World Applied Sciences Journal 24 (12): 1558-1571, 2013
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2013.24.12.2972

Pyrosequencing-A Next Generation Sequencing Technology

[1]*Fakruddin,* [2]*Reaz Mohammad Mazumdar,* [1]*Abhijit Chowdhury,*
[1]*Nur Hossain,* [1]*Suman Mahajan and* [1]*Sumaiya Islam*

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

Submitted: Jul 13, 2013; **Accepted:** Aug 24, 2013; **Published:** Sep 22, 2013

Abstract: Several shortcomings of Sanger di-deoxy nucleotide sequencing method pave the way for development of more advanced and high-throughput sequencing methods those are collectively termed as next generation sequencing methods (NGS). Pyrosequencing is the first alternative to the conventional Sanger method for *de novo* DNA sequencing. Pyrosequencing is a DNA sequencing technology based on the sequencing-by-synthesis principle. Pyrosequencing has the potential advantages of accuracy, flexibility, parallel processing and can be easily automated. Furthermore, the technique dispenses with the need for labeled primers, labeled nucleotides and gel-electrophoresis. The pyrosequencing method has emerged as a versatile DNA sequencing technology suitable for numerous applications in the field of modern biology. Pyrosequencing was earlier restricted to sequencing and analysis of SNPs and short stretches of DNA. However, it has now gone through phenomenal developments and improvements both in chemistry and instrumentation. More importantly, there are still opportunities for further breakthroughs in this area as only very few research groups have focused on this method. In future, pyrosequencing technology is expected to achieve longer read length, to reduce the sequencing time frame, to decrease the sample quantity and to make further improvements in automation. Though all current NGS technologies has shorter read length and higher error rate than those of Sanger sequence reads, further developments in the area of NGS can be expected in the near future, reducing the costs of genome sequencing even more.

Key words: Pyrosequencing · Next generation sequencing · DNA sequencing · Typing · Enzymatic

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]. Novel techniques were needed that overcame current limitations of Sanger sequencing with respect to throughput and costs [4] and in the last decade, a number of different methods were developed that not only have revolutionized the field of genome sequencing but also can be applied to other biological questions not previously addressed by sequencing-based approaches [5]. Recently, pyrosequencing has emerged as a new sequencing methodology [1].

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 [6-9], parallel signature sequencing based on ligation and cleavage [10] and pyrosequencing [11-12].

Corresponding Author: Fakruddin, Scientific Officer Institute of Food Science and Technology (IFST) Bangladesh Council of Scientific and Industrial Research (BCSIR) Dhaka, Bangladesh.
Tel: +8801717684750.

Pyrosequencing technology is a novel DNA sequencing technology, developed at the Royal Institute of Technology (KTH) and is the first alternative to the conventional Sanger method for *de novo* DNA sequencing. This method relies on the luminometric detection of pyrophosphate that is released during primer-directed DNA polymerase catalyzed nucleotide incorporation. It is suited for DNA sequencing of up to one hundred bases and it offers a number of unique advantages [13]. This technique is a widely applicable, alternative approach for the detailed characterization of nucleic acids. Pyrosequencing has potential advantages of accuracy, flexibility, parallel processing and can be easily automated. Furthermore, the technique avoids the need for labeled primers, labeled nucleotides and gel-electrophoresis. Pyrosequencing has been successful for both confirmatory sequencing and *de novo* sequencing [1].

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. (7). Level of sensitivity (generally estimated at 10-20%) insufficient for detection of clinically relevant low-level mutant alleles or organisms. (8) *cis* or *trans* orientation of heterozygous positions may be difficult to resolve during data analysis. (9) Not readily scalable to achieve a throughput capable of efficiently analyzing complex diploid genomes at low cost. (10) de novo genome assembly is difficult [14-15].

Alternative Methods of Sequencing: Next-generation DNA sequencing technologies that permit massive sequencing with a much higher throughput than the Sanger method have become recently available. 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 [4]. The first commercially available NGS platform was introduced in 2005 and as of then NGS methods have revolutionized the field of genomic analysis. Commercially available NGS technologies such as Roche/454 (http://www.454.com/), Solexa/ Illumina (http://www.illumina.com/), AB SOLiD (http://www.appliedbiosystems.com/AB_Home/applicationstechnologies/ SOLiD System Sequencing/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 [16, 17] (Table 1).

These platforms share a common technologic characteristic of performing massively parallel sequencing, either of clonally PCR-amplified products (the first three platforms) or even of single DNA molecule (the last two), based on physical separation in a flow cell surface. This feature diverges from the classic Sanger technique that is based on electrophoretic separation of chain-termination products.

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/) [18], 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 [19]. 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 [20].

Pyrosequencing Chemistry: Pyrosequencing technique is based on sequencing-by-synthesis principle [21-22] and on the detection of released pyrophosphate (PPi) during DNA synthesis [1]. It employs a series of four enzymes to accurately detect nucleic acid sequences during the synthesis. In Pyrosequencing [23] the sequencing primer is hybridized to a single-stranded DNA biotin-labeled template and mixed with the enzymes; DNA polymerase, ATP sulfurylase, luciferase and apyrase and the substrates adenosine 5' phosphosulfate (APS) and luciferin [3].

