nature biotechnology

The development and impact of 454 sequencing

Jonathan M Rothberg[1] & John H Leamon[2]

The 454 Sequencer has dramatically increased the volume of sequencing conducted by the scientific community and expanded the range of problems that can be addressed by the direct readouts of DNA sequence. Key breakthroughs in the development of the 454 sequencing platform included higher throughput, simplified all *in vitro* sample preparation and the miniaturization of sequencing chemistries, enabling massively parallel sequencing reactions to be carried out at a scale and cost not previously possible. Together with other recently released next-generation technologies, the 454 platform has started to democratize sequencing, providing individual laboratories with access to capacities that rival those previously found only at a handful of large sequencing centers. Over the past 18 months, 454 sequencing has led to a better understanding of the structure of the human genome, allowed the first non-Sanger sequence of an individual human and opened up new approaches to identify small RNAs. To make next-generation technologies more widely accessible, they must become easier to use and less costly. In the longer term, the principles established by 454 sequencing might reduce cost further, potentially enabling personalized genomics.

There has been a desire to increase the throughput of DNA sequencing ever since Nobel laureates Frederick Sanger and Walter Gilbert began sequencing by chain termination or fragmentation techniques coupled with electrophoretic size separation[1–3]. To date, the demand for sequencing has mirrored that for computing, with increases in sequencing capacity being quickly absorbed by ever-larger experiments. Several novel approaches were explored to replace Sanger as the dominant provider of sequencing technologies, including sequencing by hybridization[4], direct imaging of DNA sequence by atomic force microscopy[5], mass spectrometry resolution[6], a number of different approaches for sequencing by synthesis[7–11], and several large efforts to apply microfluidics to sequencing[12]. However, three more evolutionary improvements allowed sequencing to meet the challenges of the first human genome project[13]: the use of fluorescent tags instead of radioactive labels to detect the terminated ladders; the use of capillary electrophoresis in place of slab gels; and the development of paired-end sequencing protocols incorporating hierarchical template sizes (plasmids, fosmids and bacterial artificial chromosomes (BACs)) to provide sequence context and orientation beyond the constraints of the actual sequence read-length. Parallel efforts in liquid-handling robotics moved the established methods of library preparation (plasmid vectors and growth in bacterial colonies) from test tubes to microtiter plates with a significant drop in cost and labor to prepare samples for sequencing[14].

As the first next-generation technology to reach the market, the development of the 454 Life Sciences (454; Branford, CT, USA; now Roche, Basel) sequencing platform (the 454 Sequencer) provides a compelling case study for the establishment of a new disruptive technology[15]. 454 initiated the next-generation movement by pioneering solutions to the three bottlenecks—library preparation, template preparation and sequencing—that the research community faced. One indication of the benefits inherent in the solutions 454 provided is that in one form or another, each of them has been adopted in all the next-generation technologies that followed 454 sequencing to market.

First-mover status permitted 454 sequencing to have a more direct impact on the sequencing community than its next-generation competitors, at least on the basis of more than 570 citations of the technology's 2005 introductory paper (Leamon, Rothberg and colleagues[16]) and over 100 peer-reviewed publications using 454 sequencing to solve problems in human genetics, metagenomics, ecology, evolution and paleobiology. 454 sequencing was the first technology other than Sanger's to sequence and assemble bacterial genomes *de novo* (Leamon, Rothberg and colleagues[16]) and the first non-Sanger technology to sequence an individual human (Rothberg and colleagues[17]). Other notable studies conducted by 454 included work as diverse as uncovering the potential cause of the disappearance of the honeybee[18], revealing the complexity of rearrangements between individual human genomes[19], providing new approaches to understand infectious diseases[20] and sequencing the first million base pairs of a Neanderthal[21–23].

In this article, we discuss the development of the 454 sequencing system, including the conceptual motivations for the project as well as the technological innovations that lead to a successful product launch. The diverse set of scientific areas affected by 454 sequencing and the status of related sequencing technologies are also examined. Lastly, future directions for sequencing technology and applications are explored.

The influence of Moore's law on 454 sequencing

The development of 454 sequencing was prompted not by a desire to miniaturize Sanger sequencing, but by a picture of the new Pentium chip and the pronouncement of yet another marvel enabled by Moore's law[24]. It was clear that routine human sequencing would require increases in sequencing throughput similar to the improvements in processing power and speed in the computer industry. These advances were possible only after vacuum tubes were replaced by transistors, enabling the development of the integrated circuit at the heart of the computer industry. Attempts to routinely sequence a human genome by increasing the speed and throughput of the existing capillary array electrophoresis

[1]Rothberg Institute for Childhood Diseases, 530 Whitfield Street Guilford, Connecticut 06437, USA. [2]Ion Torrent Systems, 37 Soundview Road, Guilford, Connecticut 06437, USA. Correspondence should be addressed to J.M.R. (jrothberg@childhooddiseases.org).

Published online 9 October 2008; doi:10.1038/nbt1485

units were analogous to trying to make an integrated circuit out of vacuum tubes. Instead, the sequencing equivalent of the transistor and an approach equivalent to the 'monolithic idea' (integrating all the parts of an electronic circuit in a single 'monolithic' block of semiconductor material) were combined to enable the first application of Moore's law to DNA sequencing.

The concept of sequencing by synthesis, although at the time not yet successful, provided the basis for miniaturization and scaling. Two general approaches had been proposed. On the one hand were cyclic reversible termination technologies, based on sequencing by sequential addition of a labeled base, followed by fluorescent detection and cleavage of that fluorescent base. On the other were technologies that sequenced by detecting pyrophosphate release with an enzymatic cascade ending in luciferase and detection of the emitted light. The latter approach was chosen for the 454 platform because direct incorporation of natural nucleotides seemed more efficient than repeated cycles of incorporation, detection and cleavage. This premise has been shown to be correct, with competing next-generation technologies facing a current limit between 25 and 50 bases (ref. 25). 454 sequencing was based on miniaturizing a pyrosequencing reaction and moving both the template preparation step and the pyrosequencing chemistry to the solid phase[26,27].

