Applications and Case Studies of the Next-Generation Sequencing Technologies in Food, Nutrition and Agriculture

George E. Liu*

USDA, ARS, ANRI, Bovine Functional Genomics Laboratory, Beltsville Agricultural Research Center (BARC) - East, 10300 Baltimore Avenue, Beltsville, MD, 20705, USA

Received: September 25, 2008; Accepted: October 16, 2008; Revised: October 21, 2008

Abstract: The next-generation sequencing technologies are able to produce millions of short sequence reads in a high-throughput, cost-effective fashion. The emergence of these technologies has not only facilitated genome sequencing but also started to change the landscape of life sciences. Here, I survey their major applications ranging from whole-genome sequencing and resequencing, single nucleotide polymorphism (SNP) and structural variation discovery, to mRNA and noncoding RNA profiling and protein-nucleic acid interaction assay. These case studies in structural, functional and comparative genomics, metagenomics, and epigenomics are providing a more complete picture of the genome structures and functions. In the near future, we will witness broad impacts of these next-generation sequencing technologies for solving the complex biological problems in food, nutrition and agriculture. In this article, recent patents based information is also included.

Keywords: Sequencing, resequencing, structural genomics, functional genomics, comparative genomics, metagenomics, epigenomics

INTRODUCTION

Over the last several years, the advance in sequencing technologies has delivered the next-generation sequencing machines, which employ massively parallel approaches to produce millions of short DNA sequence reads in a single run [1]. These next-generation sequencers have significantly accelerated whole-genome *de novo* sequencing and re-sequencing. Their high-throughput capacity of sequencing DNA is also rapidly changing the landscape of life sciences and providing many novel biological applications to explore previously unanswered questions [2]. Currently, three next-generation DNA sequencing systems are available on the market: Roche 454 GS FLX Genome Analyzer, Illumina Genome Analyzer II system, and Applied Biosystems' solid system. These next-generation sequencing platforms and their relevant patents have been reviewed before [3]. This review will showcase the impact that these new sequencers had and will have on a wide variety of research fields, especially for solving the complex biological problems in food, nutrition and agriculture.

GENERAL FEATURES OF THE NEXT-GENERATION SEQUENCERS

Compared to the traditional Sanger capillary sequencer, these next-generation sequencers are capable of massively parallel sequencing of millions of amplified DNA molecules in a single run. And they do not require the conventional cloning and amplification in bacteria. Despite their different configurations, the next-generation sequencers share many common features (see Table **1**): (1) relatively small amount of starting DNA (a few micrograms) is needed; (2) fragmented DNA templates are ligated to specific adaptors at both ends; (3) multiple PCR amplification cycles are performed; (4) amplified DNA templates are attached to a solid support in a reaction chamber or a flow cell; (5) during the extension cycles, sequencing reagents are repetitively applied and washed away; and (6) the number of the extension cycles is often limited, thus producing shorter read lengths of 35-250 bases as compared to the read length of 650-800 bases in the Sanger capillary sequencing. Because of their high-throughput capacities, these next-generation sequencers are better suited for the studies dealing with the whole genomes, replacing the Sanger sequencing in many situations. However, currently these next-generation sequencers are still substantially more expensive than the Sanger sequencers. Enormous amount of sequence data generated by these sequencers also require complex bioinformatics analysis, which often imposes an additional cost. The following sections showcase published or funded studies that have used these next-generation sequencers for novel biological applications, with special emphases on food, nutrition and agriculture.

