J Med Biochem 2013; 32 (4) DOI: 10.2478/jomb-2014-0004

UDK 577.1 : 61 **ISSN 1452-8258**

J Med Biochem 32: 301–312, 2013

Review article
Pregledni članak

THE HISTORY OF DNA SEQUENCING

ISTORIJAT SEKVENCIRANJA DNK

Miodrag Gužvić[1,2]

[1]*Chair of Experimental Medicine and Therapy Research, University of Regensburg, Regensburg, Germany*
[2]*European University – Faculty of Pharmacy, Novi Sad, Serbia*

Summary

During the last decade, the cost of DNA sequencing technologies has decreased several orders of magnitude, with the proportional increase of speed and throughput. Human Genome Project took almost 15 years to complete the sequence of the human genome. With the second and third generation technologies, this can be done in the matter of days or hours. This progress and availability of sequencing instruments to virtually every researcher leads to replacing of many techniques with DNA sequencing and opens new venues of research. DNA sequencing is used to investigate basic biological phenomena, and is probably going to be increasingly used in the context of health care (preimplantation diagnostics, oncology, infectious diseases). Current trends are aiming towards the price of 1000$ for sequencing of one human genome. Without any doubt, we can expect improvement of existing and the development of fourth generation technologies in the coming years.

Keywords: DNA sequencing; genome; molecular biology

Kratak sadržaj

Tokom poslednjih 10 godina, cena sekvenciranja DNK se drastično smanjila, dok se, suprotno tome, brzina sekvenciranja i veličina genoma koja se može sekvencirati u jedinici vremena, povećala. Sekvenciranje kompletnog humanog genoma u okviru projekta »Humani Genom« trajalo je skoro 15 godina. Sa današnjim stepenom razvoja, tehnologijama sekvenciranja druge i treće generacije, isti projekat bio bi realizovan za nekoliko dana, pa čak i sati. Napredak u tehnologiji sekvenciranja i dostupnost novih tipova instrumenata za sekvenciranje gotovo svakoj laboratoriji rezultirala je mnogo masovnijom upotrebom ove metodologije u analizi DNK kao i otvaranjem novih pravaca u istraživanjima. Sekvenciranje DNK se koristi za istraživanje osnovnih bioloških fenomena, i verovatno će sve više nalaziti upotrebu u zaštiti zdravlja (preimplantaciona dijagnostika, onkologija, infektivne bolesti). Trenutno, osnovni trend u ovoj oblasti je obaranje cene sekvenciranja po pojedinačnom humanom genomu na 1000$. Nema nikakve sumnje da će se u narednim godinama pojaviti tehnologije četvrte generacije.

Ključne reči: sekvenciranje DNK; genom; molekularna biologija

Address for correspondence:

Miodrag Gužvić
Lehrstuhl für Experimentelle Medizin und Therapieverfahren
Universität Regensburg
Franz-Josef-Strauss-Allee 11
93053 Regensburg, Germany
Fax: +49 941 944 67 19
e-mail: miodrag.guzvic@klinik.uni-regensburg.de

List of abbreviations: bp, base pair; ds, double strand(ed); NGS, next generation sequencing; PAA, polyacrylamide; SBH, sequencing by hybridization; SBL, sequencing by ligation; SBS, sequencing by synthesis; SMS, single molecule sequencing; SNP, single nucleotide polymorphism; ss, single strand(ed).

Introduction

»....sequencing DNA now is one of the easiest jobs you could have besides sloppin' burgers« Kary Mullis (1)

During the last decade we have witnessed rapid development of new sequencing technologies and their applications. Interestingly enough, this progress practically started after the completion of the sequencing of the human genome. The first sequence of the human genome was obtained using so called »first generation« sequencing technology. In the following years, »second« or »next generation« sequencing (NGS) technologies were developed, characterized by massive parallelization, improved automation and speed, and, most importantly, greatly reduced price. While »next generation« approaches were being improved, the »third« or »next-next generation« technologies appeared, this time characterized by single-molecule sequencing (SMS). Very often, expressions like »deep sequencing«, »massive parallel sequencing« (MPS), »high throughput sequencing« (HTS or HT-Seq) are used for »second« and »third« generation technologies. Some of these technologies were abandoned, some are still being developed or improved, while other are now routinely used in the laboratories around the world. The aim of this article is to provide overview of the development and description of key sequencing technologies, as well as the explanation of key terms.

First generation sequencing

The first two widely-known methods for DNA sequencing appeared in 1977. One, known as »chemical cleavage« sequencing, was published in February by Maxam and Gilbert (2). The second, known as »chain terminator« sequencing, »dideoxy« sequencing or »Sanger« sequencing, was published in December by Sanger and collaborators (3). Interestingly, these works were culmination of both groups' active work on determination of the sequence of nucleic acids (4–6). Both approaches rely on the separation of the mixture of DNA fragments of various sizes on polyacrylamide (PAA) slab gels.

»Chemical cleavage« sequencing

In this approach (2), DNA fragment that is to be sequenced could be obtained by, for example, restriction digestion of plasmid and subsequent electrophoretic separation and isolation of desired fragment. Fragment is 5' or 3' terminally labelled with ^{32}P using alkaline phosphatase and polynucleotide kinase (5' labelling) or terminal transferase (3' labelling). Since both 5' or 3' termini are labelled, double stranded fragment is denatured, single strands separated on PAA gel and isolated, thereby enabling sequencing of each strand (*Figure 1.A*).

