an appropriate sequencing platform for particular types of experiments is an important consideration, and requires a detailed understanding of the technologies available; including sources of error, error rate, as well as the speed and cost of sequencing. We review the relevant concepts and compare the issues raised by the current high-throughput DNA sequencing technologies.

INTRODUCTION

High-throughput sequencing is the process of identifying the sequence of millions of short DNA fragments in parallel. In this paper, we will discuss applications and analyses of high-throughput sequencing done on the Illumina platform. The main advantage of this technology is that it allows a very high-throughput; currently up to 1.6 billion DNA fragments can be sequenced in parallel in a single run, to produce a total of 320Gbp (HiSeq 2000, version 3 kits). One challenge with this technology, however, is that the sequenced fragments are relatively short – currently up to 150bp (MiSeq instrument) or 100bp (HiSeq 2000 instrument) –

www.ignited.in

though double this can be produced using the paired-end option (see below).

We operate a service unit in a University setting providing high-throughput sequencing (henceforth, HTS) sample preparation, sequencing and initial bioinformatics analysis. Based upon our experiences over the past two years we provide the following notes. We do not aim to provide a complete picture of all of the innumerable resources available for any one of the described applications. Rather, our goal is to provide a basic overview of the opportunities and challenges that HTS represents. The field is clearly changing rapidly and so the details are to be taken with caution as they will surely need revision as new algorithms and technology emerge.

While many applications are supported by HTS, the actual input to the instrument is the same: libraries comprised of billions of DNA strands of roughly the same length (typically 300bp) with particular sequences (linkers) on either end. “Sample preparation” is the process by which an initial sample arrives at this highly ordered state. When genomic DNA is the starting material, it is fragmented and then size-selected for the tight size distribution. If the starting material is RNA, often times it is polyA-selected to limit the sequencing to mRNA. The RNA is reverse transcribed to DNA and then also size-selected. Irrespective of the application, linker DNA molecules of particular sequences are ligated to the ends of the strands. These consist of two fragments: adaptors and indices. The adaptors hybridize the DNA fragments to the flowcell on which they are sequenced. The indices are 6-7bp sequences tagging different samples within the same library that will be sequenced together. Importantly there is a PCR amplification step in many of the sample preparation protocols which has implications for the structure of the data: identical sequences may be a result of the amplification or reflect recurrence in the original sample of DNA.

Figure 1 indicates the anatomy of an insert. The following are additional basic definitions important for high-throughput sequencing:

- **Insert** – the DNA fragment that is used for sequencing.
- **Read** – the part of the insert that is sequenced.
- **Single Read (SR)** – a sequencing procedure by which the insert is sequenced from one end only.
- **Paired End (PE)** – a sequencing procedure by which the insert is sequenced from both ends.
- **Flowcell** – a small glass chip on which the DNA fragments are attached and sequenced. The flowcell is covered by probes that allow hybridization of the adaptors that were ligated to the DNA fragments.
- **Lane** – the flowcell consist of 8 physically separated channels called lanes. The sequencing is done in parallel on all lanes.
- **Multiplexing / Demultiplexing** – sequencing a few samples on the same lane is called multiplexing. The separation of reads that were sequenced on one lane to different samples is called demultiplexing and is done by a script that recognizes the index of each read and compares it to the known indices of each sample.
- **Pipeline** – a series of computational processes.

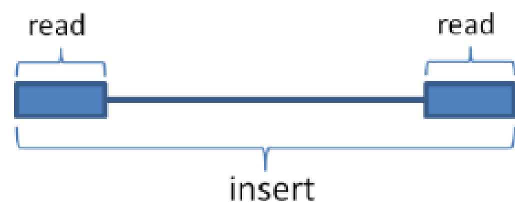

Figure 1. Schematic of a paired-end read.

