ISSN : 0974 - 7532 Volume 6 Issue 9

Trade Science Inc.

Research & Reviews in

BioSciences

Review

RRBS, 6(9), 2012 [240-247]

# Next generation sequencing technologies - Principles and prospects

**Md.Fakruddin[1]*, Khanjada Shahnewaj Bin Mannan[2]**

**[1]Institute of Food Science and Technology (IFST), Bangladesh Council of Scientific and Industrial Research (BCSIR), Dhaka, (BANGLADESH)**

**[2]Center for Food & Waterborne Disease (CFWD), icddr'b, Dhaka, (BANGLADESH)**

**E-mail : fakruddinmurad@gmail.com**

*Received: 4th July, 2012 ; Accepted: 13th September, 2012*

## ABSTRACT

DNA sequencing is one of the most important platforms for the study of biological systems today. The high-throughput - next generation sequencing (HT-NGS) technologies are currently the hottest topic in the field of human and animals genomics researches, which can produce over 100 times more data compared to the most sophisticated capillary sequencers based on the Sanger method. New generation of sequencing technologies, from Illumina/Solexa, ABI/SOLiD, 454/Roche, and Helicos, has provided unprecedented opportunities for high-throughput functional genomic research. The next-generation sequencing technologies offer novel and rapid ways for genome-wide characterisation and profiling of mRNAs, small RNAs, transcription factor regions, structure of chromatin and DNA methylation patterns, microbiology and metagenomics. However, unlike traditional Sanger dideoxy sequencing, these methods have lower accuracy and shorter read lengths than the dideoxy gold standard. An astounding potential exists for these technologies to bring enormous change in genetic and biological research and to enhance our fundamental biological knowledge. 

## KEYWORDS

Sequencing;
Next generation;
Challenges;
Prospects;
Sanger.

## INTRODUCTION

DNA sequencing is one of the most important platforms for the study of biological systems today. Sequence determination is most commonly performed using di-deoxy chain termination technology[1]. The chain termination sequencing method, also known as Sanger sequencing, was developed by Frederick Sanger and colleagues[2], has been the most widely used sequencing method since its advent in 1977 and still is in use after more than 29 years. Despite all the advantages, there are limitations in this method, which could be complemented with other techniques[3]. Many research groups around the world have made effort to develop alternative principles of DNA sequencing. Three methods that hold great promise are sequencing by hybridization[4-7] parallel signature sequencing based on ligation and cleavage[8] and pyrosequencing[9,10].

## SANGER SEQUENCING

The method is based on the DNA polymerase-dependent synthesis of a complementary DNA strand in the presence of natural 2′-deoxynucleotides (dNTPs)

and 2′, 3′ - dideoxynucleotides (ddNTPs) that serve as nonreversible synthesis terminators[2]. Sanger dideoxy sequencing[2] and its modifications[11-13] dominated the DNA sequencing field for nearly 30 years and in the past 10 years the length of Sanger sequence reads has increased from 450 bases to more than 1 kb.

## LIMITATIONS OF SANGER SEQUENCING

The Sanger sequencing method[2] has been the workhorse technology for DNA sequencing since it's invent. Though Sanger method is still considered by the research community as the gold standard for sequencing, it has several limitations such as- (1) A great limitation of the Sanger sequencing method for larger sequence output is the need for gels or polymers used as sieving separation media for the fluorescently labeled DNA fragments. (2) Relatively low number of samples could be analyzed in parallel. (3) Total automation of the sample preparation methods is difficult. (4) DNA fragments need to be cloned into bacteria for larger sequences. (5) High cost of sequencing. (6) Sequencing errors & Level of sensitivity (generally estimated at 10-20%) insufficient for detection of clinically relevant low-level mutant alleles or organisms. (7) *cis* or *trans* orientation of heterozygous positions may be difficult to resolve during data analysis. (8) Not readily scalable to achieve a throughput capable of efficiently analyzing complex diploid genomes at low cost. (9) de novo genome assembly is difficult[14, 15].

