Genomics 93 (2009) 105–111

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Review

Generations of sequencing technologies

Erik Pettersson *, Joakim Lundeberg, Afshin Ahmadian

Department of Gene Technology, Royal Institute of Technology (KTH), AlbaNova University Center, Roslagstullsbacken 21, SE - 106 91 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 2 September 2008
Accepted 2 October 2008
Available online 21 November 2008

Keywords:
DNA sequencing
Next generation

ABSTRACT

Advancements in the field of DNA sequencing are changing the scientific horizon and promising an era of personalized medicine for elevated human health. Although platforms are improving at the rate of Moore's Law, thereby reducing the sequencing costs by a factor of two or three each year, we find ourselves at a point in history where individual genomes are starting to appear but where the cost is still too high for routine sequencing of whole genomes. These needs will be met by miniaturized and parallelized platforms that allow a lower sample and template consumption thereby increasing speed and reducing costs. Current massively parallel, state-of-the-art systems are providing significantly improved throughput over Sanger systems and future single-molecule approaches will continue the exponential improvements in the field.

Contents

Introduction

The ability to swiftly and accurately gain knowledge of nucleic acid composition is essential to many of the biological sciences. As the pace of progress is high and we are moving towards an era of synthetic genomics and personalized medicine, the demand for highly efficient sequencing technologies is obvious, where effortless deciphering of genetic sequences will shed light on novel biological functions and phenotypic differences. Metagenomic endeavors [1–3] are providing new tools in the art of genetic engineering, thereby enabling the design of artificial life in the service of humanity [4]. These future synthetic organisms may produce petrol substitutes or provide systems for mopping up excessive carbon dioxide in the atmosphere [5–7]. Perhaps even more captivating is the possibility of resequencing larger and larger fractions of human genomes at an ever decreasing cost, an effort that will elucidate phenotypic variants, extending the comprehension of disease susceptibility and pharmacogenomics, permitting personalized medicine.

Although we have not yet reached the long envisioned $1000 genome [8], novel approaches and refinements of existing methods are reducing the cost per base by the day while increasing the throughput. The establishment of a reference genome in the beginning of this decade [9,10] is now permitting cost-effective resequencing of ever larger fractions of human genomes. The Advanced Sequencing Technology Development Awards initiated by the National Human Genome Research Institute (NHGRI) in 2004 [11] are beginning to show results. Advancements for the next generation sequencing methods include not only current state-of-the-art systems from 454 [12,13], Illumina [14,15] and Applied Biosystems [16] but also single-molecule detection approaches, capable of recognizing incorporation or hybridization events on single molecules. Further into the future lies more direct recognition of unamplified material, i.e. nano pores or nano edges relying on physical recognition of the bases in an unmodified DNA strand, rather than detecting chemical incorporation.

* Corresponding author. Fax: +46 8 5537 8481.
E-mail address: eriq@kth.se (E. Pettersson).

0888-7543/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2008.10.003

The drop in cost has led to the initiation of several sequencing projects aiming at elucidating the variation not covered by SNP arrays. In the Personal Genome Project [17–19], the exon regions of ten genomes are to be sequenced and compared. Researchers at the Beijing Genomics Institute (BGI) [20] are determined to sequence 100 individuals of Han Chinese origin during the upcoming three years in the Yanhuang Project and recently, an international consortium announced the "1000 Genomes Project" where the sequence of 1000 individuals will provide "A catalogue of human genetic variation" [21]. The improvements in sequencing technology and reduction in cost have allowed the first personal genomics company [22] to begin the sequencing of customers' genomes.

To allow for a further reduction in cost the X PRIZE Foundation in Santa Monica, CA, has introduced the Archon X PRIZE for Genomics [23] and will award a sum of $10 million to the first team that can design a system capable of sequencing 100 human genomes in 10 days. Additional requirements are an error rate of no more than one in 100,000 bases, a coverage of at least 98% and a cost of no more than $10,000 for each sequenced genome. Representatives from many of the different sequencing categories are represented in the Archon X PRIZE challenge and the research world is closing in on the $1000 genome. The race is on.

Present generation of DNA sequencing technologies

There are many factors to consider in DNA sequencing such as read length, bases per second and raw accuracy. All the work in the field has led to an exponential reduction in cost per base. Sanger sequencing has been one of the most influential innovations in biological research since it was first presented in 1977. A little more than 20 years later, a bioluminescence sequencing-by-synthesis approach saw the light of day [24]. Today, Pyrosequencing has evolved at 454 Life Sciences, generating about five hundred million bases of raw sequence in just a few hours [12]. This throughput, although heavily refined and improved during the years, is something Sanger sequencing in its current form cannot easily match. However, during the last year, Illumina and Applied Biosystems have introduced sequencing systems offering even higher throughput than the systems provided by 454, generating billions of bases in a single run. These novel methods all rely on parallel, cyclic interrogation of sequences from spatially separated clonal amplicons. Although with shorter read lengths and a slower sequence extraction from individual features as compared to the Sanger method, the parallelized process offers a much higher total throughput and reduces cost significantly by generating thousands of bases per second. By shearing the template and parallel sequencing of single fragments, over sampling may provide improved coverage and the possibility of stitching together the original sequence while increasing total accuracy. Already today these high-throughput methods are expanding our knowledge, also in the related fields of transcriptome and proteome research. Gene expression analysis with whole-transcriptome sequencing is possible and furthermore, in proteome research, by sequencing DNA extracted by antibodies targeting DNA-binding proteins (ChIP-Seq), transcription factor binding sites and chromatin modifications can be investigated [25,26].