The human genome sequence was completed in draft form in 2001. Shortly thereafter, the genome sequences of several model organisms were determined. These feats were accomplished with Sanger DNA sequencing, which was limited in throughput and high cost; indeed, the first human genome sequence was estimated to cost 0.5–1 billion dollars. These limitations reduced the potential of DNA sequencing for other applications, such as personal genome sequencing. Following the release of the ‘‘finished’’ human genome, the National Human Genome Research Institute (NGHRI) created a 70 million dollar DNA sequencing technology initiative aimed at achieving a $1,000 human genome in 10 years (Schloss, 2008), and a flurry of high-throughput sequencing (HTS) technologies emerged.

To put this initiative in perspective, improvements to traditional Sanger sequencing had decreased the per base cost by around 100-fold by the completion of the Human Genome Project. To reach the $1,000 dollar genome threshold, however, an additional leap of five orders of magnitude was necessary. Much of this divide has been traversed—the cost of a genome sequence (without interpretation) is presently less than $2,000. The road to this milestone involved many commercial HTS platforms, which differ in their details but typically follow a similar general paradigm: template preparation, clonal amplification, followed by

cyclical rounds of massively parallel sequencing. The specific strategy employed by each platform determines the quality, quantity, and biases of the resulting sequence data and the platform's usefulness for particular applications.

Several excellent reviews have covered HTS platform strategies in great depth (Morey et al., 2013). Many important platforms are not covered here, including Roche/454's pyrophosphate Genome Sequencer and Helicos' single-molecule Heliscope sequencer as well as the Polonator,

ABI's SOLiD, and Complete Genomics' DNA nano-array sequencer (Drmanac et al., 2010). Instead, we focus on the most commonly used platforms today as well as more recent developments. We also provide an overview of the growing array of HTS applications and highlight their use by the genomics community to illuminate previously intractable topics in biology. Finally, we discuss the limitations of current platforms and challenges to clinical sequencing.

It is now almost exactly a decade since the announcement of the completion of the first drafts of the complete human genome sequence. Despite the excitement and enthusiasm at that time, ten years later the human genome sequence has not yet delivered its promise to "revolutionise the diagnosis, prevention and treatment of most, if not all, human disease". However, the race for a complete human genome has had an enormous impact on scientific research, especially in bioinformatics and in development of DNA sequencing technologies. The opportunities offered by these technological advances are at least as important for plant science and microbiology as they are for biomedical research. New sequencing methods have enabled the advent of personal genomics; that is, moving away from the concept of the human genome (singular) to genomes (plural). The 1,000 Genomes Project Consortium has now completed its pilot stage, and in 2010 the Wellcome Trust announced the UK10K project, which aims to generate genome-wide sequence data for thousands of individuals from the United Kingdom. Not to be outdone, the Arabidopsis thaliana genomics community launched the 1,001 Genomes Project for this model plant. However, the recent advances in highthroughput sequencing run deeper and broader than genome sequencing; they have already been applied to transcriptomics, epigenetics, and discovery of protein-binding sites in plants and microbes. They also have applications in diagnostics and functional genomics, e.g., simultaneously screening every gene in a microbial genome for essentiality. This manuscript reviews some applications of these high-throughput sequencing methods that are relevant to phytopathology, with emphasis on the associated computational and bioinformatics challenges and their solutions. Necessarily, many of the examples cited in this review come from fields outside of phytopathology. This is because some of the cutting-edge technologies

andmethods have been driven by better-funded biomedical fields, and there is a lag before their full adoption in phytopathology. However, our field is quickly catching up, as exemplified by the rapid proliferation of complete genome sequences for many strains of the plantpathogenic bacterium Pseudomonas syringae.

In 1977 the first genome, that of the 5,386 nucleotide (nt), single-stranded bacteriophage wX174, was completely sequenced using a technology invented just a few years earlier.

Since then the sequencing of whole genomes as well as of individual regions and genes has become a major focus of modern biology and completely transformed the field of genetics.

Over the last decade, alternative sequencing strategies have become available which force us to completely redefine ''high-throughput sequencing.'' These technologies outperform the older Sanger-sequencing technologies by a factor of 100–1,000 in daily throughput, and at the same time reduce the cost of sequencing one million nucleotides (1 Mb) to 4–0.1% of that associated with Sanger sequencing. To reflect these huge changes, several companies, researchers, and recent reviews use the term ''next-generation sequencing'' instead of high-throughput sequencing, yet this term itself may soon be outdated considering the speed of ongoing developments.