Pyrosequencing has been available to the scientific community since the mid-1990s as a genotyping tool, but was not considered powerful enough for standard sequencing needs because of the short read-lengths it generated, relegating it to the role of single-nucleotide polymorphism (SNP)-based genotyping. At that time, the technology was being used in microtiter plates to process up to 96 genotypes in a sequential fashion[28], in relatively low throughput at costs of about 20 cents per sample[29]. Pyrosequencing had not yet shown the capability for *de novo* sequencing—which requires accurate *de novo* identification of nucleotides at every incorporation rather than the simple confirmation of bases at known positions as used in SNP-based genotyping—and longer read-lengths beyond the few bases required for genotyping[30].

However, because pyrosequencing is based on the detection of light produced whenever a nucleotide is incorporated (**Fig. 1**), it does not require a physical separation process like electrophoresis to resolve the next base in the DNA strand. This means that unlike electrophoresis, where the physical length required for achieving precise resolution of distinct fragments limits miniaturization, pyrosequencing could be reduced to any reaction volume that generates detectable levels of light. The release of light also enables sequencing to be performed in parallel, although until the 454 format, it was usually conducted sequentially. Just as with the early transistors, which were viewed with skepticism as they couldn't handle the same current as vacuum tubes, the short-read

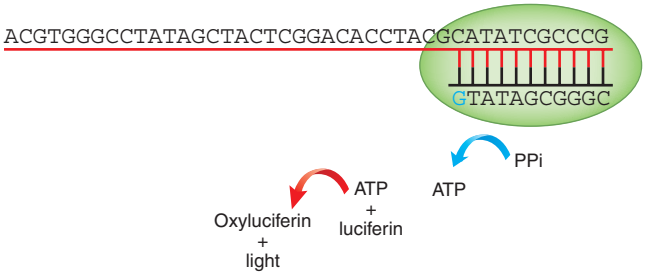

Figure 1 Diagram of the pyrosequencing process. The template strand is represented in red, the annealed primer is shown in black and the DNA polymerase is shown as the green oval. Incorporation of the complementary base (the blue "G") generates inorganic pyrophosphate (PPi), which is converted to ATP by the sulfurylase (blue arrow). Luciferase (red arrow) uses the ATP to convert luciferin to oxyluciferin, producing light.

pyrosequencing reactions had the potential to operate in parallel at high densities. Provided that the sequencing template and required enzymes could be immobilized on solid supports instead of occurring free in solution, the reactions could be miniaturized, parallelized and placed on a single suitable substrate—just as transistors could be miniaturized and placed in great number on a single substrate to produce integrated circuits. Additionally, the physical separation obtained by the discrete wells of a microtiter plate can be replaced by tightly packed reactions on a single solid support, where the reaction rates of strand polymerization and light generation can be carefully balanced against the diffusion of both reagents and products.

A new way to conduct experiments in parallel

In developing a system based on a high-throughput and highly parallel format, one of the key problems was how to supply reagents to the large quantities of simultaneous reactions required for projects on the scale of human genome sequencing. This presents a challenge because each reaction requires a recurring source of fresh reagents with each base addition, and the automated injection processes originally in use could not efficiently supply sequential rounds of reagents to large numbers of individual reactions. Again the importance of creating a parallel and highly dense process became essential to enable the technique to achieve the necessary scale. The solution to this problem was simple and elegant: at high reaction densities, diffusion from a laminar flow stream itself is sufficient to bring fresh reagents to each reaction and to wash both by-products and unincorporated reagents away. This laminar flow concept has been adopted by each subsequent next-generation platform.

To separate the individual reactions, early simulations of reactions on a flat surface showed that the densities would be limited to hundreds to thousands of reactions per cm^2. However, the goal was to sequence tens or hundreds of thousands of reactions per cm^2 to conserve the expensive reagents that would need to flow over the reactions and enable efficient imaging. Higher densities were accomplished by placing the reactions in wells, with the well depth further isolating the individual reactions from each other. Although this approach greatly reduced the area needed to carry out the reactions, the sheer number of desired reactions still required a relatively large area (60 mm × 60 mm) for real-time imaging during sequencing.

The imaging problem was solved by adapting commercially available astrological grade cameras that employed fiber-optic bundles physically glued to the surface of a charge-coupled device (CCD)—these bundles are tapered to allow a large image area to be projected onto a smaller CCD surface. Slides with high-density wells were made in two steps. After slicing the fiber-optic bundles into thin disposable slides that resembled microscope slides, wells were etched on one side of these slides by removing the glass between the fiber cladding with acid. Protocols for this acid etching procedure were based on work that had been done on small fiber bundles used as biosensors[31–33].

By etching the thin slides from a fiber-optic block, 454 was able to manufacture a plate with millions of wells. These wells would serve as individual reactors to allow individual enzymatic reactions to take place in each of the separate wells. Once placed in a flow cell to allow delivery of reagents, this fiber-optic slide would serve as a monolithic substrate for many reactions in parallel, provide individual wells to limit diffusion—and hence allow the individual sequencing reactions to occur without interference from their neighbors—and further allow for the light from the reaction to be captured by placing the slide on top of a fiber-optic bundle permanently attached to a high-end imaging CCD system (**Fig. 2**). The wells also would facilitate the addition of the other reagents needed for the light-producing reactions. Instead of having the enzymes in solution, as in the wells of a 96-well plate used in the

pyrosequencing instrument available at that time[30], they were deposited on beads along with the DNA template. This immobilization step ensured that the reaction was local to the well where the light would be captured by the fiber-optic bundle and also reduced costs, as enzyme replenishment would not be necessary under flow conditions.

Optimizing the sequencing process alone was insufficient to enable truly high-throughput genome sequencing. A large portion of the $3 billion cost of the human genome project was dedicated to a complex sample preparation strategy. Even in its most simplistic form, the sample preparation process required shotgun cloning of fragments of the genome of interest into bacteria, plating bacteria onto tens of thousands of plates, and robotic colony picking and transfer to 96-well plates. These steps were followed by time-consuming and expensive amplification of the individual clones, and cleanup of templates—all before a single sequence reaction could be run.

