

Next Generation Sequencing Technologies and Its Application in Plant Science

Gayatri Gouda¹, Vadde Ramakrishna², Santosh Kumar Behera³, Lambodar Behera⁴,

Manoj Kumar Gupta^{5*}

Ph.D. Student, Crop Improvement Division, National Rice Research Institute (Formely CRRRI), Cuttack, Odisha, India¹

Assistant Professor, Department of Biotechnology & Bioinformatics, Yogi Vemana University, Kadapa,

Andhra Pradesh, India²

Scientist -II, Biomedical Informatics Centre, Regional Medical Research Centre (ICMR), Bhubaneswar, Odisha, India³

Principal Scientist, Crop Improvement Division, National Rice Research Institute (Formely CRRRI), Cuttack, Odisha,

India⁴

Ph.D. Student, Department of Biotechnology & Bioinformatics, Yogi Vemana University, Kadapa, Andhra Pradesh,

India⁵

* Corresponding Author

ABSTRACT: All disciplines which are DNA sequence data dependent, have shown rapid advancement in the last few years because of the development of development and emergence of next-generation sequencing (NGS) technologies. Earlier to NGS technology, for almost two decades, the automated Sanger method had conquered the industry and led to a number of enormous achievement, including the completion of the human genome sequence. The arrival of NGS technologies in the marketplace brought change in the way we dwell on scientific approaches in basic, applied and clinical research. This review article explain how different next generation sequencing works, what are the different types of Next-Generation Sequencing Systems and application of Next-Generation Sequencing in plant science.

KEYWORDS: High-Throughput Nucleotide Sequencing, Polymerase Chain Reaction, DNA

I. INTRODUCTION

All disciplines which are DNA sequence data dependent have shown rapid advancement in the last few years because of the development and emergence of next-generation sequencing (NGS) technologies. Earlier to NGS technology, for almost two decades, the automated Sanger method had conquered the industry and led to a number of enormous achievement, including the completion of the human genome sequence[1]–[3]. The automated Sanger method, which is believed as a "first-generation" technology, and present methods are referred to as next-generation sequencing (NGS). These newer technologies are constituted of various strategies that rely on a combination of template preparation, followed by sequencing and imaging and genome alignment and assembly methods. The arrival of NGS technologies in the marketplace brought change in the way we dwell on scientific approaches in basic, applied and clinical research [3]–[6]. In some respects, the potential of NGS is similar to the first phase of PCR, with one's thoughts being the primary limitation to its use. The major advantages offered by NGS is the ability to generate a massive volume of data cheaply in some cases in excess of one billion short reads per instrument run. This feature increases the monarchy of experimentation beyond just shaping the order of bases. For example, microarrays studies for gene-expression are now being replaced by seq-based methods, which are capable of identifying and quantifying rare transcripts without prior knowledge of a particular gene and can also provide information about alternative splicing and sequence variation in

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

identified genes. The ability to sequence the whole genome of many related organisms, by NGS technologies, has led to the large-scale comparative and evolutionary studies to be executed that were unthinkable just a few years ago [1], [7]–[9]. The broadest use of NGS may be the re-sequencing of human genomes to boost our understanding of how genetic differences affect health and disease. In agriculture, during the initial stages of NGS development, NGS technology was mainly used by researchers like plant breeders who works with model organisms of economic importance. A comparative study using four commonly used short read alignment tools like Maq, BowTie, Novoalign and SOAP2 with their latest approach called coverage-based consensus calling (CbCC) for SNP discovery as a case study in chickpea, *Cicer arietinum L.*, a crop lacking a reference genome was done in 2012 [6]. Maq was found to be most accurate and sensitive, even at low read depth. Greater accuracy was demonstrated by all four tools at higher read depth, and SNPs predicted by three or four tools were more likely to be correct. SNP prediction accuracy normally increased with the increase in read depth. The results acquired in that study are applicable for NGS-based SNP discovery in a few other plant species that does not have a reference genome. In addition, 4543 putative SNPs have been identified in chickpea that will be useful for advancing chickpea genetics research and breeding applications [6]. 22 EST libraries were generated by Lai et al., from a variety of tissues for 11 weeds in the sunflower family using Illumina sequencing and Sanger 454. They also compared the coverage and quality of sequence assemblies. They too have developed NimbleGen microarrays for expression analyses in five of them [10]. In the present review paper, we will discuss different existing NGS technologies and their use in plant biology.