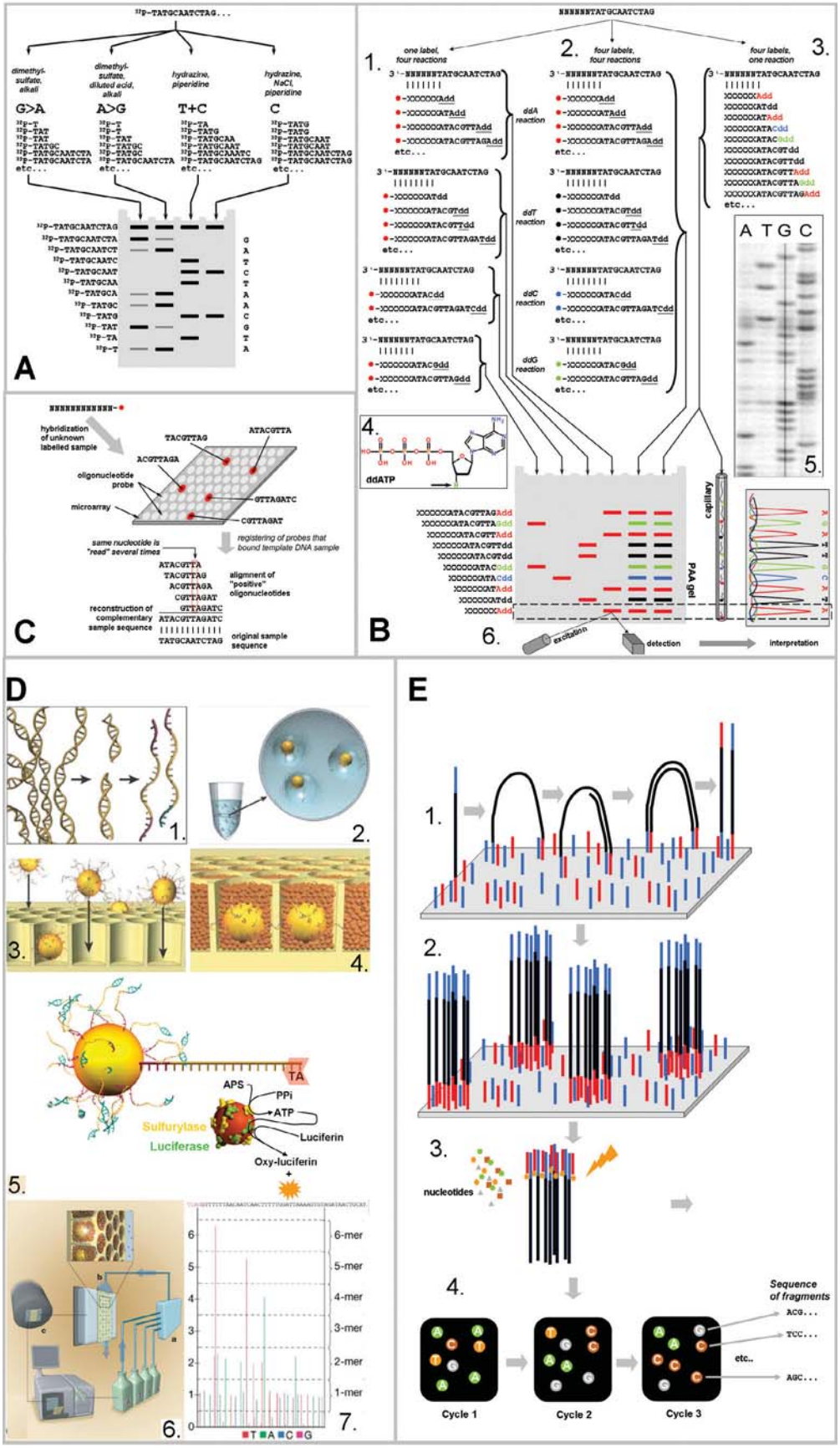
For each strand, aliquots of the four cleavage reactions are prepared. Single-stranded (ss)DNA is first chemically modified, causing creation of abasic site(s), and exposing the backbone that is subsequently easily broken by alkali treatment. For chemical modification, base- or pyrimidine-/purine-specific reagents are used. Chemical reactions are optimized to ensure random modification and cleavage along the ssDNA (i.e. the conditions are mild), thereby generating a mixture of single-stranded fragments of various sizes. Finally, cleaved ^{32}P-ssDNA is separated on denaturing PAA gel. The sequence is read from an autoradiograph of the gel. Each fragment ends with the site of cleaved base, which is being read from bottom to the top of the gel. In the original publication, each set of four cleavage reactions was run at least twice on the PAA gel. One run was short, enabling reading of short fragments, while additional, longer run, enabled reading of longer fragments. The length of sequenced fragment depended on the size of the PAA gel, and authors reported that they could sequence fragments up to 100 bp.

Although the authors have devised ways to perform base-specific cleavage, the final protocol suggests using conditions that modify two bases (albeit not with the same preference). This approach generates redundant fragments in different chemical cleavage reactions, which serves as in-experiment control. For example, dimethyl-sulfate causes methylation of adenine (A) and guanine (G), with G being methylated faster. Methylation destabilizes N-glycosidic bond, and the DNA strand can be broken upon treatment with alkali. After separation on the PAA gel, a pattern of dark and light bands can be seen. Dark bands arise from breakage of G, and light bands from breakage of A. However, gentle treatment of methylated DNA with diluted acid releases preferentially A, leading to appearance of easily distinguishable dark A bands and light G bands. Similarly, treatment with hydrazine and piperidine releases thymines (T) and cytosines (C). Alternatively, the addition of NaCl into the reaction would cause release of C only (*Figure 1.A*).

This sequencing approach was not widely used, due to two major drawbacks. One is the use of hazardous chemicals (e.g. hydrazine is neurotoxin). The other is the fact that a lot of DNA is needed, which, with the need to have four or even eight reactions, could pose significant problem. Still, this approach could be a method of choice for sequencing of very short fragments, or when design of adequate primer(s) for sequencing is problematic.

»Sanger« sequencing

In 1977. the team of F. Sanger published the almost complete sequence of the genome of bacteriophage ΦX174 (7), using previously developed »plus and minus« sequencing (8). Later that year, sim-



A »Chemical cleavage« sequencing. Each template is aliquoted in 4 reactions. Each reaction is chemically treated in such way to cause breaking of DNA strand where particular base is located. Mild conditions of chemical treatment ensure that a range of different length fragments is generated. Electrophoretic separation of four reactions in adjacent lanes of PAA gel enables reading of the sequence. B. »Sanger« sequencing. 1), 2) and 3) depict different possibilities of labelling. If one label is used, four reactions are separated in adjacent lanes of PAA gel and sequence determined by »reading« of the gel. Picture 5) shows the example of such gel. If four labels are used 3), it is possible to run fragments labelled with different dyes in the same lane of the PAA gel or in the 6) capillary. Picture 4) shows example of ddNTP (ddATP). C. Sequencing by hybridization. Labelled template DNA is hybridized on microarray, spotted with all possible octamer oligonucleotides. Knowing the sequence of the overlapping octamers that anneal to the template, it is possible to align them and reconstruct the sequence of the template. D. »454« sequencing. 1) Template DNA is fragmented and adapters are attached. After generation of ssDNA library, single strands are attached to the beads. 2) Beads are enclosed in PCR microreactors in water/oil emulsion, thus enabling clonal amplification of bound fragment. 3), 4) Beads carrying clonally amplified template ssDNA are deposited in the wells of the PicoTiterPlate, together with beads carrying enzymes needed for pyrosequencing, depicted on Figure 5). 6) PicoTiterPlate is placed within flow chamber, attached to the optical detection system. 7) Incorporation of nucleotides is measured as light signal, and intensity of the signal corresponds to the number of incorporated nucleotides. E. »Illumina« sequencing. 1) Fragments with attached adapters are ligated to the surface of the flow cells. Flow cell is coated with short oligonucleotides, complementary to the adapter sequences. This enables annealing of the free terminus to the oligonucleotide and creation of so-called bridge. By the process of bridge-amplification, the second strand is synthesized. 2) This cycle is repeated, until thousands of newly synthesized strands have created a PCR colony (polony). 3) This cluster is then used for sequencing. Incorporation of labelled nucleotide is registered as fluorescent image of the flow cells. 4) These images are recorded during each successive cycle of nucleotide flushing. *In silico* superimposition of these images reconstructs the sequence of each fragment that generated cluster. All images created by M. Gužvić, except D (from (43), distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence).