These costs can be reduced by implementing new approaches for preparing libraries by fragmentation, creating individual templates through limiting dilution and amplifying templates through compartmentalized enzymatic amplification—hence creating 'clones' without bacterial cloning. This would provide a complete sequencing process covering all aspects from genome of interest to finished sequence that would be completely *in vitro*, massively parallel and scalable as improvements to density and read-length were achieved on the fiber-optic well plate (Leamon, Rothberg and colleagues[16]).

Invention to innovation

With the conceptual framework for the technology in place, initial development of the 454 format was focused on two major areas: first, development of the etched fiber-optic slide, and second, conversion of the pyrosequencing reaction to solid supports and its optimization—solid-phase pyrosequencing, template and library preparation—to allow sequencing and *de novo* assembly of long DNA reads.

Solid-phase, long-read pyrosequencing in wells of an etched fiber-optic slide. The etching process was quickly refined and permitted reproducible etching of wells with 55 µm depth and 44 µm width on 75 mm × 75 mm slides. The goals of developing the solid-phase sequencing methods, and optimizing read-lengths in the microwells were closely intertwined, in that the drive to immobilize the reaction in small wells contributed substantially to improved accuracy and read-length. As reactants quickly diffused from the wells, the synthesis cycle could be quickly repeated. Instead of removing or degrading unincorporated nucleotides with apyrase[27], long read-lengths (100–500 base-reads) were obtained by the rapid diffusion of unincorporated nucleotides and reaction by-products from the picoliter-volume wells into the laminar flow stream of the flow cell. Along with the efficient removal of unincorporated bases, the efficiency of polymerase extensions in each cycle was greatly facilitated by the fast diffusion and removal of residual concentrations of inorganic phosphates. Complete polymerase extensions enabled acquisition of long, accurate sequencing reads (99.5% accurate at 200 bases). By reducing the concentration of residual, unincorporated nucleotides remaining in the wells, rapid diffusion limited the extent to which the residual nucleotides inhibited polymerase activity or introduced 'carry-forward' errors, where loss of sequencing synchronicity occurred because of leftover unincorporated nucleotides. The advances in read-length and accuracy also stemmed from innovative improvements in fluidics, surface chemistry and enzymology (Leamon and Rothberg[34]), including identifying superior polymerases, optimizing the sequencing reaction at higher temperatures, and replacing and rebalancing the components in the enzyme cascade (**Fig. 1**).

Other options to improve read-length and accuracy were also investigated, but not commercialized. These included the use of reversible terminators to increase accuracy across homopolymers, strategies for

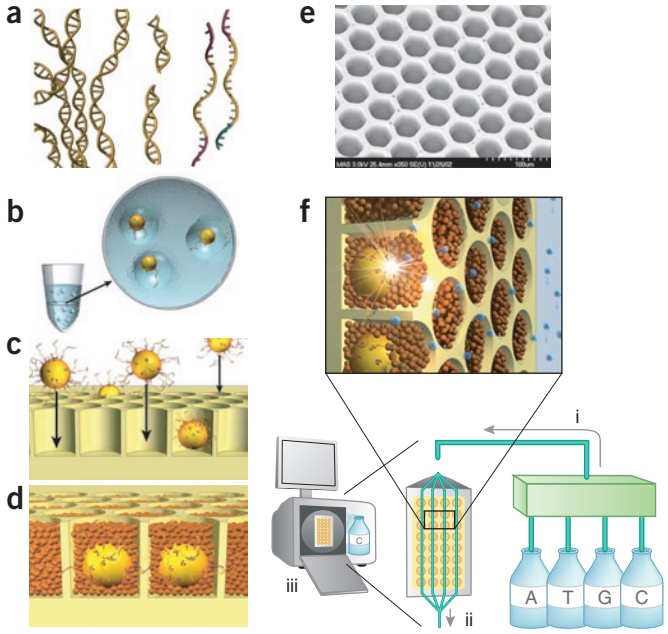

Figure 2 Overview of the 454 sequencing technology. (**a**) Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands. (**b**) Fragments are bound to beads under conditions that favor one fragment per bead, the beads are isolated and compartmentalized in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template. (**c**) The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA templates are enriched (not shown) and deposited into wells of a fiber-optic slide. (**d**) Smaller beads carrying immobilized enzymes required for a solid phase pyrophosphate sequencing reaction are deposited into each well. (**e**) Scanning electron micrograph of a portion of a fiber-optic slide, showing fiber-optic cladding and wells before bead deposition. (**f**) The 454 sequencing instrument consists of the following major subsystems: a fluidic assembly (object i), a flow cell that includes the well-containing fiber-optic slide (object ii), a CCD camera-based imaging assembly with its own fiber-optic bundle used to image the fiber-optic slide (part of object iii), and a computer that provides the necessary user interface and instrument control (part of object iii).

Table 1 Applications introduced through collaborations with 454

Application	Collaborative project	Subsequent publication references
Bacterial sequencing and comparative genomics	*Mycobacterium tuberculosis*[44]	45–47,49
Paleobiology and ancient DNA	Sequencing Neanderthal DNA[22]	21,23,63,64
Small RNA	*Arabidopsis thaliana* DICER function[54]	50,55–58,59–62
Metagenomics	Microbes in honeybee colony collapse[18]	40,66–69
Genome structure	Variation in human genome[19]	
Whole genome human sequencing	Sequencing of James Watson genome (Rothberg and colleagues[17])	

double-ended sequencing where both strands could be sequenced from the same template and alternative enzyme-immobilization methods[35]. These improvements were never incorporated, in part because the development of increased accuracy, longer read-length and an efficient paired-end library process rendered them unnecessary.

Template preparation. A completely *in vitro* and massively parallel template preparation method was needed to provide low cost templates for high-throughput sequencing. The emulsion-based method that was eventually implemented was based on compartmentalized enzyme evolution[36]. However, because of the difficulty of maintaining stable droplets through a thermocycling regime, the first clonal amplification methods attempted were isothermal[37], with clonality achieved by the combination of the limited dilution of the original template and the confinement of individual wells.