Here we review the five sequencing technologies currently available on the market (capillary sequencing, pyrosequencing, reversible terminator chemistry, sequencing-by-ligation, and virtual terminator chemistry), discuss the intrinsic limitations of each, and provide an outlook on new technologies on the horizon. We explain how the vast increases in throughput are associated with both new and old types of problems in the resulting sequence data, and how these limit the potential applications and pose challenges for data analysis.

During the last 25 years DNA sequencing has completely changed our vision of biology and particularly plant biology. It has been possible to characterize a large number of genes by their nucleotide sequences, thus providing a shortcut to the corresponding protein sequences and their functions. Information on gene polymorphisms has facilitated genetic mapping, gene cloning and the understanding of evolutionary relationships and has allowed for the initiation of biodiversity studies.

The most popular sequencing method has been the Sanger method , described in all textbooks. Since its conception, the method has been continually improved. When combined with the use of robots and with concomitant progress in cloning strategies and physical mapping, the method has allowed for sequencing larger DNA fragments and, finally, complete genomes. As a result, a series of landmark genomes were obtained: Haemophilus influenzae, Saccharomyces cerevisiae, Escherichia coli, Caenorabditis elegans, Drosophila melanogaster, Arabidopsis thaliana and, finally, Homo sapiens and rice. The deciphering of these genomes led to the era of functional genomics and completely modified biological investigation. It also demonstrated that with enough money and human resources, the genomes of other species could be obtained. However, this technology remained tedious and expensive.

Sequencing of each of the above eukaryotic genomes cost several million US dollars and mobilized hundreds of scientists all over the world. These limiting factors have prompted the development of new technologies that allow for many more samples to be analysed at the same time, without prior cloning or mapping work, and at much lower cost. This development has opened new avenues in biology to solve important questions that could not be answered with classical sequencing. The driving force to improve sequencing technology has been human biology and personalized medicine, with the objective being the sequencing of an individual's human genome costing a mere $1000 or less. This goal has not yet been achieved but seems close.

DNA SEQUENCING: HISTORICAL PERSPECTIVE

The genomics age began some years before the announcement of the draft human genomes in 2000. As early as the mid-1970s, the first DNA and RNA genomes were fully sequenced from bacteriophages. In 1995, the bacterium Haemophilus influenzae became the first free-living organism to have its genome completely sequenced. This was soon followed by complete sequences from several other bacteria and archaea, heralding the age of comparative genomics, at least for prokaryotes. Early on, several bacterial pathogens of plants were fully sequenced and compared, yielding new insights into the evolution and mechanisms of pathogenesis in these organisms. The first eukaryotic chromosome sequence was published in 1992, with the complete genome sequence available from the same yeast being published in 1996. Since then, many eukaryotes have been fully sequenced, including several plants and their pathogens (fungi and oomycetes). Until approximately 2006, almost all genome sequencing activity relied on the same Sanger chemistry as that deployed on the bacteriophage sequencing projects in the 1970s, albeit with a series of innovations, such as replacement of gels by capillaries as well as steadily increasing efficiency and throughput while decreasing per-nucleotide costs. Then came several massively parallel sequencing platforms that perform tens of thousands (or more) of sequencing reactions in a single test tube. These technologies, including Illumina's Solexa platform, Roche's 454 pyrosequencing, and ABI's SOLiD platform, have come to be known as second-generation sequencing. A third generation of sequencing technologies is also poised to make an impact.

Fifty-five years after publishing the structure of the double helix, James Watson published the complete DNA sequence of his own genome. The Watson genome sequence was generated using the 454 pyrosequencing platform at approximately one percent of the cost of the original Human Genome Project.