## ALTERNATIVE METHODS OF SEQUENCING

The term NGS is used to collectively describe technologies other than Sanger sequencing that have the potential to sequence the human genome in coming years for US$1000[16]. Commercially available NGS technologies such as Roche/454 (http://www.454.com/), Solexa/ Illumina (http://www.illumina.com/), AB SOLiD (http://www3.appliedbiosystems.com/ AB_Home/ applicationstechnologies/ SOLiDSystemSequencing/ index.html) and Helicos Biosciences (Amplification-independent (single molecule) sequencing methods Single molecule sequencing (SMS)) (http://www.helicosbio.com/) have already demonstrated the potential to circumvent the limiting factors of Sanger sequencing[17, 18].

Apart from the commercially available tSMS (true SMS) launched by Helicos Biosciences (http://www.helicosbio.com/), SMS (Single molecule sequencing) development is underway at several academic laboratories and companies such as Biotage (http://www.biotage.com/), Li-COR Biosciences (http://www.licor.com/)[19], Nanogen (http://www.nanagen.com/), Network Biosystems (http://www.networkbiosystems.com/) and Visi-Gen Biotechnologies Inc. (http://visigenbio.com/). Pacific Biosciences (http://www.pacificbiosciences.com/) has recently reported real-time sequencing[20]. It is noteworthy that all NGS technologies are constantly improving, with the goal to reduce error rates and to increase the sequence read length and read number[21].

### Roche (454) GS FLX pyrosequencer

The first NGS system to become commercially available was the Genome Sequencer from 454 Life Sciences (Branford, CT, USA) (later acquired by Roche) in 2005[17]. This sequencer works on the principle of 'pyrosequencing'[9], which uses the pyrophosphate molecule released on nucleotide incorporation by DNA polymerase to fuel a downstream set of reactions that ultimately produces light from the cleavage of oxyluciferin by luciferase[22, 23].

Roche (454) GS FLX pyrosequencer circumvents the cloning requirement of Sanger sequencing by taking advantage of a highly efficient in vitro DNA amplification method known as emulsion PCR[24]. In emulsion PCR, individual DNA fragment-carrying streptavidin beads, obtained through shearing the DNA and attaching the fragments to the beads using adapters, are captured into separate emulsion droplets. The droplets act as individual amplification reactors, producing ~$10^7$ clonal copies of a unique DNA template per bead[22]. Each template-containing bead is subsequently transferred into a well of a picotiter plate and the clonally related templates are analyzed using a pyrosequencing reaction. The use of the picotiter plate allows hundreds of thousands of pyrosequencing reactions to be carried out in parallel, massively increasing the sequencing throughput[19].

This method has significant advantages over Sanger sequencing because it requires no electrophoresis step to separate extension products and base incorporation

Review

can be detected in real time[25]. The precision of Roche/ 454 sequencing technology in handling homopolymers (short stretches of the same contiguous nucleotides) suffers in comparison with other NGS technologies.

The current 454 instrument, the GS-FLX, produces an average read length of ~250 bp per sample (per bead), with a combined throughput of ~100 Mb of sequence data per 7 h run. By contrast, a single ABI 3730 programmed to sequence 24 X 96-well plates per day produces ~440 kb of sequence data in 7 h, with an average read length of 650 bp per sample[26]. The next upgrade 454 FLX Titanium will quintuple the data output from 100 Mb to about 500 Mb, and the new picotiter plate in the device uses smaller beads about 1 mm diameter[14].

**Illumina/ Solexa genome analyzer**

Introduced in 2006, the Illumina Genome Analyzer is based on the concept of 'sequencing by synthesis' (SBS) to produce sequence reads of ~32–40 bp from tens of millions of surface amplified DNA fragments simultaneously[19]. The Illumina/Solexa approach[27] achieves cloning-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a single-molecule array, or flowcell, and conducting solid-phase bridge amplification of single-molecule DNA templates (Illumina, Inc.). In this process, one end of single DNA molecule is attached to a solid surface using an adapter; the molecules subsequently bend over and hybridize to complementary adapters (creating the "bridge"), thereby forming the template for the synthesis of their complementary strands[28]. After the amplification step, a flow cell with more than 40 million clusters is produced, wherein each cluster is composed of approximately 1000 clonal copies of a single template molecule[29]. The templates are sequenced in a massively parallel fashion using a DNA sequencing-by-synthesis approach that employs reversible terminators with removable fluorescent moieties and special DNA polymerases that can incorporate these terminators into growing oligonucleotide chains. The terminators are labeled with fluors of four different colors to distinguish among the different bases at the given sequence position and the template sequence of each cluster is deduced by reading off the color at each successive nucleotide addition step. A base-calling algorithm assigns sequences and associated quality values to each read and a quality checking pipeline evaluates the Illumina data from each run, removing poor-quality sequences[26].

