DNA sequencing

DNA sequencing is the process of determining the nucleic acid sequence – the order of nucleotides in DNA. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine. The advent of rapid DNA sequencing methods has greatly accelerated biological and medical research and discovery.[1][2]

Knowledge of **DNA sequences** has become indispensable for basic biological research, DNA Genographic Projects and in numerous applied fields such as medical diagnosis, biotechnology, forensic biology, virology and biological systematics. Comparing healthy and mutated DNA sequences can diagnose different diseases including various cancers,[3] characterize antibody repertoire,[4] and can be used to guide patient treatment.[5] Having a quick way to sequence DNA allows for faster and more individualized medical care to be administered, and for more organisms to be identified and cataloged.[4]

The rapid advancements in DNA sequencing technology have played a crucial role in sequencing complete genomes of various life forms, including humans, as well as numerous animal, plant, and microbial species.

The first DNA sequences were obtained in the early 1970s by academic researchers using laborious methods based on two-dimensional chromatography. Following the development of fluorescence-based sequencing methods with a DNA sequencer,[6] DNA sequencing has become easier and orders of magnitude faster.[7][8]

Applications

DNA sequencing may be used to determine the sequence of individual genes, larger genetic regions (i.e. clusters of genes or operons), full chromosomes, or entire genomes of any organism. DNA sequencing is also the most efficient way to indirectly sequence RNA or proteins (via their open reading frames). In fact, DNA sequencing has become a key technology in many areas of biology and other sciences such as medicine, forensics, and anthropology.

An example of the results of automated chain-termination DNA sequencing.

Molecular biology

Sequencing is used in molecular biology to study genomes and the proteins they encode. Information obtained using sequencing allows researchers to identify changes in genes and noncoding DNA (including regulatory sequences), associations with diseases and phenotypes, and identify potential drug targets.

Evolutionary biology

Since DNA is an informative macromolecule in terms of transmission from one generation to another, DNA sequencing is used in evolutionary biology to study how different organisms are related and how they evolved. In February 2021, scientists reported, for the first time, the sequencing of DNA from animal remains, a mammoth in this instance, over a million years old, the oldest DNA sequenced to date.[9][10]

Metagenomics

The field of metagenomics involves identification of organisms present in a body of water, sewage, dirt, debris filtered from the air, or swab samples from organisms. Knowing which organisms are present in a particular environment is critical to research in ecology, epidemiology, microbiology, and other fields. Sequencing enables researchers to determine which types of microbes may be present in a microbiome, for example.
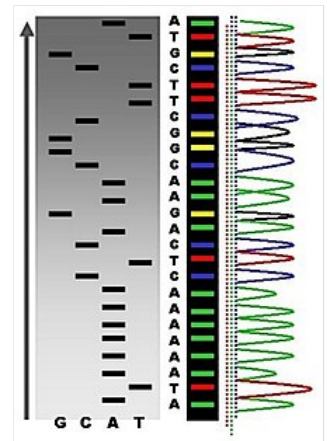
Virology

As most viruses are too small to be seen by a light microscope, sequencing is one of the main tools in virology to identify and study the virus.[11] Viral genomes can be based in DNA or RNA. RNA viruses are more time-sensitive for genome sequencing, as they degrade faster in clinical samples.[12] Traditional Sanger sequencing and next-generation sequencing are used to sequence viruses in basic and clinical research, as well as for the diagnosis of emerging viral infections, molecular epidemiology of viral pathogens, and drug-resistance testing. There are more than 2.3 million unique viral sequences in GenBank.[11] Recently, NGS has surpassed traditional Sanger as the most popular approach for generating viral genomes.[11]

During the 1997 avian influenza outbreak, viral sequencing determined that the influenza sub-type originated through reassortment between quail and poultry. This led to legislation in Hong Kong that prohibited selling live quail and poultry together at market. Viral sequencing can also be used to estimate when a viral outbreak began by using a molecular clock technique.[12]

Medicine

Medical technicians may sequence genes (or, theoretically, full genomes) from patients to determine if there is risk of genetic diseases. This is a form of genetic testing, though some genetic tests may not involve DNA sequencing.

As of 2013 DNA sequencing was increasingly used to diagnose and treat rare diseases. As more and more genes are identified that cause rare genetic diseases, molecular diagnoses for patients become more mainstream. DNA sequencing allows clinicians to identify genetic diseases, improve disease management, provide reproductive counseling, and more effective therapies.[13] Gene sequencing panels are used to identify multiple potential genetic causes of a suspected disorder.[14]

Also, DNA sequencing may be useful for determining a specific bacteria, to allow for more precise antibiotics treatments, hereby reducing the risk of creating antimicrobial resistance in bacteria populations.[15][16][17][18][19][20]

Forensic investigation

DNA sequencing may be used along with DNA profiling methods for forensic identification[21] and paternity testing. DNA testing has evolved tremendously in the last few decades to ultimately link a DNA print to what is under investigation. The DNA patterns in fingerprint, saliva, hair follicles, etc. uniquely separate each living organism from another. Testing DNA is a technique which can detect specific genomes in a DNA strand to produce a unique and individualized pattern.

DNA sequencing may be used along with DNA profiling methods for forensic identification and paternity testing, as it has evolved significantly over the past few decades to ultimately link a DNA print to what is under investigation. The DNA patterns in fingerprint, saliva, hair follicles, and other bodily fluids uniquely separate each living organism from another, making it an invaluable tool in the field of forensic science. The process of DNA testing involves detecting specific genomes in a DNA strand to produce a unique and individualized pattern, which can be used to identify individuals or determine their relationships.

The advancements in DNA sequencing technology have made it possible to analyze and compare large amounts of genetic data quickly and accurately, allowing investigators to gather evidence and solve crimes more efficiently. This technology has been used in various applications, including forensic identification, paternity testing, and human identification in cases where traditional identification methods are unavailable or unreliable. The use of DNA sequencing has also led to the development of new forensic techniques, such as DNA phenotyping, which allows investigators to predict an individual's physical characteristics based on their genetic data.

In addition to its applications in forensic science, DNA sequencing has also been used in medical research and diagnosis. It has enabled scientists to identify genetic mutations and variations that are associated with certain diseases and disorders, allowing for more accurate diagnoses and targeted treatments. Moreover, DNA sequencing has also been used in conservation biology to study the genetic diversity of endangered species and develop strategies for their conservation.

Furthermore, the use of DNA sequencing has also raised important ethical and legal considerations. For example, there are concerns about the privacy and security of genetic data, as well as the potential for misuse or discrimination based on genetic information. As a result, there are ongoing debates about the need for regulations and guidelines to ensure the responsible use of DNA sequencing technology.

Overall, the development of DNA sequencing technology has revolutionized the field of forensic science and has far-reaching implications for our understanding of genetics, medicine, and conservation biology.

The four canonical bases

The canonical structure of DNA has four bases: thymine (T), adenine (A), cytosine (C), and guanine (G). DNA sequencing is the determination of the physical order of these bases in a molecule of DNA. However, there are many other bases that may be present in a molecule. In some viruses (specifically, bacteriophage), cytosine may be replaced by hydroxy methyl or hydroxy methyl glucose cytosine.[22] In mammalian DNA, variant bases with methyl groups or phosphosulfate may be found.[23][24] Depending on the sequencing technique, a particular modification, e.g., the 5mC (5 methyl cytosine) common in humans, may or may not be detected.[25]

In almost all organisms, DNA is synthesized in vivo using only the 4 canonical bases; modification that occurs post replication creates other bases like 5 methyl C. However, some bacteriophage can incorporate a non standard base directly.[26]

In addition to modifications, DNA is under constant assault by environmental agents such as UV and Oxygen radicals. At the present time, the presence of such damaged bases is not detected by most DNA sequencing methods, although PacBio has published on this.[27]
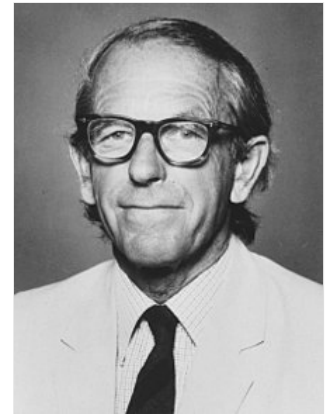
History

Discovery of DNA structure and function

Deoxyribonucleic acid (DNA) was first discovered and isolated by Friedrich Miescher in 1869, but it remained under-studied for many decades because proteins, rather than DNA, were thought to hold the genetic blueprint to life. This situation changed after 1944 as a result of some experiments by Oswald Avery, Colin MacLeod, and Maclyn McCarty demonstrating that purified DNA could change one strain of bacteria into another. This was the first time that DNA was shown capable of transforming the properties of cells.

In 1953, James Watson and Francis Crick put forward their double-helix model of DNA, based on crystallized X-ray structures being studied by Rosalind Franklin. According to the model, DNA is composed of two strands of nucleotides coiled around each other, linked together by hydrogen bonds and running in opposite directions. Each strand is composed of four complementary nucleotides – adenine (A), cytosine (C), guanine (G) and thymine (T) – with an A on one strand always paired with T on the other, and C always paired with G. They proposed that such a structure allowed each strand to be used to reconstruct the other, an idea central to the passing on of hereditary information between generations.[28]

The foundation for sequencing proteins was first laid by the work of Frederick Sanger who by 1955 had completed the sequence of all the amino acids in insulin, a small protein secreted by the pancreas. This provided the first conclusive evidence that proteins were chemical entities with a specific molecular pattern rather than a random mixture of material suspended in fluid. Sanger's success in sequencing insulin spurred on x-ray crystallographers, including Watson and Crick, who by now were trying to understand how DNA directed the formation of proteins within a cell. Soon after attending a series of lectures given by Frederick Sanger in October 1954, Crick began developing a theory which argued that the arrangement of nucleotides in DNA determined the sequence of amino acids in proteins, which in turn helped determine the function of a protein. He published this theory in 1958.[29]

RNA sequencing

RNA sequencing was one of the earliest forms of nucleotide sequencing. The major landmark of RNA sequencing is the sequence of the first complete gene and the complete genome of Bacteriophage MS2, identified and published by Walter Fiers and his coworkers at the University of Ghent (Ghent, Belgium), in 1972[30] and 1976.[31] Traditional RNA sequencing methods require the creation of a cDNA molecule which must be sequenced.[32]

Frederick Sanger, a pioneer of sequencing. Sanger is one of the few scientists who was awarded two Nobel prizes, one for the sequencing of proteins, and the other for the sequencing of DNA.

Traditional RNA Sequencing Methods

Traditional RNA sequencing methods involve several steps:

1) **Reverse Transcription**: The first step is to convert the RNA molecule into a complementary DNA (cDNA) molecule using an enzyme called reverse transcriptase.

2) **cDNA Synthesis**: The cDNA molecule is then synthesized through a process called PCR (Polymerase Chain Reaction), which amplifies the cDNA to produce multiple copies.

3) **Sequencing**: The amplified cDNA is then sequenced using a technique such as Sanger sequencing or Maxam-Gilbert sequencing.

Challenges and Limitations

Traditional RNA sequencing methods have several limitations. For example:

They require the creation of a cDNA molecule, which can be time-consuming and labor-intensive. They are prone to errors and biases, which can affect the accuracy of the sequencing results. They are limited in their ability to detect rare or low-abundance transcripts.

Advances in RNA Sequencing Technology

In recent years, advances in RNA sequencing technology have addressed some of these limitations. New methods such as next-generation sequencing (NGS) and single-molecule real-time (SMRT) sequencing have enabled faster, more accurate, and more cost-effective sequencing of RNA molecules. These advances have opened up new possibilities for studying gene expression, identifying new genes, and understanding the regulation of gene expression.