SNP DISCOVERY

One of the central themes in genomics is to study genome differences or variations, including single nucleotide polymorphism (SNP), in order to understand the relationship between the genotype and phenotype. For example, based on the draft bovine genome assembly, a major effort was made to study SNP through the bovine HapMap project. In the past, human SNP discovery relied on PCR amplification of targeted regions, followed by capillary sequencing and *in silico* sequence alignment. However, this approach was impractical for other species because of the higher labor and reagent costs. The next-generation sequencing platforms, which can produce millions of short reads significantly faster and cheaper, fit in perfectly and have started to change the processes for SNP discovery. A recent publication [4] reported "an economical, efficient, single-step method for

*Address correspondence to this author at the Bovine Functional Genomics Laboratory, USDA-ARS, Building 200, Room 124B, BARC-East, Beltsville, MD 20705, USA; Tel: (301) 504-9843; Fax: (301) 504-8414; E-mail: George.Liu@ars.usda.gov

1876-1429/09 $100.00+.00

Table 1. Comparison of the Next-Generation DNA Sequencing Platforms

Platforms	Roche (454) GS-FLX	Illumina Genome Analyzer II system	ABI SOLiD
Starting DNA (μg)	3-5	0.1-1	0.1-20
Amplification	Emulsion PCR	Bridge PCR	Emulsion PCR
Sequencing method	Pyrosequencing	Sequencing by synthesis	Sequencing by ligation
Read length (bases)	250	32-40	35
Throughput capability (Gb per run)	0.1	1.3	4
Reagent cost per run (list prices)	$8,500	$3,000	$3,400
Run time	7.5 h	3 d	7 d
Paired reads/Span	Yes/3 kb	Yes/200-400 bp	Yes/3-20 kb

SNP discovery, validation and characterization", combining both the next-generation sequencing and reduced representation libraries (RRLs) technologies. Using an Illumina Genome Analyzer, the authors generated millions of short reads from 66 cattle. After mapping and filtering, over 62,000 putative SNPs were discovered with their allele frequencies determined. Despite the limitations imposed by the short read lengths, genotype data validated 92% of nearly 24,000 selected genome-wide SNPs, revealing a high genotypic and sequence allele frequency correlation. More importantly, as the authors pointed out: "This approach for simultaneous *de novo* discovery of high-quality SNPs and population characterization of allele frequencies can be applied to many other agricultural animals, which only have a partially sequenced draft genome".

MESSENGER RNA AND NONCODING RNA PROFILING

Aside from SNP discovery, expressed regions of a genome can be probed using the next-generation sequencing. The tissue-specific mRNA expression will provide important annotations about gene function. Sonstegard *et al*. (http://cris.csrees.usda.gov/, date of access October 20, 2008) is using the Solexa/Illumina sequencing platform to construct a Bovine Gene Atlas (BGA) database. This database will include transcript profiles from over 100 different tissues of L1 Dominette 01449 (the cow used to generate the draft genome sequence) and her offspring. The next-generation sequencing platforms are capable of identifying expression levels of nearly all genes, including those rare and species-specific transcripts. A similar approach can be applied to large genomes, which are often associated with crop plants. For example, wheat has a 16-Gb hexaploid genome. Although extending the next-generation sequencing from a 3-Gb diploid genome to a 16-Gb hexaploid genome would result in loss of mappability of the reads produced, these large genomes can at least be partially probed at the transcriptional level using expressed sequence tags (EST).

Noncoding RNA, including micro RNA (miRNA), is a broad class of regulatory RNA molecules which form another fundamental layer of gene regulation. The next-generation sequencing instruments are ideal for genome-wide noncoding RNA discovery because of their short lengths [5]. Multiple studies have discovered new and different noncoding RNA classes in model organisms and humans [6-8]. Using the next-generation sequencing, Matukumalli *et al.* (http://cris.csrees.usda.gov/, date of access October 20, 2008) have planned to screen the noncoding RNA fraction from the same pool of multiple, diverse cattle tissues developed for the BGA. These studies will facilitate the investigations of regional transcriptional control, tissue phylogeny and interconnected gene networks. They will also expand our understanding of gene expression, and provide insight into the impact of gene regulation on animal development, health, and production.