II. RELATED WORK

Various paper have reported about different existing NGS technologies and their application in various field of medical science as well as agriculture. Metzker (2010) had reported review of preparation of template, sequencing, and imaging, followed by genome alignment and assembly approaches. That review article also outlines the problems and applications for NGS technologies[3]. Liu et al. (2012) have reviewed about the different NGS technologies and discussed the advantages and specifics associated with each sequencing system [11]. Egan et al. (2012) presented a brief outline of different existing NGS technologies and their application in plant science[6]. Morozova et al (2008) have reviewed about different NGS technologies and their application to functional genomics research, predominantly focusing on transcriptomics and epigenomics. This review article have reported that 454 sequencing correlated well with the already established gene expression profiling technologies such as microarray results and moderately with SAGE[12]. Radford et al (2012) have reported about various NGS technologies methods and discussed different increasing application in virology, especially in virus discovery, genome sequencing and transcriptomics [13]. Shen et al (2015) have summarized the account of massively parallel next-generation sequencing (NGS) in the context of cancer genomics. They have also highlighted the importance and possibility of complete or 100% genome sequencing, i.e., the ability to sequence highly repetitive non-coding sequences which are beyond the reach of existing NGS technologies [14]. Berglund et al (2011) have reviewed about technical aspects of next generation sequencing technology ,i.e., the preparation of DNA templates, methods for imaging and base calling, quality control, and bioinformatics approaches for sequence alignment, variant calling, and assembly[15].

III. NEXT-GENERATION SEQUENCING TECHNOLOGIES

Sequencing technologies are composed of a number of methods which are grouped broadly as template preparation, sequencing and imaging, and data analysis. The exclusive combination of explicit protocols distinguishes one technology from another and determines the kind of data produced from each platform. These differences in data output present challenges while comparing those platforms based on data quality and cost. Although each manufacturer provides quality scores and accuracy estimates, there is no accord that a "quality base" from one platform is equivalent to that from another platform.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

1. Template preparation

The need for dynamic methods that generate a representative, non-biased source of nucleic acid material from the genome under analysis cannot be overemphasized. Existing methods generally comprise of randomly breaking genomic DNA into minor sizes from which either fragment templates or mate-pair templates are formed. A common design among NGS technologies involves attachment or immobilization of the template to a solid surface or support. The immobilization of spatially separated template sites permits thousands to billions of sequencing reactions to be executed concurrently [2], [3], [9], [16], [17].

(A) Clonally amplified templates.

Most imaging systems are unable to detect single fluorescent events, thus amplified templates are required. The two most general methods for amplifying templates are emulsion PCR (emPCR) and solid-phase amplification. In emPCR, sequencing templates are prepared in a cell-free system, which avoids the random loss of genomic sequences, a problem that is intrinsic in bacterial cloning methods. A library of fragment or mate-pair targets is formed, and adaptors containing universal priming sites gets ligated to the target ends, allowing amplification of complex genomes with common PCR primers. After ligation, the DNA separate into single strands and captured onto beads under conditions which favor one DNA molecule per bead. Subsequently successful amplification and fortification of emPCR beads, millions can be immobilized in a poly-acrylamide gel on a standard microscope slide (Polonator), which gets chemically cross-linked to an amino-coated glass surface (Life/APG; Polonator) or else deposited into respective PicoTiterPlate (PTP) wells (Roche/454) where the NGS chemistry can be performed [2], [3], [6], [8], [18], [19].

(B) Single-molecule templates.