Early DNA sequencing methods

The first method for determining DNA sequences involved a location-specific primer extension strategy established by Ray Wu, a geneticist, at Cornell University in 1970.[33] DNA polymerase catalysis and specific nucleotide labeling, both of which figure prominently in current sequencing schemes, were used to sequence the cohesive ends of lambda phage DNA.[34][35][36] Between 1970 and 1973, Wu, scientist Radha Padmanabhan and colleagues demonstrated that this method can be employed to determine any DNA sequence using synthetic location-specific primers.[37][38][8]

Walter Gilbert and Allan Maxam at Harvard also developed sequencing methods, including one for "DNA sequencing by chemical degradation".[39][40] In 1973, Gilbert and Maxam reported the sequence of 24 basepairs using a method known as wandering-spot analysis.[41] Advancements in sequencing were aided by the concurrent development of recombinant DNA technology, allowing DNA samples to be isolated from sources other than viruses.

In 1977, Frederick Sanger then adopted a primer-extension strategy to develop more rapid DNA sequencing methods at the MRC Centre, Cambridge, UK. This technique was similar to his "Plus and Minus" strategy, however, it was based upon the selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication.[42][43][44] Sanger published this method in the same year. [45]

Sequencing of full genomes

The first full DNA genome to be sequenced was that of bacteriophage φX174 in 1977.[46] Medical Research Council scientists deciphered the complete DNA sequence of the Epstein-Barr virus in 1984, finding it contained 172,282 nucleotides. Completion of the sequence marked a significant turning point in DNA sequencing because it was achieved with no prior genetic profile knowledge of the virus.[47][8]

A non-radioactive method for transferring the DNA molecules of sequencing reaction mixtures onto an immobilizing matrix during electrophoresis was developed by Herbert Pohl and co-workers in the early 1980s.[48][49] Followed by the commercialization of the DNA sequencer "Direct-Blotting-Electrophoresis-System GATC 1500" by GATC Biotech, which was intensively used in the framework of the EU genome-sequencing programme, the complete DNA sequence of the yeast *Saccharomyces cerevisiae* chromosome II.[50] Leroy E. Hood's laboratory at the California Institute of Technology announced the first semi-automated DNA sequencing machine in 1986.[51] This was followed by Applied Biosystems' marketing of the first fully

automated sequencing machine, the ABI 370, in 1987 and by Dupont's Genesis 2000[52] which used a novel fluorescent labeling technique enabling all four dideoxynucleotides to be identified in a single lane. By 1990, the U.S. National Institutes of Health (NIH) had begun large-scale sequencing trials on *Mycoplasma capricolum*, *Escherichia coli*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* at a cost of US$0.75 per base. Meanwhile, sequencing of human cDNA sequences called expressed sequence tags began in Craig Venter's lab, an attempt to capture the coding fraction of the human genome.[53] In 1995, Venter, Hamilton Smith, and colleagues at The Institute for Genomic Research (TIGR) published the first complete genome of a free-living organism, the bacterium *Haemophilus influenzae*. The circular chromosome contains 1,830,137 bases and its publication in the journal Science[54] marked the first published use of whole-genome shotgun sequencing, eliminating the need for initial mapping efforts.

By 2001, shotgun sequencing methods had been used to produce a draft sequence of the human genome.[55][56]

The 5,386 bp genome of bacteriophage φX174. Each coloured block represents a gene.

High-throughput sequencing (HTS) methods

Several new methods for DNA sequencing were developed in the mid to late 1990s and were implemented in commercial DNA sequencers by 2000. Together these were called the "next-generation" or "second-generation" sequencing (NGS) methods, in order to distinguish them from the earlier methods, including Sanger sequencing. In contrast to the first generation of sequencing, NGS technology is typically characterized by being highly scalable, allowing the entire genome to be sequenced at once. Usually, this is accomplished by fragmenting the genome into small pieces, randomly sampling for a fragment, and sequencing it using one of a variety of technologies, such as those described below. An entire genome is possible because multiple fragments are sequenced at once (giving it the name "massively parallel" sequencing) in an automated process.

History of sequencing technology[57]

NGS technology has tremendously empowered researchers to look for insights into health, anthropologists to investigate human origins, and is catalyzing the "Personalized Medicine" movement. However, it has also opened the door to more room for error. There are many software tools to carry out the computational analysis of NGS data, often compiled at online platforms such as CSI NGS Portal, each with its own algorithm. Even the parameters within one software package can change the outcome of the analysis. In addition, the large quantities of data produced by DNA sequencing have also required development of new methods and programs for sequence analysis. Several efforts to develop standards in the NGS field have been attempted to address these challenges, most of which have been small-scale efforts arising from individual labs. Most recently, a large, organized, FDA-funded effort has culminated in the BioCompute standard.

On 26 October 1990, Roger Tsien, Pepi Ross, Margaret Fahnestock and Allan J Johnston filed a patent describing stepwise ("base-by-base") sequencing with removable 3' blockers on DNA arrays (blots and single DNA molecules).[58] In 1996, Pål Nyrén and his student Mostafa Ronaghi at the Royal Institute of Technology in Stockholm published their method of pyrosequencing.[59]

On 1 April 1997, Pascal Mayer and Laurent Farinelli submitted patents to the World Intellectual Property Organization describing DNA colony sequencing.[60] The DNA sample preparation and random surface-polymerase chain reaction (PCR) arraying methods described in this patent, coupled to Roger Tsien et al.'s "base-by-base" sequencing method, is now implemented in Illumina's Hi-Seq genome sequencers.

In 1998, Phil Green and Brent Ewing of the University of Washington described their phred quality score for sequencer data analysis,[61] a landmark analysis technique that gained widespread adoption, and which is still the most common metric for assessing the accuracy of a sequencing platform.[62]

Lynx Therapeutics published and marketed massively parallel signature sequencing (MPSS), in 2000. This method incorporated a parallelized, adapter/ligation-mediated, bead-based sequencing technology and served as the first commercially available "next-generation" sequencing method, though no DNA sequencers were sold to independent laboratories.[63]

Basic methods

Maxam-Gilbert sequencing

Allan Maxam and Walter Gilbert published a DNA sequencing method in 1977 based on chemical modification of DNA and subsequent cleavage at specific bases.[39] Also known as chemical sequencing, this method allowed purified samples of double-stranded DNA to be used without further cloning. This method's use of radioactive labeling and its technical complexity discouraged extensive use after refinements in the Sanger methods had been made.

Maxam-Gilbert sequencing requires radioactive labeling at one 5' end of the DNA and purification of the DNA fragment to be sequenced. Chemical treatment then generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A+G, C, C+T). The concentration of the modifying chemicals is controlled to introduce on average one modification per DNA molecule. Thus a series of labeled fragments
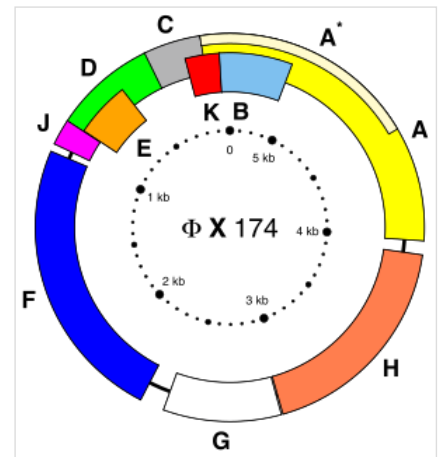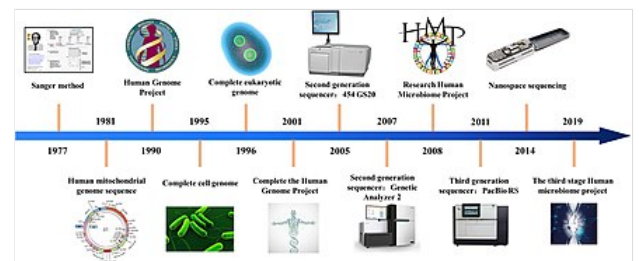
is generated, from the radiolabeled end to the first "cut" site in each molecule. The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels for size separation. To visualize the fragments, the gel is exposed to X-ray film for autoradiography, yielding a series of dark bands each corresponding to a radiolabeled DNA fragment, from which the sequence may be inferred.[39]

This method is mostly obsolete as of 2023.[64]

Chain-termination methods

The chain-termination method developed by Frederick Sanger and coworkers in 1977 soon became the method of choice, owing to its relative ease and reliability.[45][65] When invented, the chain-terminator method used fewer toxic chemicals and lower amounts of radioactivity than the Maxam and Gilbert method. Because of its comparative ease, the Sanger method was soon automated and was the method used in the first generation of DNA sequencers.

Sanger sequencing is the method which prevailed from the 1980s until the mid-2000s. Over that period, great advances were made in the technique, such as fluorescent labelling, capillary electrophoresis, and general automation. These developments allowed much more efficient sequencing, leading to lower costs. The Sanger method, in mass production form, is the technology which produced the first human genome in 2001, ushering in the age of genomics. However, later in the decade, radically different approaches reached the market, bringing the cost per genome down from $100 million in 2001 to $10,000 in 2011.[66]

Sequencing by synthesis

The objective for sequential sequencing by synthesis (SBS) is to determine the sequencing of a DNA sample by detecting the incorporation of a nucleotide by a DNA polymerase. An engineered polymerase is used to synthesize a copy of a single strand of DNA and the incorporation of each nucleotide is monitored. The principle of real-time sequencing by synthesis was first described in 1993[67] with improvements published some years later.[68] The key parts are highly similar for all embodiments of SBS and includes (1) amplification of DNA (to enhance the subsequent signal) and attach the DNA to be sequenced to a solid support, (2) generation of single stranded DNA on the solid support, (3) incorporation of nucleotides using an engineered polymerase and (4) real-time detection of the incorporation of nucleotide The steps 3-4 are repeated and the sequence is assembled from the signals obtained in step 4. This principle of real-time sequencing-by-synthesis has been used for almost all massive parallel sequencing instruments, including 454, PacBio, IonTorrent, Illumina and MGI.

Large-scale sequencing and *de novo* sequencing

Large-scale sequencing often aims at sequencing very long DNA pieces, such as whole chromosomes, although large-scale sequencing can also be used to generate very large numbers of short sequences, such as found in phage display. For longer targets such as chromosomes, common approaches consist of cutting (with restriction enzymes) or shearing (with mechanical forces) large DNA fragments into shorter DNA fragments. The fragmented DNA may then be cloned into a DNA vector and amplified in a bacterial host such as *Escherichia coli*. Short DNA fragments purified from individual bacterial colonies are individually sequenced and assembled electronically into one long, contiguous sequence. Studies have shown that adding a size selection step to collect DNA fragments of uniform size can improve sequencing efficiency and accuracy of the genome assembly. In these studies, automated sizing has proven to be more reproducible and precise than manual gel sizing.[69][70][71]
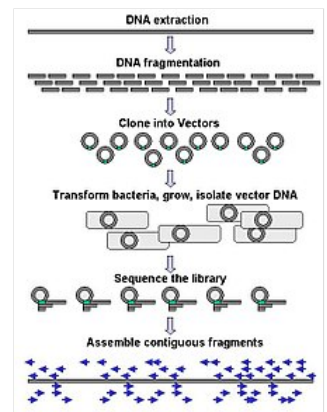
The term "*de novo* sequencing" specifically refers to methods used to determine the sequence of DNA with no previously known sequence. *De novo* translates from Latin as "from the beginning". Gaps in the assembled sequence may be filled by primer walking. The different strategies have different tradeoffs in speed and accuracy; shotgun methods are often used for sequencing large genomes, but its assembly is complex and difficult, particularly with sequence repeats often causing gaps in genome assembly.

Most sequencing approaches use an *in vitro* cloning step to amplify individual DNA molecules, because their molecular detection methods are not sensitive enough for single molecule sequencing. Emulsion PCR[72] isolates individual DNA molecules along with primer-coated beads in aqueous droplets within an oil phase. A polymerase chain reaction (PCR) then coats each bead with clonal copies of the DNA molecule followed by immobilization for later sequencing. Emulsion PCR is used in the methods developed by Marguilis et al. (commercialized by 454 Life Sciences), Shendure and Porreca et al. (also known as "polony sequencing") and SOLiD sequencing, (developed by Agencourt, later Applied Biosystems, now Life Technologies).[73][74][75] Emulsion PCR is also used in the GemCode and Chromium platforms developed by 10x Genomics.[76]

Genomic DNA is fragmented into random pieces and cloned as a bacterial library. DNA from individual bacterial clones is sequenced and the sequence is assembled by using overlapping DNA regions.