Table 1: Comparison of second and third HT NGS platforms

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

Cycles of four deoxynucleotide triphosphates (dNTPs) are separately added to the reaction mixture iteratively. The cascade starts with a nucleic acid polymerization reaction in which inorganic PPi is released as a result of nucleotide incorporation by polymerase. Each nucleotide incorporation event is followed by release of inorganic pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. Subsequently The released PPi is quantitatively converted to ATP by ATP sulfurylase in the presence of APS. The generated ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin, producing visible light in amounts that are proportional to the amount of ATPs. The light in the luciferase-catalyzed reaction with a maximum of 560 nanometer wavelength is then detected by a photon detection device such as a charge coupled device (CCD) camera or photomultiplier. Apyrase is a nucleotide-degrading enzyme, which continuously degrades ATP and non-incorporated dNTPs in the reaction mixture. There is a certain time interval (usually 65 seconds) between each nucleotide dispensation to allow complete degradation. For this reason, dNTP addition is performed one at a time [3]. Because the added nucleotide is known, the sequence of the template can be determined [1]. A schematic representation of pyrosequencing is shown in figure 1.

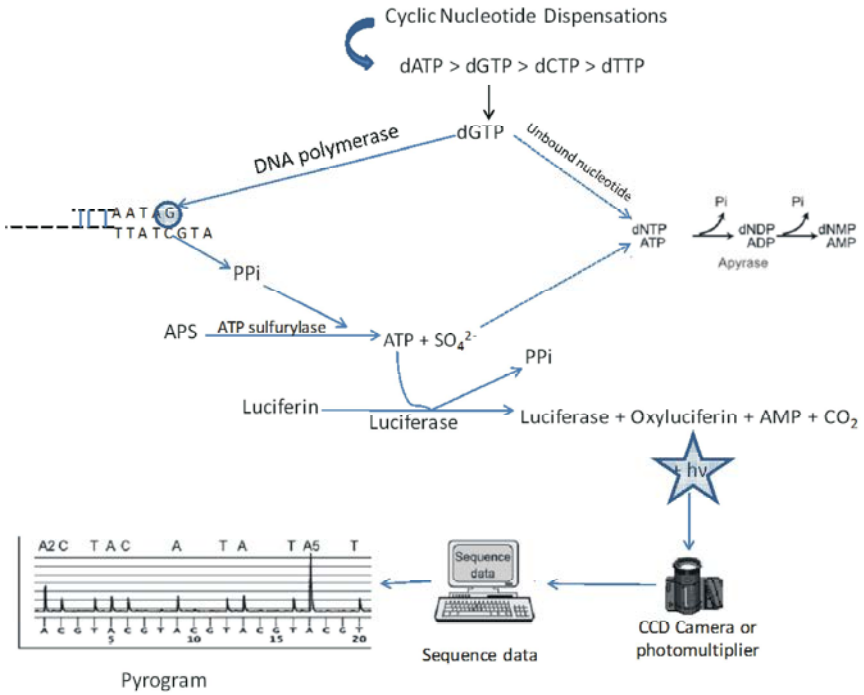

Fig. 1: Schematic representation of pyrosequencing.

The generated light is observed as a peak signal in the pyrogram (corresponding to electropherogram in dideoxy sequencing) proportional to the number of nucleotides incorporated (a triple dGTP incorporation generates a triple higher peak) [3]. During this synthesis process, the DNA strand is extended by complementary nucleotides and the DNA sequence is demonstrated by the pyrogram on a screen. The slope of the ascending curve in a pyrogram displays the activities of DNA polymerase and ATP sulfurylase, the height of the signal shows the activity of luciferase and the slope of the descending curve demonstrates the nucleotide degradation [13]. Base-callings are performed with integrated software, which has many features for related SNP and sequencing analysis [3]. The overall reaction from polymerization to light detection takes place within 3–4 sec at room temperature [1]. ATP sulfurylase converts PPi to ATP in approximately 1.5 seconds and the generation of light by luciferase takes place in less than 0.2 seconds [24].

Standard pyrosequencing uses the Klenow fragment of *Escherichia coli* DNA Pol I [25]. The ATP sulfurylase used in pyrosequencing is a recombinant version from the yeast *Saccharomyces cerevisiae* [26] and the luciferase is from the American firefly *Photinus pyralis* [1]. The apyrase is from *Solanum tuberosum (*Pimpernel variety) [27-28].

Commercialization of Pyrosequencing: The first NGS system to become commercially available was the Genome Sequencer from 454 Life Sciences (Branford, CT, USA) (later acquired by Roche; *Roche (454) GS FLX Pyrosequencer*) in 2005 [29]. This sequencer works on the principle of 'pyrosequencing' [30], 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 [31-32].

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 [33]. 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 [31]. 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 [34].

This method has significant advantages over Sanger sequencing because it requires no electrophoresis step to separate extension products and base incorporation can be detected in real time [35]. 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 [18]. 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 [36].

Application of Pyrosequencing: Pyrosequencing has opened up new possibilities for performing sequence-based DNA analysis [1]. Pyrosequencing is well suited for *de novo* sequencing and resequencing [1]. Currently, pyrosequencing method is broadly being used in many applications such as Single Nucleotide Polymorphism (SNP) genotyping [37-39], identification of bacteria [13, 40-41], fungal [42-44] and viral typing [42, 45-48]. Moreover, the method has demonstrated the ability to determine difficult secondary structures [49] and perform mutation detection [50-51], DNA methylation analysis [52-54], multiplex sequencing [55-56], tag sequencing of cDNA library [57] and clone checking [28]. Another highly significant application is whole genome sequencing [31]. Some of the potential applications of pyrosequencing have been described here.