Terminating chains

Since 1977, a total nucleic acid polymer of approximately 10^{11} bases has been determined with Sanger's chain termination sequencing method [27]. By halting the elongation with a labeled, and thereby identifiable, dideoxyribonucleotide triphosphate (ddNTP), the length of the fragment can be utilized for interrogating the base identity of the terminating base [28]. In its current form, fluorescently labeled ddNTPs [29,30] are mixed with regular, non labeled, non terminating nucleotides in a cycle sequencing reaction [31,32] rendering elongation stops at all positions in the template. Capillary electrophoresis can then be applied for separating sequences by length and providing subsequent interrogation of the terminating base (see Fig. 1A).

Initially at a high cost, refinements and automation have improved cost effectiveness significantly. In 1985, $10 allowed reading one single base, while the same amount of money rendered 10,000 bases 20 years later [8,27]. Current instruments provided by Applied Biosystems deliver read lengths of up to 1000 bases, high raw accuracy and allow for 384 samples to be sequenced in parallel generating 24 bases per instrument second. Projects of multiplexing and miniaturization in order to reduce reagent volumes, lower consumable costs and increase throughput are being pursued [33,34].

Hybridization to tiling arrays

The concept of allele-specific hybridization (ASH) has been used for resequencing and genotyping purposes by expanding a probe set, targeting a specific position in the genome, to include interrogation of each of the four possible nucleotides [35]. A tiling array can be fabricated with probe sets targeting each position in the reference genome. Read length is given by the probe length (often 25 bp) and base calling is performed by examining the signal intensities for the different probes of each set. Accuracy is an issue and is dependent on the ability of the assay to discriminate between exact matches and those with a single base difference. Performance may vary significantly due to different base compositions (different thermal annealing properties) of different regions, resulting in problems with false positives as well as with large inaccessible regions composed of repetitive sequence stretches [36,37].

The throughput is an obvious benefit, since all bases are interrogated simultaneously and the concept has been applied to resequencing the human chromosome 21 by Perlgen [37] and HIV [36]. By representing all possible sequences for a given probe length, de novo sequencing can be performed and overlapping sequences used for sequence assembly [38]. In a recent report, the genome of Bacteriophage λ and *Escherichia coli* were resequenced by "shotgun sequencing by hybridization" with an accuracy of 99.93% and a raw throughput of 320 Mbp/day [39].

Parallelized Pyrosequencing

The Genome Sequencer FLX by 454 Life Sciences [13] and Roche depends on an emulsion PCR followed by parallel and individual Pyrosequencing of the clonally amplified beads in a PicoTiterPlate (see Fig. 1B). Emulsion PCR is a clonal amplification performed in an oil-aqueous emulsion. Unlike when digesting a genome with restriction endonucleases, shearing will provide randomly fragmented pieces of more or less similar length. By the addition of general adaptor sequences to the fragments, only one primer pair is required for amplification. In the emulsion PCR, a primer-coated bead, a DNA fragment and other necessary components for PCR (including the second general primer) are isolated in a water micro-reactor, favoring a 1:1 bead to fragment ratio. Once the emulsion is broken, beads not carrying any amplified DNA are removed in an enrichment process [12,40]. The amplified and enriched beads are then distributed on the PicoTiterPlate, where a well (44 μm in diameter) allows fixation of one bead (28 μm in diameter) [12]. However, out of the 1.6 million wells, not all will contain a bead and not all of those that do will give a useful sequence.

Following the distribution of the DNA-carrying beads to the PicoTiterPlate Pyrosequencing will be performed. Pyrosequencing is a sequencing-by-synthesis method where a successful nucleotide incorporation event is detected as emitted photons [41]. Since the single-stranded DNA fragments on the beads have been amplified with general tags, a general primer is annealed permitting an elongation