Shotgun sequencing

Shotgun sequencing is a sequencing method designed for analysis of DNA sequences longer than 1000 base pairs, up to and including entire chromosomes. This method requires the target DNA to be broken into random fragments. After sequencing individual fragments using the chain termination method, the sequences can be reassembled on the basis of their overlapping regions.[77]
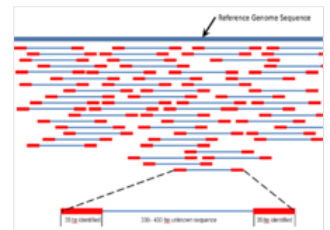
High-throughput methods

High-throughput sequencing, which includes next-generation "short-read" and third-generation "long-read" sequencing methods,[nt 1] applies to exome sequencing, genome sequencing, genome resequencing, transcriptome profiling (RNA-Seq), DNA-protein interactions (ChIP-sequencing), and epigenome characterization.[78]

The high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies that parallelize the sequencing process, producing thousands or millions of sequences concurrently.[79][80][81] High-throughput sequencing technologies are intended to lower the cost of DNA sequencing beyond what is possible with standard dye-terminator methods.[82] In ultra-high-throughput sequencing as many as 500,000 sequencing-by-synthesis operations may be run in parallel.[83][84][85] Such technologies led to the ability to sequence an entire human genome in as little as one day.[86] As of 2019, corporate leaders in the development of high-throughput sequencing products included Illumina, Qiagen and ThermoFisher Scientific.[86]

Multiple, fragmented sequence reads must be assembled together on the basis of their overlapping areas.

Comparison of high-throughput sequencing methods[87][88]

Method	Read length	Accuracy (single read not consensus)	Reads per run	Time per run	Cost per 1 billion bases (in US$)	Advantages	Disadvantages
Single-molecule real-time sequencing (Pacific Biosciences)	30,000 bp (N50); maximum read length >100,000 bases[89][90][91]	87% raw-read accuracy[92]	4,000,000 per Sequel 2 SMRT cell, 100–200 gigabases[89][93][94]	30 minutes to 20 hours[89][95]	$7.2-$43.3	Fast. Detects 4mC, 5mC, 6mA.[96]	Moderate throughput. Equipment can be very expensive.
Ion semiconductor (Ion Torrent sequencing)	up to 600 bp[97]	99.6%[98]	up to 80 million	2 hours	$66.8-$950	Less expensive equipment. Fast.	Homopolymer errors.
Pyrosequencing (454)	700 bp	99.9%	1 million	24 hours	$10,000	Long read size. Fast.	Runs are expensive. Homopolymer errors.
Sequencing by synthesis (Illumina)	MiniSeq, NextSeq: 75–300 bp; MiSeq: 50–600 bp; HiSeq 2500: 50–500 bp; HiSeq 3/4000: 50–300 bp; HiSeq X: 300 bp	99.9% (Phred30)	MiniSeq/MiSeq: 1–25 Million; NextSeq: 130-00 Million; HiSeq 2500: 300 million – 2 billion; HiSeq 3/4000 2.5 billion; HiSeq X: 3 billion	1 to 11 days, depending upon sequencer and specified read length[99]	$5 to $150	Potential for high sequence yield, depending upon sequencer model and desired application.	Equipment can be very expensive. Requires high concentrations of DNA.
Combinatorial probe anchor synthesis (cPAS- BGI/MGI)	BGISEQ-50: 35-50bp; MGISEQ 200: 50-200bp; BGISEQ-500, MGISEQ-2000: 50-300bp[100]	99.9% (Phred30)	BGISEQ-50: 160M; MGISEQ 200: 300M; BGISEQ-500: 1300M per flow cell; MGISEQ-2000: 375M FCS flow cell, 1500M FCL flow cell per flow cell.	1 to 9 days depending on instrument, read length and number of flow cells run at a time.	$5– $120		
Sequencing by ligation (SOLiD sequencing)	50+35 or 50+50 bp	99.9%	1.2 to 1.4 billion	1 to 2 weeks	$60–130	Low cost per base.	Slower than other methods. Has issues sequencing palindromic sequences.[101]
Nanopore Sequencing	Dependent on library preparation, not the device, so user chooses read length (up to 2,272,580 bp reported[102]).	~92–97% single read	dependent on read length selected by user	data streamed in real time. Choose 1 min to 48 hrs	$7–100	Longest individual reads. Accessible user community. Portable (Palm sized).	Lower throughput than other machines, Single read accuracy in 90s.
GenapSys Sequencing	Around 150 bp single-end	99.9% (Phred30)	1 to 16 million	Around 24 hours	$667	Low-cost of instrument ($10,000)	
Chain termination (Sanger sequencing)	400 to 900 bp	99.9%	N/A	20 minutes to 3 hours	$2,400,000	Useful for many applications.	More expensive and impractical for larger sequencing projects. This method also requires the time-consuming step of plasmid cloning or PCR.

Long-read sequencing methods

Single molecule real time (SMRT) sequencing

SMRT sequencing is based on the sequencing by synthesis approach. The DNA is synthesized in zero-mode wave-guides (ZMWs) – small well-like containers with the capturing tools located at the bottom of the well. The sequencing is performed with use of unmodified polymerase (attached to the ZMW bottom) and fluorescently labelled nucleotides flowing freely in the solution. The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected. The fluorescent label is detached from the nucleotide upon its incorporation into the DNA strand, leaving an unmodified DNA strand. According to Pacific Biosciences (PacBio), the SMRT technology developer, this methodology allows detection of nucleotide modifications (such as cytosine methylation). This happens through the observation of polymerase kinetics. This approach allows reads of 20,000 nucleotides or more, with average read lengths of 5 kilobases.[93][103] In 2015, Pacific Biosciences announced the launch of a new sequencing instrument called the Sequel System, with 1 million ZMWs compared to 150,000 ZMWs in the PacBio RS II instrument.[104][105] SMRT sequencing is referred to as "third-generation" or "long-read" sequencing.

Nanopore DNA sequencing

The DNA passing through the nanopore changes its ion current. This change is dependent on the shape, size and length of the DNA sequence. Each type of the nucleotide blocks the ion flow through the pore for a different period of time. The method does not require modified nucleotides and is performed in real time. Nanopore sequencing is referred to as "third-generation" or "long-read" sequencing, along with SMRT sequencing.

Early industrial research into this method was based on a technique called 'exonuclease sequencing', where the readout of electrical signals occurred as nucleotides passed by alpha(α)-hemolysin pores covalently bound with cyclodextrin.[106] However the subsequent commercial method, 'strand sequencing', sequenced DNA bases in an intact strand.

Two main areas of nanopore sequencing in development are solid state nanopore sequencing, and protein based nanopore sequencing. Protein nanopore sequencing utilizes membrane protein complexes such as α-hemolysin, MspA (*Mycobacterium smegmatis* Porin A) or CssG, which show great promise given their ability to distinguish between individual and groups of nucleotides.[107] In contrast, solid-state nanopore sequencing utilizes synthetic materials such as silicon nitride and aluminum oxide and it is preferred for its superior mechanical ability and thermal and chemical stability.[108] The fabrication method is essential for this type of sequencing given that the nanopore array can contain hundreds of pores with diameters smaller than eight nanometers.[107]

The concept originated from the idea that single stranded DNA or RNA molecules can be electrophoretically driven in a strict linear sequence through a biological pore that can be less than eight nanometers, and can be detected given that the molecules release an ionic current while moving through the pore. The pore contains a detection region capable of recognizing different bases, with each base generating various time specific signals corresponding to the sequence of bases as they cross the pore which are then evaluated.[108] Precise control over the DNA transport through the pore is crucial for success. Various enzymes such as exonucleases and polymerases have been used to moderate this process by positioning them near the pore's entrance.[109]

Short-read sequencing methods

Massively parallel signature sequencing (MPSS)

The first of the high-throughput sequencing technologies, massively parallel signature sequencing (or MPSS, also called next generation sequencing), was developed in the 1990s at Lynx Therapeutics, a company founded in 1992 by Sydney Brenner and Sam Eletr. MPSS was a bead-based method that used a complex approach of adapter ligation followed by adapter decoding, reading the sequence in increments of four nucleotides. This method made it susceptible to sequence-specific bias or loss of specific sequences. Because the technology was so complex, MPSS was only performed 'in-house' by Lynx Therapeutics and no DNA sequencing machines were sold to independent laboratories. Lynx Therapeutics merged with Solexa (later acquired by Illumina) in 2004, leading to the development of sequencing-by-synthesis, a simpler approach acquired from Manteia Predictive Medicine, which rendered MPSS obsolete. However, the essential properties of the MPSS output were typical of later high-throughput data types, including hundreds of thousands of short DNA sequences. In the case of MPSS, these were typically used for sequencing cDNA for measurements of gene expression levels.[63]

Polony sequencing

The polony sequencing method, developed in the laboratory of George M. Church at Harvard, was among the first high-throughput sequencing systems and was used to sequence a full *E. coli* genome in 2005.[110] It combined an in vitro paired-tag library with emulsion PCR, an automated microscope, and ligation-based sequencing chemistry to sequence an *E. coli* genome at an accuracy of >99.9999% and a cost approximately 1/9 that of Sanger sequencing.[110] The technology was licensed to Agencourt Biosciences, subsequently spun out into Agencourt Personal Genomics, and eventually incorporated into the Applied Biosystems SOLiD platform. Applied Biosystems was later acquired by Life Technologies, now part of Thermo Fisher Scientific.

454 pyrosequencing

A parallelized version of pyrosequencing was developed by 454 Life Sciences, which has since been acquired by Roche Diagnostics. The method amplifies DNA inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picoliter-volume wells each containing a single bead and
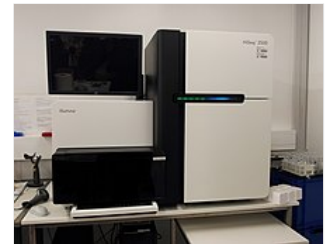
sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA, and the combined data are used to generate sequence reads.[73] This technology provides intermediate read length and price per base compared to Sanger sequencing on one end and Solexa and SOLiD on the other.[82]

Illumina (Solexa) sequencing

Solexa, now part of Illumina, was founded by Shankar Balasubramanian and David Klenerman in 1998, and developed a sequencing method based on reversible dye-terminators technology, and engineered polymerases.[111] The reversible terminated chemistry concept was invented by Bruno Canard and Simon Sarfati at the Pasteur Institute in Paris.[112][113] It was developed internally at Solexa by those named on the relevant patents. In 2004, Solexa acquired the company Manteia Predictive Medicine in order to gain a massively parallel sequencing technology invented in 1997 by Pascal Mayer and Laurent Farinelli.[60] It is based on "DNA clusters" or "DNA colonies", which involves the clonal amplification of DNA on a surface. The cluster technology was co-acquired with Lynx Therapeutics of California. Solexa Ltd. later merged with Lynx to form Solexa Inc.

In this method, DNA molecules and primers are first attached on a slide or flow cell and amplified with polymerase so that local clonal DNA colonies, later coined "DNA clusters", are formed. To determine the sequence, four types of reversible terminator bases (RT-bases) are added and non-incorporated nucleotides are washed away. A camera takes images of the fluorescently labeled nucleotides. Then the dye, along with the terminal 3' blocker, is chemically removed from the DNA, allowing for the next cycle to begin. Unlike pyrosequencing, the DNA chains are extended one nucleotide at a time and image acquisition can be performed at a delayed moment, allowing for very large arrays of DNA colonies to be captured by sequential images taken from a single camera.

