A STUDY ON HIGH THROUGHPUT DNA SEQUENCING AND ITS IMPLICATION IN PLANT SCIENCE RESEARCH

Apparao Thota*

Abstract

Due to certain limitations such as low throughput, higher cost of sequencing per base. Sanger method has been partially supplanted by several "next-generation sequencing" or "second generation sequencing" technologies that offer dramatic increase in cost-effective high throughput sequencing, albeit at the expense of sequence lengths. Several next-generation technologies are commercially available today as provided in Table 1. However, for a review of the history of DNA sequencing, the readers are referred to Hutchison.

Keywords: Sequencing Technology, Plant Research.

INTRODUCTION

The living organisms consist of cells which contain the genetic material that is popularly known as chromosome. Further, the chromosomes are made up of the DNA and protein. This DNA is a linear arrangement of genes made up of four basic units. i.e. adenine (A), thymine (T), guanine (G), cytosine (C), in different combinations. Sequencing of part or full genome always remains; a central focus of biotechnological research, therefore development of faster, cost- effective sequencing technology remains a/challenge to the scientists. Significant achievements have been made in sequencing from a simple short- length manual sequencing during 1977 to millions of bases per-day presently. The advancement of sequencing technology within 3 çlecades can be segmented in 3 different phases which are known as first generation, second or next generation and third generation. Next generation or second generation sequencing: Due to certain limitations such as low throughput, higher cost of sequencing per base. Sanger method has been partially supplanted by several "next-generation sequencing" or "second generation sequencing" technologies that offer dramatic increase in cost-effective high throughput sequencing, albeit at the expense of sequence lengths. Several next-generation technologies are commercially available today as provided in Table 1. However, for a review of the history of DNA sequencing, the readers are referred to Hutchison (2007).

NEED OF SEQUENCING

Although the DNA of any organism in this universe is made up of four basic units or their derivatives, one may wonder as to how diverse organisms, starting from tiny bacteria to giant dinosaurs or small grasses to big banyan trees, are created by nature. This is because each of them has different genes in their DNA or in more simple sense, it is the difference in sequence of their genome.

CHEMISTRY OF DIFFERENT SEQUENCING TECHNOLOGIES

Although different sequencing techniques use different chemistry yet there is commonality in the process of sequencing. First, DNAs of interest are broken into small pieces, sequenced and then these small sequences are aligned or joined one by one to form contig or bigger piece of DNA molecule. Finally, the contigs are joined together to get entire long sequence of the genome.

APPLICATIONS

*Research Scholar, Sunrise University, Alwar, Rajasthan

Although compiling the various applications of high throughput sequencing is the beyond the scope of this article and readers may refer to the recent review for this purpose, yet an introductory, account is given below.

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED TRANSCRIPTS

When the primary objective of a biological study is gene expression profiling between two samples, mRNA sequencing (RNAseq) is the most appropriate approach for number of reasons. This sort of analysis is particularly relevant for detecting the difference of mRNA population between the wild-type and mutant strains, between treated versus untreated cells, cancer versus normal, and so on.

Although this can be studied in a number of ways such as SSH (selective subtractive hybridization), cDNA-AFLP (amplified fragment length polymorphism), MPSS (massive parallel sequence signature), SAGE (serial analysis of gene expression) and microarray, yet each of them is embodied with merits and demerits. Nevertheless, newer techniques become always superior to older one. Thus, besides larger dynamic range and sensitivity of RNA-seq, several additional factors have contributed to the rapid acceptance of sequencing for differential expression analysis. For an example, microarrays are simply not available for many non-model organisms (for example. Affymetrix offers microarrays for approximately 30 organisms). Additionally, sequencing gives unprecedented detail about transcriptional features that arrays cannot, such as novel transcribed regions, allele-specific expression, RNA editing and a comprehensive capability to capture alternative splicing.

DE NOVO SEQUENCING OF CROP SPECIES WITHOUT REFERENCE GENOME SEQUENCE

This is particularly important for sequencing of a crop for the first time. For years, people were doing in little different way which is known as 'shot gun' technique based on bacterial cloning. However, shot gun technique is slow, less efficient than this de novo sequencing or high throughput sequencing. It is noteworthy to mention here that recently pigeon pea and neem have been sequenced for the first time by Indian scientists.