HIGH-THROUGHPUT SEQUENCING APPLICATIONS

HTS applications can be divided into two main categories: 'reading' and 'counting'. In reading applications the focus of the experiment is the sequence itself, for example for finding genomic variants or assembling the sequence of an unknown genome. Counting applications are based on the ability to count amounts of reads and compare these counts, for example to assess gene expression levels. Table 1 shows some of the main applications enabled by high-throughput sequencing. These represent but a sampling of the main HTS applications. It should be noted that one can invoke HTS in practically any experiment that produces DNA fragments. What should be considered and planned before the sequencing however is the method by which the analysis of the sequenced fragments will be done to extract the meaning from the experiment. As an example of a unique HTS experiment, chromatin interactions can be identified by PE sequencing. This procedure includes capturing interacting loci in the genome by immune-precipitating cross-linked fragments of DNA and proteins from fixed cells. There are many others, published at a rate of about one per day.

	Application	Goal	Experiment details	Basic analysis summary
Reading	**Re-sequencing**	Find variants in a given sample relative to reference genome.	Extract DNA from the relevant cells, conduct sample preparation consisting of DNA fragmentation and sequencing.	Mapping of the sequenced fragment to the reference genome and identifying variants relative to the reference genome by summarizing the differences of the fragments from the genomic loci to which they map.
	Target-enriched sequencing	Target enrichment sequencing is a specific form of re-sequencing that is focused only on certain genomic loci. This is useful for organisms with large genomes where enrichment increases the coverage on the loci of interest thereby reducing costs	After the DNA is extracted from the cells and undergoes sample preparation, an enrichment process is done to capture the relevant loci. Target enrichment can be done on specific regions of the genome using "tailored" target-enrichment probes, or by using available kits such as exome-enrichment kits.	Same as in resequencing.
	De-novo assembly	Identify a genomic sequence without any additional reference.	Same as in re-sequencing.	The assembly process relies on overlaps of DNA fragments. These overlaps are merged into consensus sequences called contigs and scaffolds.
Counting	**ChIP-Seq/ RIP-Seq**	Find the binding locations of RNA- or DNA-binding proteins.	First, the ChIP/RIP experiment is done: proteins are bound to the DNA/ RNA and are cross-linked to it. The DNA/RNA is then fragmented. The proteins are pulled down by an immuno precipitation process and are then the cross-linking is reversed. The DNA/RNA fragments that are enriched in the protein binding sites locations are then sequenced.	The sequenced fragments are mapped to the genome. The enriched locations in the genome are found by detecting "peaks" of mapped fragments along the genome. These peaks should be significantly higher than the mapped fragments in the surrounding loci, and significantly higher compared to a control sample – usually the input DNA of the ChIP experiments or another sample of immuno-precipitation done by a non-specific antibody.
	RNA-Seq	Detecting and comparing gene expression levels.	Total RNA is extracted from the cells. In a sample preparation process the mRNA is pulled down and fragmented. The mRNA fragments are then reversed transcribed to cDNA. The cDNA fragments are sequenced.	The cDNA fragments are mapped to the reference genome. The fragments that map to each gene are counted and normalized to allow comparisons between different genes and different samples. Un-annotated genes and transcripts can be found in an RNA-Seq experiment by detecting bundles of fragments that are mapped to the genome in an un-annotated region.
Reading/ Counting	**microRNA-Seq**	Detect and count microRNAs.	Total RNA is extracted from the cells, and the microRNA is isolated by recognizing the natural structure common to most known microRNA molecules. The microRNA fragments are then reversed transcribed and sequenced.	The sequenced fragments are mapped to the genome. The microRNA can then be detected and counted.

Table 1. HTS applications.

THE NEXT-GENERATION SEQUENCING TECHNOLOGIES: BASIC PRINCIPLES

NGSTs are evolving rapidly and have been described in several recent reviews, including in the plant domain . Therefore this section presents only the basic principles. More details are found in these reviews and at the websites of the companies marketing the technologies and instruments. All NGSTs have benefited, to various extent, from new developments in imaging, automation, microprocessing and nanotechnologies, domains that have developed independently of biology. Various ways of reducing costs include avoiding cloning, miniaturizing reactions, use of newchemical procedures, and use of massively parallel sequencing.