An Illumina HiSeq 2500 sequencer

Decoupling the enzymatic reaction and the image capture allows for optimal throughput and theoretically unlimited sequencing capacity. With an optimal configuration, the ultimately reachable instrument throughput is thus dictated solely by the analog-to-digital conversion rate of the camera, multiplied by the number of cameras and divided by the number of pixels per DNA colony required for visualizing them optimally (approximately 10 pixels/colony). In 2012, with cameras operating at more than 10 MHz A/D conversion rates and available optics, fluidics and enzymatics, throughput can be multiples of 1 million nucleotides/second, corresponding roughly to 1 human genome equivalent at 1x coverage per hour per instrument, and 1 human genome re-sequenced (at approx. 30x) per day per instrument (equipped with a single camera).[114]

Illumina NovaSeq 6000 flow cell

Combinatorial probe anchor synthesis (cPAS)

This method is an upgraded modification to combinatorial probe anchor ligation technology (cPAL) described by Complete Genomics[115] which has since become part of Chinese genomics company BGI in 2013.[116] The two companies have refined the technology to allow for longer read lengths, reaction time reductions and faster time to results. In addition, data are now generated as contiguous full-length reads in the standard FASTQ file format and can be used as-is in most short-read-based bioinformatics analysis pipelines.[117]

The two technologies that form the basis for this high-throughput sequencing technology are DNA nanoballs (DNB) and patterned arrays for nanoball attachment to a solid surface.[115] DNA nanoballs are simply formed by denaturing double stranded, adapter ligated libraries and ligating the forward strand only to a splint oligonucleotide to form a ssDNA circle. Faithful copies of the circles containing the DNA insert are produced utilizing Rolling Circle Amplification that generates approximately 300–500 copies. The long strand of ssDNA folds upon itself to produce a three-dimensional nanoball structure that is approximately 220 nm in diameter. Making DNBs replaces the need to generate PCR copies of the library on the flow cell and as such can remove large proportions of duplicate reads, adapter-adapter ligations and PCR induced errors.[117]

An Illumina MiSeq sequencer

The patterned array of positively charged spots is fabricated through photolithography and etching techniques followed by chemical modification to generate a sequencing flow cell. Each spot on the flow cell is approximately 250 nm in diameter, are separated by 700 nm (centre to centre) and allows easy attachment of a single negatively charged DNB to the flow cell and thus reducing under or over-clustering on the flow cell.[115]

Sequencing is then performed by addition of an oligonucleotide probe that attaches in combination to specific sites within the DNB. The probe acts as an anchor that then allows one of four single reversibly inactivated, labelled nucleotides to bind after flowing across the flow cell. Unbound nucleotides are washed away before laser excitation of the attached labels then emit fluorescence and signal is captured by cameras that is converted to a digital output for base calling. The attached base has its terminator and label chemically cleaved at completion of the cycle. The cycle is repeated with another flow of free, labelled nucleotides across the flow cell to allow the next nucleotide to bind and have its signal captured. This process is completed a number of times (usually 50 to 300 times) to determine the sequence of the inserted piece of DNA at a rate of approximately 40 million nucleotides per second as of 2018.

A BGI MGISEQ-2000RS sequencer

SOLiD sequencing

Applied Biosystems' (now a Life Technologies brand) SOLiD technology employs sequencing by ligation. Here, a pool of all possible oligonucleotides of a fixed length are labeled according to the sequenced position. Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position. Each base in the template is sequenced twice, and the resulting data are decoded according to the 2 base encoding scheme used in this method. Before sequencing, the DNA is amplified by emulsion PCR. The resulting

beads, each containing single copies of the same DNA molecule, are deposited on a glass slide.[118] The result is sequences of quantities and lengths comparable to Illumina sequencing.[82] This sequencing by ligation method has been reported to have some issue sequencing palindromic sequences.[101]

Ion Torrent semiconductor sequencing

Ion Torrent Systems Inc. (now owned by Life Technologies) developed a system based on using standard sequencing chemistry, but with a novel, semiconductor-based detection system. This method of sequencing is based on the detection of hydrogen ions that are released during the polymerisation of DNA, as opposed to the optical methods used in other sequencing systems. A microwell containing a template DNA strand to be sequenced is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence, multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal.[119]

Library preparation for the SOLiD platform

DNA nanoball sequencing

DNA nanoball sequencing is a type of high throughput sequencing technology used to determine the entire genomic sequence of an organism. The company Complete Genomics uses this technology to sequence samples submitted by independent researchers. The method uses rolling circle replication to amplify small fragments of genomic DNA into DNA nanoballs. Unchained sequencing by ligation is then used to determine the nucleotide sequence.[120] This method of DNA sequencing allows large numbers of DNA nanoballs to be sequenced per run and at low reagent costs compared to other high-throughput sequencing platforms.[121] However, only short sequences of DNA are determined from each DNA nanoball which makes mapping the short reads to a reference genome difficult.[120]

Two-base encoding scheme. In two-base encoding, each unique pair of bases on the 3' end of the probe is assigned one out of four possible colors. For example, "AA" is assigned to blue, "AC" is assigned to green, and so on for all 16 unique pairs. During sequencing, each base in the template is sequenced twice, and the resulting data are decoded according to this scheme.

Heliscope single molecule sequencing

Heliscope sequencing is a method of single-molecule sequencing developed by Helicos Biosciences. It uses DNA fragments with added poly-A tail adapters which are attached to the flow cell surface. The next steps involve extension-based sequencing with cyclic washes of the flow cell with fluorescently labeled nucleotides (one nucleotide type at a time, as with the Sanger method). The reads are performed by the Heliscope sequencer.[122][123] The reads are short, averaging 35 bp.[124] What made this technology especially novel was that it was the first of its class to sequence non-amplified DNA, thus preventing any read errors associated with amplification steps.[125] In 2009 a human genome was sequenced using the Heliscope, however in 2012 the company went bankrupt.[126]

Microfluidic Systems

There are two main microfluidic systems that are used to sequence DNA; droplet based microfluidics and digital microfluidics. Microfluidic devices solve many of the current limitations of current sequencing arrays.
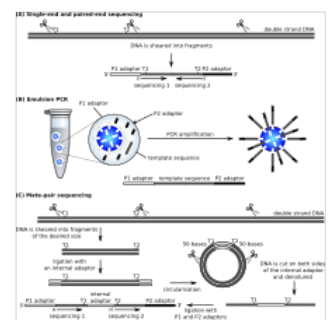
Abate et al. studied the use of droplet-based microfluidic devices for DNA sequencing.[4] These devices have the ability to form and process picoliter sized droplets at the rate of thousands per second. The devices were created from polydimethylsiloxane (PDMS) and used Forster resonance energy transfer, FRET assays to read the sequences of DNA encompassed in the droplets. Each position on the array tested for a specific 15 base sequence.[4]

Sequencing of the TAGGCT template with IonTorrent, PacBioRS and GridION

Fair et al. used digital microfluidic devices to study DNA pyrosequencing.[127] Significant advantages include the portability of the device, reagent volume, speed of analysis, mass manufacturing abilities, and high throughput. This study provided a proof of concept showing that digital devices can be used for pyrosequencing; the study included using synthesis, which involves the extension of the enzymes and addition of labeled nucleotides.[127]
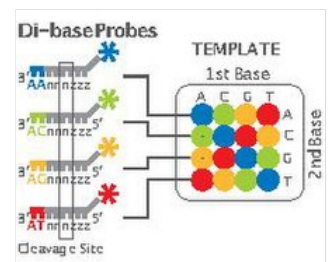
Boles et al. also studied pyrosequencing on digital microfluidic devices.[128] They used an electro-wetting device to create, mix, and split droplets. The sequencing uses a three-enzyme protocol and DNA templates anchored with magnetic beads. The device was tested using two protocols and resulted in 100% accuracy based on raw pyrogram levels. The advantages of these digital microfluidic devices include size, cost, and achievable levels of functional integration.[128]

DNA sequencing research, using microfluidics, also has the ability to be applied to the sequencing of RNA, using similar droplet microfluidic techniques, such as the method, inDrops.[129] This shows that many of these DNA sequencing techniques will be able to be applied further and be used to understand more about genomes and transcriptomes.
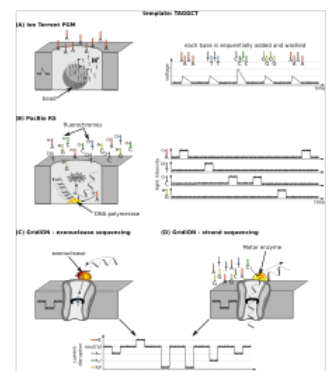
Methods in development

DNA sequencing methods currently under development include reading the sequence as a DNA strand transits through nanopores (a method that is now commercial but subsequent generations such as solid-state nanopores are still in development),[130][131] and microscopy-based techniques, such as atomic force microscopy or transmission electron microscopy that are used to identify the positions of individual nucleotides within long DNA fragments (>5,000 bp) by nucleotide labeling with heavier elements (e.g., halogens) for visual detection and recording.[132][133] Third generation technologies aim to increase throughput and decrease the time to result and cost by eliminating the need for excessive reagents and harnessing the processivity of DNA polymerase.[134]

Tunnelling currents DNA sequencing

Another approach uses measurements of the electrical tunnelling currents across single-strand DNA as it moves through a channel. Depending on its electronic structure, each base affects the tunnelling current differently,[135] allowing differentiation between different bases.[136]

The use of tunnelling currents has the potential to sequence orders of magnitude faster than ionic current methods and the sequencing of several DNA oligomers and micro-RNA has already been achieved.[137]

Sequencing by hybridization

Sequencing by hybridization is a non-enzymatic method that uses a DNA microarray. A single pool of DNA whose sequence is to be determined is fluorescently labeled and hybridized to an array containing known sequences. Strong hybridization signals from a given spot on the array identifies its sequence in the DNA being sequenced.[138]

This method of sequencing utilizes binding characteristics of a library of short single stranded DNA molecules (oligonucleotides), also called DNA probes, to reconstruct a target DNA sequence. Non-specific hybrids are removed by washing and the target DNA is eluted.[139] Hybrids are re-arranged such that the DNA sequence can be reconstructed. The benefit of this sequencing type is its ability to capture a large number of targets with a homogenous coverage.[140] A large number of chemicals and starting DNA is usually required. However, with the advent of solution-based hybridization, much less equipment and chemicals are necessary.[139]

Sequencing with mass spectrometry

Mass spectrometry may be used to determine DNA sequences. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, or MALDI-TOF MS, has specifically been investigated as an alternative method to gel electrophoresis for visualizing DNA fragments. With this method, DNA fragments generated by chain-termination sequencing reactions are compared by mass rather than by size. The mass of each nucleotide is different from the others and this difference is detectable by mass spectrometry. Single-nucleotide mutations in a fragment can be more easily detected with MS than by gel electrophoresis alone. MALDI-TOF MS can more easily detect differences between RNA fragments, so researchers may indirectly sequence DNA with MS-based methods by converting it to RNA first.[141]

The higher resolution of DNA fragments permitted by MS-based methods is of special interest to researchers in forensic science, as they may wish to find single-nucleotide polymorphisms in human DNA samples to identify individuals. These samples may be highly degraded so forensic researchers often prefer mitochondrial DNA for its higher stability and applications for lineage studies. MS-based sequencing methods have been used to compare the sequences of human mitochondrial DNA from samples in a Federal Bureau of Investigation database[142] and from bones found in mass graves of World War I soldiers.[143]

Early chain-termination and TOF MS methods demonstrated read lengths of up to 100 base pairs.[144] Researchers have been unable to exceed this average read size; like chain-termination sequencing alone, MS-based DNA sequencing may not be suitable for large *de novo* sequencing projects. Even so, a recent study did use the short sequence reads and mass spectroscopy to compare single-nucleotide polymorphisms in pathogenic *Streptococcus* strains.[145]

Microfluidic Sanger sequencing

In microfluidic Sanger sequencing the entire thermocycling amplification of DNA fragments as well as their separation by electrophoresis is done on a single glass wafer (approximately 10 cm in diameter) thus reducing the reagent usage as well as cost.[146] In some instances researchers have shown that they can increase the throughput of conventional sequencing through the use of microchips.[147] Research will still need to be done in order to make this use of technology effective.