DISCOVERY OF SNP

One of the central themes in genomics is to study allelic differences or variations which can be detected by discovering biallelic markers such as single nucleotide polymorphism (SNP) or haplotype i.e. group of SNPs that are associated in particular trait. Short sequences can be compared to a reference genome sequence to detect the variants. These SNPs are of very good molecular marker for developing the genetic map which is useful information for plant breeder.

WIDE CROSSES AND ALIEN INTROGRESSION

Pre-breeding is an important component in plant improvement particularly for crops that have a narrow genetic base. Sequencing of wild races is anticipated to have a profound effect because additional molecular markers could be rapidly developed on a genome-wide scale that help to target more narrowly defined genome regions to trace introgression and selection cycles.

MAPPING OF DNA-BINDING PROTEINS AND CHROMATIN

Chip-on-chip Chip (chromatin immune-precipitation) using microarrays) is a key approach to map in viva binding sites of various DNA-binding proteins across the genome. Here DNA-protein complex is precipitated with the protein specific antibody after which DNA was isolated with protease K digestion for sequence. This approach is termed as 'Chipset', which should produce a huge windfall, in particular for studies of multi-cellular eukaryotes where whole genorne coverage has generally required the use of several arrays.

EPIGENETIC MODIFICATIONS

Epigenetic is different from genetics which means changes of phenotype without any change in genetic material that is normally associated to methylation in cytosine bases of DNA or post-translational modification of amino acid present in his tone tails. They play a key role in gene expression and in plant development under stress. Genome-wide analysis of epigenetic modifications of DNA is an active area of research.

SMALL NON-CODING RNA PROFILING AND THE DISCOVERY OF NOVEL SMALL RNAs

Small RNAs play important role in eukaryote gene expression. A related application of next-generation sequencing technologies to the analysis of transcriptomes is small noncoding RNA (ncRNA) discovery and profiling. These ncRNAs are RNA molecules that are not translated into a protein product This class of RNAs includes transfer RNA (tRNA), ribosomal RNA (rRNA), small nuclear and small nucleolus RNA, and micro RNA and small interfering RNA (miRNA and siRNA).

CONCLUSION

Next-generation-sequencing of aDNA is associated with a number of problems and requires modification of existing protocols. Nevertheless, the large amounts of data produced by the various

NGS instruments together with the short read length makes this new technology ideal for aDNA research. This is evidenced by the massive increase of available sequence data from long-dead organisms since the invention of NGS. Due to the small amount of endogenous DNA and high background contamination characteristic for aDNA extracts, shotgun sequencing of aDNA is still only of limited use, and is mainly used for complete genome sequencing projects. Therefore, aDNA is likely to especially profit from new developments in barcoded and targeted sequencing. Capture methods in particular are a very promising approach. As the technology improves it is likely that even whole genome capture will be possible in the not too distant future, which will greatly reduce the costs and time for whole genome sequencing from extinct organisms. Next-generation-sequencing techniques carry the promise to bring aDNA research into the center of evolutionary biology and make it a crucial part of modern genetics. After all, what could help understand evolution better than following its progress in real time?

REFERENCES

1. Higuchi, R.; Bowman, B.; Freiberger, M.; Ryder, O.A.; Wilson, A.C. DNA-sequences from the quagga, an extinct member of the horse family. *Nature* **1984**, *312*, 282-284.
2. Saiki, R.K.; Scharf, S.; Faloona, F.; Mullis, K.B.; Horn, G.T.; Erlich, H.A.; Arnheim, N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anemia. *Science* **1985**, *230*, 1350-1354.
3. Mullis, K.B.; Faloona, F.A. Specific synthesis of DNA invitro via a polymerase-catalyzed chain- reaction. *Methods Enzymol.* **1987**, *155*, 335-350.
4. Saiki, R.K.; Gelfand, D.H.; Stoffel, S.; Scharf, S.J.; Higuchi, R.; Horn, G.T.; Mullis, K.B.; Erlich, H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **1988**, *239*, 487-491.
5. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.T.; *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376-380.