Although clonally amplified methods are advantageous than bacterial cloning, some of the protocols are burdensome for implementing and need a huge amount of genomic DNA material (3-20 μg). The preparation of single-molecule templates is more simple and requires less starting material (<1 μg). Apart from this, these methods do not require PCR, which may cause mutations in clonally amplified templates that pretend to be as sequence variants. Amplification bias in product yield is seen in AT-rich and GC-rich target sequences, which results in their under-representation in genome alignments and assemblies. Quantitative applications, such as RNA-seq, perform more efficiently with non-amplified template sources, which do not modify the representational abundance of mRNA molecules. Prior to the NGS reaction is carried out, single molecule templates are generally immobilized on solid supports by means of, either of three different methods. In the first method, spatially distributed individual primer molecules gets covalently attached to the solid support. The template is generated by randomly breaking the starting material into small sizes (for example, ~200-250 bp) and the addition of common adaptors to the fragment ends, which is hybridized to the immobilized primer. In the second method, spatially distributed single-molecule templates are made covalently attached to the solid support via priming and extending single-stranded, single-molecule templates from immobilized primers. After that, a common primer is hybridized to the template. In either approach, DNA polymerase gets bind to the immobilized primed template which initiates the NGS reaction. Both of the above methods are used by Helicos BioSciences. In a third method, spatially dispersed single polymerase molecules becomes attached to the solid support, to which a primed template molecule is bound. This approach is utilized by Pacific Biosciences. this technique use larger DNA molecules (up to tens of thousands of base pairs). Contrast to the first two approaches, the third approach can be used with real-time methods, which results in potentially longer read lengths [3], [20]–[24]

Template immobilization strategies are depicted in Figure 1. In emulsion PCR (emPCR) (a), a reaction mixture consisting of an oil-aqueous emulsion is created to encapsulate bead-DNA complexes into single aqueous droplets. PCR amplification is performed within these droplets to create beads containing several thousand copies of the same template sequence. EmPCR beads can be chemically attached to a glass slide or deposited into PicoTiterPlate wells. Solid-phase amplification (b) is composed of two basic steps: initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. Three approaches are shown for immobilizing single-molecule templates to a solid support: immobilization by a primer (c); immobilization by a template (d); and immobilization of a polymerase (e). dNTP, 2'-deoxyribonucleoside triphosphate.

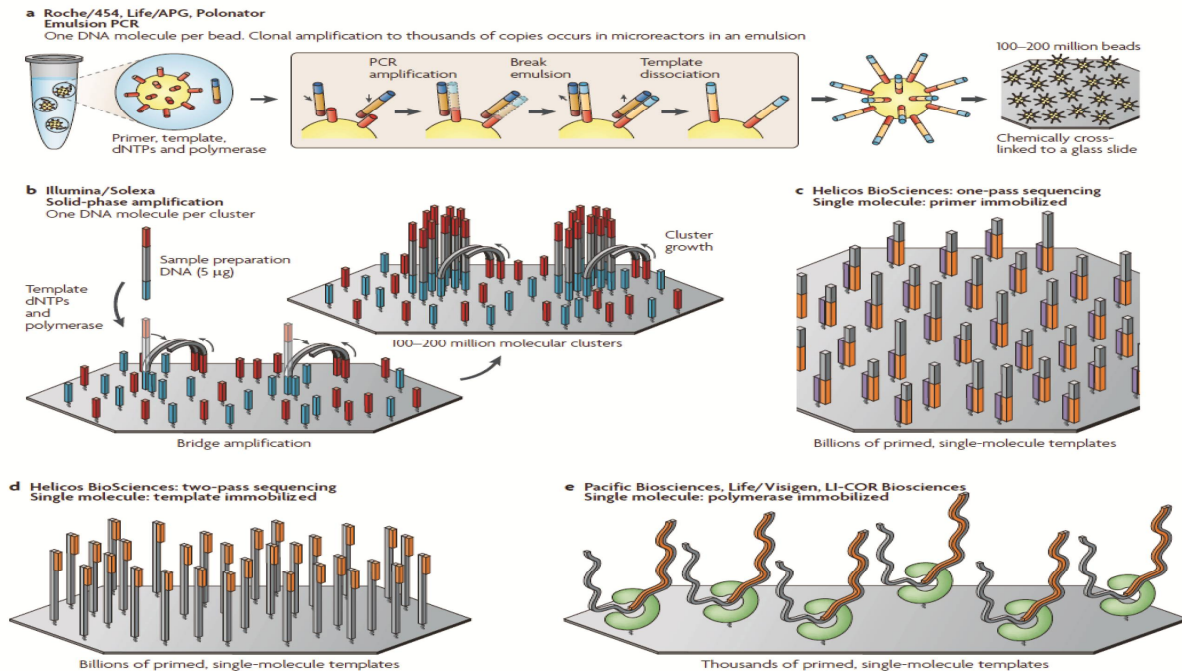


Figure 1 : Template immobilization strategies. This picture is adapted from (Metzker 2010)