Figure 1 Overview of DNA sequencing approaches, part I.

pler and more rapid »chain terminator« sequencing method was published (3), and, shortly after, the complete sequence of 5,368 bp long ΦX174 genome was determined using the new method (9), representing the first DNA genome ever sequenced.

Unlike »chemical cleavage« method, in Sanger's approach DNA was synthesized using DNA polymerase I (Klenow fragment)-mediated extension of oligonucleotide primer, with ssDNA being a template. Single-stranded DNA was divided in 4 aliquots. Each reaction contained DNA polymerase, a primer, and mixture of four deoxyribonucleotide-triphosphates (dNTPs), of which three (e.g. ^{32}P-dATP, dCTP, dGTP) were in equimolar concentration and fourth (e.g. dTTP) in 20 times lower concentration. In addition, the 2',3'-dideoxy analogue of the fourth dNTP (e.g. ddTTP) was added in excess. Since ddNTPs lack 3'-OH group, the synthesis of DNA chain is terminated upon incorporation of ddNTP (hence the name of the method; *Figure 1.B.4*). The presence of both dNTP and ddNTP enables synthesis of chains of different lengths (*Figure 1.B.1,6*). Separation of four reactions, containing denatured ^{32}P-ssDNA, on the adjacent lanes of denaturing PAA gel and subsequent analysis of autoradiograph enabled reading of the sequence, similar to Maxam-Gilbert method. Initially, authors were able to read the sequence of up to 300 bp.

This method also had its difficulties. Short mononucleotide repeats were hard to interpret, as the bands were often not separated and correct reading of the sequence mostly depended on the experience of the operator. Furthermore, sometimes artefacts appeared on the PAA gel, most likely originating from the conformers of the ssDNA. These difficulties were alleviated by introduction of thinner PAA gels (10), enabling runs at higher voltages (and thereby generating higher temperatures) and requiring less material, finally resulting in sharper bands. Later introduction of silver staining of PAA gels (11–13) and fluorescent dyes (see below) eliminated the use of radioactive isotopes.

Due to its simplicity and reliability, »chain terminator« sequencing quickly became dominant method for sequencing. Its automated adaptation was used in all major sequencing enterprises, as, for example, sequencing of human mitochondrial genome (14), the first bacterial genome, *Haemophilus influenzae* (15), the first eukaryotic genome, *Saccharomyces cerevisiae* (16), the first multicellular animal genome, *Caenorhabditis elegans* (17), the first plant genome, *Arabidopsis thaliana* (18), and culminating with the sequencing of genomes of *Drosophila melanogaster* (19, 20) and human (21–24). Even nowadays, automated »chain terminator« sequencing is frequently used for small-scale targeted resequencing.

Automation of »Sanger« sequencing

As mentioned before, the potential of »Sanger« sequencing was quickly realized by the research community. This motivated a number of researchers to work on automation of the procedure. The most prominent were L. Hood and M. Hunkapiller, associated with Applied Biosystems, Inc. (ABI). In the first report (25), they offered two major improvements – labelling of DNA with fluorescent dyes and computer-based acquisition and analysis of the sequence data. The instrument that was developed (and that would soon become model ABI 370) contained PAA gel with 16 lanes. Instead of radioactive isotopes, DNA primers were labelled with fluorescent dyes. Initially, one dye was used for all reactions, which required running of four reactions in adjacent lanes of the gel (*Figure 1.B.5,6*). Later, different dye for each of the four sequencing reactions was used (*Figure 1.B.2,6*). This enabled running all reactions in one lane of the gel, thus greatly improving throughput. Excited by laser, the fluorescence was detected, in real time, close to the bottom of the gel. In both »one dye« and »four dyes« configuration, four electropherograms or chromatograms were obtained, that, when superimposed, produced the sequence of the analyzed fragment (*Figure 1.B.6*). Using one dye, authors could sequence around 400 bp, and with four dyes around 200 bp. In order to ensure the reliability of sequence determination, authors suggested sequencing of both strands.

The problem with »one dye« configuration was that different mobility of bands in adjacent lanes made it difficult for computer to assemble the sequence (25). This problem was partially alleviated by use of four dyes and running the reactions in the same lane. Some mobility shifts between different dyes (1 – 1.25 bp) were observed, and were later compensated by adding linkers between the dye and the primer, and by modification of the analysis software (25). Further improvements of the automation came from introducing of dye-labelled ddNTPs (26), which made possible to perform four reactions in one tube (*Figure 1.B.3*), and from introducing capillaries (27) instead of slab gels (*Figure 1.B.6*).

Over the next two decades, ABI sequencers were significantly improved (28–33). The number of lanes in PAA gel-based models increased from 16 (ABI 370A) to 96 (ABI 377). The first capillary sequencer was model ABI 310, with one capillary, and the model 3730xl had 96 capillaries. At the same time, the length of the reads increased from 350 (ABI 370A) to over 900 (ABI 3730xl), while the run times decreased from 18h to 3h. The introduction of model 3730 increased the speed of the sequencing of the human genome.

The major drawbacks are the price per base and the problems related to cloning and sequencing of regions containing repetitive sequences (34). Interestingly, in the dawn of »next generation« sequencing, scientist from the J. Craig Venter Institute used automated »dideoxy« sequencing to sequence the diploid genome of a human individual (J.C. Venter himself), using the ABI sequencer 3730xl (24), most

likely the last human genome ever to be sequenced using »Sanger« sequencing.

Sequencing by hybridization (SBH)

SBH can be considered as a sequencing approach between »first« and »second« generation, because the core of the idea included massive parallelization. It was developed by a team of Serbian scientists in the early 1990's (35, 36). The idea was to determine the oligonucleotide content of the template DNA chain by hybridizing it with short (5-25 bp) oligonucleotides (37). By aligning of overlapping oligonucleotides (of known sequence) that hybridized with the template DNA, one could reconstruct its sequence. The method was envisaged to have two formats: first, where the template DNA is immobilized on solid surface, while labelled short oligonucleotides of known sequence are used to query the template, and second, where short oligonucleotides of known sequence are arrayed onto solid surface, while labelled test-DNA is offered as a probe (*Figure 1.C*). Stringent hybridization conditions are essential factor in the procedure, since no mismatches are allowed. Obviously, the read length corresponds to the length of the used oligonucleotides. Each queried base is interrogated by several overlapping oligonucleotides, providing internal control of sequencing. Using this approach, authors were able to reliably sequence fragments of 100 and 340 bp (38, 39).