Microscopy-based techniques

This approach directly visualizes the sequence of DNA molecules using electron microscopy. The first identification of DNA base pairs within intact DNA molecules by enzymatically incorporating modified bases, which contain atoms of increased atomic number, direct visualization and identification of individually labeled bases within a synthetic 3,272 base-pair DNA molecule and a 7,249 base-pair viral genome has been demonstrated.[148]

RNAP sequencing

This method is based on use of RNA polymerase (RNAP), which is attached to a polystyrene bead. One end of DNA to be sequenced is attached to another bead, with both beads being placed in optical traps. RNAP motion during transcription brings the beads in closer and their relative distance changes, which can then be recorded at a single nucleotide resolution. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types, similarly to the Sanger method.[149] A comparison is made between regions and sequence information is deduced by comparing the known sequence regions to the unknown sequence regions.[150]

In vitro virus high-throughput sequencing

A method has been developed to analyze full sets of protein interactions using a combination of 454 pyrosequencing and an *in vitro* virus mRNA display method. Specifically, this method covalently links proteins of interest to the mRNAs encoding them, then detects the mRNA pieces using reverse transcription PCRs. The mRNA may then be amplified and sequenced. The combined method was titled IVV-HiTSeq and can be performed under cell-free conditions, though its results may not be representative of *in vivo* conditions.[151]

Market share

While there are many different ways to sequence DNA, only a few dominate the market. In 2022, Illumina had about 80% of the market; the rest of the market is taken by only a few players (PacBio, Oxford, 454, MGI)[152]

Sample preparation

The success of any DNA sequencing protocol relies upon the DNA or RNA sample extraction and preparation from the biological material of interest.

- A successful DNA extraction will yield a DNA sample with long, non-degraded strands.
- A successful RNA extraction will yield a RNA sample that should be converted to complementary DNA (cDNA) using reverse transcriptase—a DNA polymerase that synthesizes a complementary DNA based on existing strands of RNA in a PCR-like manner.[153] Complementary DNA can then be processed the same way as genomic DNA.

After DNA or RNA extraction, samples may require further preparation depending on the sequencing method. For Sanger sequencing, either cloning procedures or PCR are required prior to sequencing. In the case of next-generation sequencing methods, library preparation is required before processing.[154] Assessing the quality and quantity of nucleic acids both after extraction and after library preparation identifies degraded, fragmented, and low-purity samples and yields high-quality sequencing data.[155]

Development initiatives

In October 2006, the X Prize Foundation established an initiative to promote the development of full genome sequencing technologies, called the Archon X Prize, intending to award $10 million to "the first Team that can build a device and use it to sequence 100 human genomes within 10 days or less, with an accuracy of no more than one error in every 100,000 bases sequenced, with sequences accurately covering at least 98% of the genome, and at a recurring cost of no more than $10,000 (US) per genome."[156]

Each year the National Human Genome Research Institute, or NHGRI, promotes grants for new research and developments in genomics. 2010 grants and 2011 candidates include continuing work in microfluidic, polony and base-heavy sequencing methodologies.[157]
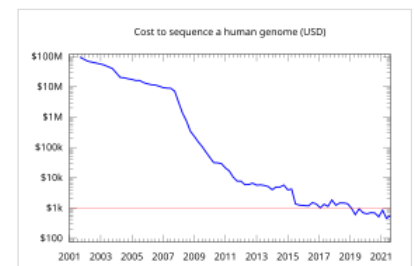
Total cost of sequencing a human genome over time as calculated by the NHGRI.

Computational challenges

The sequencing technologies described here produce raw data that needs to be assembled into longer sequences such as complete genomes (sequence assembly). There are many computational challenges to achieve this, such as the evaluation of the raw sequence data which is done by programs and algorithms such as Phred and Phrap. Other challenges have to deal with repetitive sequences that often prevent complete genome assemblies because they occur in many places of the genome. As a consequence, many sequences may not be assigned to particular chromosomes. The production of raw sequence data is only the beginning of its detailed bioinformatical analysis.[158] Yet new methods for sequencing and correcting sequencing errors were developed.[159]

Read trimming

Sometimes, the raw reads produced by the sequencer are correct and precise only in a fraction of their length. Using the entire read may introduce artifacts in the downstream analyses like genome assembly, SNP calling, or gene expression estimation. Two classes of trimming programs have been introduced, based on the window-based or the running-sum classes of algorithms.[160] This is a partial list of the trimming algorithms currently available, specifying the algorithm class they belong to:

Read Trimming Algorithms

Name of algorithm	Type of algorithm
Cutadapt[161]	Running sum
ConDeTri[162]	Window based
ERNE-FILTER[163]	Running sum
FASTX quality trimmer	Window based
PRINSEQ[164]	Window based
Trimmomatic[165]	Window based
SolexaQA[166]	Window based
SolexaQA-BWA	Running sum
Sickle	Window based

Ethical issues

Human genetics have been included within the field of bioethics since the early 1970s[167] and the growth in the use of DNA sequencing (particularly high-throughput sequencing) has introduced a number of ethical issues. One key issue is the ownership of an individual's DNA and the data produced when that DNA is sequenced.[168] Regarding the DNA molecule itself, the leading legal case on this topic, *Moore v. Regents of the University of California* (1990) ruled that individuals have no property rights to discarded cells or any profits made using these cells (for instance, as a patented cell line). However, individuals have a right to informed consent regarding removal and use of cells. Regarding the data produced through DNA sequencing, *Moore* gives the individual no rights to the information derived from their DNA.[168]

As DNA sequencing becomes more widespread, the storage, security and sharing of genomic data has also become more important.[168][169] For instance, one concern is that insurers may use an individual's genomic data to modify their quote, depending on the perceived future health of the individual based on their DNA.[169][170] In May 2008, the Genetic Information Nondiscrimination Act (GINA) was signed in the United States, prohibiting discrimination on the basis of genetic information with respect to health insurance and employment.[171][172] In 2012, the US Presidential Commission for the Study of Bioethical Issues reported that existing privacy legislation for DNA sequencing data such as GINA and the Health Insurance Portability and Accountability Act were insufficient, noting that whole-genome sequencing data was particularly sensitive, as it could be used to identify not only the individual from which the data was created, but also their relatives.[173][174]

In most of the United States, DNA that is "abandoned", such as that found on a licked stamp or envelope, coffee cup, cigarette, chewing gum, household trash, or hair that has fallen on a public sidewalk, may legally be collected and sequenced by anyone, including the police, private investigators, political opponents, or people involved in paternity disputes. As of 2013, eleven states have laws that can be interpreted to prohibit "DNA theft".[175]

Ethical issues have also been raised by the increasing use of genetic variation screening, both in newborns, and in adults by companies such as 23andMe.[176][177] It has been asserted that screening for genetic variations can be harmful, increasing anxiety in individuals who have been found to have an increased risk of disease.[178] For example, in one case noted in *Time*, doctors screening an ill baby for genetic variants chose not to inform the parents of an unrelated variant linked to dementia due to the harm it would cause to the parents.[179] However, a 2011 study in *The New England Journal of Medicine* has shown that individuals undergoing disease risk profiling did not show increased levels of anxiety.[178] Also, the development of Next Generation sequencing technologies such as Nanopore based sequencing has also raised further ethical concerns.[180]

See also

- Bioinformatics – Computational analysis of large, complex sets of biological data
- Cancer genome sequencing
- Circular consensus sequencing
- DNA computing – Computing using molecular biology hardware
- DNA field-effect transistor – transistor which uses the field-effect due to the partial charges of DNA
- DNA sequencing theory – Biological theory
- DNA sequencer – A scientific instrument used to automate the DNA sequencing process
- Genographic Project – Citizen science project
- Genome project – Scientific endeavours to determine the complete genome sequence of an organism
- Genome sequencing of endangered species – DNA testing for endangerment assessment
- Genome skimming – Method of genome sequencing
- IsoBase – Database for identifying functionally related proteins
- Linked-read sequencing
- Jumping library
- Nucleic acid sequence – Succession of nucleotides in a nucleic acid
- Multiplex ligation-dependent probe amplification
- Personalized medicine – Medical model that tailors medical practices to the individual patient
- Protein sequencing – Sequencing of amino acid arrangement in a protein
- Sequence mining
- Sequence profiling tool

- Sequencing by hybridization – method for determining the constituent nucleotides of a fixed size in a strand of DNA
- Sequencing by ligation – DNA sequencing method that uses the enzyme DNA ligase to identify the nucleotide present at a given position in a DNA sequence
- TIARA (database) – Database of personal genomics information
- Transmission electron microscopy DNA sequencing – Single-molecule sequencing technology

Notes

1. "Next-generation" remains in broad use as of 2019. For instance, Straiton J, Free T, Sawyer A, Martin J (February 2019). "From Sanger Sequencing to Genome Databases and Beyond" (https://doi.org/10.2144%2Fbtn-2019-0011). *BioTechniques*. **66** (2): 60–63. doi:10.2144/btn-2019-0011 (https://doi.org/10.2144%2Fbtn-2019-0011). PMID 30744413 (https://pubmed.ncbi.nlm.nih.gov/30744413). "Next-generation sequencing (NGS) technologies have revolutionized genomic research. (opening sentence of the article)"