2. Sequencing and imaging

There are basic differences in the product of sequencing clonally amplified and single-molecule templates. Clonal amplification generates a population of identical templates, each member of which has undergone the sequencing reaction. On imaging, the observed signal represents the consensus of the nucleotides or probes which are added to the similar templates for a given cycle. This may lead to the decrease in the efficiency of the addition process, and incomplete extension of the template assembly may result in lagging-strand dephasing. Leading-strand dephasing can also occur due to the addition of multiple nucleotides or probes in a given cycle, resulting in. Signal dephasing causes an increase in fluorescence noise, which further cause base-calling errors and shorter reads. Because dephasing is not generated with single-molecule templates, the requirement for cycle efficiency is undisturbed. Single molecules, on the other hand, are predisposed to multiple nucleotide or probe additions in any given cycle. Here, deletion errors will occur due to quenching effects between adjacent dye molecules or no signal will be detected owing to the integration of dark nucleotides or probes [3], [16], [25].

3. Genome alignment and assembly

Subsequently, after generation of NGS reads, they are aligned to a known reference sequence or assembled *de novo*. The choice to use either approach is based on the planned biological application as well as cost, effort and time considerations, e. g. in identifying and classifying genetic variation in multiple strains of highly related genomes, like those found in specific species of bacteria, *C. elegans* and *Arabidopsis thaliana*, which can be achieved by aligning NGS reads to their reference genomes. This approach is significantly cheaper and quicker than Sanger sequencing. *De novo* assemblies have been reported for the studies of bacterial genomes and mammalian bacterial artificial chromosomes only [1], [3].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

IV. COMPARISON OF NEXT-GENERATION SEQUENCING SYSTEMS

Following the human genome project, 454 launched 454 GS/FLX in 2005, and followed by the release of Genome Analyzer by Solexa, followed by (Sequencing by Oligo Ligation Detection) SOLiD which is provided by Agencourt, which are three important massively parallel sequencing systems in the NGS that shared good performance on throughput, accuracy, and cost compared with Sanger sequencing (shown in Table 1). The reagent cost of 454 GS FLX Titanium is calculated based on the sequencing of 400 bp; the reagent cost of HiSeq 2000 is calculated based on the sequencing of 200 bp; the reagent cost of SOLiDv4 is calculated based on the sequencing of 85 bp. HiSeq 2000 is more flexible in sequencing types like 50SE, 50PE, or 101PE. SOLiD has high accuracy especially when coverage is more than 30x, so it is widely used in detecting variations in resequencing, targeted resequencing, and transcriptome sequencing. Lanes can be independently run to reduce cost. Advantages and specifics associated with each sequencing system are summarized in Table 1, 2 and 3, which have been adapted from Liu et al. 2012 [11].

TABLE-1 : ADVANTAGE AND MECHANISM OF SEQUENCERS

SEQUENCERS	454 GS/FLX	HiSeq 2000	SOLiDv4	SANGER 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two base coding	Dideoxy-chain termination
Read length	700bp	50SE, 50PE, 101PE	50+35bp or 50+50bp	400~900bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	100.00%
Reads	1M	3G	1200~1400	-
Output data/run	0.7Gb	600GB	120GB	1.9~84 KB
Time/run	24 Hours	3~10 days	7 days for SE, 14 days for	20 mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length

TABLE-2 : COMPONENTS AND COST OF SEQUENCERS

Sequencers	454 GS/FLX	HiSeq 2000	SOLiDv4	SANGER 3730xl
Instrument price	Instrument \$500000, \$7000 per turn	Instrument \$690000, \$6000/(30x) human genome	Instrument \$495000, \$15000/100Gb	Instrument \$95000, about\$4/800bp reaction
CPU	2*Intel Xeon X5675	2*Intel Xeon X5560	8*processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1GB
Hard disk	1.1 TB	3 TB	10 TB	280 TB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REME System	cBot System	EZ beads system	No
Cost/million base	\$10	\$0.07	\$0.13	\$2400