Although the method was suggested to have the potential to be used for large-scale sequencing, there are some drawbacks that most likely prevented this approach to be widely used. The number of oligonucleotides required for complete sequencing depends on their length and can be very high (4^n for n-mer, e.g. for octamer $4^8 = 65356$). Furthermore, hybridization and wash conditions that retain only perfectly matched hybridizations are not easy to optimize, with, for example, end-mismatched duplexes being the particular problem. In addition, when short oligonucleotides were used to sequence longer fragments, there was a high chance of getting branching points in the final assembly, leading to multiple possible output sequences. Finally, this approach also has no good solution for dealing with repetitive sequences. On the other hand, SBH can be used for targeted resequencing of short fragments, or high-throughput typing of known mutations or SNPs (40). The potential of this approach is also illustrated by the fact that SBH inspired development of novel sequencing methodologies in the last few years (41, 42).

Second generation sequencing

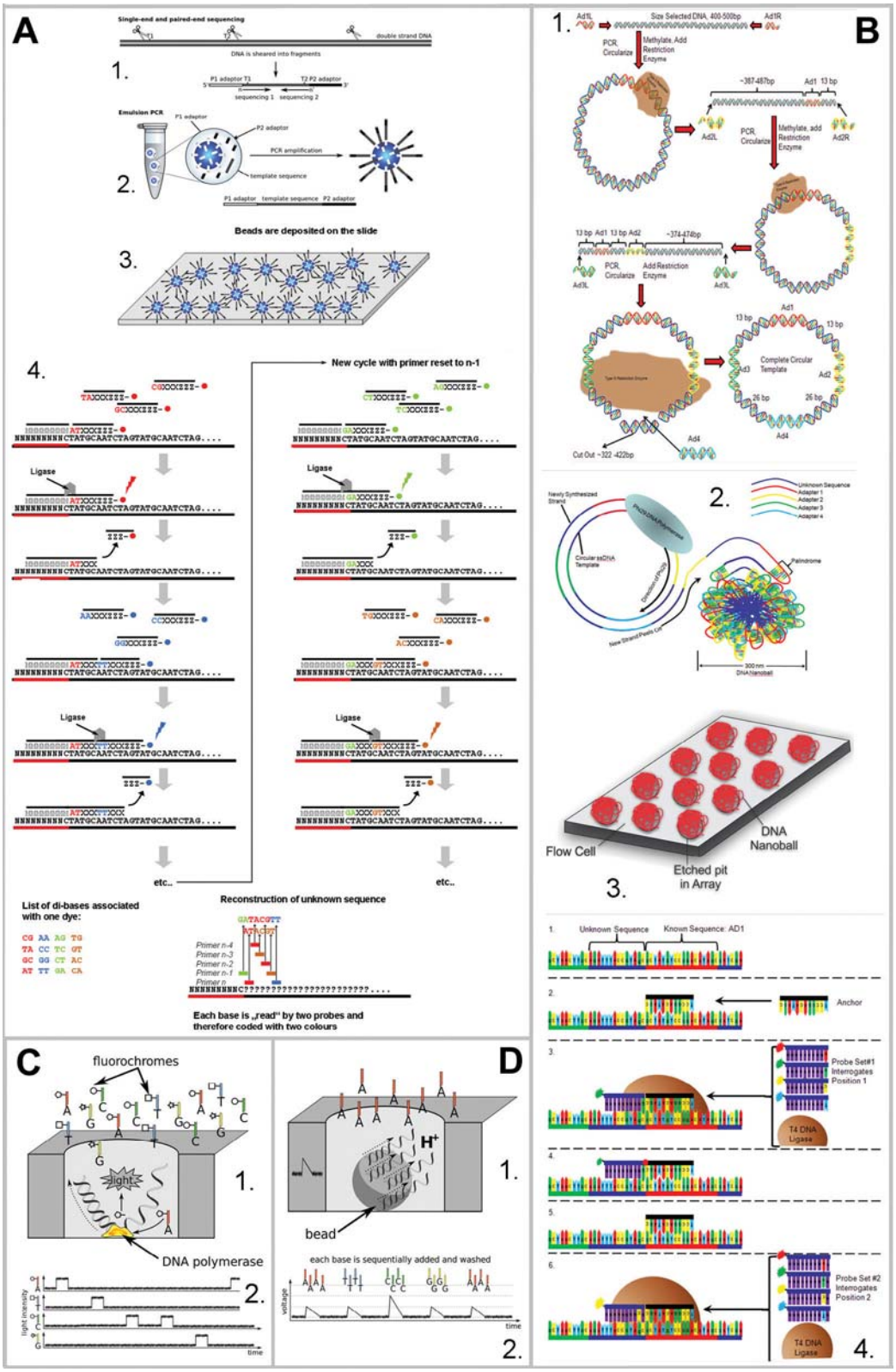
The key feature of »next generation« sequencing methodologies is parallelization of high number of sequencing reactions. This was achieved by miniaturization of sequencing reactions. In addition, detection systems were greatly improved. The time needed to determine the Gbp-sized sequences was reduced to hours or days, with the accompanying price reduction, compared to »Sanger« sequencing. It all started in 2005. with the report from 454 Life sciences.

»454« sequencing

»454« sequencing by synthesis (SBS) was developed by 454 Life Sciences, which was later acquired by Roche. This technology is used in Genome Sequencer series of instruments. »454« sequencing (43) relies on pyrosequencing (44–46), miniaturized and massively parallelized using PicoTiterPlates (47).

Pyrosequencing (*Figure 1.D.5*) is based on the detection of light signal upon incorporation of nucleotide by polymerase. Incorporation of dNTP releases pyrophosphate (PPi). In the presence of ATP sulfurylase and adenosine 5'-phosphosulfate (APS), PPi is converted to ATP. Luciferase-ATP mediates conversion of luciferin to oxyluciferin, which generates a light signal. Amount of light is proportional to the number of ATP molecules, which is proportional to the number of released pyrophosphates, or, in other words, incorporated nucleotides.