Genotyping of Single-Nucleotide Polymorphisms: For analysis of single-nucleotide polymorphisms (SNPs) by pyrosequencing, the 3'-end of a primer is designed to hybridize one or a few bases before the polymorphic position. In a single tube, all the different variations can be determined as the region is sequenced. A striking feature of pyrogram readouts for SNP analysis is the clear distinction between the various genotypes; each allele combination (homozygous or heterozygous) will give a specific pattern compared to the two other variants [37, 58-60]. This feature makes typing extremely accurate and easy. Relative standard deviation values for the ratio between key peaks of the respective SNPs and reference counterparts are =0.1 [58]. Simple manual comparison of predicted SNP patterns and the raw data obtained can score an SNP, especially as no editing is needed. Because specific patterns can be readily achieved for individual SNPs, it is also be possible to automatically score the allelic status by pattern recognition software [1].

As pyrosequencing signals are very quantitative, it is possible to use this strategy for the studies of allelic frequency in large population. This system allows >5000 samples to be analyzed in 8 h. Furthermore, pyrosequencing enables determination of the phase of SNPs when they are in the vicinity of each other allowing the detection of haplotypes [50].

Resequencing: Pyrosequencing is currently the fastest method for sequencing a PCR product. Because pyrosequencing generates an accurate quantification of the mutated nucleotides, the resequencing of PCR-amplified disease associated genes for mutation scanning will be one of the interesting applications. Using this technique for resequencing results in longer read length than de novo sequencing because nucleotide delivery can be specified according to the order of the sequence [1]. Programmed dispensing generates a signal for each addition in a pyrogram, therefore variation in the pattern indicates the appearance of a mutation. This strategy has been used for resequencing of the p53 tumor suppressor gene, where mutations were successfully determined and quantified [50].

Tag Sequencing: The sequence order of nucleotides determines the nature of the DNA. Theoretically, eight or nine nucleotides in a row should define a unique sequence for every gene in the human genome. However, it has been found that, to uniquely identify a gene of a complex organism such as human, a longer sequence of DNA is needed. In a pilot study, 98% of genes in a human cDNA library could be uniquely identified by sequencing a length of 30 nucleotides. Pyrosequencing was used to sequence this length for gene identification and the results were in complete agreement with longer sequence data obtained by Sanger DNA sequencing.

Pyrosequencing offers high-throughput analysis of cDNA libraries because 96 samples can be analyzed in less than one hour. Like Sanger DNA sequencing, pyrosequencing also has the advantage of library screening as the original cDNA clone is directly available for further analysis [1].

Analysis of Difficult Secondary Structures: Hairpin structures are common features in genomic materials and have been proposed to have regulatory functions in gene transcription and replication. However, analyzing these sequences by conventional DNA sequencing usually gives rise to DNA sequence ambiguities seen as "run-off" or compressions. These problems have been associated with gel electrophoresis [1]. Pyrosequencing was successfully applied to decipher the sequence of such regions [49].

Microbial and Viral Typing: Near instantaneous detection of pathogens from clinical materials is important for diagnosis, treatment and prophylactic measurements. The Microbial and viral threats have been altered in the course of history as a substantial number of pathogens have been eradicated or controlled. New pathogens are emerging and some are developing resistance. Accurate and specific typing of microbial and viral pathogens is of utmost importance in clinical diagnosis [59]. Many microorganisms usually lack adequate morphological details for easy identification. Furthermore, the development of appropriate therapies/ vaccines requires implementation of sufficient parameters for microbial and viral detection and a reliable and robust genotyping method is necessary for accurate follow-up during clinical trials and serveillance of treatment [60]. Pyrosequencing technique, due to its many advantages, has been turned out to be an useful implement for microbial and viral typing. It is currently being applied for rapid typing of a large number of bacteria, yeasts and viruses [1].

Fungal Identification: Pyrosequencing appears to be a good diagnostic tool for detection and identification of fungal pathogens. It can be utilized in typing of fungi clinically isolated from immune-compromised patients suffering from proven invasive fungal infections. For typing, the DNA has to be amplified by general consensus primers [61] having complementary to a highly conserved region within the 18S rRNA gene allowing the amplification of a broad range of fungal species. The amplified DNA fragments are then sequenced up to 40 bases to identify the samples. All the samples can be accurately identified by the pyrosequencing technique. The reproducibility of the technique has been confirmed by sequencing the amplicons several times with identical sequence data each time for every sample. The sequence data obtained by pyrosequencing technique suggested that 18 to 32 bases are sufficient for identification of *C. albicans, C. glabrata, C. krusei, C. parapsilosis, C. tropicalis* and *Aspergillus* spp. [62].

Bacterial Identification: DNA markers used for typing normally contain both conserved and variable regions. A DNA primer complementary to the conserved or semiconserved region is usually employed to sequence the variable region. In bacteria, 16s rRNA gene is commonly used to identify different species and strains. By analyzing a sequence between 20–100 nucleotides on 16S rRNA gene, it is possible to group different bacteria taxonomically and, in many cases to get information about strains. Further development of bacterial identification by pyrosequencing could be brought about in order for the method to be suitable for regions other than 16s rRNA of different bacteria [62].

Advantages of Pyrosequencing: Pyrosequencing has emerged as an alternative method of sequencing. Although it has read-length limitations compared with di-deoxy sequencing, it is a fast method with real-time read-out that is highly suitable for sequencing short stretches of DNA [3].

Pyrosequencing employs co-operativity of several enzymes to monitor DNA synthesis. Parameters such as stability, fidelity, specificity, sensitivity, K_M and *k*cat are mandatory for the optimal performance of the enzymes used in the sequencing reaction. The kinetics of the enzymes can be studied in realtime [62].