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

TABLE-3 : COMPONENTS AND COST OF SEQUENCERS (adapted from (Liu et al. 2012))

Sequencers	454 GS/FLX	HiSeq 2000	SOLiDv4	SANGER 3730xl
Resequencing	No	Yes	Yes	No
De novo	Yes	Yes	No	Yes
Cancer	Yes	Yes	Yes	No
Array	Yes	Yes	Yes	Yes
High GC sample	Yes	Yes	Yes	No
Bacterial	Yes	Yes	Yes	No
Large genome	Yes	Yes	No	No

V. APPLICATION OF NEXT-GENERATION SEQUENCING SYSTEMS IN PLANT BIOLOGY

Next-generation sequencing technologies are revolutionizing scientific discovery at a rapid pace. Since genome sequencing is easily available, genomics is becoming an integral part of every branch of the life sciences including plant biology [6]. Below are some of the examples of the use of NGS technology in plant biology.

A. Molecular marker development :

Single nucleotide polymorphisms (SNPs) denote changes in a single base in DNA relative to that expected at that position. They can be used for determining genetic variation, associate SNPs to phenotypic variants, for example, disease associations in humans and traits for breeding in crop plants or construct genetic linkage maps. four commonly used short read alignment tools (SOAP2, Novoalign, BowTie, and Maq) was compared by Azam et al. with their new approach called coverage-based consensus calling (CbCC) for the discovery of SNP as a case study in chickpea, *Cicer arietinum* L. There is no reference genome for chickpea. Maq was found to be most accurate and sensitive. All four tools confirmed better accuracy at higher read depth, and SNPs predicted by three or four tools. With increasing read depth, the accuracy of SNP prediction generally increased. The results obtained in this study are appropriate for NGS-based SNP discovery in any other plant species, which do not have a reference genome [26].

B. Hybridization and introgression :

Weeds are economically significant, yet have been acknowledged little attention in the development of their genome resources. 22 EST libraries from a variety of tissues for 11 was generated by Lai et al. in the sunflower family by means of Sanger, 454, and Illumina sequencing, compared the coverage and quality of sequence assemblies and developed NimbleGen microarrays for expression analyses in five of them [10].

C. Transcriptome investigations:

NGS method that sequences the transcriptome (all RNA transcripts) is called RNA-seq. It is practical in non-model plant species mainly because a reference genome is not required. Common reed (*Phragmites australis* L.) is one of the most common angiosperms, which like many vascular plants, including important hay, forage, pasture crops, and wild grain species, which uses rhizomes as propagation organ and the primary energy storage. Comparative transcriptomic and proteomic study of developing rhizomes in common reed is done by He et al. for identification of candidate genes and proteins which are involved in rhizome growth, development, metabolism, and invasiveness. A Large number of phytohormone-related transcripts/genes which are important for the growth and development of rhizome are expressed in the developing tips of reed rhizomes. In addition, a number of putative transcription factors and regulators were also identified that may play important roles in development and function of the rhizome [27].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

D. Phylogenetic and ecological studies:

Steele et al. had examined the use of NGS for the study of phylogeny of two lineages of monocots, the Asparagales, and the grasses, using Illumina data (80 - 120-bp reads). They confirm that even low-coverage data provide genomic sequences of high-copy sufficiently well to provide high-quality assemblies. Adequate quantity of phylogenetically informative characters are provided by these data, to produce robust phylogenies, even of closely related and recently diverged taxa that were unmanageable to phylogenetic analysis before. Their results were not dependent on genome size or amount of plastid present in the total genomic DNA template or relatedness of available reference sequences for assembly. Costs are cheaper for generating the data, and much laboratory time is even saved [28].

VI. CONCLUSION

With the continuous improvement in NGS technologies, their scope and application will equally expand within and across scientific disciplines. Plant biology is getting benefited from increasing our technological capability in genomics, with applications reaching from plant breeding to evolutionary studies. With respect to comparative genomics, the increasing number of fully sequenced plant genomes will facilitate greater understanding of genetic, genomic, developmental, and evolutionary processes which create the diversity of plant life on earth. The improvement and application of NGS technologies will paint a bright future for plant biology along with other areas of life science research, as illustrated above.