The first step in sample preparation is mechanical fragmentation (nebulization) of DNA (*Figure 1.D.1*). Next, double-stranded adapters, »A« and »B«, are ligated to the termini of each fragment (*Figure 1.D.1*). Adapter »A« contains binding site for the sequencing primer. One strand of adapter »B« is biotinylated and serves to attach the fragment on streptavidin-beads. This mixture of fragments is incubated with streptavidin-beads that bind all fragments that contain adapter »B«. Fragments that are ligated to two »A« adapters are lost upon first washing step, while fragments that are ligated to two »B« adapters bind beads with both termini. In the next step, DNA on the beads is denatured, unbiotinylated strands are released from the beads and collected, thus generating ssDNA library. This library is then mixed with the excess of capture beads coated with oligonucleotides complementary to adapter »B«, aiming to bind maximally one fragment per bead. Then, beads are mixed with water/oil emulsion, containing PCR reagents and primers complementary to adapters, thus generating thousands discrete PCR microreactors (48), small droplets containing only one bead (*Figure 1.D.2*). This mixture is then subjected to thermal cycling (emulsion PCR, emPCR), which leads to clonal amplification of the fragment bound to the capture bead. Each newly generated strand is, upon denaturation during PCR, bound by free oligonucleotides of the capture bead. This procedure produces millions of new DNA molecules. Once the clonal amplification is completed, microreactors are broken, and DNA-containing beads are enriched and subjected to denaturation that leaves only ssDNA bound to beads. In the



A. SOLiD sequencing. 1) DNA template is fragmented and adapters attached to its termini. 2) Fragments are attached to the beads and, similar to »454« technology, clonally amplified using emulsion PCR. 3) Beads carrying clonally enriched fragments are deposited onto the glass slide. 4) Sequencing is performed by successive addition of pools of four probes, with each probe interrogating two bases. Probes annealing to the template are ligated. Please see the text for details. B. DNA nanoball sequencing. 1) Four adapters are introduced into the template fragment by successive steps of circularization and opening of the circular molecule. 2) Using rolling circle amplification, a concatemer of a single stranded template is made. This long molecule is coiled into the DNA nanoball, assisted by interactions of adapters containing palindromic sequences. 3) Nanoballs are deposited into the pits of a microarray. 4) Sequencing by ligation takes place in 5′ and 3′ direction from each adapter. Probes complementary to the adapter and labelled degenerate probes interrogating one base at a time within 10 bases of the adapter-fragment boundary are sequentially added. In case of recognition, two probes are ligated, and fluorescence recorded. C. Single molecule real time sequencing. 1) DNA polymerase is attached to the bottom of the well. Single ssDNA molecule is coupled with DNA polymerase. Incorporation of fluorescently labelled nucleotides is registered in real time. D. Ion Torrent sequencing. 1) Release of H^+ during incorporation of nucleotide changes pH in the well. 2) This change is registered as electronic signal. All images created by M. Gužvić, except A.1), A.2), C, D, (author Philippe Hupé) and B (author Spencer Martin), taken from http://commons.wikimedia.org, distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence).

Figure 2 Overview of DNA sequencing approaches, part II.

final step, the sequencing primer is attached to ssDNA fragments.

These beads are then deposited into the wells of the PicoTiterPlate, one bead per well (*Figure 1.D.3*). In addition, enzyme-beads, coated with sulfurylase and luciferase, are also deposited into the wells (*Figure 1.D.4*). The plate is loaded into the instrument's flow chamber, through which the sequencing reagents flow (*Figure 1.D.6*). The bottom of the plate is in optical contact with charge-coupled device (CCD) sensor, which captures emitted light from individual wells. During sequencing, the plate is flushed with each nucleotide separately. Incorporation of the dNTPs releases PPi molecules, which are converted to light signals by sulfurylase and luciferase attached to the enzyme-beads. After incorporation of dNTPs and detection of light signals, the reagents are washed away and flow cell is treated with apyrase, which removes any remaining dNTPs. Incorporation of 2 or 3 dNTPs of the same type is registered as a stronger light signal (*Figure 1.D.7*). Depending on the format, the plate contains few hundred thousand to almost two million wells. In the first publication, the read length was more than 100 bp, and later it improved to up to 1000 bp (49). The latest instrument is claimed to be able to produce 700 million bp per 23 h run, with 99.997% accuracy.

The major drawbacks of this technology are the reagent cost and high error rates in homopolymer repeats (50), in particular that it currently does not allow efficient detection of single-base insertions or deletions in homopolymers. Sequencing of the genome of J.D. Watson (of the Watson and Crick fame) by the same group (51) was performed at the same coverage (~7.5x) like sequencing of the genome of J.C Venter (24), but, reportedly, for approximately one hundredth of the price. Furthermore, compared to the Human Genome Project that took almost 15 years, the sequencing of J. D. Watson's genome took 2 months, using three instruments.

»Illumina« sequencing

»Illumina« sequencer is able to produce larger volume of data, compared to 454 sequencer. With the read length of 100 bp, and run time of 11 days, the current instrument can produce 600 Gbp of data (52). This is great improvement, compared to the original report (53), where, while obtaining reads of 35 bp, 2 Gbp of sequence was produced.

This technology is also based on sequencing by synthesis. Using bridge amplification, polonies (PCR colonies) of clonally enriched template DNA are generated. Similar to »454« sequencing, template DNA is fragmented, and two different sets of adapters are attached to the termini of the fragments – one set needed for binding the flow cell and second needed for sequencing. These fragments are denatured and single strands are then injected into the flow cell, the surface of which is covered with the dense lawn of primers, complementary to the sequence of adapters (*Figure 1.E.1*). Template strands are attached via adapter sequence onto the surface of the flow cell. Complementarity of attached primers and adapter on the free end of single-stranded template DNA leads to creation of bridges, i.e. annealing of adapters to the primers (*Figure 1.E.1*). Upon addition of PCR reagents, the second strand is generated. Denaturation leads to dissociation of bridges, and the new annealing step creates new bridges by each strand (*Figure 1.E.1*). Repeating the process of bridge amplification, small cluster of around 1000 clonally amplified fragments is generated (*Figure 1.E.2*). Reverse strands are cleaved and washed away. Remaining clusters of single strands are templates for subsequent sequencing. The procedure involves reversible terminator nucleotides, each labelled with different fluorescent dye. This enables addition of all four dNTPs at the same time (*Figure 1.E.3*). Upon addition, complementary nucleotides are incorporated, and unincorporated dNTPs are washed away. The chemical group on dNTPs prevents extension of the chain. Next, an image of the flow cell is recorded, with fluorescent signal in those places where clusters are (*Figure 1.E.4*). Chemical treatment of the flow cell removes the fluorescent dye from incorporated nucleotides and recovers 3'-OH group. The flow cell is then flushed with the new cycle of dNTPs, that are incorporated, and image of the flow cell is recorded. By stacking and superimposing images obtained during all cycles, the software is able to reconstruct sequence of each template cluster (*Figure 1.E.4*). The described procedure is termed single-read sequencing. For paired-read sequencing, i.e. sequencing of the second strand, the original clusters are converted to dsDNA, and already sequenced strands are removed. Since different adapters are on the termini of each fragment, the choice of strand to be sequenced can be determined by the complementarity of the sequencing primer with one or the other adapter. Finally, if such need exists, one could perform mate-pair sequencing. The termini of template dsDNA are biotinylated and the fragments are circularized. Circular molecules are fragmented and biotinylated fragments (i.e. fused termini of the original template) are enriched. Adapters are then attached to these hybrid fragments, and sequencing is performed as described. This approach enables sequencing of the short termini of the 2–5 Kbp long fragments, while ignoring the portion in the middle and can be used, for example, for analysis of structural variants, e.g. translocations or inversions.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)