References

1. "Introducing 'dark DNA' – the phenomenon that could change how we think about evolution" (https://theconversation.com/introducing-dark-dna-the-phenomenon-that-could-change-how-we-think-about-evolution-82867). 24 August 2017.
2. Behjati S, Tarpey PS (December 2013). "What is next generation sequencing?" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3841808). *Archives of Disease in Childhood: Education and Practice Edition*. **98** (6): 236–8. doi:10.1136/archdischild-2013-304340 (https://doi.org/10.1136%2Farchdischild-2013-304340). PMC 3841808 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3841808). PMID 23986538 (https://pubmed.ncbi.nlm.nih.gov/23986538).
3. Chmielecki J, Meyerson M (14 January 2014). "DNA sequencing of cancer: what have we learned?" (https://doi.org/10.1146%2Fannurev-med-060712-200152). *Annual Review of Medicine*. **65** (1): 63–79. doi:10.1146/annurev-med-060712-200152 (https://doi.org/10.1146%2Fannurev-med-060712-200152). PMID 24274178 (https://pubmed.ncbi.nlm.nih.gov/24274178).
4. Abate AR, Hung T, Sperling RA, Mary P, Rotem A, Agresti JJ, et al. (December 2013). "DNA sequence analysis with droplet-based microfluidics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4090915). *Lab on a Chip*. **13** (24): 4864–9. doi:10.1039/c3lc50905b (https://doi.org/10.1039%2Fc3lc50905b). PMC 4090915 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4090915). PMID 24185402 (https://pubmed.ncbi.nlm.nih.gov/24185402).
5. Pekin D, Skhiri Y, Baret JC, Le Corre D, Mazutis L, Salem CB, et al. (July 2011). "Quantitative and sensitive detection of rare mutations using droplet-based microfluidics". *Lab on a Chip*. **11** (13): 2156–66. doi:10.1039/c1lc20128j (https://doi.org/10.1039%2Fc1lc20128j). PMID 21594292 (https://pubmed.ncbi.nlm.nih.gov/21594292).
6. Olsvik O, Wahlberg J, Petterson B, Uhlén M, Popovic T, Wachsmuth IK, Fields PI (January 1993). "Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in Vibrio cholerae O1 strains" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC262614). *J. Clin. Microbiol.* **31** (1): 22–25. doi:10.1128/JCM.31.1.22-25.1993 (https://doi.org/10.1128%2FJCM.31.1.22-25.1993). PMC 262614 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC262614). PMID 7678018 (https://pubmed.ncbi.nlm.nih.gov/7678018).
7. Pettersson E, Lundeberg J, Ahmadian A (February 2009). "Generations of sequencing technologies" (https://doi.org/10.1016%2Fj.ygeno.2008.10.003). *Genomics*. **93** (2): 105–11. doi:10.1016/j.ygeno.2008.10.003 (https://doi.org/10.1016%2Fj.ygeno.2008.10.003). PMID 18992322 (https://pubmed.ncbi.nlm.nih.gov/18992322).
8. Jay E, Bambara R, Padmanabhan R, Wu R (March 1974). "DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC344020). *Nucleic Acids Research*. **1** (3): 331–53. doi:10.1093/nar/1.3.331 (https://doi.org/10.1093%2Fnar%2F1.3.331). PMC 344020 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC344020). PMID 10793670 (https://pubmed.ncbi.nlm.nih.gov/10793670).
9. Hunt, Katie (17 February 2021). "World's oldest DNA sequenced from a mammoth that lived more than a million years ago" (https://www.cnn.com/2021/02/17/world/mammoth-oldest-dna-million-years-ago-scn/index.html). CNN. Retrieved 17 February 2021.
10. Callaway, Ewen (17 February 2021). "Million-year-old mammoth genomes shatter record for oldest ancient DNA – Permafrost-preserved teeth, up to 1.6 million years old, identify a new kind of mammoth in Siberia" (https://doi.org/10.1038%2Fd41586-021-00436-x). *Nature*. **590** (7847): 537–538. Bibcode:2021Natur.590..537C (https://ui.adsabs.harvard.edu/abs/2021Natur.590..537C). doi:10.1038/d41586-021-00436-x (https://doi.org/10.1038%2Fd41586-021-00436-x). PMID 33597786 (https://pubmed.ncbi.nlm.nih.gov/33597786).
11. Castro, Christina; Marine, Rachel; Ramos, Edward; Ng, Terry Fei Fan (2019). "The effect of variant interference on de novo assembly for viral deep sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7306937). *BMC Genomics*. **21** (1): 421. bioRxiv 10.1101/815480 (https://doi.org/10.1101%2F815480). doi:10.1186/s12864-020-06801-w (https://doi.org/10.1186%2Fs12864-020-06801-w). PMC 7306937 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7306937). PMID 32571214 (https://pubmed.ncbi.nlm.nih.gov/32571214).
12. Wohl, Shirlee; Schaffner, Stephen F.; Sabeti, Pardis C. (2016). "Genomic Analysis of Viral Outbreaks" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210220). *Annual Review of Virology*. **3** (1): 173–195. doi:10.1146/annurev-virology-110615-035747 (https://doi.org/10.1146%2Fannurev-virology-110615-035747). PMC 5210220 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210220). PMID 27501264 (https://pubmed.ncbi.nlm.nih.gov/27501264).
13. Boycott, Kym M.; Vanstone, Megan R.; Bulman, Dennis E.; MacKenzie, Alex E. (October 2013). "Rare-disease genetics in the era of next-generation sequencing: discovery to translation". *Nature Reviews Genetics*. **14** (10): 681–691. doi:10.1038/nrg3555 (https://doi.org/10.1038%2Fnrg3555). PMID 23999272 (https://pubmed.ncbi.nlm.nih.gov/23999272). S2CID 8496181 (https://api.semanticscholar.org/CorpusID:8496181).
14. Bean, Lora; Funke, Birgit; Carlston, Colleen M.; Gannon, Jennifer L.; Kantarci, Sibel; Krock, Bryan L.; Zhang, Shulin; Bayrak-Toydemir, Pinar (March 2020). "Diagnostic gene sequencing panels: from design to report—a technical standard of the American College of Medical Genetics and Genomics (ACMG)" (https://doi.org/10.1038%2Fs41436-019-0666-z). *Genetics in Medicine*. **22** (3): 453–461. doi:10.1038/s41436-019-0666-z (https://doi.org/10.1038%2Fs41436-019-0666-z). ISSN 1098-3600 (https://search.worldcat.org/issn/1098-3600). PMID 31732716 (https://pubmed.ncbi.nlm.nih.gov/31732716).
15. Schleusener V, Köser CU, Beckert P, Niemann S, Feuerriegel S (2017). "*Mycobacterium tuberculosis* resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7365310). *Sci Rep*. **7**: 46327. Bibcode:2017NatSR...746327S (https://ui.adsabs.harvard.edu/abs/2017NatSR...746327S). doi:10.1038/srep46327 (https://doi.org/10.1038%2Fsrep46327). PMC 7365310 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7365310). PMID 28425484 (https://pubmed.ncbi.nlm.nih.gov/28425484).

16. Mahé P, El Azami M, Barlas P, Tournoud M (2019). "A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in *Mycobacterium tuberculosis*" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6500375). *PeerJ*. **7**: e6857. doi:10.7717/peerj.6857 (https://doi.org/10.7717%2Fpeerj.6857). PMC 6500375 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6500375). PMID 31106066 (https://pubmed.ncbi.nlm.nih.gov/31106066).

17. Mykrobe predictor –Antibiotic resistance prediction for S. aureus and M. tuberculosis from whole genome sequence data (https://innovation.ox.ac.uk/wp-content/uploads/2015/04/Mykrobe-predictor-Poster.pdf)

18. Bradley, Phelim; Gordon, N. Claire; Walker, Timothy M.; Dunn, Laura; Heys, Simon; Huang, Bill; Earle, Sarah; Pankhurst, Louise J.; Anson, Luke; de Cesare, Mariateresa; Piazza, Paolo; Votintseva, Antonina A.; Golubchik, Tanya; Wilson, Daniel J.; Wyllie, David H.; Diel, Roland; Niemann, Stefan; Feuerriegel, Silke; Kohl, Thomas A.; Ismail, Nazir; Omar, Shaheed V.; Smith, E. Grace; Buck, David; McVean, Gil; Walker, A. Sarah; Peto, Tim E. A.; Crook, Derrick W.; Iqbal, Zamin (21 December 2015). "Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4703848). *Nature Communications*. **6** (1): 10063. Bibcode:2015NatCo...610063B (https://ui.adsabs.harvard.edu/abs/2015NatCo...610063B). doi:10.1038/ncomms10063 (https://doi.org/10.1038%2Fncomms10063). PMC 4703848 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4703848). PMID 26686880 (https://pubmed.ncbi.nlm.nih.gov/26686880).

19. "Michael Mosley vs the superbugs" (https://web.archive.org/web/20201124174040/https://www.tvo.org/transcript/115187X/michael-mosley-vs-the-superbugs). Archived from the original (https://www.tvo.org/transcript/115187X/michael-mosley-vs-the-superbugs) on 24 November 2020. Retrieved 21 October 2019.

20. *Mykrobe* (https://github.com/Mykrobe-tools/mykrobe), Mykrobe-tools, 24 December 2022, retrieved 2 January 2023

21. Curtis C, Hereward J (29 August 2017). "From the crime scene to the courtroom: the journey of a DNA sample" (https://theconversation.com/from-the-crime-scene-to-the-courtroom-the-journey-of-a-dna-sample-82250). *The Conversation*.

22. Moréra S, Larivière L, Kurzeck J, Aschke-Sonnenborn U, Freemont PS, Janin J, Rüger W (August 2001). "High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding". *Journal of Molecular Biology*. **311** (3): 569–77. doi:10.1006/jmbi.2001.4905 (https://doi.org/10.1006%2Fjmbi.2001.4905). PMID 11493010 (https://pubmed.ncbi.nlm.nih.gov/11493010).

23. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C (April 1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC320645). *Nucleic Acids Research*. **10** (8): 2709–21. doi:10.1093/nar/10.8.2709 (https://doi.org/10.1093%2Fnar%2F10.8.2709). PMC 320645 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC320645). PMID 7079182 (https://pubmed.ncbi.nlm.nih.gov/7079182).

24. Ehrlich M, Wang RY (June 1981). "5-Methylcytosine in eukaryotic DNA". *Science*. **212** (4501): 1350–7. Bibcode:1981Sci...212.1350E (https://ui.adsabs.harvard.edu/abs/1981Sci...212.1350E). doi:10.1126/science.6262918 (https://doi.org/10.1126%2Fscience.6262918). PMID 6262918 (https://pubmed.ncbi.nlm.nih.gov/6262918).

25. Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, et al. (November 2011). "Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3646335). *Nature Methods*. **9** (1): 75–7. doi:10.1038/nmeth.1779 (https://doi.org/10.1038%2Fnmeth.1779). PMC 3646335 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3646335). PMID 22101853 (https://pubmed.ncbi.nlm.nih.gov/22101853).

26. Czernecki, Dariusz; Bonhomme, Frédéric; Kaminski, Pierre-Alexandre; Delarue, Marc (5 August 2021). "Characterization of a triad of genes in cyanophage S-2L sufficient to replace adenine by 2-aminoadenine in bacterial DNA" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8342488). *Nature Communications*. **12** (1): 4710. Bibcode:2021NatCo..12.4710C (https://ui.adsabs.harvard.edu/abs/2021NatCo..12.4710C). doi:10.1038/s41467-021-25064-x (https://doi.org/10.1038%2Fs41467-021-25064-x). PMC 8342488 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8342488). PMID 34354070 (https://pubmed.ncbi.nlm.nih.gov/34354070). S2CID 233745192 (https://api.semanticscholar.org/CorpusID:233745192).

27. "Direct detection and sequencing of damaged DNA bases" (https://www.pacb.com/publications/direct-detection-and-sequencing-of-damaged-dna-bases/). *PacBio*. Retrieved 31 July 2024.

28. Watson JD, Crick FH (1953). "The structure of DNA". *Cold Spring Harb. Symp. Quant. Biol.* **18**: 123–31. doi:10.1101/SQB.1953.018.01.020 (https://doi.org/10.1101%2FSQB.1953.018.01.020). PMID 13168976 (https://pubmed.ncbi.nlm.nih.gov/13168976).

29. Marks, L. "The path to DNA sequencing: The life and work of Frederick Sanger" (http://www.whatisbiotechnology.org/exhibitions/sanger/path). *What is Biotechnology?*. Retrieved 27 June 2023.

30. Min Jou W, Haegeman G, Ysebaert M, Fiers W (May 1972). "Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein". *Nature*. **237** (5350): 82–8. Bibcode:1972Natur.237...82J (https://ui.adsabs.harvard.edu/abs/1972Natur.237...82J). doi:10.1038/237082a0 (https://doi.org/10.1038%2F237082a0). PMID 4555447 (https://pubmed.ncbi.nlm.nih.gov/4555447). S2CID 4153893 (https://api.semanticscholar.org/CorpusID:4153893).

31. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (April 1976). "Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature*. **260** (5551): 500–7. Bibcode:1976Natur.260..500F (https://ui.adsabs.harvard.edu/abs/1976Natur.260..500F). doi:10.1038/260500a0 (https://doi.org/10.1038%2F260500a0). PMID 1264203 (https://pubmed.ncbi.nlm.nih.gov/1264203). S2CID 4289674 (https://api.semanticscholar.org/CorpusID:4289674).

32. Ozsolak F, Milos PM (February 2011). "RNA sequencing: advances, challenges and opportunities" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3031867). *Nature Reviews Genetics*. **12** (2): 87–98. doi:10.1038/nrg2934 (https://doi.org/10.1038%2Fnrg2934). PMC 3031867 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3031867). PMID 21191423 (https://pubmed.ncbi.nlm.nih.gov/21191423).

33. "Ray Wu Faculty Profile" (https://web.archive.org/web/20090304121126/http://www.mbg.cornell.edu/faculty-staff/faculty/wu.cfm). Cornell University. Archived from the original (http://www.mbg.cornell.edu/faculty-staff/faculty/wu.cfm) on 4 March 2009.