Although emulsion-based techniques for clonal production of templates without the need for bacteria were of interest from the start, suitable surfactants were not known at the time to allow the emulsions to survive thermal cycling with the stability required for subsequent sequencing[38]. For this reason isothermal approaches were explored. Although amplification yields from rolling circle amplification (RCA) reactions were extremely high (able to generate solid masses of amplified DNA in the wells), the majority of the RCA products was not accessible to the sequencing primers. Still determined to create a system that didn't need bacteria and made use of 'limiting dilution', the amplification focus shifted from isothermal- to PCR-based methods. As with the RCA experiments, early work obtained clonality through loading a limited dilution of templates into the wells of the fiber-optic slide. Amplification was achieved by sealing the surface of the fiber-optic slide with rubber gaskets, and amplifying the reaction on a traditional flat-topped PCR machine (Leamon, Rothberg and colleagues[39]). This method was successful, but inefficient: the thermal mass in the glass fiber-optic plate and its clamping mechanism (referred to internally as the 'coffin') demanded lengthy PCR cycle times, and the limiting template dilution used less than 10% of the available wells (Leamon, Rothberg and colleagues[39]). Contamination resulting from well-to-well diffusion was also a concern. Nonetheless, the process was able to amplify templates from a whole-genome library for subsequent sequencing resulting in the first *de novo* sequencing of any genome using a non-Sanger, non-Gilbert sequencing method, as well as the first sequencing of a genome (adenovirus) with complete *in vitro* sample preparation (GenBank AY370909).

The thermostability of large droplets was finally obtained with the incorporation of surfactants used in the manufacture of explosives, where thermostable segregation of diesel oil from ammonium nitrate is required. The success of this formulation dramatically increased the effectiveness of emulsion-based PCR (emPCR; Leamon, Rothberg and colleagues[16]). The emPCR system proved highly effective and scalable, able to produce templates from genomes ranging in size from the 30-kb Adenovirus genome to the multi-megabase *Streptococcus pneumoniae* genomes (Leamon, Rothberg and colleagues[16]).

Along with the increases in sequencing quality and read-length, continued improvements in emulsion stability enabled 454 to effectively extend beyond bacterial sequencing and into

Box 1 Next-generation sequencing formats

As of September 2008, three additional next-generation technologies have come to market. The first technology to follow 454 to market, Illumina's (Hayward, CA, USA) Genome Analyzer, was developed by Solexa (Cambridge, UK) and brought to market by Illumina; the second technology was based on the work of Church, Shendure and colleagues[78], subsequently refined at Agencourt (Beverly, MA, USA) and brought to market by Applied Biosystems (Foster City, CA, USA) as the SOLiD system; the third system from Helicos Biosciences (Cambridge, MA, USA) does not require PCR amplification of template material, thereby enabling true single-molecule sequencing. Another system from Pacific BioSciences (Menlo Park, CA, USA) is under active development and could soon emerge on the marketplace. These systems are described in more detail in **Table 2**.

The short read-lengths of both Solexa and SOLiD, coupled with the decreased sequencing costs afforded by the high-read densities make these two technologies ideal for counting applications, such as sequence-based expression analysis and promoter binding site studies. These technologies will also be applicable in cases in which short sequences, such as miRNAs, are probed. Whereas the throughput and cost of these two techniques would be appealing for whole-genome sequencing, the read-lengths do not allow them to perform *de novo* sequencing assembly, although both companies have stated that the use of paired-end reads may improve this. Even after obtaining sequencing data that represents 30-fold coverage, preliminary genome sequencing projects of complex organisms only cover 90% of the complete reference genome (Leamon, Rothberg and colleagues[84]). 454 sequencing was able to achieve 98.7% coverage of human genomic DNA using sequencing data that represents only 7.4-fold coverage (Rothberg and colleagues[17]). These results suggest that for many applications the Solexa and SOLiD platforms might require fourfold more sequencing data to achieve genome coverage comparable to that derived using long-read technologies. In addition, identification of protein-coding regions in infectious disease and other metagenomic projects usually requires longer reads derived from traditional Sanger or 454 Sequencers. One area that does show promise for the short-read technologies is resequencing of exons after enrichment[85] and the identification of chromosomal rearrangements[86]. However, all novel insertions will be missed by the short reads, and at least fourfold more oversampling might be required to achieve coverage and accuracy similar to that of long-read techniques. The relative costs per base and read-lengths provided by the SOLiD, Solexa and 454 platforms, and the segmentation of the next-generation sequencing market based on those factors is shown in **Figure 3**.

It will be interesting to monitor these technologies over time as they attempt to increase read-length and throughput while reducing cost per base. Whereas the long-read technologies achieve 10–20 more bases per read, the Solexa and SOLiD platforms sequence ~100 times the number of discrete templates per cm^2 compared to that of the 454 system. Further density increases with either Solexa or SOLiD may be challenging. Because Solexa currently controls template density through limiting dilution and *in situ* amplification, it is not immediately obvious how much higher the density can be increased without negatively affecting sequence quality owing to overlapping or mixed template populations. The SOLiD system avoids template overlap by employing the 1-μm beads, but the degree to which those beads can be reduced in size and still possess sufficient amplified material to enable sequencing by ligation is unknown. It will be interesting to see if the density of the 454 system can continue to increase by use of smaller beads, while the tenfold advantage in read-length and time per base extension is retained.

truly large and complex sequencing projects, including Neanderthal[21–23], *Homo sapiens* (Rothberg and colleagues[17]) and entire ecological communities[40].

Library preparation. The library preparation methodology was fairly well determined from the conception of the project, with the generation of a randomly fragmented library from a sample genome followed by ligation of distinct adapters to opposing ends of the template for subsequent amplification. The 454 sample preparation process differs substantively from the whole-genome shotgun sequencing method popularized by Craig Venter and colleagues[41] in that the 454 process relies upon limiting dilutions and physically compartmentalizing clonal amplification to completely remove the bacterial cloning step (**Fig. 2**). Removal of the bacterial cloning step dramatically increased the speed and efficiency of the 454 method while sidestepping the potential loss of sequence coverage due to bacteria-induced bias. This methodology would also prove to be ideal for metagenomic analysis and ancient DNA studies. A paired-end library preparation method was developed to enable the *de novo* assembly of complex genomes, span repetitive regions and allow the systematic study of genome structure, including duplications and other rearrangements[19]. The development of this paired-end library protocol was inspired by the work of Bender's group[42,43] on jumping libraries in *Drosophila*—the technique brought together distant ends of carefully sized, randomly fragmented genomic templates to span repeats and facilitate whole-genome assembly.