SOLiD is next-generation sequencer instrument, marketed by ABI. It is based on sequencing by liga-

tion (SBL). This procedure involves sequential annealing of probes to the template and their subsequent ligation, and is, therefore, related to SBH. Sample preparation is somewhat similar to that of »454« sequencing (54). Template dsDNA is fragmented, and two different adapters attached to the termini of the fragments (*Figure 2.A.1*). These fragments are then mixed with the excess of beads. Single-fragment-bound beads are mixed with PCR reagents and each bead is enclosed within a microreactor droplet in water/oil emulsion (*Figure 2.A.2*). Subsequent PCR clonally amplifies fragment on the bead. In the next step, beads are deposited onto the glass slide (*Figure 2.A.3*). Then, the bases are read by probing the beads with mixtures of 5′ fluorescently labelled octamer probes. Last two 3′ bases of the octamer are known, while the rest is degenerate. With 16 possible combinations of two bases, there is limited number of fluorescent dyes that can be used. Therefore, probes containing certain combinations of two 3′ (interrogatory) bases (di-base) are labelled with the same dye: for example, for each di-base, the reverse sequence is in the same colour, as well as the complementary di-base and its reverse sequence (e.g. AC, CA, GT and TG), and so on. The sequencing begins by adding oligonucleotide primer complementary to the adapter, with its last 3′ base annealed to the last base of adapter i.e. adapter-template junction. Then, a pool of probes labelled in the same colour is added. If the two interrogatory bases (together with 6 degenerate bases) find complementary sequence, DNA ligase ligates the probe to the primer (*Figure 2.A.4*). The fluorescent signal is recorded. Then, last three 5′ bases (including terminal base with fluorescent dye) of the probe are cleaved, followed by addition of the next pool of probes, labelled with the second dye. In case of successful annealing, the second probe is ligated to the first probe. The process is repeated, followed by probing with the pools of probes labelled with the third and fourth dye. After 4 cycles, two nucleotides are »read« and three are skipped, per each probe (e.g. + + – – – + + – – – + + – – –...). The addition of 4 pools of probes can be done *n* times. Then, the ligated probes and oligonucleotide primer are removed and the new one is annealed. The new primer is reset, i.e. located 1 base away from adapter-template junction towards 5′ end (*n*-1). The cycles of probe annealing and ligation are repeated (*Figure 2.A.4*). Since the new primer is shifted one base towards 5′ end, the first of the two bases with the interrogated previous primer, the 5′ base, will be interrogated again, along with the new 5′ base. The reader should keep in mind that at this level, it is not possible to know which base is »recognized« in first or second step, since there are four probes labelled with the same dye. Resetting of the primer is done 4 times (from *n*-1 to *n*-4), thus enabling interrogations of all bases, including those covered by degenerate bases. Each base is interrogated twice.

To decipher the bases behind two fluorescent signals registered for each interrogated position, there is an artificial four-colour code (not to be mixed with the four dyes used for labelling of the probes). For example, red fluorescence followed by red fluorescence is blue, yellow fluorescence followed by yellow fluorescence is blue, red fluorescence followed by yellow fluorescence is green, etc. At the same time, each colour encodes the first and the second »recognized« base. Simply by the order of four colours, computer is able to reconstruct the sequence of bases, since the identity (and colour) of bases #1 and #2 influence the identity and colour of bases #2 and #3, etc. (*Figure 2.A.4*). Using this approach, 50 bp reads can be obtained. The key advantage of this method is that each base is interrogated twice. Major disadvantages are read length and run times.

Third generation sequencing – emerging technologies

Third generation technologies are driven by the goal to reduce the price of sequencing and to simplify the procedure. One way to achieve this is to avoid preparation and amplification of samples by sequencing single molecules, thereby reducing the cost of reagents. Furthermore, optical systems for detection of incorporation events have inherent drawbacks, and another tendency of third generation methods is to use non-optical detection systems.

Single molecule real time (SMRT) sequencing

Most prominent technology for single molecule sequencing (SMS) is marketed by Pacific Biosciences (55). To achieve SMS, the authors have developed a chip with wells in the zeptoliter (10^{-21} L) volume range. On the bottom of each well, one molecule of DNA polymerase is immobilized, using biotin-streptavidin system. Once the template ssDNA is coupled with immobilized DNA polymerase, fluorescently labelled dNTPs are added. Although the fluorescence of free-floating dNTPs is not quenched, the special detection system, due to the design of the well, is able to detect only the fluorescence of the nucleotide being incorporated in the growing strand of DNA (*Figure 2.C*). Since the fluorescent dye is attached to the phosphate group, incorporation releases fluorescent signal from the nucleotide and the detection of signal (i.e. pulse) stops. Therefore, during synthesis of the second strand, each incorporation is measured in real time. Since all four nucleotides are added simultaneously, and measurement is made in real time, the speed of sequencing is increased compared to technologies where individual nucleotides are flushed sequentially. Reported accuracy is 99.3%, and it most likely could be improved by circularizing the template and sequencing it several times. Read length is more than 900 bp (50).

DNA nanoball sequencing

DNA nanoball sequencing (DNB) was developed by the inventor of SBH (42). It also belongs to sequencing by ligation approaches. DNA template is fragmented into fragments of 400 – 500 bp. Next, two halves of adapter 1 are ligated onto the termini of each fragment (*Figure 2.B.1*). Upon circularization of the molecule, adapter #1 site is created. Then, this circular molecule is nicked and two halves of adapter #2 are ligated, which upon circularization create adapter 2. In the same way, adapters #3 and #4 are introduced into the fragment, with the final step being circularization that creates adapter #4 site, along with removal of the portion of template sequence (*Figure 2.B.1)*. Four adapters are located roughly at the positions 3, 6, 9 and 12 o'clock. In the next step, this circular molecule is subjected to many cycles of rolling circle amplification, generating ssDNA concatemer. Four adapter sequences contain palindromic sequences, which interact and cause coiling of the long ssDNA into tight ball of DNA, roughly 300 nm in size (*Figure 2.B.2*). DNBs are deposited onto the arrayed flow cell, with wells accommodating only one nanoball, enabling high density of DNBs to be sequenced (*Figure 2.B.3*). The bases of the template are read in 5' and 3' direction from each adapter, up to ten bases in each direction. Since only short sequences, adjacent to adapters, are read, this sequencing format resembles a multiplexed form of mate-pair sequencing approach. The actual sequencing procedure involves probing nanoballs with anchor probe, complementary to adapter sequence and pool of four fluorescently labelled probes with degenerate nucleotides in all but one position, which is actually interrogated (similar to SOLiD sequencing). Upon annealing, anchor and degenerate probe will be adjacent, and ligated. Unbound probes are washed away and fluorescence recorded (*Figure 2.B.4*). Ligated probes are removed, and a new pool of probes is added, specific for different interrogated position. The cycle of annealing, ligation, washing and image recording is repeated for all ten positions adjacent to one terminus of one adapter. This process is repeated for all seven remaining adapter termini.