34. Padmanabhan R, Jay E, Wu R (June 1974). "Chemical synthesis of a primer and its use in the sequence analysis of the lysozyme gene of bacteriophage T4" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC388489). *Proceedings of the National Academy of Sciences of the United States of America*. **71** (6): 2510–4. Bibcode:1974PNAS...71.2510P (https://ui.adsabs.harvard.edu/abs/1974PNAS...71.2510P). doi:10.1073/pnas.71.6.2510 (https://doi.org/10.1073%2Fpnas.71.6.2510). PMC 388489 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC388489). PMID 4526223 (https://pubmed.ncbi.nlm.nih.gov/4526223).

35. Onaga LA (June 2014). "Ray Wu as Fifth Business: Demonstrating Collective Memory in the History of DNA Sequencing". *Studies in the History and Philosophy of Science*. Part C. **46**: 1–14. doi:10.1016/j.shpsc.2013.12.006 (https://doi.org/10.1016%2Fj.shpsc.2013.12.006). PMID 24565976 (https://pubmed.ncbi.nlm.nih.gov/24565976).

36. Wu R (1972). "Nucleotide sequence analysis of DNA". *Nature New Biology*. **236** (68): 198–200. doi:10.1038/newbio236198a0 (https://doi.org/10.1038%2Fnewbio236198a0). PMID 4553110 (https://pubmed.ncbi.nlm.nih.gov/4553110).

37. Padmanabhan R, Wu R (1972). "Nucleotide sequence analysis of DNA. IX. Use of oligonucleotides of defined sequence as primers in DNA sequence analysis". *Biochem. Biophys. Res. Commun*. **48** (5): 1295–302. doi:10.1016/0006-291X(72)90852-2 (https://doi.org/10.1016%2F0006-291X%2872%2990852-2). PMID 4560009 (https://pubmed.ncbi.nlm.nih.gov/4560009).

38. Wu R, Tu CD, Padmanabhan R (1973). "Nucleotide sequence analysis of DNA. XII. The chemical synthesis and sequence analysis of a dodecadeoxynucleotide which binds to the endolysin gene of bacteriophage lambda". *Biochem. Biophys. Res. Commun*. **55** (4): 1092–99. doi:10.1016/S0006-291X(73)80007-5 (https://doi.org/10.1016%2FS0006-291X%2873%2980007-5). PMID 4358929 (https://pubmed.ncbi.nlm.nih.gov/4358929).

39. Maxam AM, Gilbert W (February 1977). "A new method for sequencing DNA" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392330). *Proc. Natl. Acad. Sci. USA*. **74** (2): 560–64. Bibcode:1977PNAS...74..560M (https://ui.adsabs.harvard.edu/abs/1977PNAS...74..560M). doi:10.1073/pnas.74.2.560 (https://doi.org/10.1073%2Fpnas.74.2.560). PMC 392330 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC392330). PMID 265521 (https://pubmed.ncbi.nlm.nih.gov/265521).

40. Gilbert, W. DNA sequencing and gene structure (http://nobelprize.org/nobel_prizes/chemistry/laureates/1980/gilbert-lecture.pdf). Nobel lecture, 8 December 1980.

41. Gilbert W, Maxam A (December 1973). "The Nucleotide Sequence of the lac Operator" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC427284). *Proc. Natl. Acad. Sci. U.S.A*. **70** (12): 3581–84. Bibcode:1973PNAS...70.3581G (https://ui.adsabs.harvard.edu/abs/1973PNAS...70.3581G). doi:10.1073/pnas.70.12.3581 (https://doi.org/10.1073%2Fpnas.70.12.3581). PMC 427284 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC427284). PMID 4587255 (https://pubmed.ncbi.nlm.nih.gov/4587255).

42. Deharvengt, Sophie J.; Petersen, Lauren M.; Jung, Hou-Sung; Tsongalis, Gregory J. (2020). "Nucleic acid analysis in the clinical laboratory". *Contemporary Practice in Clinical Chemistry*. pp. 215–234. doi:10.1016/B978-0-12-815499-1.00013-2 (https://doi.org/10.1016%2FB978-0-12-815499-1.00013-2). ISBN 978-0-12-815499-1.

43. Heather, James M.; Chain, Benjamin (January 2016). "The sequence of sequencers: The history of sequencing DNA" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727787). *Genomics*. **107** (1): 1–8. doi:10.1016/j.ygeno.2015.11.003 (https://doi.org/10.1016%2Fj.ygeno.2015.11.003). PMC 4727787 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727787). PMID 26554401 (https://pubmed.ncbi.nlm.nih.gov/26554401).

44. Elsayed, Fadwa A.; Grolleman, Judith E.; Ragunathan, Abiramy; Buchanan, Daniel D.; van Wezel, Tom; de Voer, Richarda M.; Boot, Arnoud; Stojovska, Marija Staninova; Mahmood, Khalid; Clendenning, Mark; de Miranda, Noel; Dymerska, Dagmara; Egmond, Demi van; Gallinger, Steven; Georgeson, Peter; Hoogerbrugge, Nicoline; Hopper, John L.; Jansen, Erik A.M.; Jenkins, Mark A.; Joo, Jihoon E.; Kuiper, Roland P.; Ligtenberg, Marjolijn J.L.; Lubinski, Jan; Macrae, Finlay A.; Morreau, Hans; Newcomb, Polly; Nielsen, Maartje; Palles, Claire; Park, Daniel J.; Pope, Bernard J.; Rosty, Christophe; Ruiz Ponte, Clara; Schackert, Hans K.; Sijmons, Rolf H.; Tomlinson, Ian P.; Tops, Carli M.J.; Vreede, Lilian; Walker, Romy; Win, Aung K. (December 2020). "Monoallelic NTHL1 Loss-of-Function Variants and Risk of Polyposis and Colorectal Cancer". *Gastroenterology*. **159** (6): 2241–2243.e6. doi:10.1053/j.gastro.2020.08.042 (https://doi.org/10.1053%2Fj.gastro.2020.08.042). hdl:2066/228713 (https://hdl.handle.net/2066%2F228713). PMID 32860789 (https://pubmed.ncbi.nlm.nih.gov/32860789).

45. Sanger F, Nicklen S, Coulson AR (December 1977). "DNA sequencing with chain-terminating inhibitors" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). *Proc. Natl. Acad. Sci. USA*. **74** (12): 5463–77. Bibcode:1977PNAS...74.5463S (https://ui.adsabs.harvard.edu/abs/1977PNAS...74.5463S). doi:10.1073/pnas.74.12.5463 (https://doi.org/10.1073%2Fpnas.74.12.5463). PMC 431765 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765). PMID 271968 (https://pubmed.ncbi.nlm.nih.gov/271968).

46. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (February 1977). "Nucleotide sequence of bacteriophage phi X174 DNA". *Nature*. **265** (5596): 687–95. Bibcode:1977Natur.265..687S (https://ui.adsabs.harvard.edu/abs/1977Natur.265..687S). doi:10.1038/265687a0 (https://doi.org/10.1038%2F265687a0). PMID 870828 (https://pubmed.ncbi.nlm.nih.gov/870828). S2CID 4206886 (https://api.semanticscholar.org/CorpusID:4206886).

47. Marks, L. "The next frontier: Human viruses" (https://www.whatisbiotechnology.org/index.php/exhibitions/sanger/sequencing). *What is Biotechnology?*. Retrieved 27 June 2023.

48. Beck S, Pohl FM (1984). "DNA sequencing with direct blotting electrophoresis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC557787). *EMBO J*. **3** (12): 2905–09. doi:10.1002/j.1460-2075.1984.tb02230.x (https://doi.org/10.1002%2Fj.1460-2075.1984.tb02230.x). PMC 557787 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC557787). PMID 6396083 (https://pubmed.ncbi.nlm.nih.gov/6396083).

49. United States Patent 4,631,122 (1986)

50. Feldmann H, et al. (1994). "Complete DNA sequence of yeast chromosome II" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395553). *EMBO J*. **13** (24): 5795–809. doi:10.1002/j.1460-2075.1994.tb06923.x (https://doi.org/10.1002%2Fj.1460-2075.1994.tb06923.x). PMC 395553 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395553). PMID 7813418 (https://pubmed.ncbi.nlm.nih.gov/7813418).

51. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE (12 June 1986). "Fluorescence Detection in Automated DNA Sequence Analysis". *Nature*. **321** (6071): 674–79. Bibcode:1986Natur.321..674S (https://ui.adsabs.harvard.edu/abs/1986Natur.321..674S). doi:10.1038/321674a0 (https://doi.org/10.1038%2F321674a0). PMID 3713851 (https://pubmed.ncbi.nlm.nih.gov/3713851). S2CID 27800972 (https://api.semanticscholar.org/CorpusID:27800972).

52. Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K (16 October 1987). "A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides". *Science*. **238** (4825): 336–41. Bibcode:1987Sci...238..336P (https://ui.adsabs.harvard.edu/abs/1987Sci...238..336P). doi:10.1126/science.2443975 (https://doi.org/10.1126%2Fscience.2443975). PMID 2443975 (https://pubmed.ncbi.nlm.nih.gov/2443975).

53. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF (June 1991). "Complementary DNA sequencing: expressed sequence tags and human genome project". *Science*. **252** (5013): 1651–56. Bibcode:1991Sci...252.1651A (https://ui.adsabs.harvard.edu/abs/1991Sci...252.1651A). doi:10.1126/science.2047873 (https://doi.org/10.1126%2Fscience.2047873). PMID 2047873 (https://pubmed.ncbi.nlm.nih.gov/2047873). S2CID 13436211 (https://api.semanticscholar.org/CorpusID:13436211).

54. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM (July 1995). "Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*". *Science*. **269** (5223): 496–512. Bibcode:1995Sci...269..496F (https://ui.adsabs.harvard.edu/abs/1995Sci...269..496F). doi:10.1126/science.7542800 (https://doi.org/10.1126%2Fscience.7542800). PMID 7542800 (https://pubmed.ncbi.nlm.nih.gov/7542800).

55. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (February 2001). "Initial sequencing and analysis of the human genome" (https://deepblue.lib.umich.edu/bitstream/2027.42/62798/1/409860a0.pdf) (PDF). *Nature*. **409** (6822): 860–921. Bibcode:2001Natur.409..860L (https://ui.adsabs.harvard.edu/abs/2001Natur.409..860L). doi:10.1038/35057062 (https://doi.org/10.1038%2F35057062). PMID 11237011 (https://pubmed.ncbi.nlm.nih.gov/11237011).

56. Venter JC, Adams MD, et al. (February 2001). "The sequence of the human genome" (https://doi.org/10.1126%2Fscience.10580 40). *Science*. **291** (5507): 1304–51. Bibcode:2001Sci...291.1304V (https://ui.adsabs.harvard.edu/ab s/2001Sci...291.1304V). doi:10.1126/science.1058040 (https://d oi.org/10.1126%2Fscience.1058040). PMID 11181995 (https://p ubmed.ncbi.nlm.nih.gov/11181995).

57. Yang, Aimin; Zhang, Wei; Wang, Jiahao; Yang, Ke; Han, Yang; Zhang, Limin (2020). "Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA" (http s://www.ncbi.nlm.nih.gov/pmc/articles/PMC7498545). *Frontiers in Bioengineering and Biotechnology*. **8**: 1032. doi:10.3389/fbioe.2020.01032 (https://doi.org/10.3389%2Ffbioe. 2020.01032). PMC 7498545 (https://www.ncbi.nlm.nih.gov/pmc/ articles/PMC7498545). PMID 33015010 (https://pubmed.ncbi.nl m.nih.gov/33015010).

58. "Espacenet – Bibliographic data" (http://worldwide.espacenet.co m/publicationDetails/biblio?FT=D&date=19910516&DB=EPOD OC&locale=en_EP&CC=WO&NR=9106678A1&KC=A1&ND=4). *worldwide.espacenet.com*.

59. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P (1996). "Real-time DNA sequencing using detection of pyrophosphate release". *Analytical Biochemistry*. **242** (1): 84–89. doi:10.1006/abio.1996.0432 (https://doi.org/10.1006%2Fabi o.1996.0432). PMID 8923969 (https://pubmed.ncbi.nlm.nih.gov/ 8923969).