Although the Illumina approach is more effective at sequencing homopolymeric stretches than pyrosequencing, it produces shorter sequence reads and hence cannot resolve short sequence repeats[18]. In addition, due to the use of modified DNA polymerases and reversible terminators, substitution errors have been noted in Illumina sequencing data[30]. Typically, the 1G genome analyzer from Illumina, Inc., is capable of generating 35-bp reads and producing at least 1 Gb of sequence per run in 2–3 days.

In 2008 Illumina introduced an upgrade, the Genome Analyzer II that triples output compared to the previous Genome Analyzer instrument. A paired-end module for the sequencer was introduced, and with new optics and camera components that allow the system to image DNA clusters more efficiently over larger areas, the new instrument triples the output per paired-end run from 1 to 3 Gb[14]. The system generates at least 1.5 Gb of single-read data per run, at least 3 Gb of data in a paired-end run, recording data from more than 50 million reads per flow cell. The run time for a 36- cycle run was decreased to two days for a single-read run, and four days for a paired-end run[31].

**Applied biosystems SOLiD sequencer**

Applied Biosystems SOLiD sequencer, which achieved commercial release in October 2007, uses a unique sequencing process catalyzed by DNA ligase[26].

DNA fragments are ligated to adapters then bound to beads[14]. SOLiD instrument begins with an emulsion PCR single-molecule amplification step similar to that used in the 454 technique. The amplification products are transferred onto a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotide combinations labeled by four different fluorescent dyes (each dye used to label four dinucleotides). Using the four dye encoding scheme, each position is effectively probed twice, and the identity of the nucleotide is determined by analyzing the color that results from two successive ligation reactions. Significantly, the two base encoding scheme enables the distinction between a sequencing error and a sequence polymorphism: an error would be detected in only one particular ligation reaction, whereas a polymorphism would be detected in both[18].

The achieved sequence reading length is at present about 35 bases. Because each base is determined with a different fluorescent label, error rate is reduced. Sequences can be determined in parallel for more than 50 million bead clusters, resulting in a very high throughput of the order of Gigabases per run. Applied Biosystems produced an updated version in 2008, the SOLiD 2.0 platform, which may increase the output of the instrument from 3 to 10 Gb per run. This change will reduce the overall run time of a fragment library on the new system to 4.5 days from 8.5 days on the existing machine[14].

## The Helicos single-molecule sequencing device, HeliScope

One of the first techniques for sequencing from a single DNA molecule was described by the team of S. Quake[32], and licensed by Helicos Biosciences. Helicos introduced the first commercial single-molecule DNA sequencing system in 2007. The nucleic acid fragments are hybridized to primers covalently anchored in random positions on a glass cover slip in a flow cell. The primer, polymerase enzyme and labelled nucleotides are added to the glass support. The next base incorporated into the synthesised strand is determined by analysis of the emitted light signal, in the sequencing-by-synthesis technique. This system also analyses many millions of single DNA fragments simultaneously, resulting in sequence throughput in the Gigabase range[33].

Read lengths averaged about 23 bases. There were still some limitations in the single-molecule technology, on the basis of the first generation of the chemistry. In the homopolar regions, multiple fluorophore incorporations could decrease emissions, sometimes below the level of detection; when errors did occur, most were deletions. Helicos announced that it has recently developed a new generation of 'one-base-at-a-time' nucleotides which allow more accurate homopolymer sequencing, and lower overall error rates[14].

## Other future techniques

Developments of novel DNA sequencing techniques are taking place in many groups worldwide.

The laboratory of Church[34] has developed a sequencing technology similar to sequencing-by synthesis principle combining with multiple polony technology. It has the advantage of the capacity to analyze increased magnitude of samples simultaneously and reduce reaction volume, reduced reagent requirement and lower cost[35].