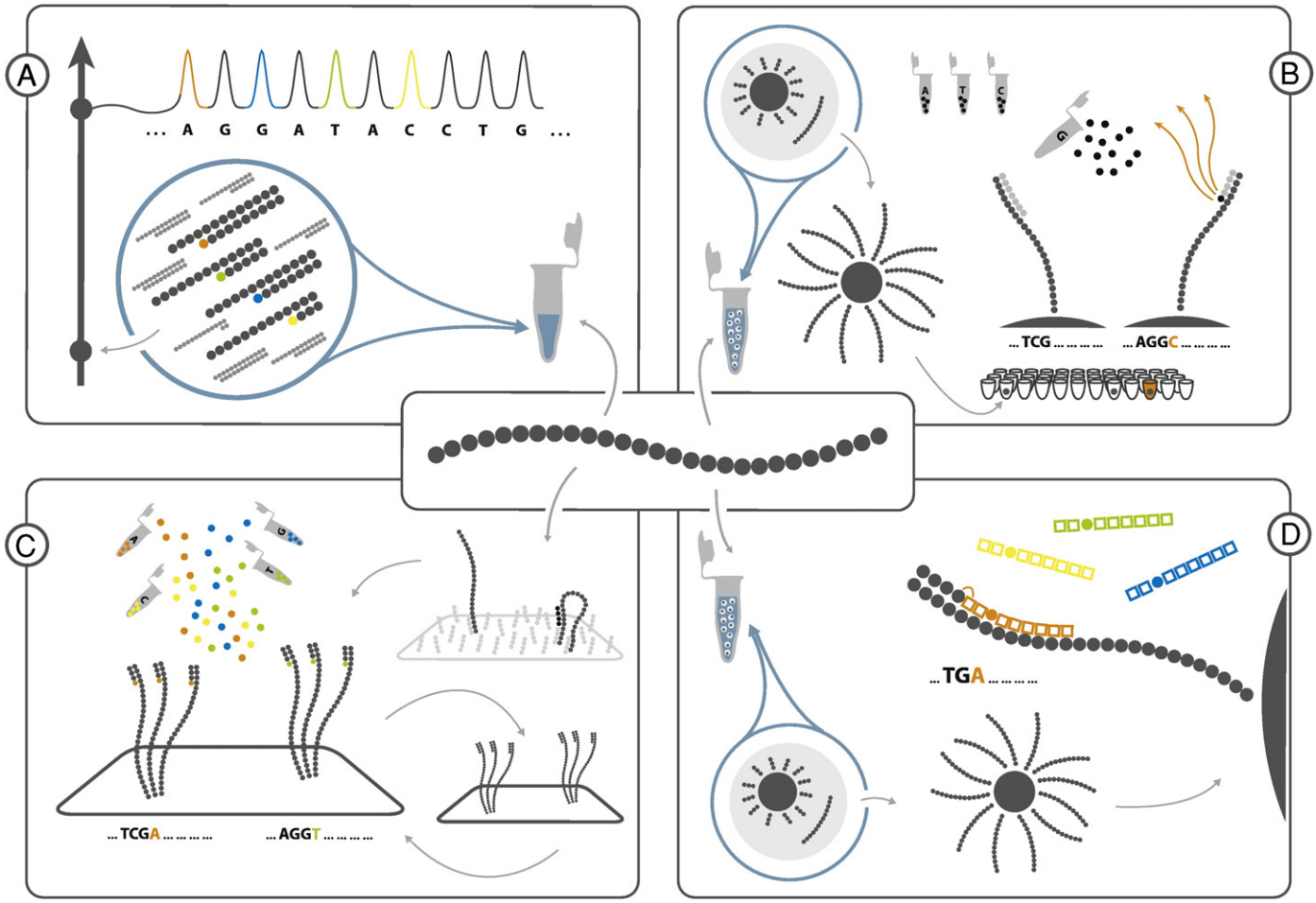

Fig. 1. (A) Chain Termination. A DNA sequence of choice is prepared using a sequencing reaction where regular deoxynucleotides are combined with terminating dideoxynucleotides. Each chain terminating nucleotide is labeled with a base-specific color. The sequencing reaction generates fragments of all lengths and separation can be made using a gel or capillary electrophoresis where the labeled bases reveal the sequence information at each position. Read length is approximately 700 bases. (B) Miniaturized Pyrosequencing. Highly parallel and miniaturized Pyrosequencing reactions are achieved by first performing a water-in-oil emulsion PCR that permits generation of hundreds of thousands of single-clone amplified beads. The beads are then single fitted into the wells of a PicoTiterPlate where individual Pyrosequencing reactions are taking place. A sequential addition of the four bases is performed in a cyclic fashion and upon successful incorporation of each base, an enzyme cascade generates light which can be detected. Read lengths of 400 bases are now possible allowing for a total of 500 Mbp from a PicoTiterPlate in each run. (C) Reverse Termination. The DNA fragments of choice are bridge-amplified using a solid-phase PCR on a surface generating spatially separated colonies of approximately thousand fragments each. A cyclic sequence interrogation procedure is performed using fluorescently labeled, reversibly terminating nucleotides. All four bases are added in each cycle and following incorporation and stringent washing procedures, the color of each colony is detected. The dye is then removed and the termination reversed allowing for interrogation of the following base in each colony. At 30–35 bases, the error rate is becoming high thereby limiting the read-length. (D) Sequencing-by-Ligation. Single-clone beads are amplified in an emulsion PCR and immobilized in a gel. By utilizing previously introduced general tag sequences, anchor primers can be annealed next to unknown sequence regions. Hybridizing and ligating degenerated nonamers, where only one base and position in the primer is specific and un-degenerated, reveals the base at the position in question since each primer and base is correlated to a particular color. The specificity of the ligase permits sequencing of six or seven bases depending on the ligation direction (5′–3′ and 3′–5′ respectively). In the illustration, the third base (a filled circle) is interrogated at one bead and the color of ligated probe indicates an "A" at the third position. Note that squares indicate degenerated bases). After each cycle, the ligated products are removed and a new round of ligation is performed by shifting the position of the specific base in the nonamers.

towards the bead. The emission of photons upon incorporation depends on a series of enzymatic steps. Incorporation of a nucleotide by a polymerase releases a diphosphate group (PPi), which catalyzed by ATP sulphurylase forms adenosine triphosphate (ATP) by the use of adenosine phosphosulphate (APS). Finally, the enzyme luciferase (together with D-luciferin and oxygen) can use the newly formed ATP to emit light. Another enzyme, apyrase, is used for degradation of unincorporated dNTPs as well as to stop the reaction by degrading ATP [41].