New sequencing technologies can be grouped into several classes: sequencing by hybridization, sequencing by synthesis from amplified molecules distributed in microarrays and sequencing single molecules. Fig. 2 presents a flow chart of the different methods. However, sequencing by hybridization is a re-sequencing technology that now seems of limited value as compared with the other technologies and will not be described here.

Sequencing by synthesis after amplification –

In this approach, DNA fragments are amplified in clusters, denatured and distributed on microarrays or in microtiter plates that are introduced into a flow cell where the sequencing reactions take place. A primer is extended cyclically by one or a few nucleotides at a time, and the sequence is read at each step of the DNA synthesis.

This strategy differs from the Sanger method whereby a whole range of partial copies of the DNA molecules are first synthesized and then analysed. The various methods differ in the strategy used to amplify the sequences, the chemistry used, and the length of the reads. The methods have in common the possibility of sequencing up to several million DNA fragments in parallel.

Pyrosequencing and 454 technology - Roche's 454 technology was the first to be marketed in 2004. It introduced several important innovations. Small DNA fragments (300–800 bp) are ligated to adapters, separated into single strands and bound to small DNA capture beads under conditions favouring a single fragment per bead. The fragments are then amplified by a technique dubbed "PCR-emulsion", whereby each bead is isolated within a droplet of a PCR reaction mixture – in oil emulsion. At the end of amplification, each bead carries several million copies of a unique DNA fragment. This step avoids the cloning of each DNA fragment and the tedious work of colony picking and preparing DNA templates.

Then, the emulsion is broken, the DNA is denatured and the beads are deposited in the wells of a PicoTiterPlate. The plate contains millions of wells that are individual reactors for the sequencing reactions that are catalysed by the Bacillus stearothermophilus (Bst) DNA-polymerase. This device is placed in a flow cell into which reagents are injected. The diameter of the wells is manufactured so that only one bead can be accepted in each well. When a nucleotide is added to the primer by the DNA-polymerase, a pyrophosphate molecule is released. This pyrophosphate is converted into ATP by a sulfurylase and the ATP is used to produce a chemiluminescent signal by the luciferase reaction. This method was therefore named "pyrosequencing". First a nucleotide is injected into the flow cell. A chemiluminescent signal is detected in the microwells where the nucleotide has been incorporated and a camera records the signal. Unincorporated nucleotides are then washed away and are replaced by a second nucleotide and the ordered incorporation/recording/washing cycle is repeated

with the four nucleotides until the primer has been sufficiently extended.

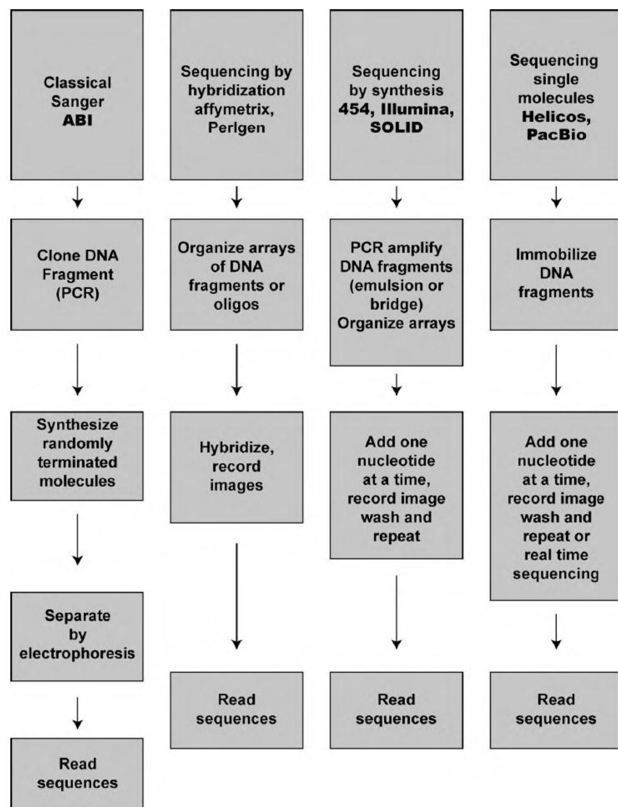

Fig. 2. A flow chart of the different types of sequencing methods.