Applications of the technology

With the realization that the greatest scientific impact would be made by demonstrating the utility of sequencing in as wide a range as applications as possible (**Table 1**), 454 set up collaborations in which the company would sequence and analyze samples with both industrial and academic researchers. These collaborations were further designed to demonstrate the utility of the specific technologies or strategies developed at 454, such as the use of paired-end libraries to understand genome structure, or the power of the emPCR to capture DNA fragments independent of the ability to culture or clone the DNA from the sample of interest.

Bacterial sequencing and comparative genomics. To demonstrate the potential for whole-genome sequencing, and with methods and software both in place, one of the earliest collaborations was a project in which four genomes were sequenced to determine the mechanism of resistance to the drug R207910 in *Mycobacterium tuberculosis*[44] using one R207910-resistant *M. tuberculosis* (4 Mb) strain, two R207910-resistant *Mycobacterium smegmatis* (6 Mb) strains and one parental *M. smegmatis* (6 Mb) strain. This project clearly illustrated the advantages of 454 sequencing both in terms of speed and accuracy; using traditional Sanger technologies, sequencing one 4-Mb genome and three 6-Mb genomes would require several months, whereas with the 454 sequencing system, the entire process, including sample preparation, amplification and sequencing of four genomes, was conducted by a single technician in ~1 week. The 454 system also avoided possible biases introduced by the traditional cloning process, generating high-quality data that permitted accurate identification of the two missense mutations that conferred R207910 resistance[44]. The *M. tuberculosis* study resulted in the identification of the first tuberculosis-specific drug candidate in 40 years and also underscored the value of the 454 Sequencer for bacterial sequencing applications, leading to comparative genomic studies[45], *de novo* sequencing and assembly of the genome of a highly pathogenic strain of *Campylobacter jejuni*[46], an evolutionary study of *Helicobacter pylori* during progression of chronic gastritis[47], the discovery and sequencing of a novel ice-binding protein from an Antarctic sea ice bacterium[48], and determination of the pathogenic content of the bacteria responsible for pneumonia, meningitis and urinary tract infections[49].

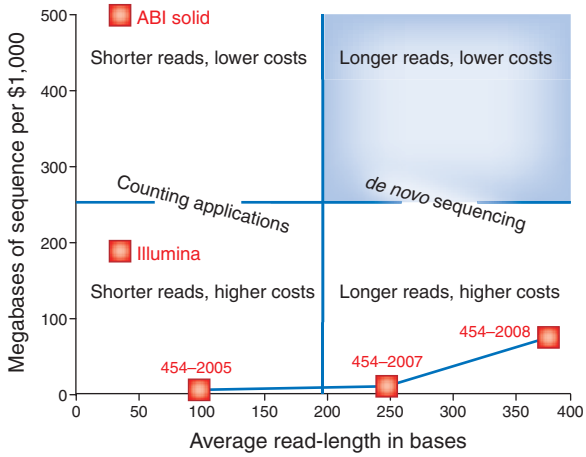

Figure 3 Next-generation market segmentation. Sequencing application segmentation as a function of cost per megabase, and read-length. Data for costs per run and read-lengths for Solexa and SOLiD were obtained from a variety of sources, including respective company websites and a recent technology review[74]. Data for 454 Life Sciences reflects improving read-lengths and throughput from those reported in the initial publication (data point '454 – 2005'; Leamon, Rothberg and colleagues[16]), those obtained with the launch of the 454 FLX system and during the Watson sequencing project (data point '454 – 2007'; Rothberg and colleagues[17]), and those released in the first half of 2008 (data point '454 – 2008'; ref. 75). Note the run times for 454 data points are 5, 8 and 10 h, respectively, and those for Solexa and SOLiD are 72 h.

The lack of cloning bias inherent in 454 sequencing that enabled the mutation detection in the *M. tuberculosis* study has been documented by others as well[50–53]. Additionally, recent human sequencing with the 454 system revealed roughly 29 Mb of sequence that failed to map to the *H. sapiens* reference genome build-36 but is thought to be euchromatin and missing from the reference sequence (Rothberg and colleagues[17]). It should also be noted, however, that there have been reports of some assembly bias resulting from stretches of repetitive DNA[52], as well as sequence bias arising from samples prepared by nebulization of small template fragments[51].

Small RNAs. The explosion of interest in small RNAs, including microRNAs (miRNAs), in 2005 was perfectly timed with the commercial availability of the 454 sequencing system and the ability to sequence hundreds of thousands of templates simultaneously. With no need for traditional cloning, and read-lengths able to sequence the entire 21-bp miRNAs, 454 was able to play a pivotal role in the understanding and elucidation of the role of miRNAs in a wide range of biological processes, as shown by the following examples. An early pivotal study was a collaborative effort in which miRNAs in *Arabidopsis thaliana* were investigated[54]. This was closely followed by another 454 collaboration in which a novel class of small RNAs, the Piwi-interacting RNAs (piRNAs)[55], was identified in mice. These studies paved the way for additional studies on small RNAs in human, chimpanzee[56], zebrafish[57] and tumor cell lines[58]. The ease with which the 454 system was able to use small RNAs as a substrate spawned numerous applications such as analysis of the transcriptome[50,59] expressed sequence tags[60], 5′-transcript ends (5′-RATE)[61] as well as transcriptome-based SNP discovery[62].