In the 454 system, the Pyrosequencing technology is adapted as follows. The enzymes luciferase and ATP sulphurylase are immobilized on smaller beads surrounding the larger amplicon carrying beads. All other reagents are supplied through a flow allowing reagents to diffuse to the templates in the PicoTiterPlate. Polymerase and one exclusive dNTP per cycle generate one or more incorporation events and the emitted light is proportional to the number of incorporated nucleotides. Photons are detected by a CCD camera and after each round, apyrase is flowed through in order to degrade excess nucleotides. The washing procedure for the removal of by-products permits read lengths of over 400 bp (250 bp in the GS FLX system and over 400 bp in the recently upgraded instrument, the GS FLX Titanium). This limitation is due to negative frame shifts (incorporation of nucleotides in each cycle is not 100% complete) and positive frame shifts (the population of nucleotides that is not fully degraded by the apyrase and can therefore be incorporated after the next nucleotide) that eventually will generate high levels of noise. Approximately 1.2 million wells will give one unique sequence of 400 bp, on average generating less than 500 million bases (Mb) in one single run. Whole-genome sequencing has been performed on bacterial genomes in single runs [12]. An oversampling of 20× permits the identification of PCR-introduced errors and to call homopolymeric errors [42].

454 Life Sciences is a competitor in the Archon X PRIZE and by moving the parallel Pyrosequencing technology onto a microchip [23] they believe the system will achieve the "scalability it needs to win".

Reverse termination

The Illumina 1G Genome Analyzer is relying on clonal bridge amplification on a flow cell surface generating 10 million single-molecule clusters per square centimeter. Bridge amplification is performed after immobilization of oligonucleotides complementary to the adaptor sequences on a surface [15,43,44]. Sheared and adaptor-ligated sample DNA fragments can be attached to the solid support and due to the dense lawn of adaptor complementary sequences on the surface, each will anneal to a nearby primer. A double stranded bridge will form after elongation, and denaturing will free the two strands, both now fixed on the surface. Repeated cycles will form colony like local clusters, each containing approximately 1000 copies and with a diameter of about 1 μm (see Fig. 1C). Sequencing is then carried out with fluorescently labeled nucleotides that are also reversible terminators. One base is incorporated and interrogated at a time since further elongation of the chain is prevented [14]. When all colonies are scanned at the end of a cycle and the base determined for each colony, the fluorophores are cleaved off and terminating bases are activated, allowing another round of nucleotide incorporation (see Fig. 1C). The presence of and competition among all four nucleotides is claimed to reduce the chance of misincorporation. Incomplete incorporation of nucleotides and insufficient removal of reverse terminators or fluorophores may be the explanation for the relatively short read length of 35 bases. Although shorter read lengths than the 454 system, the throughput is much higher and, as of February 2008, 1.5 Gbp are generated in each run, which takes approximately 3 days. The use of paired-end libraries will generate about 3 Gbp in a single run. The raw accuracy is said to be at 98.5% and the consensus (3× coverage) at 99.99%. The cost per base is approximately 1% of the cost for Sanger sequencing [15,45]. A variant of Illumina's sequencing by synthesis chemistry was recently reported where a hybrid of sequencing by synthesis and Sanger method promises longer reads [46].

Ligating degenerated probes

Strategies for sequencing-by-ligation have been presented in the form of Massively Parallel Signature Sequencing (MPSS) and Polony sequencing [40,47]. MPSS was demonstrated as signature sequencing of expression libraries of in vitro cloned microbeads, i.e. beads carrying multiple copies of a single DNA sequence [48]. Signature sequencing was carried out by restriction enzyme mediated exposure of four nucleotides in each cycle followed by ligation of an interrogator probe. This process was repeated for 4–5 cycles, i.e. querying 16–20 bases in total. An overhang of four bases would require 256 different complementary probes and just as many fluorophores for immediate recognition. Instead, the use of 16 (4×4) probes, each with a unique decoder binding site, has enabled single dye detection.

Resequencing of a bacterial genome was used to demonstrate the Polony sequencing method [40]. A mate-paired library was clonally amplified with an emulsion PCR on 1 μm beads and subsequently immobilized in a polyacrylamide gel. Each DNA-carrying bead (polony) represented two 17–18 bp genomic sequences flanked by different universal sequences. Due to the nature of the mate-pair construction, the two genomic sequences were separated by approximately 1 kb in the genome. Sequencing-by-ligation (see Fig. 1D) could then be performed using degenerate nonamers, where each known nucleotide was associated with one of four fluorophores. By using four different anchor primers, degenerate sequencing-by-ligation could be performed from each end of the tags. 7 bases could be obtained when sequencing in the 5′ to 3′ direction and 6 bases from 3′ to 5′. Ligated primers were removed after each round rendering information of 26 bases from each amplicon in a pattern of: 7 bases, a gap of 4–5 bases, 6 bases, then a gap of approximately 1 kb (mate-paired constructed) and then another 7 bases, a gap of 4–5 bases, followed finally by 6 bases.

These two methods have spawned the development of the commercial SOLiD system (Sequencing by Oligonucleotide Ligation and Detection) from Applied Biosystems where clonal amplicons on 1 μm beads are generated by an emulsion PCR, either from fragments or mate-paired libraries. The beads are enriched, so that 80% of them generate signals, and attached on a glass surface forming a very high-density random array. Sequencing-by-ligation is performed by ligating 3′-degenerated and 5′-labeled probes to the amplicons and detecting the color. Accuracy is improved by implementing a two-base encoding system that leads to interrogation of each base twice. A sequencing run takes 6–10 days and the output is high, approximately 3–6 Gbp per run given a read length of 25–35 bases per clonally amplified bead [16].