DE NOVO SEQUENCING AND RE-SEQUENCING

Metagenomics is defined as the application of genomics techniques to directly study communities of microbial organisms without isolation and cultivation of individual species [9]. It involves the characterization of the genomes in these communities, as well as their mRNA, protein and metabolic products. Its major motivations came from the larger-than-expected diversity of microbial organisms and viral agents in ecosystems, such as the human body, global soil, deep mine and the ocean [10]. The main goal of metagenomics is to assess the extent and the role of microbial biodiversity in these systems. An earlier survey of 16S ribosomal RNA (rRNA) genes revealed that cultivation based methods find less than 1% of the species in a sample [11]. The next-generation sequencing technologies have moved metagenomics from a single organism type in isolation to the studies of whole communities. By removing the cloning and culture steps, the main biases in metagenomics, the next-generation sequencing strategies are straight forward in that (1) deep sequencing of DNA fragments is conducted on an uncultured sample; (2) short reads are compared against databases of known sequences using bioinformatics tool like MEGAN [12] with or without assembly; and (3) these data are then used to compute and explore their contents to infer relative abundances. Therefore, the next-generation sequencing of microbial pathogens directly from food, clinical, agricultural and environmental specimens provide unprecedented opportunities for pathogen discovery and surveillance.

Food Safety - *E.coli* and *Salmonella* outbreak: Between 1996 and 2006, there were 24 reported incidents of foodborne illness caused by contaminated fresh-cut produce in the United States. In fall 2006, an outbreak of *E. coli* O157:H7 linked to the consumption of bagged spinach resulted in 205 confirmed illnesses and three deaths. Using the product codes on the bags and conventional DNA fingerprinting on the bacteria from the bags, the investigators performed longtime and painstaking detective work. Finally, the investigators successfully identified the environmental risk factors and the areas that were most likely involved in the outbreak. But they were unable to definitely determine how the contamination originated (http://www.fda.gov/bbs/topics/NEWS/2007/NEW01593.html, date of access October 20, 2008). In spring 2008, a consumer warning was issued nationwide that an outbreak of *Salmonella* serotype Saintpaul is linked to consumption of certain types of raw tomatoes and peppers. This uncommon type *Salmonella* of the same genetic fingerprint has been identified in hundreds of infected persons in over 32 states (http://www.cdc.gov/*Salmonella*/saintpaul/, and http://www.fda.gov/oc/opacom/hottopics/tomatoes.html, date of access October 20, 2008). Based on multiple lines of evidence, the investigators indicated that *Salmonella* Saintpaul originated in Mexico. In such food safety outbreaks, the focus is on discovering the source of contamination and removing it from the shelves to prevent further illness. In the event of bioterrorist attacks, it is crucial to quickly pinpoint the genetic basis for pathogen virulence or drug resistance to find a cure. Therefore, the rapid identification of the strain-specific sequences as well as gene polymorphisms, gain, loss, and modifications, is critical for emergency response decision-making. It has been demonstrated that the next-generation sequencing and comparative genomics could be used to effectively identify unique features of a known bacterial pathogen strain in a matter of weeks [13]. This strategy will drastically reduce the time needed for accurately identifying unique genetic properties of a potential outbreak strain. However, the process of using genomics information to identify, develop and test a drug for treatment of the illness or pathogen is still far from being a developed technology.

Animal and Human Health - Sequencing Clinical Isolates: Many pathogens evolve continually by mutation and by exchanging gene segments among their population (reassortment). For example, the influenza viruses constantly mutate by reassortment, making the existing vaccines ineffective, and requiring a new vaccine to be made based on new strains. Less-frequent major changes, known as antigenic shift, create new strains against which the human population has little protective immunity, thereby causing the worldwide flu pandemics [14,15]. Understanding the evolution of influenza viruses in birds and mammals is important for surveillance and vaccine strain selection. Using the Sanger sequencing, Salzberg *et al.* sequenced and analyzed the complete genomes of 36 recent influenza A (H5N1) viruses collected from birds in Europe, northern Africa, and southeastern Asia [16]. The next-generation sequencers can provide rapid data about antibiotic susceptibility and/or resistance, and other virulence markers. Another clear benefit of the next-generation platforms is that rare variants in the clinical strain population can be detected by virtue of the depth of sampling obtained. Several studies using the 454 platform have been published on pathogens such as HIV viruses [17] and *Mycobacterium tuberculosis* [18]. This knowledge, in turn, should lead to improved diagnostics, monitoring and treatments of diseases.