Sequencing from single DNA molecules -

In contrast to the previous technologies, this group of techniques does not require amplification ofDNAfragment libraries and can be used to sequence any DNA molecule directly. They are sometimes named third-generation sequencing methods.

Helicos technology - The Helicos platform (http://helicosbio.com) was the first to propose direct sequencing of a single DNA molecule. This breakthrough is essentially due to improved chemistry with highly fluorescent modified nucleotides, which act as reversible terminators.

Tails of poly(A) are added to the DNA fragments, which are captured on the surface of an array coated with oligo(dT). Billions of fragments are thus randomly bound to a surface divided into 50 channels and can be directly sequenced by synthesis with a DNA-polymerase when the array is introduced into a flow cell. At each cycle, a given fluorescent nucleotide is injected into the flow cell to allow for template-dependent single-nucleotide extensions of the primers on a fraction of the fragments. The image is recorded, excess unincorporated nucleotides are washed away and the fluorescent moiety of the incorporated nucleotide is removed.

This cycle is repeated with each of the four nucleotides until the sequence read is about 30 nt. A machine can produce 30–40 Gbp per 8 day run. The first genomes to be sequenced by this method were M13 bacteriophage and E. coli.

Real-time sequencing with Pacific Biosciences technology - Pacific Biosciences (http://pacificbiosciences.com) developed the single molecule real-time (SMRTTM) sequencing technology, involving a chip with several thousand nanoscale wells whereby a single _29-phage DNA-polymerase is immobilized and bound to a single primedDNAtemplate. Hexaphosphate nucleotides labeled on the phosphate moiety with four distinct fluorochromes are added. The nucleotide residence time on the polymerase active site is of the order of a millisecond and allows for a fluorescent pulse to be generated and recorded in real time. The geometry of the nanowell determines a limited detection zone with a volume of 10−21 L, conferring both high sensitivity and low-back ground.

HIGH-THROUGHPUT DNA SEQUENCING TO BENIGN HEMATOLOGY

Hematology has a long history of being a discipline at the forefront of applying novel technology to understanding and diagnosing disease. Maxwell Wintrobe pointed out in his classic text, “Blood,Pure and Eloquent,” that important developments in the field of hematology were often driven by technology. For instance, early observations of blood and marrow morphology were enabled by the advances in microscopy. Similarly, quantitation of the different cellular elements of blood were made possible by development of the hemocytometer.

Hematology continued to lead the way into the molecular era, with the description of sickle cell anemia as the first molecular disease,2 solution of the hemoglobin molecule by x-ray crystallography (the first multisubunit protein to be understood at the molecular level3), and determination of the molecular basis of sickle cell disease at the amino acid level.4 As Wintrobe’s classic text was being published, breakthrough technologies in the fledgling field of molecular genetics were leading to groundbreaking developments.

Studying sickle cell disease and the thalassemia syndromes, investigators linked genetic polymorphisms to human disease, identified disease-causing mutations at the DNA level, and developed strategies for prenatal diagnosis (reviewed in Sankaran and Nathan5). A few years later, the first human disease gene isolated by positional cloning, the chronic granulomatous disease gene CYBB, was identified.6 Recently, advances in genomic

technologies have led to numerous discoveries in hematology, detailed below.

As we make our way through the 21st century, technology continues to rapidly evolve. The use of next-generation DNA sequencing has dramatically advanced the way we assess gene expression, protein-DNA interactions, long-range DNA interactions, and both normal and pathologic DNA variation. This latter area is the focus of this review, which is one of a series of reviews on the application of high-throughput sequencing approaches to hematology. In this review, we discuss the application of these approaches to benign hematology, including red blood cell, neutrophil, and other white blood cell disorders. Malignant hematologic disorders and bleeding disorders, including abnormalities in platelets, are covered in other reviews. A review from Deborah Nickerson provides an overview of the high-throughput DNA sequencing technology.