An open source implementation [49] of the Polony sequencing technology is the Danaher Motion Polonator G.007 where 200 such modules will be used by a team competing in the Archon X PRIZE race. They are hoping to reach the $10K per genome during 2008 by further improvements and optimizations of the technology.

Future generation of DNA sequencing technologies

The initial sequencing and mapping of the human genome is estimated to have cost about $3 billion [9,10]. The genome of Craig Venter, determined a year ago [50], cost around $70 million [51]. Resequencing a human genome with the Sanger sequencing method would today cost approximately $10 million [14,50] while the 454 system enables a 10-fold reduction in cost and about 20-fold reduction in time [52]. Illumina claims to be able to sequence a human genome with the 1G Analyzer for approximately $100,000 [45]. Neither 454 nor Illumina has shown data describing the exact workload and reagent cost but this important information will hopefully be revealed to the scientific community soon. Although the progresses in the last few years have shown a significant reduction in sequencing cost, it is still too early and too expensive to use these platforms to routinely sequence human genomes at a larger scale. The realization of the $1000 genome requires novel approaches and there is an immense activity in the field.

As mentioned above, in 2004 NHGRI initiated the "Advanced Sequencing Technology Development Projects" where grants were approved for some 20 novel ideas and approaches to develop cutting-edge, low cost sequencing for the future. Today around 35 projects in industry and academia have been granted a total of $56 million for technology development in the quest for the $1000 genome [53]. A key feature among most contenders is to look at single molecules. Although it is challenging to sequence single DNA fragments, there are advantages such as improved read length, since molecules are not getting out of phase, and a significant drop in reagent cost. A number of routes to the future are pursued, such as sequencing-by-synthesis approaches like 454 and Solexa, without the prior amplification step, and indirect approaches using physical recognition of the DNA strand and the investigation of bases using nano pores or equivalents.

Single-molecule sequencing

The concept of sequencing-by-synthesis without a prior amplification step i.e. single-molecule sequencing is currently pursued by a number of companies. Helicos Biosciences [54] has an instrument, the HeliScope™, with a claimed throughput of 1.1 Gpb per day (as of October 2008 [54]). Single fragments are labeled with Cy3 for localization of template strands on an array and a predefined, Cy5-labeled nucleotide (for instance "A") are incorporated, detected by a fluorescent microscope and cleaved off in each cycle [55,56]. Four

cycles, one for each nucleotide, constitutes a "quad" and multiple quad runs are claimed to produce read lengths of up to ~55 bases (see Fig. 2A). At 20 bases or longer, 86% of the strands are available and at 30 bases, around 50%. The first order of a HeliScope instrument was announced in the beginning of February 2008 and the company claims its machine to be able to sequence a human genome for $72,000.

A different, although very promising, approach is taken by Pacific Biosciences [57]. The technology, denoted Single-molecule Real Time Sequencing-by-synthesis (SMRT™), has in a proof-of-concept study shown read lengths of single DNA fragments of over 1500 bases in 3000 parallel reactions. The heart of the technology is so called zero-mode waveguides (ZMW) [58] which essentially are nanometer scale wells with a diameter of 70 nm (see Fig. 2B). Light bulges inward at the opening, permitting illumination of a detection volume of 20 zl (10^{-21}) where a single DNA polymerase is immobilized. Nucleotides, fluorescently labeled at the terminal phosphate, are incorporated by the polymerase and thereby exposing its base-specific fluorophore for a few milliseconds which is enough for detection. Benefits are long read lengths of thousands of bases in one stretch and high speed (10 bases per second and molecule). It is still at the proof-of-concept stage and no commercial instrument is ready. Thousands of ZMWs in parallel may in a future instrument (no sooner than 2010) generate 100 gigabases per hour. A second generation instrument capable of sequencing a human genome for $1000 is an additional number of years in the future. The Menlo Park based company has received grants from the NHGRI but is not signed up for the X PRIZE race so far.

Unlike Pacific Biosciences, a contender in the X PRIZE is Visigen Biotechnologies [59], Houston, TX, which platform consists of an engineered polymerase and modified nucleotides for single-molecule detection. An immobilized polymerase on a surface, modified with a fluorescence resonance energy transfer (FRET) donor incorporates nucleotides modified with different acceptors, allowing base-specific and real time detection of incorporation events (see Fig. 2C). A theoretical throughput of 1 million bases per instrument second has