Unlike Sanger sequencing, which lays a reading gap of roughly 20-30 bases from the sequencing primer, pyrosequencing can generate sequence signals immediately downstream of the primer. As sequencing starts with the first base next to the annealed primer, making primer design becomes more flexible in this method. Sample and single-strand DNA preparation process is also relatively rapid (about 15 min), while sample preparation takes approximately 4 hours for Sanger sequencing (60 min for PCR cleanup, 3-4 hours for cyclic amplification and 15 minutes for dye cleanup). The reagent costs are considerably lower for sequencing short stretches of DNA compared to currently available methods [3].

The pyrosequencing technology has many unique advantages over other DNA sequencing technologies. One advantage is that the order of nucleotide dispensation can be easily programmed and alterations in the pyrogram pattern reveal mutations, deletions and insertions. Moreover, this technique is carried out in real-time, as nucleotide incorporations and base callings can be observed continuously for each sample. In addition, the Pyrosequencing method can be automated for large-scale screenings [62].

As other NGS technologies, pyrosequencing differ from automated Sanger sequencing in that it does not require cloning of template DNA into bacterial vectors. Apart from being less labor-intensive, this has the distinct advantage that cloning biases, e.g., due to difficulties with cloning of AT-rich regions or genes that are toxic for *Escherichia coli*, are avoided [63].

Another advantage of NGS technologies like pyrosequencing is the high degree of parallelization, in which millions to billions of sequencing reactions take place at the same time in small reaction volumes, thereby allowing a much higher throughput than automated Sanger sequencing [34].

A significant improvement in pyrosequencing is the ability to sequence both ends from a DNA fragment (paired-end sequencing) [64].

Multiplexing can be achieved by labeling individual samples using bar-coded primers during preparation of the sequencing library [65].

Technical Limitations of Pyrosequencing:

- Earlier pyrosequencing was earlier limited to sequencing of short stretches of DNA, due to the inhibition of apyrase. The natural dATP was a substrate for luciferase, resulting in false sequence signals. dATP was substituted by dATP-α-S [11]. The dATP- α-S consisted of two isomers, Sp and Rp. The Rp isomer was not incorporated in the DNA template as it was not a substrate for DNA polymerase and its presence in the sequencing reaction simply inhibits apyrase activity. By introducing the dATP- α-S Sp isomer, substantial longer reads were achieved. This improvement had a major impact on pyrosequencing read length and allowed sequencing of up to one hundred bases [66] and opened up avenues for numerous applications [3].
- Homopolymer Ts (more than 3-4) are a challenge in Pyrosequencing. Homopolymer string (mainly homopolymeric T) regions can influence synchronized extension and synthesis of the DNA strand causing non-uniform sequence peak heights, affecting the read-length and possibly causing sequence errors. Studies have shown that the incorporation of dATP-α-S in T homopolymeric regions results in uneven sequence signals and reduced sequence quality directly downstream of such homopolymers [66-67]. As mentioned earlier, during the development of pyrosequencing technique, the natural dATP was replaced by dATP- α-S to increase the signal-to-noise ratio. Inefficient dATP-α-S incorporation by the exonuclease-deficient Klenow DNA polymerase causes the template extensions to go out of the phase making sequence peak signals asynchronous and ambiguous that could be demonstrated in the pyrograms as uniformly reduced sequence signal peak-height. Therefore, efforts were made to compensate this in the reading of those sections. By employing Sequenase [68], an exonuclease deficient T7 DNA polymerase, the poly-T homopolymer string reads were significantly improved by generation of significantly more synchronized sequence and uniform signal peaks after homopolymeric T regions [3].
- An important factor in pyrosequencing is primer design for PCR and sequencing. Sequencing primers should be checked for self-looping, primer-dimer (primer-primer hybridizations) and cross-hybridization (when more than one sequencing primer is used). Single-stranded DNA binding protein (SSB) [69] is inhibitory to primer-template complications in pyrosequencing. Furthermore, it is highly recommended that sequencing primers be designed before PCR primers in order to biotin-label the suitable DNA strand. This facilitates avoiding possible difficult homopolymers or regions that can interfere with efficient sequencing and base-callings [3].
- An inherent problem with the described method is *de novo* sequencing of polymorphic regions in heterozygous DNA material. In most cases, it will be possible to detect the polymorphism. If the polymorphism is a result of substitution, it is possible to obtain a synchronized extension after

the substituted nucleotide. If the polymorphism is due to deletion or insertion of the same kind as the adjacent nucleotide on the template, the sequence after the polymorphism will be synchronized [1].

However, if the polymorphism is because a deletion or insertion of another type, the sequencing reaction can become out of phase, making the interpretation of the subsequent sequence difficult. If the polymorphism is known, it is always possible to use programmed nucleotide delivery to keep the extension of different alleles synchronized after the polymorphic region. It is also possible to use a bidirectional approach [49] whereby the complementary strand is sequenced to decipher the sequence flanking the polymorphism [1].

- Another inherent problem is the difficulty in determining the number of incorporated nucleotides in homopolymeric regions, due to the nonlinear light response following incorporation of more than 5–6 identical nucleotides. The polymerization efficiency over homopolymeric regions has been investigated and the results indicate that it is possible to incorporate $\leq$10 identical adjacent nucleotides in the presence of apyrase [69]. However, to elucidate the correct number of incorporated nucleotides, it may be necessary to use specific software algorithms that integrate the signals. For resequencing, it is possible to add the nucleotide twice for a homopolymeric region to ensure complete polymerization [1].