Because a comprehensive review is impractical, we provide vignettes that illustrate how high-throughput DNA sequencing is impacting hematology. These highlight applications of these approaches for disease diagnosis, gene discovery, and a better understanding of the genetic basis of complex traits. Prognostic and therapeutic implications of advances driven by sequencing technology are discussed and future applications highlighted.

CONCLUSION

Current high-throughput sequencing technologies provide a huge variety of sequencing applications to many researchers and projects. Given the immense diversity, we have not discussed these applications in depth here; other reviews with a stronger focus on specific applications and data analysis are available. The discussed technologies make it possible for even single research groups to generate large amounts of sequence data very rapidly and at substantially lower costs than traditional Sanger sequencing.

High-throughput DNA sequencing technology, which has already had an impact upon hematologic disease diagnosis and gene discovery, holds significant promise for the future. In the coming years, genomic studies will permit discovery of new disease genes and modifier alleles and provide important insights into disease pathobiology.

The advent of relatively cheap, massively high throughput nucleotide sequencing is already making a huge impact in many areas of biology, especially in the study of microbial evolution, mechanisms of pathogenesis, and phylogeography. On a more practical level, it promises to replace hybridization-based methods for transcriptome profiling and is playing an increasingly high-profile role in diagnostics and epidemiology.

This review has illustrated some of the fascinating progress in DNA sequencing during the last 5 years. In addition to the current commercialized technologies, others in development make use of new advances in nanotechnologies and imaging to further improve the power of the technology. Sequencing now costs almost nothing; however, machines are expensive and preparing samples and analyzing the data still require a very significant effort as well as highly qualified human resources.

REFERENCES

Agarwal M., Shrivastava N., Padh H. (2008). Advances in molecular marker techniques and their applications in plant sciences. Plant Cell. Rep. 27: pp. 617–31.

Al Rwahnih M, Daubert S, Golino D, Rowhani A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387: pp. 395–401.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, pp. 78–81.

GreubG, Kebbi-Beghdadi C., Bertelli C., Collyn F., Riederer B. M., et al. (2009). High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. PLoS ONE 4: p. 8423

Harris TD, Buzby PR, Babcock H, et al. 2008. Single-molecule DNA sequencing of a viral genome. Science 320: pp. 106–9.

International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome, Nature 436, pp. 793–800.

J. Shendure, H. Li. (2008). Next generation DNA sequencing, Nat. Biotechnol. 26. 1135–1145. M.L. Metzker, Sequencing technologies. The next generation, Nat. Rev. Genet. 11 (2010) pp. 31–46.

Kambara H., Takahashi S. (1993). Multiplesheathflow capillary array DNA analyser. Nature 361: pp. 565–6.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 10, R25.

Margulies M, Egholm M, Altman WE, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: pp. 376–80.

McCarthy J.J., McLeod H.L., Ginsburg G.S. (2013). Genomic medicine: a decade of successes, challenges, and opportunities. Sci Transl Med. 5(189): pp. 189-184.

Morey, M., Ferna´ ndez-Marmiesse, A., Castin˜ eiras, D., Fraga, J.M., Couce, M.L., and Cocho, J.A. (2013). A glimpse into past, present, and future DNA sequencing. Mol. Genet. Metab. 110, pp. 3–24.

Sanger F, Nicklen S, Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: pp. 5463–7.

Schatz, M.C., Witkowski, J. and McCombie, W.R. (2012). Current challenges in de novo plant genome sequencing and assembly. Genome Biol, 13, p. 243.

Schloss, J.A. (2008). How to get genomes at one ten-thousandth the cost. Nat. Biotechnol. 26, pp. 1113–1115.

Shendure J, Lieberman Aiden E. The expanding scope of DNA sequencing. Nat Biotechnol. 2012; 30 (11): pp. 1084-1094.

Zagursky RJ, McCormick RM. (1990). DNA sequencing separations in capillary gels on a modified commercial DNA sequencing instrument. Bio techniques 9: pp. 74–9.

Corresponding Author

Shaiqua Imam*

Research Scholar, SSSUTMS, Sehore

E-Mail – chairman.iab@gmail.com