Table 2 Details of selected next-generation sequencing platforms under commercialization

Platform (company)	Description	Performance issues
454 Sequencer (454)	The first commercially available next-generation sequencer, the GS-20, released in 2005. Template DNA is nebulized and size-selected to produce a population of double-stranded fragments ranging from 400 to 600 bases. Two distinct oligonucleotide adapters are ligated onto the fragments, providing priming sites for subsequent amplification and sequencing. One of the adapters is biotinylated, permitting collection of single-stranded templates. The templates are amplified and immobilized by compartmentalizing individual template molecules and 28-μm DNA capture beads within droplets of an emulsion. PCR reactions conducted inside the droplets amplify the template molecules and complementary primers covalently attached to the DNA capture immobilize the product on the bead surface. Template-covered DNA capture beads are loaded into individual wells etched into the surface of a fiber-optic slide. The sequencing process uses an enzymatic cascade to generate light from inorganic pyrophosphate (PPi) molecules released by the incorporation of nucleotides as a polymerase replicates the template DNA (Leamon, Rothberg and colleagues[16]). Individual nucleotides are provided to the open wells by flowing them over the fiber-optic slide. The number of photons generated by the cascade is proportional to the number of nucleotides incorporated by the polymerase and the release of the PPi generated by the individual sequencing reactions.	The initial system generated ~20 megabases of 110 base-reads per 8-h run (Leamon, Rothberg and colleagues[16]), a subsequent product release generated an average of 100 megabases of 250 base-reads. With current work using higher density fiber-optic plates, base-reads in excess of 500 are expected to generate between 400 and 600 megabases per run. The relatively low throughput per sequencing run results in the highest cost per base of any of the next-generation systems. For some projects, such as counting-based and some resequencing applications, the 454 system's increased read-length is unnecessary, leading to utilization of the lower cost, higher output, shorter read-length systems.
Genome Analyzer (Illumina)	The Solexa system operates via a sequencing-by-synthesis process that incorporates fluorescently labeled nucleotides into immobilized template strands. Amplification is conducted *in situ* via bridge amplification[76], starting with a limiting dilution of template. Sequence information is obtained by interrogating the flow cell with a laser and recording the fluorescent signal at each location after every incorporation, after which the fluorescent label is cleaved from the incorporated nucleotides and the process repeated for the next nucleotide.	Sample density on the Solexa platform is currently 100 million samples per cm^2 with a sequencing output of 1 gigabase, higher than the 454 system per run but comparable in terms of total bases per hour. As the size of the flow cell exceeds the field of the scanning microscope, multiple images are required to cover the entire sequencing area. The requirement for multiple data acquisition events for every nucleotide incorporation cycle increases the time required for a sequencing run beyond that of the 454 system (72 h versus 8 h), and the difficulty of both incorporating fluorescent bases and cleaving the fluorescent moiety efficiently has limited sequencing read-lengths to ~35 bases. Also, as the amplification occurs directly on the slide, it is not possible to control the density of amplified templates on the slide through enrichment of amplified beads (as with 454 and SOLiD) and deposition into spatially controlled locations (as with 454).
SOLiD system (Applied Biosystems)	SOLiD platform employs a template preparation and amplification process similar to those developed by 454. Template DNA is fragmented and added at limiting dilutions to an emulsion-based PCR system[77]. As in the 454 process, templates are clonally amplified and retained on DNA capture beads although the SOLiD process uses a smaller 1-μm bead. Beads possessing amplified template are then enriched using a larger polystyrene bead passing through a glycerol cushion and subsequently covalently anchored to the surface of a glass slide. The SOLiD sequencing process differs substantially from the 454 and Solexa methods in that it relies upon sequencing by ligation[78], wherein fluorescently labeled 8-base oligonucleotides are sequentially ligated onto the sequencing primer. Each of the four probe types carries a distinctive 3′ fluorophore representing the template sequence complementary to the 4^{th} and 5^{th} bases in the oligonucleotide probe. The fluorescent signal generated by the hybridized probe is then detected and recorded through a laser scanner and a data acquisition process similar to that of the Solexa system. The ligated probe is then cleaved between the 5th and 6th bases, removing the fluorescent moiety, and repeated rounds of probe ligation, interrogation and cleavage are conducted until read-lengths of up to 35 bases are achieved.	As with the Solexa system, the data acquisition process is lengthy, but as with the Solexa system, each run produces on the order of 1 gigabase of data generated from ~100 million reads per flow cell in a 3- to 5-day process.
HeliScope (Helicos Bioscience)	Based on studies sequencing individual molecules of DNA[79], the recently released HeliScope differs from the previously described sequencing processes in that the template DNA is not amplified before sequencing.	Limited information on the system available via the company website (http://www.helicosbio.com/) and the company's presentation at the Advances in Genome Biology and Technology (AGBT) meeting where the results from sequencing of a 194-Kb BAC over a 14-day period were reported. Low coverage in TA-rich regions and high error rates required over 7.5 gigabases of data to be generated in this 14-day period to sequence this 194,000 base DNA fragment[80]. More recently, the company published the results from sequencing the M13 genome[81], using ~150× oversampling to completely sequence the 7-kb genome. Read-lengths were 23 bases on average, although the average length was expanded to 27 bases by using software to deconvolute the homopolymers. Average error rates per sequenced base ranged from 0.1% to 0.3%, for per read accuracies of 92–97% (0.1% or 0.3% times 27 bases per read).
Single-molecule sequencing (Pacific Biosciences)	Based on zero wave guide technology[82], the system is described as capable of generating data from individual templates at a rate of 10 bases/second. The data released at the AGBT meeting in 2008, however, did not demonstrate single-molecule sequencing, but instead showed the ability to see base incorporation using 1,000 copies of a single template[83]. Company has plans for a commercial release in 2010 (ref. 82).	

Paleobiology and ancient DNA. Sequencing the Neanderthal genome with conventional platforms was highly problematic because of the fragmented nature of the ancient DNA and its availability only in minute quantities. Initial tests on less valuable, more available ancient DNA samples demonstrated technological compatibility with the 454 platform; thus, although at the time 454 could only sequence 100 base-reads, Neanderthal DNA was composed mostly of 40- to 90-bp fragments anyway, and the newly developed emPCR was designed to capture minute quantities consisting of individual molecules of template. A collaboration to sequence 38,000-year-old DNA from a Neanderthal generated several papers[21–23] and a great deal of popular interest, but also essentially established the field of paleogenomics, with sequencing conducted on woolly mammoth[63,64] and Pleistocene wolves[65].