REFERENCES

- [1] Roy Somak, LaFramboise William A., Nikiforov Yuri E., Nikiforova Marina N., Routbort Mark J., Pfeifer John, Nagarajan Rakesh, Carter Alexis B., and Pantanowitz Liron, "Next-Generation Sequencing Informatics," *Arch. Pathol. Lab. Med.*, Feb. 2016.
- [2] C. I. N. Unamba, A. Nag, and R. K. Sharma, "Next Generation Sequencing Technologies: The Doorway to the Unexplored Genomics of Non-Model Plants," *Front. Plant Sci.*, vol. 6, Dec. 2015.
- [3] M. L. Metzker, "Sequencing technologies—the next generation," *Nat. Rev. Genet.*, vol. 11, no. 1, pp. 31–46, 2010.
- [4] J. Fu and G. Lin, "[Application of molecular diagnostic techniques in precision medicine of personalized treatment for colorectal cancer].," *Zhonghua Wei Chang Wai Ke Za Zhi Chin. J. Gastrointest. Surg.*, vol. 19, no. 1, pp. 22–26, 2016.
- [5] J. Zhang and H. Xing, "[Application of Next-generation Sequencing Techniques in the Dynamics of HIV-1 Quasispecies].," *Bing Xue Bao Chin. J. Virol. Bian Ji Bing Xue Bao Bian Ji Wei Yuan Hui*, vol. 31, no. 5, pp. 573–578, Sep. 2015.
- [6] A. N. Egan, J. Schlueter, and D. M. Spooner, "Applications of next-generation sequencing in plant biology," *Am. J. Bot.*, vol. 99, no. 2, pp. 175–185, 2012.
- [7] M. Okuno, R. Kajitani, R. Ryusui, H. Morimoto, Y. Kodama, and T. Itoh, "Next-generation sequencing analysis of lager brewing yeast strains reveals the evolutionary history of interspecies hybridization," *DNA Res.*, p. dsv037, 2016.
- [8] C. Angelini, M. Savarese, M. Fanin, and V. Nigro, "Next generation sequencing detection of late onset pompe disease," *Muscle Nerve*, vol. 53, no. 6, pp. 981–983, 2016.
- [9] B. Suter, X. Zhang, C. G. Pesce, A. R. Mendelsohn, S. P. Dinesh-Kumar, and J.-H. Mao, "Next-Generation Sequencing for Binary Protein–Protein Interactions," *Syst. Biol.*, p. 346, 2015.
- [10] Z. Lai, N. C. Kane, A. Kozik, K. A. Hodgins, K. M. Dlugosch, M. S. Barker, M. Matvienko, Q. Yu, K. G. Turner, S. A. Pearl, and others, "Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression," *Am. J. Bot.*, vol. 99, no. 2, pp. 209–218, 2012.
- [11] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, and M. Law, "Comparison of next-generation sequencing systems," *BioMed Res. Int.*, vol. 2012, 2012.
- [12] O. Morozova and M. A. Marra, "Applications of next-generation sequencing technologies in functional genomics," *Genomics*, vol. 92, no. 5, pp. 255–264, Nov. 2008.
- [13] A. D. Radford, D. Chapman, L. Dixon, J. Chantrey, A. C. Darby, and N. Hall, "Application of next-generation sequencing technologies in virology," *J. Gen. Virol.*, vol. 93, no. Pt 9, pp. 1853–1868, Sep. 2012.
- [14] T. Shen, S. H. Pajaro-Van de Stadt, N. C. Yeat, and J. C.-H. Lin, "Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes," *Front. Genet.*, vol. 6, Jun. 2015.
- [15] E. C. Berglund, A. Kiialainen, and A.-C. Syvänen, "Next-generation sequencing technologies and applications for human genetic history and forensics," *Investig. Genet.*, vol. 2, p. 23, 2011.
- [16] A. B. Mazumdar and S. Chattopadhyay, "Sequencing, De novo Assembly, Functional Annotation and Analysis of Phyllanthus amarus Leaf Transcriptome Using the Illumina Platform," *Front. Plant Sci.*, vol. 6, 2015.
- [17] G. P. Wright, D. W. Chesla, and M. H. Chung, "Using next-generation sequencing to determine potential molecularly guided therapy options for patients with resectable pancreatic adenocarcinoma," *Am. J. Surg.*, vol. 211, no. 3, pp. 506–511, Mar. 2016.
- [18] M. I. Lefterova, P. Shen, J. I. Odegaard, E. Fung, T. Chiang, G. Peng, R. W. Davis, W. Wang, M. Kharrazi, I. Schrijver, and others, "Next-generation molecular testing of newborn dried blood spots for cystic fibrosis," *J. Mol. Diagn.*, vol. 18, no. 2, pp. 267–282, 2016.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 8, August 2016