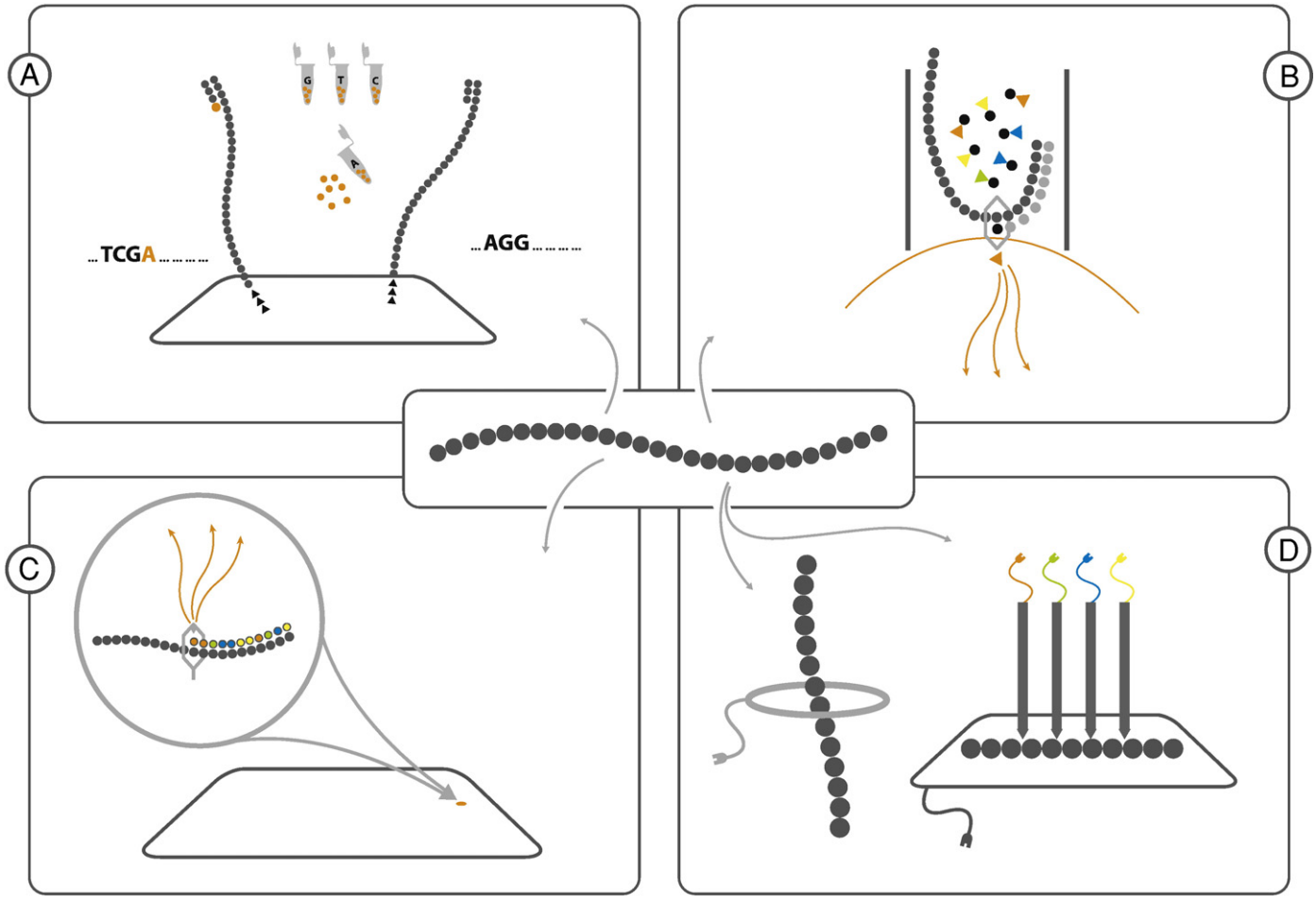

Fig. 2. (A) Helicos Biosciences. True single-molecule sequencing (tSMS™) is achieved by initially adding a poly A sequence to the 3′-end of each fragment, which allows hybridization to complementary poly T sequences in a flow cell. After hybridization, the poly T sequence is extended and a complementary sequence is generated. In addition, the template is fluorescently labeled at the 3′-end and thus, illumination of the surface reveals the location of each hybridized template. This process allows generation of a map of the single-molecule landscape before the labeled template is removed. Fluorescently labeled nucleotides are added, one in each cycle, followed by imaging. A cleavage step removes the fluorophore and permits nucleotide incorporation in the next round. (B) Pacific Biosciences. A zero-mode waveguide contains a single polymerase macromolecule immobilized at the bottom (hexagon), nucleotides (circles) that are fluorescently labeled at the triphosphates (colored triangles) and a DNA strand which permits single-molecule real time (SMRT™) sequencing. Single incorporation events are possible to detect with this design since an excitation beam penetrates the lower 20–30 nm of the waveguide, i.e. approximately a volume of 20 zl. This volume is sufficient to detect the incorporated nucleotide while avoiding excitation of unincorporated nucleotides, thereby reducing the noise. (C) Visigen Biotechnologies. A slightly different approach for single-molecule real time sequencing is to immobilize a modified polymerase (hexagon) on a glass surface. The polymerase is engineered to carry a fluorescent donor molecule and by coding the four different nucleotides with different acceptors, base-specific FRET emission upon incorporation will reveal the sequence. (D) Nano Pores and Nano-Knife Edge Probes. Nano pores and nano-knife edge probes are two approaches for physical and direct recognition of bases. Bases in a DNA strand can be recognized either by threading through a nano pore (left), measuring a change in conductivity, or using an array of nano-knife edge probes (right) tuned to recognize each base in a stretched, immobilized DNA strand by detecting the unique electron tunneling characteristics for each base.

been given, although no proof-of-concept study has been presented. Applied Biosystems has completed an equity investment as of December 2005.