Challenges of next Generation Sequencing Technologies: Pyrosequencing as well as other 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:

- 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 [70].
- There is a great need for quantitative studies and analysis tools that help investigators optimally design NG sequencing experiments to address specific goals [71].
- 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 [72].
- 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.
- 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.

Bioinformatics Challenges of NGS: NGS can be used for a wide variety of highly interesting and promising experimental approaches and additional NGS-based applications will certainly be devised within the coming years. However, the most significant hurdle that researchers face with any NGS approach is the analysis of the huge amounts of data generated by these experiments. This starts with the fact that there are currently no unified data formats, moves on to problems associated with *de novo* assembly or mapping of millions to billions of reads, assembly of spliced transcripts, or quantification of RNA-seq data and at some point leads to the questions of where and how to store the large files associated with different steps of NGS analysis [5].

Data Formats: Different variants of FASTA format has been developed for different NGS technologies. As most software packages currently accept only a limited number of input file formats, the first step in dealing with NGS data is often the conversion of the original files into a format that can be used as input for downstream processing, e.g., assembly, mapping, etc. A number of freely available software tools have been developed to deal with this problem (e.g., http://bioinf.comav.upv.es/sff_extract/index.html or http://maq.sourceforge.net/fq_all2std.pl); however, the constant development of new technologies accompanied by new file formats as well as the lack of standard formats for downstream applications requires that users pay close attention to data format changes. However, there are efforts under way to standardize mapping data formats, thereby making them more interchangeable among programs [73].

De Novo Assembly: The *de novo* assembly of genomes from NGS reads is not a trivial task and assembly programs that were developed for Sanger reads usually are not suitable for this purpose. Genome assemblies from shorter reads of NGS require a much higher coverage.

The short read length, however, causes difficulties that cannot be overcome by increasing the number of reads alone because repeat regions, which are widespread especially in the genomes of higher eukaryotes, cannot be assembled from reads that are shorter than the lengths of the repeats.

During the last years, a number of programs were developed specifically for the *de novo* assembly of NGS reads. These programs incorporate new types of algorithms that are able to deal with the huge number of reads. Most of them allow combinations of reads from different sequencing platforms and incorporate paired-end data. This makes the assembly of large, complex genomes from NGS data possible [74].

Mapping of Reads to a Reference Genome, Quantification and Detection of Sequence Variants: NGS applications usually require mapping of reads to a reference genome prior to downstream analyses. Similar to *de novo* assembly, mapping NGS reads requires algorithms that can deal with the huge number of short reads.

Therefore, a number of novel algorithms specifically adapted to NGS read mapping were developed in the last years. These programs are designed to deal with a number of issues that have to be considered when mapping NGS reads.

One of the major aims of genome resequencing is to detect genomic variability between individuals or different strains/ lines. This requires the ability to detect differences between the reference genome and the resequenced DNA [75].

Data Storage: However, one problem for all researchers conducting NGS projects is how to store the resulting data. The large number of sequence reads delivered by NGS platforms results in file sizes, e.g., for FASTQ files, that are well in the gigabyte range and downstream applications like mapping yield files of a similar size so that the amount of data even from a single project often reaches the terabyte range. Therefore, especially for individual researchers, storage and backup schemes that were quite sufficient for Sanger sequencing projects can quickly run out of storage space with even a few NGS-based experiments [76].

The International Nucleotide Sequence Database Collabo-ration (INSDC) (http://www.insdc.org/), which includes the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EBI) and the DNA Data Bank of Japan (DDBJ), has established a database for NGS data, the SRA (Sequence Read Archive). The SRA contains not only the actual sequences and quality scores but also metadata about the sequencing project to which individual data files belong. However, the large number of reads makes searching those data with established algorithms like BLAST impractical in terms of search time and would in most cases not yield informative results anyway due to the short read lengths. Therefore, researchers can download the data, but the convenient online searches and other tools that we have become accustomed to are currently not available for NGS data. One major challenge is therefore how to make NGS data accessible to the research community [77].

Future of Pyrosequencing: Genome sequencing in general provides bulk of information that can be used in different areas of biology [1]. Future applications require more robust and efficient DNA sequencing techniques for sequence determination. The Pyrosequencing method has already shown evidence of high accuracy in DNA sequencing and analysis of polymorphic DNA

fragments in many clinical and research settings. It is a relatively straightforward and user-friendly method possessing unique methodological characteristics and this technique is currently being used in multidisciplinary fields in academic, clinical and industrial settings. By increasing the read length to higher scores and by shortening the sequence reaction time per base calling, pyrosequencing may take over many broad areas of DNA sequencing applications as the trend is directed to analysis of fewer amounts of specimens and large-scale settings, with higher throughput and lower cost [3].

The massively parallel pyrosequencing of emulsified PCR-based templates holds great promise to revolutionize high-throughput sequencing. However, there is concern over the potentially high degree of error [78].

Pyrosequencing has shown excellent accuracy in analysis of polymorphic DNA fragments. This technology has also been used for quantification of allelic frequency in populations. While the variations are characterized, correlation of variation to phenotype can be performed. Pyrosequencing will have a large impact in that area because a large number of samples can be pooled in one pyrosequencing reaction. A high throughput version of this technology can potentially be used for resequencing of genomes. Pyrosequencing technology is relatively new and there lies ample room for versatile developments in both chemistry and instrumentation. This technology is already time- and cost-competitive as compared to the most conventional sequencing methods. Work is underway to further improve the chemistry, to measure the sequencing efficiency at elevated temperatures and to run the reaction in miniaturized formats [1].