VisiGen Biotechnologies (http://visigenbio.com/) has developed another promising approach using real-time single molecule DNA sequencing principle. This method uses a special DNA polymerase with a donor fluorescent dye and four types of modified nucleotides with different acceptor dye. When correct nucleotide is found, donor dye of polymerase comes into close proximity with acceptor dye on the nucleotides and energy transfer between them produces fluorescent resonant energy transfer (FRET) light signal. Through this technology, an entire human genome can be sequenced in less than a day for under $1000. It is expected to generate around 4Gb of data per day. This principle obviates the requirement of cloning and amplification and thus offer lower cost. Read length is expected to be around 1Kb[36].

Pacific Biosciences (http://www.pacificbiosciences.com/index.php), a US company has developed a next-generation DNA sequencing instrument that will eventually be able to produce 100 Gb of sequence data per hour, i.e. a diploid human genome at onefold coverage in about 4 min. This method uses single molecule real-time (SMRT) technology which is based on zero mode waveguides (ZMWs). Expected read length is about 1500 bases with a rate of 10 bases/s, and able to analyse up to 3000 ZMWs in parallel[37].

A new single-molecule DNA sequencing approach using RNA polymerase (RNAP) has been described by Greenleaf and Block[38]. In the planned method, RNAP is attached to one polystyrene bead, whilst the distal end of a DNA fragment is attached to another bead. Each bead is placed in an optical trap and the pair of optical traps levitates the beads. The RNAP interacts with the DNA fragment and the transcriptional motion of RNAP along the template changes the length of the DNA between the two beads. This leads to displacement of the two beads that can be registered with precision in the Angstrom range, resulting in single-base resolution on a single DNA molecule. By aligning four displacement records, each with a lower concentration of one of the four nucleotides, in a role analogous to the primers used in Sanger sequencing, and using for calibration the known sequences flanking the unknown fragment to be sequenced, it is possible to deduce the se-

Review

quence information. Thirty out of 32 bases were correctly identified in about 2 min. The technique demonstrates that the movement of a nucleic acid enzyme, and the very sensitive optical trap method, may allow extraction of sequence information directly from a single DNA molecule[14].

Other future techniques involved nano-DNA sequencing, Sequenom (http://www.sequenom.com), BioNanomatrix and Complete Genomics (http://bionanomatrix.com/) etc.[14].

## ADVANTAGES OF NEXT GENERATION SEQUENCING

Next generation sequencing technologies offer several advantages over traditional Sanger sequencing methods, such as-

1. Considerably higher throughput.
2. Lower cost per sequenced base[17].
3. Ability to discover novel genetic variations is very high.
4. Greater sensitivity.
5. Determination of the sequence data from amplified single DNA Fragments.
6. Avoiding the need for cloning of DNA fragments.
7. Reduced of sequencing errors[14]

## COMPARISON OF NEXT GENERATION SEQUENCING METHODS

A comprehensive comparison of different next generation sequencing methods has been presented in TABLE 1.

## CHALLENGES

Next generation sequencing technologies are still in their premature stage facing many challenges which should be resolved in future to make these methods as applicable as Sanger sequencing. The challenges are as follows:

1. The major limitation of a next generation sequencing approach is that the length of the sequence reads produced was until recently only 25-200 bases, as opposed to over a kilobase generated by conventional capillary based sequencing methods. Although short sequence reads do not limit the amount of sequence data collected, this can hamper the assembly of the short sequence reads into large contigs[40].
2. There is a great need for quantitative studies and analysis tools that help investigators optimally design NG sequencing experiments to address specific goals[41].
3. NGS data analysis: In a majority of the sequencing projects the quantum of sequencing data has outpaced computational capabilities, making NGS data analysis and management the biggest bottleneck and a field in itself for research which is still evolving. Although the cost of actual sequencing is reducing drastically, the associated bioinformatics cost for NGS data storage and analysis has grown exponentially. Unlike microarray which has multiple robust analysis solutions, NGS data analysis largely rely on non-standard open source tools and requires highly trained Bioinformaticians. Although few commercial solutions are available they are extremely expensive and not very reliable.
4. Computing infrastructure: NGS data demands sophisticated and high-end computing infrastructure. For example, to perform de novo assembly and annotation of mammalian genome, a system with atleast eight quad core processor and 512 GB RAM along with 10 terabytes (TB) of disk space is required. Additionally, highly skilled IT and bioinformatics staff is required to set up, maintain and run NGS data analysis tools.
5. Commercially unviable: The current NGS instruments available are not capable of sequencing complete human genomes on a large scale at a low price. Researchers cannot afford these high costs.