Agriculture - Honeybee colony collapse disorder: Honey bees are essential for production of over 90 food crops. Honey bee colony collapse disorder (CCD) was first reported in 2006. CCD resulted in a loss of 50 to 90% of colonies across the United States threatening the honey and pollination industries. It has the following characteristics: (1) a quick loss of adult bees; (2) excess brood, in all stages, being abandoned in the hive; (3) a lack of dead bees in or near the hive; and (4) low levels of varroa (a type of parasitic mites associated with honey bees). Researchers proposed and investigated multiple hypotheses that CCD was due to infectious agents. These hypotheses were motivated by the observation that CCD was transmissible through the reuse of equipment from CCD colonies and that such transmission could be prevented by irradiation treatment. The authors performed a metagenomic survey of microflora in CCD hives, normal hives, and imported royal jelly collected from multiple places over a period of time [19]. Multiple candidate pathogens were screened for significant association with CCD. Although no evidence was presented for causal relationships between any infectious agent and CCD, the authors have concluded that Israeli acute paralysis virus (IAPV) of bees was strongly linked to CCD.

Personalized Nutrition and The Human Microbiome Project: The microorganisms that live on and inside humans (including the skin, oral and nasal cavities, gastrointestinal tracks, and vagina) are estimated to outnumber human cells by a factor of ten to one [20]. The Human Microbiome Project (HMP) is defined as “a strategy to understand the microbal components of the human genetics and metabolic landscape and how they contribute to normal physiology and predisposition to disease” [20]. Direct and thorough examination of these microbial communities has the potential to greatly improve human development, nutrition, physiology, and immunity. The next-generation sequencing technologies have inspired several studies using 16S rRNA gene-based surveys of bacterial communities that reside on and in the human bodies [21,22]. For example, obesity and being overweight pose a major risk for human health. Interestingly, obesity was also associated with changes in the relative abundance of the two dominant bacterial divisions in both human and mouse GI tracts [20,23]. Comparing the distal gut microbiota from obese individuals to those from their lean counterparts, the authors showed that changes in the gut microbiota affected their host metabolic potential and the obese microbiome had an increased capacity to harvest energy from the diet. Furthermore, these authors demonstrated that this trait was transmissible using the mouse model: colonization of germ-free mice with an obese microbiota resulted in a significantly greater increase in total body fat than colonization with a lean microbiota. These results identified the gut microbiota as an additional contributing factor to the balance of nutrition, which might have potential therapeutic implications. Therefore, personalized nutrition incorporating the HMP results may lead to new recommendations for food production, distribution and consumption in the future. Similar type of research has been applied in mammals [24], including domesticated livestock, to improve the feed utilization and productive efficiency in the agriculturally important animals.

DEFINING PROTEIN-NUCLUIC ACID INTERACTIONS AND EPIGENOMICS

The interactions between proteins and DNA regulate chromatin packaging or DNA transcription into mRNA. How the genome instructs protein binding in a sequence- and/or structure-dependent manners is poorly studied in complex organisms. The most prevalent high-throughput approach to study protein-DNA interactions has been the combination of chromatin immunoprecipitation with DNA microarray (ChIP-chip). In contrast, ChIP-seq (chromatin immunoprecipitation and direct sequencing) is a recently developed technique, which inherits two advantages from the next-generation sequencing: (1) it is not limited by the microarray content; and (2) it does not depend on the efficiency of probe hybridization. The ChIP-seq approach was recently used to identify binding sites of two transcription factors, STAT1 and NRSF in human cells [25,26]. Both studies compared their findings with those generated by ChIP-chip, demonstrating that ChIP-seq had better resolution and required fewer replicates.

In many biological processes, the regulation of gene expression also involves epigenetic mechanisms. DNA methylation, histone modification, nucleosome positioning, chromatin packaging and other factors such as noncoding RNAs all contribute to an overall epigenome [27]. Method for methylation analysis of nucleic acid was studied by Tetzner and Lewin in WO08017411 patent [28]. Methylation can be also predicted based on distinguished methylated regions as discussed by Karger and Boyd [29]. Using ChIP-seq, several studies demonstrated the connections among chromatin packaging, histone binding and gene expression in human cells [30,31].