The major disadvantage is the short length of reads, although authors managed to sequence the whole human genome. Claimed cost of the reagents for sequencing of the whole human genome is under 5000$. The major advantage of this approach is the high density of arrays, and therefore high number DNBs (~350 million) that can be sequenced.

Nanopore sequencing

Nanopore sequencing is not yet commercially available. Nanopores are formed by pore forming proteins, such as different kinds of biological channels, e.g. α–haemolysin of *Staphylococcus aureus* (56). The idea behind sequencing using nanopores is that the conductivity of the pore for the ion currents changes when the pore is blocked by the strand of nucleic acid passing through the pore. The flow of ion current will depend on the shape of the molecule translocating through the pore. Since nucleotides have different shapes, in theory it should be possible to recognize each nucleotide by its subtle effect on the change of the ionic current (57). The key advantage of this approach is that it sequences single molecules, and therefore minimal sample preparation is necessary. In addition, as no amplification or ligation occurs, nucleotides and enzymes are not required. It is plausible that it will be possible to achieve very long read lengths (in the Kbp range). Currently, the crucial problem of the technology is reducing the speed of DNA translocation through the nanopore, in order to ensure reliable measurement of the current change.

Ion torrent

Ion Torrent technology was developed by the inventor of »454« sequencing (58), and it is marketed by Life Technologies. Sample preparation is very similar to the one used for »454« sequencing. However, instead of PicoTiterPlate, a chip with ion-sensitive field-effect transistor sensor, engineered to be able to detect individual protons, is used. The instrument has no optical components. Beads containing enriched template DNA are deposited into the wells of the chip. The chip is situated within the flow cell, and is sequentially flushed with individual dNTPs. Integration of nucleotide releases H^+ that changes the pH of the surrounding solution proportional to the number of incorporated nucleotides (*Figure 2.D*). This is detected by the sensor on the bottom of each well, and converted to electronic signal, that is recorded by the system. The latest version of the instrument (Ion Proton) is claimed to have throughput of 20 Gbp, with reads of 200 bp, and run time of up to 4 h (59). The major disadvantages of the system are problems in reading homopolymer stretches and relative short read length, what is partially compensated by the number of wells (up to 11 million). The major advantage seems to be the price.

Conclusion

»Sanger« sequencing is still considered to be the »gold standard« of sequencing, due to its accuracy and read lengths. Automated »Sanger« sequencing still produces the longest reads, although »next-generation« technologies are coming closer to this mark. Still, »Sanger« sequencing is notoriously slow and expensive. On the other hand, second and third generation technologies are able to sequence whole human genome at 30x coverage within 10 days, for 10000$. Key problems of these technologies are accuracy and short reads, which makes final assembly

or alignment difficult and computationally challenging. The ideal technology of the future will have to come with increased accuracy, longer read lengths and reduced price. Single molecule sequencing technologies probably hold the potential to become the sequencing technologies of the future, especially with their potential to directly sequence RNA molecules (by replacing DNA polymerase with RNA-dependent DNA polymerase (60)). Another important venue of research will be analysis of genome and transcriptome of single cells (61, 62).

Improvement and development of new sequencing technologies is making them increasingly available and affordable to researchers. To meet the needs of scientists who do not have the ambition to routinely sequence large genomes, few companies have launched benchtop sequencers, with somewhat reduced capabilities (63). The choice of the technology depends on the needed application and advantages or disadvantages of the particular sequencing approach. For example, 454 technology produces long reads, but its sequencing volume is in hundreds of Mbp range and is produced within hours, while »Illumina« sequencing produces shorter reads, with volume of hundreds of Gbp, within days. In most cases, an average academic laboratory or core facility can purchase at most one platform. Therefore, this decision should be thoroughly contemplated, depending on the user's current, but also prospective needs.

Increasing throughput and decreasing price of sequencing are slowly replacing other methods. For example, it can be expected that expression profiling by microarrays will be abandoned. Contemporary applications of sequencing include *de novo* sequencing and resequencing of whole genomes (51), gene expression profiling (64, 65), detection of methylated regions in the genome (66), but also analysis of portions of the genome (e.g. exome (67), ChIP-Seq (68), genotyping, mutation analysis (69)), with different methods of sequence enrichment methods being increasingly available (70).

Completion of the sequencing of the human genome had little implication for routine healthcare (71). However, with the decrease of price of sequencing, many clinical diagnostic procedures (e.g. preimplantation diagnostics (72, 73)) will be replaced by sequencing. It will become possible to untangle great heterogeneity of complex genetic disorders (cancer, neuro-psychiatric disorders) (74, 75). In addition to improved health care, basic science will also benefits from novel sequencing technologies. Genomes of many viruses, bacteria, plant and animal species are being sequenced, what will have great impact on agricultural and environmental sciences.

New sequencing technologies are developed and improved very rapidly. Probably by the time this article is published, some numbers will not be up to date. It is difficult to predict the fate of existing and emerging technologies. Some technologies, like 454 or Illumina are already established in the laboratories around the world. If they are going be replaced by more recent technologies, such as Ion Torrent or single molecule sequencing, remains to be seen. Most likely, not only the existing technologies will be improved, but we can expect the rise of fourth generation sequencing technologies, all aiming to bring the 1000$ within our reach (76, 77).

Conflict of interest statement

The author stated that there are no conflicts of interest regarding the publication of this article.

References

1. Manier J Caution: Replicating The Lifestyle Of DNA Guru Kary Mullis May Be Hazardous Journal/Chicago Tribune 1998;
2. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci USA 1977; 74: 560–4.
3. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–7.
4. Brownlee GG, Sanger F. Nucleotide sequences from the low molecular weight ribosomal RNA of Escherichia coli. J Mol Biol 1967; 23: 337–53.
5. Gilbert W, Maxam A. The nucleotide sequence of the lac operator. Proc Natl Acad Sci U S A 1973; 70: 3581–4.
6. Sanger F, Donelson JE, Coulson AR, Kossel H, Fischer D. Use of DNA polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. Proc Natl Acad Sci USA 1973; 70: 1209–13.
7. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 1977; 265: 687–95.
8. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975; 94: 441–8.
9. Sanger F, Coulson AR, Friedmann T, Air GM, Barrell BG, Brown NL, et al. The nucleotide sequence of bacteriophage phiX174. J Mol Biol 1978; 125: 225–46.
10. Sanger F, Coulson AR. The use of thin acrylamide gels for DNA sequencing. FEBS Lett 1978; 87: 107–10.
11. Boulikas T, Hancock R. A highly sensitive technique for staining DNA and RNA in polyacrylamide gels using silver. J Biochem Biophys Methods 1981; 5: 219–28.
12. Somerville LL, Wang K. The ultrasensitive silver »protein« stain also detects nanograms of nucleic acids. Biochem Biophys Res Commun 1981; 102: 53–8.