## FUTURE PROSPECTS

NGS is indeed a revolutionary sequencing technology. It offers an unprecedented opportunity to informatics scientists, biologists and clinicians to collaborate and overcome the current bottlenecks and explore genomic information in-depth[19]. The availability of ultra-deep sequencing of genomic DNA will transform the biological and medical fields in the near future, in analysis of the causes of disease, development of new drugs and diagnostics. It may become a promising tool in the analysis of mental and developmental disor-

**TABLE 1: Comparison of second and third HT NGS platforms[39]**

| Companies | Roche GS FLX | llumina-Sollexa | 454 Life Technologies | Helicos Biosciences | Pacific Biosciences |
|---|---|---|---|---|---|
| Company homepage | http://www.454.com/ | http://www.solexa.com/ | http://www3. appliedbiosystems.com/ | http://www.helicosbio. com/ | http://www. pacificbiosciences.com |
| Platforms | GS FLX Tiatnium | Hiseq 2000, Genome analyzer | ABI SOLiD, SOLiD 4 | Heliscope | SMRT |
| Template Preparation | Clonal-ePCR on bead surface | Clonal bridge enzymatic amplification on glass surface | Clonal-ePCR on bead surface | Single molecule detection | Single molecule detection |
| Sample requirements | 1 µg for shotgun library, 5 µg for paired end | <1 g for single or pairedend libraries | <2 µg for shotgun library, 5–20 µg for paired end | <2 µg, single end only | Not available (NA) |
| Detection method | Light emitted from secondary reactions initiated by release of pyrophosphate | Fluorescent emission from incorporated dye-labelled nucleotides | Fluorescent emission from ligated dye-labelled oligonucleotides | Real time detection of fluorescent dye in polymerase active site during incorporation | Real time detection of fluorescent dye in polymerase active site during incorporation |
| Length of library prep/ feature generation (days) | 3-4 | 2 | 2-4.5 | 1 | NA |
| Method of feature generation | Bead-based/emulsion PCR | Isothermal 'bridge amplification' on flow cell surface | Bead-based/emulsion PCR | Single molecule sequencing | Single molecule real time sequencing by synthesis |
| Paired ends /separation | 3 kb (2×110 p) | 200 bp (2×36 bp) | 3 kb (2×25 bp) | 25–55 bp | NA |
| Chemistry | Pyrosequencing | Reversible Dye Terminators | Oligonucleotide Probe Ligation | Reversible Dye Terminators | Phospho-linked Fluorescent Nucleotides |
| Bases/template | ~400 | ~75 (35–100) | 35–50 | 35 | 800–1000 |
| Templates run | 1,000,000 | 40,000,000 | 85,000,000 | NA | NA |
| Data production/day | 400 MB/run/7.5 hr | 3,000 MB/run/6.5 days | 4,000 MB/run/6 days | 8 days | 0.02 days |
| Maximum samples | 16 regions/plate | 8 channels/flow cell | 16chambers/2 slides | NA | NA |
| Raw accuracy | 99.5% | >98.5% | 99.94% | >99% | NA |
| Sequencing method | Pyrosequencing | Reversible dye terminators | Sequencing by ligation | One base-at-a-time | Sequencing by synthesis |
| Read lengths | 400 bases | 36 bases | 35 bases | Longer than 1000 | Longer than 1000 |
| Sequencing run time | 10 h | 2-5 days | 6 days | 12 | <1 |
| Total Throughput bases/run (Gb) | 0.40–0.60 Gb, 0.035 Gb | 3–6 Gb | 10–20 Gb | 28 GB | 100 Gb per hour |
| Throughput /day (Gb) | ~1 | 1.5 | 1.7–2 | 2.5 | ~1 |
| Estimated system cost | $500,000 | ~$400,000 | $525,000 | Lower than second NGS | Lower than second NGS |
| Consumable cost per singleend run (paired-end run) | $5000 | $3000 | $4000 | Lower than second NGS | Lower than second NGS |
| Cost per run (total direct) | $8439 | $8950 | $17,447 | Lower than second NGS | Lower than second NGS |
| Cost per Mb | $84.39 | $5.97 | $5.81 | Lower than second NGS | Lower than second NGS |

ders such as schizophrenia and autism[42-44].