Metagenomics and infectious disease. With the US anthrax scares in 2001, 454 became very interested in the power of using the 454 Sequencer to sequence from complex, unknown and unculturable environmental mixtures. Two separate collaborations demonstrated the utility of the 454 sequencing system in detecting and classifying unknown organisms from complex mixtures of DNA. In the first collaboration, DNA enriched for nonhuman elements from samples from three patients (who died from unknown causes after receiving organs from a single donor in Australia) was sequenced[20]. Analysis of the 144,000 resulting reads identified fragments from 14 different genes from a virus from the Arenaviridae family. A second collaboration strongly implicated Israeli acute paralysis virus as the cause of colony collapse disorder in honeybees[18] by comparing results of a metagenomic survey of healthy and infected hives. These studies highlighted a key feature of the 454 system; as the process does not require cloning or pre-amplification before sample preparation, unknown and unculturable organisms can be readily sequenced. These benefits are also evident in the many studies of community structure and composition undertaken using 454 technology from environments as varied as subterranean mines[40], the deep sea[66,67], soil[68] and solar salterns[69].

Genomic structure. Technological advancements in the 454 system enabled new applications. 454's newly developed paired-end sequencing protocol was used to explore human structural variation[19]. The paired-end mapping process was used to sequence, identify and map structural variants from two individuals, one of African and one of purportedly European origin. By mapping over 1,000 large, 3 kb or longer structural variations back to a reference genotype, the research revealed that many more structural variations existed than previously hypothesized within the human genome, many with potentially important phenotypic consequences. This structural variation research, the work on sequencing the genome of Nobel laureate James Watson (Rothberg and colleagues[17]) and other related studies elevated human genetic variation to be named breakthrough of the year by the journal *Science*[70].

The future of next-generation sequencing

The June 2007 release of James Watson's genome sequence to GenBank represented the first time that an individual human genome had been sequenced with non-Sanger technology and released to the public. The process was completed in 2 months in the absence of bacterial cloning, employing massively parallel template preparation and sequencing at a cost of less than a million dollars, a 1,000-fold improvement over the cost of the decade-long Human Genome Project[31,33,71] and 100 times less expensive then the widely publicized Venter genome released to the web in May 2007 and published in September 2007 (ref. 72). The technological progression of 454 sequencing described in the initial publication (20 million bases per run generated from 100 base-reads at 96% accuracy; Leamon, Rothberg and colleagues[16]) compared with the performance of the technology applied to the Watson genome (100 million bases per run, 250 base reads at >99% accuracy; Rothberg and colleagues[17]), suggests the reappearance of Moore's law[24] in an unexpected place—genome sequencing.

454 sequencing and other next-generation sequencing platforms (**Box 1**, **Fig. 3** and **Table 2**) have demonstrated the power of miniaturization and parallelization, increasing throughputs and decreasing costs of sequencing. In addition to pioneering next-generation sequencing, the team at 454 developed and popularized the use of all *in vitro* library and template preparation techniques now routinely used in various forms by other next-generation instruments on the market. The cost of sequencing an individual human genome will cross the $100,000, $10,000 and $1,000 barrier, much as increases in density with concomitant performance improvements led to improvements in computing power. These advances have set the stage for an entirely new industry predicated on the prospect of sequencing individual genomes[73].

COMPETING INTERESTS STATEMENT
The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. USA* **70**, 3581–3584 (1973).
2. Sanger, F. The Croonian Lecture, 1975. Nucleotide sequences in DNA. *Proc. R. Soc. Lond. B Biol. Sci.* **191**, 317–333 (1975).
3. Sanger, F. & Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–448 (1975).
4. Khrapko, K.R. *et al.* An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* **256**, 118–122 (1989).
5. Hansma, H.G. *et al.* Progress in sequencing deoxyribonucleic acid with an atomic force microscope. *J. Vac. Sci. Technol. B* **9**, 1282–1284 (1991).
6. Koster, H. *et al.* A strategy for rapid and efficient DNA sequencing by mass spectrometry. *Nat. Biotechnol.* **14**, 1123–1128 (1996).
7. Hyman, E.D. A new method of sequencing DNA. *Anal. Biochem.* **174**, 423–436 (1988).
8. Koster, H. *et al.* Oligonucleotide synthesis and multiplex DNA sequencing using chemiluminescent detection. *Nucleic Acids Symp. Ser.* **24** 318–321 (1991).
9. Nyren, P., Pettersson, B. & Uhlen, M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem.* **208**, 171–175 (1993).
10. Brenner, S. *et al.* Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**, 630–634 (2000).
11. Melamede, R.J. Automatable process for sequencing nucleotide. US patent 4,863,849 (1985).
12. Woolley, A.T. & Mathies, R.A. Ultra-high-speed DNA sequencing using capillary electrophoresis chips. *Anal. Chem.* **67**, 3676–3680 (1995).
13. Cantor, C.R. & Smith, C. *Genomics: The Science and Technology Behind The Human Genome Project*, edn. 1 (Wiley-Interscience, Hoboken, NJ, 1999).
14. Meldrum, D. Automation for genomics, part one: preparation for sequencing. *Genome Res.* **10**, 1081–1092 (2000).
15. Christensen, C.M. The innovator's dilemma: when new technologies cause great firms to fail (Harvard Business School Press, Boston, 1997).
16. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
17. Wheeler, D. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–877 (2008).
18. Cox-Foster, D.L. *et al.* A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283–287 (2007).
19. Korbel, J.O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
20. Palacios, G. *et al.* A new arenavirus in a cluster of fatal transplant-associated diseases. *NEJM* **358**, 991–998 (2008).
21. Briggs, A.W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104**, 14616–14621 (2007).
22. Green, R.E. *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330–336 (2006).
23. Noonan, J.P. *et al.* Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118 (2006).
24. Moore, G.E. Cramming more components onto integrated circuits. *Electronics* **38**, 114–117 (1965).