Together these studies established ChIP-seq as a powerful genome-wide approach to study DNA-protein interaction and epigenomics. Since the presence of a draft genome assembly is no longer required, many organisms are now accessible for ChIP-seq based assays. Even though no relevant literature exists yet in the fields of food, nutrition and agriculture, it is expected that ChIP-seq will soon be applied to these fields.

CURRENT & FUTURE DEVELOPMENTS

The current cost of using the next-generation technologies to resequence a human or mammalian genome is still high. One reason is that shorter reads may impair their effectiveness in *de novo* sequencing and genome assembly. Short reads (e.g. 35-50 bases) will likely require 25- to 30-fold coverage of the genome to capture all the genetic information. Short reads also restrict our power to detect sequence variation, because the mapping process is based on their sequence uniqueness within the genome. However, this situation is being improved by (1) the longer read length using better extension reagents and chemistries and (2) the ability to sequence the paired ends of a given fragment, using a different library preparation process. As a front end technique, microarray sequence capture of interested genomic regions has been recently developed to extend the utility of the next-generation sequencing platforms [32,33]. Although the accuracy of the next-generation sequencing reads and associated quality values are not yet well defined or standardized, multiple efforts are underway to define them as compared to the Sanger sequencing [34]. More information about the bioinformatics issues such as data storage, sharing and management, can be found in several recent reviews [35-37]. Venter *et al.* have recently demonstrated the synthetic genomes which are used for a variety of purposes, including the generation of synthetic fuels, such as hydrogen or ethanol [38]. As the sequencing cost further decreases, it will soon be possible to resequence a large number of genomes. The 1000 Genomes Project began to resequence over 1000 ‘normal’ human genomes (http://www.1000genomes.org/index.html, date of access October 20, 2008) using the next-generation platforms to create a deep catalog of human genetic variation. This effort aims to capture not only rare alleles at the single base level, but also structural variants, as well as large and small insertions and deletions. Whole genome based genetic evaluation and selection method have been studied by Raadsma *et al.* in WO08025093 patent [39]. With the genome structures are now better characterized by the next-generation sequencing, the next goal is to understand their functions: how the genome determines the different phenotypes and traits. For example, the bovine HapMap project was a systematic effort to catalog the common variants like SNP and study their relationships within the cattle population. It reduced the complexity for the genetic variation so that whole genome association studies could be performed to correlate the inheritance of complex traits with the specific variants or specific parts of genome tagged by variants. Therefore, it is expected that whole genome association studies may soon begin to identify candidate genes that influence common traits. In the near future, we will also witness broad impacts brought in by the new biotechnologies. For example, Valentin and Mitsky discussed the gene isolation, gene analysis and the production of transgenic plants which were modified to express proteins associated with the tocopherol pathway. They also studied methods for the production of products from the tocopherol biosynthesis pathway [40]. All these developments will dramatically change the fields of food, nutrition and agriculture.

ACKNOWLEDGEMENTS

The author is supported in part by Project 1265-31000-090-00D from USDA-ARS and National Research Initiative (NRI) Grant No. 2007-35205-17869 from USDA-CSREES. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] Chi KR. The year of sequencing. Nat Methods 2008; 5: 11-14.

[2] Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008; 5: 16-18.

[3] Lin B, Wang J, Cheng Y. Recent patents and advances in the next-generation sequencing technologies. Recent Pat Biomed Eng 2008; 1: 60-67.

[4] Van Tassell CP, Smith TP, Matukumalli LK, *et al*. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 2008; 5: 247-252.

[5] Kiriakidou M, Nelson PT, Kouranov A, *et al*. A combined computational-experimental approach predicts human microRNA targets. Genes Dev 2004; 18: 1165-1178.

[6] Kasschau KD, Fahlgren N, Chapman EJ, *et al*. Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 2007; 5: e57.

[7] Ruby JG, Jan C, Player C, *et al*. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell 2006; 127: 1193-1207.