By removal of inhibitory factors and improvement of chemistry it is feasible to ensure enhanced sequence quality and read-length, which are likely to open new avenues for many new applications relating to *de novo* sequencing as well as re-sequencing, mutation detection and microbial and viral typing.

Multiple sequencing primer method has also contributed to typing of samples containing a multitude of types/ species and unspecific amplified products in clinical settings, eliminating the need for stringent PCR reactions, nested PCRs and gene cloning. This strategy has many other potential applications as automation is being more and more integrated into clinical settings and the cost of DNA sequencing is dropping.

By increasing the read length to higher scores and by shortening the sequence reaction time per base calling, pyrosequencing may take over many broad areas of DNA sequencing applications as the trend is directed to analysis of fewer amounts of specimens and large-scale settings, with higher throughput and lower cost. Pyrosequencing was earlier restricted to sequencing and analysis of SNPs and short stretches of DNA. However, it has now gone through phenomenal developments and improvements both in chemistry and instrumentation. More importantly, there are still opportunities for further breakthroughs in this area as only very few research groups have focused on this method. In future, pyrosequencing technology is expected to achieve longer read length, to reduce the sequencing time frame, to decrease the sample quantity and to make further improvements in automation.

Concluding Remarks: NGS technologies such as pyrosequencing offer a promising new avenue for discovery in many fields of research, including several that, so far, have mainly used approaches not based on sequencing. Although there is still room for improvement, especially with respect to read lengths and error rates, NGS platforms as well as analysis software tools have evolved to already allow, for example, the *de novo* sequencing and assembly of eukaryotic genomes solely from NGS reads. However, one major concern about NGS is the question of how to deal with the flood of data that NGS platforms produce. Therefore, the development of bioinformatics tools and databases to better cope with these types of data will be one of the main factors determining how useful NGS will be for a wider research community. Nevertheless, the advent of NGS is already a major breakthrough in molecular biology, genetics and beyond, as well as a great leap forward for genomics and systems biology analyses.

REFERENCES

1. Ronaghi, M., 2001. Pyrosequencing sheds light on DNA sequencing. Genome Research, 11: 3-11. doi:10.1101/gr.150601
2. Sanger, F., S. Nicklen and A.R. Coulson, 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA., 74: 5463-5467.
3. Gharizadeh, B., M. Ghaderi and P. Nyren, 2007. Pyrosequencing technology for short DNA sequencing and whole genome sequencing. Technology, 47: 129-132.

4. Service, R.F., 2006. The race for the $1000 genome. Science, 311: 1544-1546.
5. Nowrousian, M., 2010. Next-Generation Sequencing Techniques for Eukaryotic Microorganisms: Sequencing-Based Solutions to Biological Problems. Eukaryotic Cell, 9(9): 1300-1310.
6. Bains, W. and G.C. Smith, 1988. A novel method for nucleic acid sequence determination. J. Theoretical Biol., 135: 303-307.
7. Drmanac, R., I. Labat, I. Brukner and R. Crkvenjakov, 1989. Sequencing of megabase plus DNA by hybridization: Theory of the method. Genomics, 4: 114-128.
8. Khrapko, K.R., Y.P. Lysov, A.A. Khorlyn, V.V. Shick, V.L. Florentiev and A.D. Mirzabekov, 1989. An oligonucleotide hybridization approach to DNA sequencing. FEBS Lett., 256: 118-122.
9. Southern, E.M., 1989. Analysing polynucleotide sequences. US patent no. WO/10977
10. Brenner, S., S.R. Williams, E.H. Vermaas, T. Storck and K. Moon, 2000. In vitro cloning of complex mixtures of DNA on microbeads: Physical separation of differentially expressed cDNAs. Proc. Natl. Acad. Sci., 97: 1665-1670.
11. Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen and P. Nyren. 1996. Real-time DNAsequencing using detection of pyrophosphate release. Anal. Biochem., 242: 84-89.
12. Ronaghi, M., M. Uhlen and P. Nyren, 1998. A sequencing method based on real-time pyrophosphate. Science, 281: 363-365.
13. Gharizadeh, B., 2003. Method development and applications of pyrosequencing technology. Royal Institute of technology. Biotechnology.
14. Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New Biotechnol, 25(4): 195-203.
15. Hall, N., 2007. Advanced sequencing technologies and their wider impact in microbiology. The J Exp. Biol., 209: 1518-1525.
16. Berglund, E.C., A. Kiialainen and A.C. Syvänen, 2011. Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative Genetics, 2: 23.
17. Morozova, O. and M.A. Marra, 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics, 92: 255-264.
18. Mardis, E.R., 2008a. The impact of next-generation sequencing technology on genetics. Trends in Genetics, 24(3): 133-141.
19. Eid, J., *et al*., 2009. Real-time DNA sequencing from single polymerase molecules. Science, 323: 133-138.
20. Varshney, R.K., S.N. Nayak, G.D. May and S.A. Jackson, 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding.Trends in Biotechnology, 27(9): 522-530.
21. Hyman, E.D., 1988. A new method of sequencing DNA. Anal. Biochem., 174: 423-436.
22. Melamede, R.J., 1985. Automatable process for sequencing nucleotide. U.S. Patent 4863849. http://www.freepatentsonline.com/4863849.html.
23. Nyren, P. and S.E. Skarpnack, 2001. Method of sequencing DNA based on the detection of the release of pyrophosphate and enzymatic nucleotide degradation. *U.S. Patent 6258568*. http://ip.com/patent/US6258568.
24. Nyren, P. and A. Lundin, 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal. Biochem., 151: 504-509.
25. Benkovic, S.J. and C.E. Cameron, 1995. Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of Escherichia coli DNA polymerase I. Methods Enzymol., 262: 257-269.
26. Karamohamed, S., J. Nilsson, K. Nourizad, M. Ronaghi, B. Pettersson and P. Nyren, 1999. Production, purification and luminometric analysis of recombinant Saccharomyces cerevisiae MET3 adenosine triphosphate sulfurylase expressed in Escherichia coli. Protein Exp. Purif., 15: 381-388.
27. Espinosa, V., A.M. Kettlun, A. Zanocco, E. Cardemil and M.A. Valenzuela., 2003. Differences in nucleotide-binding site of isoapyrases deduced from tryptophan fluorescence. Photochemistry, 63: 7-14.
28. Nourizad, N., M. Ehn, B. Gharizadeh, S. Hober and P. Nyren, 2003. Methylotrophic yeast Pichia pastoris as a host for production of ATP-diphosphohydrolase (apyrase) from potato tubers (Solanum tuberosum). Protein Exp. Purif., 27: 229-237.
29. Berglund, E.C., A. Kiialainen and A.Ce. Syvänen, 2011. Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative Genetics, 2: 23.