Intelligent Biosystems [60] is pursuing an array-based sequencing-by-synthesis approach [61] similar to the Illumina 1G system and claims launch of an instrument by the end of 2008 that might reduce the cost of a genome to $10,000. The company received a grant in 2006 by NHGRI and is not competing for the X PRIZE.

Further in the future may lie sequencing approaches that utilize physical recognition of nucleic bases. One alternative is nano pores, where the aim is to sequence a DNA strand that is pulled electrophoretically through a synthetic or natural pore, only 1.5 nm wide, measuring changes in conductivity (See Fig. 1D). A common issue with nano pores is the sensitivity of detection. By utilizing conversion of single bases to longer Design Polymers by LingVitae [62] such problems may be circumvented.

Reveo [63], an X PRIZE contender, is developing a Personal Genome Sequencer (PGS) based on nano-knife edges permitting non-destructive detection of bases in single DNA strands by measuring electron tunneling characteristics for each base [23] (see Fig. 2D).

Remarks

Several X PRIZE attempts can be anticipated within a year and the $1000 genome may be as little as three years away.

One's destiny is a result of genes, environment, life style, behavior and luck. As we are entering an era of low-cost sequencing and preventive medicine one can discover a new meaning to the ancient inscription at the temple of Apollo. Γνωθεις εαυτον — Know thyself.

References

[1] J.C. Venter, K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D. Wu, I. Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.H. Rogers, H.O. Smith, Environmental genome shotgun sequencing of the Sargasso Sea, Science 304 (2004) 66–74.

[2] S. Yooseph, G. Sutton, D.B. Rusch, A.L. Halpern, S.J. Williamson, K. Remington, J.A. Eisen, K.B. Heidelberg, G. Manning, W. Li, L. Jaroszewski, P. Cieplak, C.S. Miller, H. Li, S.T. Mashiyama, M.P. Joachimiak, C. van Belle, J.M. Chandonia, D.A. Soergel, Y. Zhai, K. Natarajan, S. Lee, B.J. Raphael, V. Bafna, R. Friedman, S.E. Brenner, A. Godzik, D. Eisenberg, J.E. Dixon, S.S. Taylor, R.L. Strausberg, M. Frazier, J.C. Venter, The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families, PLoS Biol. 5 (2007) e16.

[3] D.B. Rusch, A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J.A. Eisen, J.M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J.E. Venter, K. Li, S. Kravitz, J.F. Heidelberg, T. Utterback, Y.H. Rogers, L.I. Falcon, V. Souza, G. Bonilla-Rosso, L.E. Eguiarte, D.M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M.R. Ferrari, R.L. Strausberg, K. Nealson, R. Friedman, M. Frazier, J.C. Venter, The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific, PLoS Biol. 5 (2007) e77.

[4] http://www.syntheticgenomics.com.

[5] http://www.ls9.com.

[6] http://www.amyrisbiotechnologies.com.

[7] http://www.aurorabiofuels.com.

[8] R.F. Service, Gene sequencing. The race for the $1000 genome, Science 311 (2006) 1544–1546.

[9] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.

[10] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.

[11] http://www.genome.gov/12513162.

[12] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380.

[13] http://www.454.com.

[14] D.R. Bentley, Whole-genome re-sequencing, Curr. Opin. Genet. Dev. 16 (2006) 545–552.

[15] http://www.illumina.com.

[16] http://solid.appliedbiosystems.com.

[17] http://www.personalgenomes.org.

[18] G.M. Church, The personal genome project, Mol. Syst. Biol. 1 (2005) 2005 0030.

[19] G.M. Church, Genomes for all, Sci. Am. 294 (2006) 46–54.

[20] http://www.genomics.org.cn.

[21] http://www.1000genomes.com.

[22] http://www.knome.com.

[23] http://www.genomics.xprize.org.

[24] P. Nyren, The history of pyrosequencing, Methods Mol. Biol. 373 (2007) 1–14.

[25] A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.

[26] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O, Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, Nature 448 (2007) 553–560.

[27] J. Shendure, R.D. Mitra, C. Varma, G.M. Church, Advanced sequencing technologies: methods and goals, Nat. Rev. Genet. 5 (2004) 335–344.

[28] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463–5467.

[29] J.M. Prober, G.L. Trainor, R.J. Dam, F.W. Hobbs, C.W. Robertson, R.J. Zagursky, A.J. Cocuzza, M.A. Jensen, K. Baumeister, A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides, Science 238 (1987) 336–341.

[30] L.M. Smith, J.Z. Sanders, R.J. Kaiser, P. Hughes, C. Dodd, C.R. Connell, C. Heiner, S.B. Kent, L.E. Hood, Fluorescence detection in automated DNA sequence analysis, Nature 321 (1986) 674–679.

[31] M.A. Innis, K.B. Myambo, D.H. Gelfand, M.A. Brow, DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 9436–9440.

[32] A.M. Carothers, G. Urlaub, J. Mucha, D. Grunberger, and L.A. Chasin, Point mutation analysis in a mammalian gene: rapid preparation of total RNA, PCR amplification of cDNA, and Taq sequencing by a novel method. Biotechniques 7 (1989) 494–496, 498–499.

[33] C.A. Emrich, H. Tian, I.L. Medintz, R.A. Mathies, Microfabricated 384-lane capillary array electrophoresis bioanalyzer for ultrahigh-throughput genetic analysis, Anal. Chem. 74 (2002) 5076–5083.