[8] Berezikov E, Thuemmler F, van Laake LW, *et al*. Diversity of microRNAs in human and chimpanzee brain. Nat Genet 2006; 38: 1375-1377.

[9] Chen K, Pachter L. Bioinformatics for whole-genome shotgun sequencing of microbial communities. PLoS Comput Biol 2005; 1: 106-112.

[10] Nee S. More than meets the eye. Nature 2004; 429: 804-805.

[11] Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 1998; 180: 4765-4774.

[12] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007; 17: 377-386.

[13] La Scola B, Elkarkouri K, Li W *et al*. Rapid comparative genomic analysis for clinical microbiology: The *Francisella tularensis* paradigm. Genome Res 2008; 18: 742-750.

[14] Ghedin E, Sengamalay NA, Shumway M, *et al*. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 2005; 437: 1162-1166.

[15] Holmes EC, Ghedin E, Miller N, *et al*. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. PLoS Biol 2005; 3: e300.

[16] Salzberg SL, Kingsford C, Cattoli G, *et al*. Genome analysis linking recent European and African influenza (H5N1) viruses. Emerg Infect Dis 2007; 13: 713-718.

[17] Wang C, Mitsuya Y, Gharizadeh B, *et al*. Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. Genome Res 2007; 17: 1195-1201.

[18] Andries K, Verhasselt P, Guillemont J, *et al*. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. Science 2005; 307: 223-227.

[19] Cox-Foster DL, Conlan S, Holmes EC, *et al*. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 2007; 318: 283-287.

[20] Turnbaugh PJ, Ley RE, Hamady M, *et al*. The human microbiome project. Nature 2007; 449: 804-810.

[21] Grice EA, Kong HH, Renaud G, *et al*. A diversity profile of the human skin microbiota. Genome Res 2008; 18: 1043-1050.

[22] Ley RE, Turnbaugh PJ, Klein S, *et al*. Microbial ecology: Human gut microbes associated with obesity. Nature 2006; 444: 1022-1023.

[23] Turnbaugh PJ, Ley RE, Mahowald MA, *et al*. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 2006; 444: 1027-1031.

[24] Ley RE, Hamady M, Lozupone C, *et al*. Evolution of mammals and their gut microbes. Science 2008; 320: 1647-1651.

[25] Robertson G, Hirst M, Bainbridge M, *et al*. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 2007; 4: 651-657.

[26] Euskirchen GM, Rozowsky JS, Wei CL, *et al*. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res 2007; 17: 898-909.

[27] Schones DE, Zhao K. Genome-wide approaches to studying chromatin modifications. Nat Rev Genet 2008; 9: 179-191.

[28] Tetzner, R., Lewin, J.: WO08017411 (**2008**).

[29] Karger, A., Boyd, V.: US2008017218A1 (**2008**).

[30] Barski A, Cuddapah S, Cui K, *et al*. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129: 823-837.

[31] Mikkelsen TS, Ku M, Jaffe DB, *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007; 448: 553-560.

[32] Hodges E, Xuan Z, Balija V, *et al*. Genome-wide *in situ* exon capture for selective resequencing. Nat Genet 2007; 39: 1522-1527.

[33] Dahl F, Stenberg J, Fredriksson S, *et al*. Multigene amplification and massively parallel sequencing for cancer mutation discovery. Proc Natl Acad Sci USA 2007; 104: 9387-9392.

[34] Holt RA, Jones SJ. The new paradigm of flow cell sequencing. Genome Res 2008; 18: 839-846.

[35] Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet 2008; 24: 133-141.

[36] Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. Trends Genet 2008; 24: 142-149.

[37] Rusk N, Kiermer V. Primer: Sequencing--the next generation. Nat Methods 2008; 5: 15.

[38] Venter, C.J., Smith, H.O.: WO08024129 (**2008**).

[39] Raadsma, H., Tier, B., Woolaston, A.F., Moser, G.C.: WO08025093 (**2008**).

[40] Valentin, H.E., Mitsky, T.A.: EP1950305 (**2008**).