25. Rusk, N. & Kiermer, V. Primer: sequencing—the next generation. *Nat. Methods* **5**, 15 (2008).
26. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84–89 (1996).
27. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
28. Ronaghi, M. Pyrosequencing for SNP genotyping. *Methods Mol. Biol.* **212**, 189–195 (2003).
29. Fakhrai-Rad, H., Pourmand, N. & Ronaghi, M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* **19**, 479–485 (2002).
30. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3–11 (2001).
31. Steemers, F.J. & Walt, D.R. Multi-analyte sensing: from site-selective deposition to randomly-ordered addressable optical fiber sensors. *Mikrochim. Acta* **131**, 99–105 (1999).
32. Pantano, P. & Walt, D.R. Ordered nanowell arrays. *Chem. Mater.* **8**, 2832–2835 (1996).
33. Ferguson, J.A., Steemers, F.J. & Walt, D.R. High-density fiber-optic DNA random microsphere array. *Anal. Chem.* **72**, 5618–5624 (2000).
34. Leamon, J.H. & Rothberg, J.M. Cramming more sequencing reactions onto microreactor chips. *Chem. Rev.* **107**, 3367–3376 (2007).
35. Chen, Y.J. *et al.* Double ended sequencing. US patent 7,244,567. (2007).
36. Tawfik, D.S. & Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **16**, 652–656 (1998).
37. Lizardi, P.M. *et al.* Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19**, 225–232 (1998).
38. Nakano, M. *et al.* Single-molecule PCR using water-in-oil emulsion. *J. Biotechnol.* **102**, 117–124 (2003).
39. Leamon, J.H. *et al.* A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24**, 3769–3777 (2003).
40. Edwards, R.A. *et al.* Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* **7**, 57 (2006).
41. Fraser, C.M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
42. Bender, W., Spierer, P. & Hogness, D.S. Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.* **168**, 17–33 (1983).
43. Spierer, P., Spierer, A., Bender, W. & Hogness, D.S. Molecular mapping of genetic and chromomeric units in *Drosophila melanogaster*. *J. Mol. Biol.* **168**, 35–50 (1983).
44. Andries, K. *et al.* A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* **307**, 223–227 (2005).
45. Hiller, N.L. *et al.* comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome. *J. Bacteriol.* **189**, 8186–8195 (2007).
46. Hofreuter, D. *et al.* Unique features of a highly pathogenic *Campylobacter jejuni* strain. *Infect. Immun.* **74**, 4694–4707 (2006).
47. Oh, J.D. *et al.* The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc. Natl. Acad. Sci. USA* **103**, 9999–10004 (2006).
48. Raymond, J.A., Fritsen, C. & Shen, K. An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol. Ecol.* **61**, 214–221 (2007).
49. Smith, M.G. *et al.* New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. *Genes Dev.* **21**, 601–614 (2007).
50. Bainbridge, M.N. *et al.* Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* **7**, 246 (2006).
51. Goldberg, S.M. *et al.* A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. *Proc. Natl. Acad. Sci. USA* **103**, 11240–11245 (2006).
52. Wicker, T. *et al.* 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* **7**, 275 (2006).
53. Swaminathan, K., Varala, K. & Hudson, M.E. Global repeat discovery and estimation of genomic copy number in a large, complex genome using a high-throughput 454 sequence survey. *BMC Genomics* **8**, 132 (2007).
54. Henderson, I.R. *et al.* Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725 (2006).
55. Girard, A. I., Sachidanandam, R., Hannon, G.J. & Carmell, M.A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
56. Berezikov, E. *et al.* Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.* **38**, 1375–1377 (2006).
57. Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**, 69–82 (2007).
58. Tarasov, V. *et al.* Differential regulation of microRNAs by p53 revealed by massively parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell Cycle* **6**, 1586–1593 (2007).
59. Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C. & Ohlrogge, J.B. Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiol.* **144**, 32–42 (2007).
60. Cheung, F. *et al.* Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. *BMC Genomics* **7**, 272 (2006).
61. Gowda, M. *et al.* Robust analysis of 5′-transcript ends (5′-RATE): a novel technique for transcriptome analysis and genome annotation. *Nucleic Acids Res.* **34**, e126 (2006).
62. Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L. & Schnable, P.S. SNP discovery via 454 transcriptome sequencing. *Plant J.* **51**, 910–918 (2007).
63. Poinar, H.N. *et al.* Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394 (2006).
64. Gilbert, M.T.P. *et al.* Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res.* **35**, 1–10 (2007).
65. Stiller, M. *et al.* Inaugural article: patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc. Natl. Acad. Sci. USA* **103**, 13578–13584 (2006).
66. Sogin, M.L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **103**, 12115–12120 (2006).
67. Angly, F.E. *et al.* The marine viromes of four oceanic regions. *PLoS Biol.* **4**, e368 (2006).
68. Leininger, S. *et al.* Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**, 806–809 (2006).
69. Krause, L. *et al.* Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics* **22**, e281–e289 (2006).
70. Pennisi, E. Breakthrough of the year: human genetic variation. *Science* **318**, 1842–1843 (2007).
71. Steemers, F.J., Ferguson, J.A. & Walt, D.R. Screening unlabeled DNA targets with randomly ordered fiber-optic gene arrays. *Nat. Biotechnol.* **18**, 91–94 (2000).
72. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
73. Hayden, E. Personalized genomes go mainstream. *Nature (News)* **250**, 11 (2007).
74. Karow, J. Next-gen sequencers improve in '07; vendors promise more gains in 2008. *In Sequence* **2** (2008).
75. Egholm, M. Why length matters in next generation sequencing. Presented at Cambridge Healthtech Institute's Exploring Next Generation Sequencing meeting, Providence, RI. October 17–18, 2007.
76. Mitra, R.D. & Church, G.M. *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* **27**, e34 (1999).
77. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA* **100**, 8817–8822 (2003).
78. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
79. Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S.R. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **100**, 3960–3964 (2003).
80. Efcavitch, W. The HeliScope single molecule sequencer: an integrated genetic analyzer for true single molecule sequencing. Presented at Advances in Genome Biology and Technology meeting, Marco Island, FL, February 7–9, 2008.
81. Harris, T.D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).
82. Levene, M.J. *et al.* Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682–686 (2003).
83. Turner, S. Harnessing nature's powerful DNA sequencing engine: single molecule real time sequencing-by-synthesis. Presented at Advances in Genome Biology and Technology meeting, Marco Island, FL, February 7–9, 2008.
84 Illumina. Illumina sequences the first African human genome (Illumina, San Diego) http://investor.illumina.com/phoenix.zhtml?c=121127&p=irol-newsArticle&ID=1105194&highlight (6 February 2008)
85 Porreca, G.J. *et al.* Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
86. Campbell, P.J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).