It is now widely being accepted that the exhaustive sequence knowledge gleaned using NGS will considerably impact the field of medicine. By studying how

Review

genes, differ amongst species and also amongst individuals within the same species, scientists are now in a position to better understand the role and functions genes play in affecting health and disease progression[17]. This improved understanding of genomes will be the new driver of medical advancements and will further facilitate the development of novel diagnostic assays, targeted therapies and better assessment of hypersensitivities/drug resistance along with improved ability to predict the onset, severity and progression of diseases. NGS is indeed a major leap forward in the direction of understanding personalized genomics and medicine[14].

Genome resequencing will likely be used to characterize strains or isolates relative to high-quality reference genomes such as *C. elegans*, *Drosophila*, and human[19]. Epigenomic variation, as an extension of genome resequencing applications, also will be investigated using next-generation sequencing approaches that enable the ascertainment of genome-wide patterns of methylation and how these patterns change through the course of an organism's development, in the context of disease, and under various other influences[26].

The novel sequencing technologies will be also useful in microbial genomics, for example in the metagenomics measuring the genetic diversity encoded by microbial life in organisms inhabiting a common environment[45]. Genomics, proteomics and medical research all benefit from recent advances and novel techniques for high-throughput analysis (e.g. DNA and protein microarrays, quantitative PCR, mass spectrometry, novel DNA sequencing techniques and others)[14].

## CONCLUSION

The sequence-based characterization of genomes is a relatively young pursuit in the biological sciences that has to date primarily enhanced model organism and human genetics, providing a substrate to discover genes and to understand genetics. This fundamental knowledge is now being enhanced by the ability to gather genome-wide sequence information more rapidly that can inform a higher-level appreciation of the functional genome, using front-end techniques combined with massively parallel throughput sequencing data production. Several early studies have shown the power of this new paradigm, and only time and ingenuity will determine its boundaries and its consequent ability to transform genetics.

## REFERENCES


[1] R.Pool; Research News, **245**, 1187-1189 **(1989)**.
[2] M.Ronaghi; Genome Res., **11**, 3-11 **(2001)**.
[3] F.Sanger, S.Nicklen, A.R.Coulson; Proc.Natl.Acad. Sci.USA, **74**, 5463-5467 **(1977)**.
[4] B.Gharizadeh, M.Ghaderi, P.Nyren; Technology, **47**, 129-132 **(2007)**.
[5] W. Bains, G.C.Smith; J.Theoretical.Biol., **135**, 303-307 **(1988)**.
[6] R.Drmanac, I.Labat, I.Brukner, R.Crkvenjakov; Genomics, **4**, 114-128 **(1989)**.
[7] K.R.Khrapko, Y.P.Lysov, A.A.Khorlyn, V.V.Shick, V.L.Florentiev *et al.*; FEBS Lett., **256**, 118-122 **(1989)**.
[8] E.M.Southern; US patent no. WO/10977 **(1989)**.
[9] S.Brenner, S.R.Williams, E.H.Vermaas, T.Storck, K.Moon; Proc.Natl.Acad.Sci., **97**, 1665-1670 **(2000)**.
[10] M.Ronaghi, S.Karamohamed, B.Pettersson, M.Uhlen, P.Nyren; Anal.Biochem., **242**, 84-89 **(1996)**.
[11] M.Ronaghi, M.Uhlen, P.Nyren; Science, **281**, 363-365 **(1998)**.
[12] J.M.Prober et al.; Science, **238**, 336-341 **(1987)**.
[13] L.M.Smith et al.; **321**, 674-679 **(1986)**.
[14] R.S.Madabhushi; Electrophoresis, **19**, 224-230 **(1998)**.
[15] W.J.Ansorge; New Biotechnology, **25(4)**, 195-203 **(2009)**.
[16] N.Hall; J.Exp.Biol., **209**, 1518-1525 **(2007)**.
[17] R.F.Service; Science, **311**, 1544-1546 **(2006)**.
[18] E.C.Berglund, A.Kiialainen, A.-C.Syvänen; Investigative Genetics, **2**, 23 **(2011)**.
[19] O.Morozova, M.A.Marra; Genomics, **92**, 255-264 **(2008)**.
[20] E.R.Mardis; Trends in Genetics, **24(3)**, 133-141 **(2008a)**.
[21] J.Eid et al.; Science, **323**, 133-138 **(2009)**.
[22] R.K.Varshney, S.N.Nayak, G.D.May, S.A.Jackson; Trends in Biotechnology, **27(9)**, 522-530 **(2009)**.
[23] M.Margulies et al.; Nature, **437**, 376-380 **(2005)**.
[24] M.Fakruddin, A.Chowdhury, M.N.Hossain, K.S.B.Mannan, R.M.Mazumdar; Int.J.of Life Sci.and Pharma Res., **2(2)**, L65-L76 **(2012)**.
[25] D.S.Tawfik, A.D.Griffiths; Nat.Biotechnol., **16**, 652-656 **(1998)**.
[26] M.Fakruddin, A.Chowdhury; American Journal of Biochemistry and Biotechnology, **8(1)**, 14-20 **(2012)**.
[27] E.R.Mardis; Annu.Rev.Genomics Hum.Genet., **9**, 387-402 **(2008b)**.