13. Bassam BJ, Caetano-Anolles G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 1991; 196: 80–3.

14. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981; 290: 457–65.

15. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995; 269: 496–512.

16. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science 1996; 274: 546, 563–47.

17. TCESC. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 1998; 282: 2012–8.

18. TAGI. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000; 408: 796–815.

19. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of Drosophila melanogaster. Science 2000; 287: 2185–95.

20. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, et al. A whole-genome assembly of Drosophila. Science 2000; 287: 2196–204.

21. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001; 291: 1304–51.

22. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860–921.

23. IHGS. Finishing the euchromatic sequence of the human genome. Nature 2004; 431: 931–45.

24. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol 2007; 5: e254.

25. Hood LE, Hunkapiller MW, Smith LM. Automated DNA sequencing and analysis of the human genome. Genomics 1987; 1: 201–12.

26. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 1987; 238: 336–41.

27. Swerdlow H, Gesteland R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. Nucleic Acids Res 1990; 18: 1415–9.

28. ABI. ABI PRISM 373 DNA sequencer with XL upgrade: User's manual. 2001.

29. ABI. ABI Prism 310 Genetic analyzer: User's manual. 2010.

30. ABI. ABI PRISM 377 DNA Sequencer For Sequencing and GeneScan Analysis Software Applications: User's manual. 2000.

31. ABI, Hitachi. ABI PRISM 3100 Genetic Analyzer: User's Manual. 2001.

32. ABI. Applied Biosystems 3730 and 3730xl DNA Analyzers: Specification Sheet. 2006.

33. Kieleczawa J. Automated DNA scanners used in sequencing laboratories. In: Kieleczawa J, editor. DNA Sequencing: Optimizing the Process and Analysis, Volume 1. Sudbury: Jones and Bartlett Publishers, Inc., 2005: 123–30.

34. Men AE, Wilson P, Siemering K, Forrest S. Sanger DNA sequencing. In: Janitz M, editor. Next-Generation Sequencing: Towards Personalized Medicine. Weinheim: Wiley-VCH Verlag GmbH, 2008: 3–11.

35. Drmanac R, Labat I, Brukner I, Crkvenjakov R. Sequencing of megabase plus DNA by hybridization: theory of the method. Genomics 1989; 4: 114–28.

36. Drmanac R, Strezoska Z, Labat I, Drmanac S, Crkvenjakov R. Reliable hybridization of oligonucleotides as short as six nucleotides. DNA Cell Biol 1990; 9: 527–34.

37. Drmanac R, Drmanac S, Chui G, Diaz R, Hou A, Jin H, et al. Sequencing by hybridization (SBH): advantages, achievements, and opportunities. Adv Biochem Eng Biotechnol 2002; 77: 75–101.

38. Strezoska Z, Paunesku T, Radosavljevic D, Labat I, Drmanac R, Crkvenjakov R. DNA sequencing by hybridization: 100 bases read by a non-gel-based method. Proc Natl Acad Sci U S A 1991; 88: 10089–93.

39. Drmanac R, Drmanac S, Strezoska Z, Paunesku T, Labat I, Zeremski M, et al. DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. Science 1993; 260: 1649–52.

40. Hacia JG. Resequencing and mutational analysis using oligonucleotide microarrays. Nat Genet 1999; 21: 42–7.

41. Pihlak A, Bauren G, Hersoug E, Lonnerberg P, Metsis A, Linnarsson S. Rapid genome sequencing with short universal tiling probes. Nat Biotechnol 2008; 26: 676–84.

42. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 2010; 327: 78–81.

43. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437: 376–80.

44. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem 1996; 242: 84–9.

45. Hyman ED. A new method of sequencing DNA. Anal Biochem 1988; 174: 423–36.

46. Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science 1998; 281: 363–5.

47. Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD, et al. A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. Electrophoresis 2003; 24: 3769–77.

48. Nakano M, Komatsu J, Matsuura S, Takashima K, Katsura S, Mizuno A. Single-molecule PCR using water-in-oil emulsion. J Biotechnol 2003; 102: 117–24.

49. 454, Roche. GS FLX+ System. Flyer. 2011.

50. Metzker ML. Sequencing technologies – the next generation. Nat Rev Genet 2010; 11: 31–46.

51. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008; 452: 872–6.

52. Illumina. HiSeq sequencing systems: Specification sheet. 2011.

53. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008; 456: 53–9.

54. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res 2008; 18: 1051–63.

55. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009; 323: 133–8.

56. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol 2009; 4: 265–70.

57. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nat Biotechnol 2008; 26: 1146–1153.

58. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011; 475: 348–52.

59. LifeTechnologies. Ion Proton: Specification sheet. 2012.

60. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, et al. Direct RNA sequencing. Nature 2009; 461: 814–8.

61. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011; 472: 90–4.

62. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 2009; 6: 377–82.

63. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012; 30: 434–9.

64. Lipson D, Raz T, Kieu A, Jones DR, Giladi E, Thayer E, et al. Quantification of the yeast transcriptome by single-molecule sequencing. Nat Biotechnol 2009; 27: 652–8.

65. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008; 18: 1509–17.

66. Korshunova Y, Maloney RK, Lakey N, Citek RW, Bacher B, Budiman A, et al. Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. Genome Res 2008; 18: 19–29.

67. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009; 461: 272–6.

68. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129: 823–37.

69. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010; 463: 191–6.

70. Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, et al. Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief Funct Genomics 2011; 10: 374–86.

71. Marshall E. Human genome 10th anniversary. Waiting for the revolution. Science 2011; 331: 526–9.

72. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A 2008; 105: 20458–63.

73. Martin J, Cervero A, Mir P, Martinez-Conejero JA, Pellicer A, Simon C. The impact of next-generation sequencing technology on preimplantation genetic diagnosis and screening. Fertil Steril 2013; 99: 1054–1061 e1053.

74. Aparicio SA, Huntsman DG. Does massively parallel DNA resequencing signify the end of histopathology as we know it? J Pathol 2010; 220: 307–15.

75. Ezewudo M, Zwick ME. Evaluating rare variants in complex disorders using next-generation sequencing. Curr Psychiatry Rep 2013; 15: 349.

76. Kedes L, Campany G. The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE competition. Nat Genet 2011; 43: 1055–8.

77. Goldberg MD. Science at the crossroads: Fact or Fiction? J Med Biochem 2011; 30: 79–92.

Received: August 15, 2013

Accepted: September 10, 2013