30. Ronaghi M., *et al*., 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem., 242: 84-89.
31. Margulies, M.M., M. Egholm, W.E. Altman, S. Attiya and J.S. Bader *et al.*, 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature, 437: 376-380.
32. Fakruddin, M. and A. Chowdhury, 2012. Pyrosequencing-An Alternative to Traditional Sanger Sequencing. American J Biochem. Biotechnol. 8(1): 14-20.
33. Tawfik, D.S. and A.D. Griffiths, 1998. Man-made cell-like compartments for molecular evolution, Nat. Biotechnol, 16: 652-656.
34. Mardis, E.R., 2008b. Next-Generation DNA Sequencing Methods. Annu. Rev. Genomics Hum. Genet, 9: 387-402.
35. Fakruddin, M., A. Chowdhury, M.N. Hossain, K.S.B. Mannan and R.M. Mazumdar 2012. Pyrosequencing-Principles and Applications. Intl. J. Life Sci. Pharma Res., 2(2): L65-L76.
36. Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. New Biotechnology, 25(4): 195-203.
37. Ahmadian, A., B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen and J. Lundeberg. 2000a. Single-nucleotide polymorphism analysis by pyrosequencing. Anal. Biochem., 280: 103-110.
38. Nordstrom, T., M. Ronaghi, L. Forsberg, U. de Faire, R. Morgenstern and P. Nyren, 2000. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. Biotechnol. Applied Biochem., 31: 107-112.
39. Milan, D., J.T. Jeon, C. Looft, V. Amarger and A. Robic, *et al*., 2000. A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle. Science, 288: 1248-1251.
40. Grahn, N., M. Olofsson, K. Ellnebo-Svedlund, H.J. Monstein and J. Jonasson, 2003. Identification of mixed bacterial DNA contamination in broadrange PCR amplification of 16S rDNA V1 and V3 variable regions by pyrosequencing of cloned amplicons. FEMS Microbiol. Lett., 219: 87-91.
41. Jonasson, J., M. Olofsson and H.J. Monstein, 2002. Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. APMIS, 110: 263-272.
42. Gharizadeh, B., M. Oggionni, B. Zheng, E. Akom and N. Pourmand, 2005. Type-specific multiple sequencing primers: A novel strategy for reliable and rapid genotyping of human papillomaviruses by pyrosequencing technology. J. Mol. Diagn., 7: 198-205.
43. Trama, J.P., E. Mordechai and M.E. Adelson, 2005a. Detection of *Aspergillus fumigatus* and a mutation that confers reduced susceptibility to itraconazole and posaconazole by real-time PCR and pyrosequencing. J. Clin. Microbiol., 43: 906-908.
44. Trama, J.P., E. Mordechai and M.E. Adelson, 2005b. Detection and identification of Candida species associated with Candida Vaginitis by real-time PCR and pyro sequencing. Mol Cell Probes, 19(2): 145-152.
45. Gharizadeh, B., M. Kalantari, C.A. Garcia, B. Johansson and P. Nyren, 2001. Typing of human papillomavirus by pyrosequencing. Lab. Invest., 81: 673-679.
46. Adelson, M.E., M. Feola, J. Trama, R.C. Tilton and E. Mordechai, 2005. Simultaneous detection of herpes simplex virus types 1 and 2 by real-time PCR and pyrosequencing. J. Clin. Virol., 33: 25-34.
47. Elahi, E., N. Pourmand, R. Chaung, A. Rofoogaran and J. Boisver, 2003. Determination of hepatitis C virus genotype by pyrosequencing. J. Virol. Methods, 109: 171-176.
48. Gharizadeh, B., M. Kaller, P. Nyren, A. Andersson, M. Uhlen, J. Lundeberg and A. Ahmadian, 2003a. Viral and microbial genotyping by a combination of multiplex competitive hybridization and specific extension followed by hybridization to generic tag array*s*. Nucl. Acids Res., 31: e146.
49. Ronaghi, M., M. Nygren, J. Lundeberg and P. Nyren, 1999. Analyses of secondary structures in DNA by pyrosequencing. Anal. Biochem., 267: 65-71.
50. Ahmadian, A., J. Lundeberg, P. Nyren, M. Uhlen and M. Ronaghi, 2000b. Analysis of the p53 tumor suppressor gene by pyrosequencing. Biotechniques, 28: 140-147.
51. Garcia, C.A., A. Ahmadian, B. Gharizadeh, J. Lundeberg, M. Ronaghi and P. Nyren, 2000. Mutation detection by pyrosequencing: Sequencing of exons 5-8 of the p53 tumour suppressor gene. Gene, 253: 249-257.
52. Neve, B., P. Froguel, L. Corset, E. Vaillant, V. Vatin and P. Boutin, 2002. Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. Biotechniques, 32: 1138-1142.