[34] L. Koutny, D. Schmalzing, O. Salas-Solano, S. El-Difrawy, A. Adourian, S. Buonocore, K. Abbey, P. McEwan, P. Matsudaira, D. Ehrlich, Eight hundred-base sequencing in a microfabricated electrophoretic device, Anal. Chem. 72 (2000) 3388–3391.

[35] D.J. Cutler, M.E. Zwick, M.M. Carrasquillo, C.T. Yohn, K.P. Tobin, C. Kashuk, D.J. Mathews, N.A. Shah, E.E. Eichler, J.A. Warrington, A. Chakravarti, High-throughput variation detection and genotyping using microarrays, Genome Res. 11 (2001) 1913–1925.

[36] R.J. Lipshutz, D. Morris, M. Chee, E. Hubbell, M.J. Kozal, N. Shah, N. Shen, R. Yang, S. P. Fodor, Using oligonucleotide probe arrays to access genetic diversity, Biotechniques 19 (1995) 442–447.

[37] N. Patil, A.J. Berno, D.A. Hinds, W.A. Barrett, J.M. Doshi, C.R. Hacker, C.R. Kautzer, D. H. Lee, C. Marjoribanks, D.P. McDonough, B.T. Nguyen, M.C. Norris, J.B. Sheehan, N. Shen, D. Stern, R.P. Stokowski, D.J. Thomas, M.O. Trulson, K.R. Vyas, K.A. Frazer, S.P. Fodor, D.R. Cox, Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21, Science 294 (2001) 1719–1723.

[38] R. Drmanac, I. Labat, I. Brukner, R. Crkvenjakov, Sequencing of megabase plus DNA by hybridization: theory of the method, Genomics 4 (1989) 114–128.

[39] A. Pihlak, G. Bauren, E. Hersoug, P. Lonnerberg, A. Metsis, S. Linnarsson, Rapid genome sequencing with short universal tiling probes, Nat. Biotechnol. 26 (2008) 676–684.

[40] J. Shendure, G.J. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M. D. Wang, K. Zhang, R.D. Mitra, G.M. Church, Accurate multiplex polony sequencing of an evolved bacterial genome, Science 309 (2005) 1728–1732.
[41] M. Ronaghi, M. Uhlen, P. Nyren, A sequencing method based on real-time pyrophosphate, Science 281 (1998) 363, 365.
[42] A. Gutin, Myriad Genetics Inc. Genetic Testing for Cancer Predisposition: Opportunities Offered by Next Generation Sequencing; Exploring Next Generation Sequencing: Applications and Case Studies. October 17–18, 2007. Renaissance Providence Hotel. Providence, Rhode Island.
[43] U.S., Patent 5,641,658.
[44] C. Adessi, G. Matton, G. Ayala, G. Turcatti, J.J. Mermod, P. Mayer, E. Kawashima, Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms, Nucleic Acids Res. 28 (2000) E87.
[45] Illumina, Analyst Day, September 15th 2007, Mandarin Oriental, New York, NY.
[46] J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D.H. Kim, M. Sano Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J.J. Russo, N.J. Turro, J. Ju, Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 9145–9150.
[47] S. Brenner, M. Johnson, J. Bridgham, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S.R. Williams, K. Moon, T. Burcham, M. Pallas, R.B. DuBridge, J. Kirchner, K. Fearon, J. Mao, K. Corcoran, Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, Nat. Biotechnol. 18 (2000) 630–634.
[48] S. Brenner, S.R. Williams, E.H. Vermaas, T. Storck, K. Moon, C. McCollum, J.I. Mao, S. Luo, J.J. Kirchner, S. Eletr, R.B. DuBridge, T. Burcham, G. Albrecht, In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 1665–1670.
[49] http://www.polonator.org.
[50] S. Levy, G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, N. Axelrod, J. Huang, E. F. Kirkness, G. Denisov, Y. Lin, J.R. Macdonald, A.W. Pang, M. Shago, T.B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S.A. Kravitz, D.A. Busam, K.Y. Beeson, T.C. McIntosh, K.A. Remington, J.F. Abril, J. Gill, J. Borman, Y.H. Rogers, M.E. Frazier, S.W. Scherer, R.L. Strausberg, J.C. Venter, The diploid genome sequence of an individual human, PLoS Biol. 5 (2007) e254.
[51] http://digitaldaily.allthingsd.com/20071019/web-20-summit-j-craig-venter-bio-engineer/.
[52] http://www.iht.com/articles/2007/06/01/america/dna.php.
[53] http://www.nytimes.com/2008/02/09/business/09genome.html.
[54] http://www.helicosbio.com.
[55] I. Braslavsky, B. Hebert, E. Kartalov, S.R. Quake, Sequence information can be obtained from single DNA molecules, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3960–3964.
[56] T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, Z. Xie, Single-molecule DNA sequencing of a viral genome, Science 320 (2008) 106–109.
[57] http://www.pacificbiosciences.com.
[58] M.J. Levene, J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb, Zero-mode waveguides for single-molecule analysis at high concentrations, Science 299 (2003) 682–686.
[59] http://www.visigenbio.com.
[60] http://www.intelligentbiosystems.com/.
[61] J. Ju, D.H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M.S. Marma, S. Shi, J. Wu, J.R. Edwards, A. Romu, N.J. Turro, Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 19635–19640.
[62] http://www.lingvitae.com.
[63] http://reveo.com.