**[28]** S.Bennett; Pharmacogenomics, **5**, 433-438 **(2004)**.
**[29]** S.T.Bennett, C.Barnes, A.Cox, L.Davies, C.Brown; Pharmacogenomics., **6**, 373-382 **(2005)**.
**[30]** D.R.Bentley; Curr.Opin.Genet.Dev., **16**, 545-552 **(2006)**.
**[31]** C.A.Hutchison; Nucleic Acids Res., **35**, 6227-6237 **(2007)**.
**[32]** S.C.Schuster et al.; Nat.Methods, **5**, 11-21 **(2008)**.
**[33]** S.R.Quake et al.; Proc.Natl.Acad.Sci.U.S.A., **105**, 16266-16271 **(2008)**.
**[34]** I.Braslavsky et al.; Proc.Natl.Acad.Sci.U.S.A., **100**, 3960-3964 **(2003)**.
**[35]** J.Shendure et al.; Science, **309**, 1728-1732 **(2005)**.
**[36]** D.G.Hert, C.P.Fredlake, A.E.Barron; Electrophoresis, **29**, 4618-4626 **(2008)**.
**[37]** L.T.C.Franca, E.Carrilho, T.B.L.Kist; Quarterly Reviews of Biophysics, **35(2)**, 169-200 **(2002)**.
**[38]** M.Nowrousian; Eukaryotic Cell, **9(9)**, 1300-1310 **(2010)**.
**[39]** W.J.Greenleaf, S.M.Block; Science, **313**, 801 **(2006)**.
**[40]** C.S.Pareek, R.Smoczynski, A.Tretyn; J.Appl.Genetics, **52**, 413-435 **(2011)**.
**[41]** S.Subramanian, L.Huynen, C.D.Millar, D.M.Lambert; BMC Evolutionary Biology, **10**, 387-397 **(2010)**.
**[42]** P.K.Wall, J.Leebens-Mack, A.S.Chanderbali, A.Barakat, E.Wolcott, H.Liang, L.Landherr, L.P.Tomsho, Y.Hu, J.E.Carlson, H.Ma, S.C.Schuster, D.E.Soltis, P.S.Soltis, N.Altman, C.W.dePamphilis; BMC Genomics, **10**, 347-366 **(2009)**.
**[43]** E.M.Morrow et al.; Science, **321**, 218-223 **(2008)**.
**[44]** D.H.Geschwind; Nature, **454**, 838-839 **(2008)**.
**[45]** J.S.Sutcliffe; Science, **321**, 208-209 **(2008)**.
**[46]** P.Hugenholtz, G.W.Tyson; Nature, **455**, 481-483 **(2008)**.