- [19]B. Vrancken, N. S. Trovão, G. Baele, E. van Wijngaerden, A.-M. Vandamme, K. van Laethem, and P. Lemey, "Quantifying Next Generation Sequencing Sample Pre-Processing Bias in HIV-1 Complete Genome Sequencing," *Viruses*, vol. 8, no. 1, Jan. 2016.
- [20]B. Xu, Y.-W. Zhang, S.-X. Zheng, X.-H. Tong, and Y.-S. Liu, "Expression Profile of microRNAs and Their Targeted Pathways in Human Ovaries Detected by Next-Generation Small RNA Sequencing," *DNA Cell Biol.*, vol. 35, no. 5, pp. 226–234, May 2016.
- [21]S. Sun and X. Yu, "HMM-Fisher: identifying differential methylation using a hidden Markov model and Fisher's exact test," *Stat. Appl. Genet. Mol. Biol.*, vol. 15, no. 1, pp. 55–67, Mar. 2016.
- [22]Y. Lin, E. A. Lewallen, E. T. Camilleri, C. A. Bonin, D. L. Jones, A. Dudakovic, C. Galeano-Garces, W. Wang, M. J. Karperien, A. N. Larson, and others, "RNA-seq analysis of clinical-grade osteochondral allografts reveals activation of early response genes," *J. Orthop. Res.*, 2016.
- [23]Á. C. Román, A. Morales-Hernández, and P. M. Fernández-Salguero, "RNA-Seq Analysis to Measure the Expression of SINE Retroelements," *Methods Mol. Biol. Clifton NJ*, vol. 1400, pp. 107–116, 2016.
- [24]S. E. P. Joosten, V. P. Retèl, V. M. H. Coupé, M. M. van den Heuvel, and W. H. van Harten, "Scenario drafting for early technology assessment of next generation sequencing in clinical oncology," *BMC Cancer*, vol. 16, no. 1, p. 1, 2016.
- [25]B. A. Talseth-Palmer, D. C. Bauer, W. Sjursen, T. J. Evans, M. McPhillips, A. Proietto, G. Otton, A. D. Spigelman, and R. J. Scott, "Targeted next-generation sequencing of 22 mismatch repair genes identifies Lynch syndrome families," *Cancer Med.*, vol. 5, no. 5, pp. 929–941, May 2016.
- [26]S. Azam, V. Thakur, P. Ruperao, T. Shah, J. Balaji, B. Amindala, A. D. Farmer, D. J. Studholme, G. D. May, D. Edwards, and others, "Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome," *Am. J. Bot.*, vol. 99, no. 2, pp. 186–192, 2012.
- [27]R. He, M.-J. Kim, W. Nelson, T. S. Balbuena, R. Kim, R. Kramer, J. A. Crow, G. D. May, J. J. Thelen, C. A. Soderlund, and D. R. Gang, "Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, *Phragmites australis* (Poaceae), reveals genes involved in invasiveness and rhizome specificity," *Am. J. Bot.*, vol. 99, no. 2, pp. 232–247, Feb. 2012.
- [28]P. R. Steele, K. L. Hertweck, D. Mayfield, M. R. McKain, J. Leebens-Mack, and J. C. Pires, "Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae," *Am. J. Bot.*, vol. 99, no. 2, pp. 330–348, Feb. 2012.