53. Rickert, A.M., A. Premstaller, C. Gebhardt and P.J. Oefner, 2002. Genotyping of SNPs in a polyploid genome by pyrosequencing. Biotechniques, 32: 592-600.
54. Uhlmann, K., A. Brinckmann, M.R. Toliat, H. Ritter and P. Nurnberg, 2002. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. Electrophoresis, 23: 4072-4079.
55. Gharizadeh, B., M. Ghaderi, D. Donnelly, B. Amini, Wallin, K.L. and P. Nyren, 2003b. Multiple-primer DNA sequencing method. Electrophoresis, 24: 1145-1151.
56. Gharizadeh B., B. Zheng, M. Akhras, M. Ghaderi and O. Jejelowo, 2006. Sentinel-base DNA genotyping using multiple sequencing primers for high-risk human papillomaviruses. Mol. Cell Probes, 20: 230-238.
57. Nordstrom, T., B. Gharizadeh, N. Pourmand, P. Nyren and M. Ronaghi, 2001. Method enabling fast partial sequencing of cDNA clones. Anal. Biochem., 292: 266-271.
58. Alderborn, A., A. Kristofferson and U. Hammerling, 2000. Determination of single nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res., 10: 1249*-1258.
59. Ekstrom, B., A. Alderborn and U. Hammerling, 2000. Pyrosequencing for SNPs. Progress Biomed. Optics, 1: 134-139.
60. Van Doorn, L.J., B. Kleter and W.G. Quint, 2001. Molecular detection and genotyping of human papillomavirus. Expert Rev. Mol. Diagn., 1: 394-402.
61. Einsele, H., H. Hebart, G. Roller, J. Loffler and I. Rothenhofer, 1997. Detection and identification of fungal pathogens in blood by using molecular probes. J. Clin. Microbiol., 35: 1353-1360.
62. Gharizadeh, B., A. Ohlin, P. Molling, A. Backman, B. Amini, P. Olcen and P. Nyren, 2003. Multiple group-specific sequencing primers for reliable and rapid DNA sequencing. Mol. Cell Probes, 17: 203-210.
63. Metzker, M.L., 2010. Sequencing technologies—the next generation. Nat. Rev. Genet, 11: 31-46.
64. Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage *et al*. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 269: 496-512.
65. Craig, D.W., J.V. Person, S. Szelinger, A. Sekar, M. Redman, J.J. Corneveaux, *et al*., 2008. Identification of genetic variants using barcoded multiplexed sequencing. Nat. Methods, 5: 887-893.
66. Gharizadeh, B., T. Nordstrom, A. Ahmadian, M. Ronaghi and P. Nyren, 2002. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. Anal. Biochem., 301: 82-90.
67. Eriksson, J., B. Gharizadeh, N. Nourizad and P. Nyren, 2004. 7-Deaza-2'-deoxyadenosine-5'-triphosphate as an alternative nucleotide for the pyrosequencing technology. Nucleosides Nucleotides Nucleic Acids, 23: 1583-1594.
68. Gharizadeh, B., J. Eriksson, N. Nourizad, T. Nordstrom and P. Nyren, 2004. Improvement in pyrosequencing technology by employing sequenase polymerase. Anal. Biochem., 330: 272-280.
69. Ronaghi, M., 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. Anal. Biochem., 286: 282-288.
70. Subramanian, S., L. Huynen, C.D. Millar and D.M. Lambert, 2010. Next generation sequencing and analysis of a conserved transcriptome of New Zealand's kiwi. BMC Evol. Biol., 10: 387-397.
71. Wall, P.K., 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 and C.W. dePamphilis, 2009. Comparison of next generation sequencing technologies for transcriptome characterization. BMC Genomics, 10: 347-366.
72. Fakruddin, M. and K.S.B. Mannan, 2012. Next Generation sequencing technologies-Principles and prospects. Research & Reviews in Biosciences, 6(9): 240-247.
73. Cock, P.J.A., C. Fields, N. Goto, M.L. Heuer and P.M. Rice, 2010. The Sanger FASTQ file format for sequences with quality scores and the Solexa/Illumina FASTQ variants. Nucleic Acids Res., 38: 1767-1771.
74. Miller, J.R., S. Koren and G. Sutton, 2010. Assembly algorithms for next-generation sequencing data. Genomics, 95: 315-327.
75. Trapnell, C. and S.L. Salzberg, 2009. How to map billions of short reads onto genomes. Nat. Biotechnol., 27: 455-457.

76. McPherson, J.D., 2009. Next-generation gap. Nat. Methods, 6: S2-S5.
77. Shumway, M., G. Cochrane and H. Sugawara, 2010. Archiving next generation sequencing data. Nucleic Acids Res., 38: D870-871.
78. Huse, S.M.S.M., J.A. Huber, H.G. Morrison, M.L. Sogin and D.M. Welch, 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol., 8: R143.