60. Kawashima, Eric H.; Laurent Farinelli; Pascal Mayer (12 May 2005). "Patent: Method of nucleic acid amplification" (https://arc hive.today/20130222020134/http://www.patentlens.net/patentlen s/patent/WO_1998_044151_A1/en/). Archived from the original (http://www.patentlens.net/patentlens/patent/WO_1998_044151 _A1/en/) on 22 February 2013. Retrieved 22 December 2012.

61. Ewing B, Green P (March 1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities" (https://doi. org/10.1101%2Fgr.8.3.186). *Genome Res*. **8** (3): 186–94. doi:10.1101/gr.8.3.186 (https://doi.org/10.1101%2Fgr.8.3.186). PMID 9521922 (https://pubmed.ncbi.nlm.nih.gov/9521922).

62. "Quality Scores for Next-Generation Sequencing" (https://www.il lumina.com/documents/products/technotes/technote_Q-Scores. pdf) (PDF). *Illumina*. 31 October 2011. Retrieved 8 May 2018.

63. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, Roth R, George D, Eletr S, Albrecht G, Vermaas E, Williams SR, Moon K, Burcham T, Pallas M, DuBridge RB, Kirchner J, Fearon K, Mao J, Corcoran K (2000). "Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays". *Nature Biotechnology*. **18** (6): 630–34. doi:10.1038/76469 (https://doi.org/10.1038%2F76469). PMID 10835600 (https://pubmed.ncbi.nlm.nih.gov/10835600). S2CID 13884154 (https://api.semanticscholar.org/CorpusID:138 84154).

64. "maxam gilbert sequencing" (https://pubmed.ncbi.nlm.nih.gov/?t erm=maxam+gilbert+sequencing&filter=years.2013-2023&sort= pubdate). *PubMed*.

65. Sanger F, Coulson AR (May 1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". *J. Mol. Biol*. **94** (3): 441–48. doi:10.1016/0022-2836(75)90213-2 (https://doi.org/10.1016%2F0022-2836%287 5%2990213-2). PMID 1100841 (https://pubmed.ncbi.nlm.nih.go v/1100841).

66. Wetterstrand, Kris. "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)" (https://www.gen ome.gov/sequencingcosts). National Human Genome Research Institute. Retrieved 30 May 2013.

67. Nyren, P.; Pettersson, B.; Uhlen, M. (January 1993). "Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay". *Analytical Biochemistry*. **208** (1): 171–175. doi:10.1006/abio.1993.1024 (ht tps://doi.org/10.1006%2Fabio.1993.1024). PMID 8382019 (http s://pubmed.ncbi.nlm.nih.gov/8382019).

68. Ronaghi, Mostafa; Uhlén, Mathias; Nyrén, Pål (17 July 1998). "A Sequencing Method Based on Real-Time Pyrophosphate". *Science*. **281** (5375): 363–365. doi:10.1126/science.281.5375.363 (https://doi.org/10.1126%2Fs cience.281.5375.363). PMID 9705713 (https://pubmed.ncbi.nlm. nih.gov/9705713). S2CID 26331871 (https://api.semanticschola r.org/CorpusID:26331871).

69. Quail MA, Gu Y, Swerdlow H, Mayho M (2012). "Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation". *Electrophoresis*. **33** (23): 3521–28. doi:10.1002/elps.201200128 (https://doi.org/10.1 002%2Felps.201200128). PMID 23147856 (https://pubmed.ncb i.nlm.nih.gov/23147856). S2CID 39818212 (https://api.semantic scholar.org/CorpusID:39818212).

70. Duhaime MB, Deng L, Poulos BT, Sullivan MB (2012). "Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method" (https://www.ncb i.nlm.nih.gov/pmc/articles/PMC3466414). *Environ. Microbiol*. **14** (9): 2526–37. Bibcode:2012EnvMi..14.2526D (https://ui.adsabs. harvard.edu/abs/2012EnvMi..14.2526D). doi:10.1111/j.1462-2920.2012.02791.x (https://doi.org/10.1111%2Fj.1462-2920.201 2.02791.x). PMC 3466414 (https://www.ncbi.nlm.nih.gov/pmc/ar ticles/PMC3466414). PMID 22713159 (https://pubmed.ncbi.nlm. nih.gov/22713159).

71. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012). "Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC336503 4). *PLOS ONE*. **7** (5): e37135. Bibcode:2012PLoSO...737135P (https://ui.adsabs.harvard.edu/abs/2012PLoSO...737135P). doi:10.1371/journal.pone.0037135 (https://doi.org/10.1371%2Fj ournal.pone.0037135). PMC 3365034 (https://www.ncbi.nlm.nih. gov/pmc/articles/PMC3365034). PMID 22675423 (https://pubme d.ncbi.nlm.nih.gov/22675423).

72. Williams R, Peisajovich SG, Miller OJ, Magdassi S, Tawfik DS, Griffiths AD (2006). "Amplification of complex gene libraries by emulsion PCR". *Nature Methods*. **3** (7): 545–50. doi:10.1038/nmeth896 (https://doi.org/10.1038%2Fnmeth896). PMID 16791213 (https://pubmed.ncbi.nlm.nih.gov/16791213). S2CID 27459628 (https://api.semanticscholar.org/CorpusID:274 59628).

73. Margulies M, Egholm M, et al. (September 2005). "Genome Sequencing in Open Microfabricated High Density Picoliter Reactors" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC14644 27). *Nature*. **437** (7057): 376–80. Bibcode:2005Natur.437..376M (https://ui.adsabs.harvard.edu/abs/2005Natur.437..376M). doi:10.1038/nature03959 (https://doi.org/10.1038%2Fnature039 59). PMC 1464427 (https://www.ncbi.nlm.nih.gov/pmc/articles/P MC1464427). PMID 16056220 (https://pubmed.ncbi.nlm.nih.go v/16056220).

74. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (2005). "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome" (https://doi.org/10.1126%2Fscience.111738 9). *Science*. **309** (5741): 1728–32. Bibcode:2005Sci...309.1728S (https://ui.adsabs.harvard.edu/ab s/2005Sci...309.1728S). doi:10.1126/science.1117389 (https://d oi.org/10.1126%2Fscience.1117389). PMID 16081699 (https://p ubmed.ncbi.nlm.nih.gov/16081699). S2CID 11405973 (https://a pi.semanticscholar.org/CorpusID:11405973).

75. "Applied Biosystems – File Not Found (404 Error)" (https://web. archive.org/web/20080516181322/http://solid.appliedbiosystem s.com/). 16 May 2008. Archived from the original (http://solid.ap pliedbiosystems.com/) on 16 May 2008.

76. Goodwin S, McPherson JD, McCombie WR (May 2016). "Coming of age: ten years of next-generation sequencing technologies" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10 373632). *Nature Reviews Genetics*. **17** (6): 333–51. doi:10.1038/nrg.2016.49 (https://doi.org/10.1038%2Fnrg.2016.4 9). PMC 10373632 (https://www.ncbi.nlm.nih.gov/pmc/articles/P MC10373632). PMID 27184599 (https://pubmed.ncbi.nlm.nih.go v/27184599). S2CID 8295541 (https://api.semanticscholar.org/C orpusID:8295541).

77. Staden R (11 June 1979). "A strategy of DNA sequencing employing computer programs" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327874). *Nucleic Acids Research*. **6** (7): 2601–10. doi:10.1093/nar/6.7.2601 (https://doi.org/10.1093%2Fnar%2F6.7.2601). PMC 327874 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC327874). PMID 461197 (https://pubmed.ncbi.nlm.nih.gov/461197).

78. de Magalhães JP, Finch CE, Janssens G (2010). "Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2878865). *Ageing Research Reviews*. **9** (3): 315–23. doi:10.1016/j.arr.2009.10.006 (https://doi.org/10.1016%2Fj.arr.2009.10.006). PMC 2878865 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2878865). PMID 19900591 (https://pubmed.ncbi.nlm.nih.gov/19900591).

79. Grada A (August 2013). "Next-generation sequencing: methodology and application" (https://doi.org/10.1038%2Fjid.2013.248). *J Invest Dermatol*. **133** (8): e11. doi:10.1038/jid.2013.248 (https://doi.org/10.1038%2Fjid.2013.248). PMID 23856935 (https://pubmed.ncbi.nlm.nih.gov/23856935).

80. Hall N (May 2007). "Advanced sequencing technologies and their wider impact in microbiology" (https://doi.org/10.1242%2Fjeb.001370). *J. Exp. Biol.* **210** (Pt 9): 1518–25. Bibcode:2007JExpB.210.1518H (https://ui.adsabs.harvard.edu/abs/2007JExpB.210.1518H). doi:10.1242/jeb.001370 (https://doi.org/10.1242%2Fjeb.001370). PMID 17449817 (https://pubmed.ncbi.nlm.nih.gov/17449817).

81. Church GM (January 2006). "Genomes for all". *Sci. Am.* **294** (1): 46–54. Bibcode:2006SciAm.294a..46C (https://ui.adsabs.harvard.edu/abs/2006SciAm.294a..46C). doi:10.1038/scientificamerican0106-46 (https://doi.org/10.1038%2Fscientificamerican0106-46). PMID 16468433 (https://pubmed.ncbi.nlm.nih.gov/16468433). S2CID 28769137 (https://api.semanticscholar.org/CorpusID:28769137).(subscription required)

82. Schuster SC (January 2008). "Next-generation sequencing transforms today's biology". *Nat. Methods*. **5** (1): 16–18. doi:10.1038/nmeth1156 (https://doi.org/10.1038%2Fnmeth1156). PMID 18165802 (https://pubmed.ncbi.nlm.nih.gov/18165802). S2CID 1465786 (https://api.semanticscholar.org/CorpusID:1465786).

83. Kalb, Gilbert; Moxley, Robert (1992). *Massively Parallel, Optical, and Neural Computing in the United States*. IOS Press. ISBN 978-90-5199-097-3.

84. ten Bosch JR, Grody WW (2008). "Keeping Up with the Next Generation" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570630). *The Journal of Molecular Diagnostics*. **10** (6): 484–92. doi:10.2353/jmoldx.2008.080027 (https://doi.org/10.2353%2Fjmoldx.2008.080027). PMC 2570630 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570630). PMID 18832462 (https://pubmed.ncbi.nlm.nih.gov/18832462).

85. Tucker T, Marra M, Friedman JM (2009). "Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725244). *The American Journal of Human Genetics*. **85** (2): 142–54. doi:10.1016/j.ajhg.2009.06.022 (https://doi.org/10.1016%2Fj.ajhg.2009.06.022). PMC 2725244 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725244). PMID 19679224 (https://pubmed.ncbi.nlm.nih.gov/19679224).

86. Straiton J, Free T, Sawyer A, Martin J (February 2019). "From Sanger sequencing to genome databases and beyond" (https://doi.org/10.2144%2Fbtn-2019-0011). *BioTechniques*. **66** (2). Future Science: 60–63. doi:10.2144/btn-2019-0011 (https://doi.org/10.2144%2Fbtn-2019-0011). PMID 30744413 (https://pubmed.ncbi.nlm.nih.gov/30744413).

87. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (1 January 2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and illumina MiSeq sequencers" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431227). *BMC Genomics*. **13** (1): 341. doi:10.1186/1471-2164-13-341 (https://doi.org/10.1186%2F1471-2164-13-341). PMC 3431227 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431227). PMID 22827831 (https://pubmed.ncbi.nlm.nih.gov/22827831).

88. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (1 January 2012). "Comparison of Next-Generation Sequencing Systems" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3398667). *Journal of Biomedicine and Biotechnology*. **2012**: 251364. doi:10.1155/2012/251364 (https://doi.org/10.1155%2F2012%2F251364). PMC 3398667 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3398667). PMID 22829749 (https://pubmed.ncbi.nlm.nih.gov/22829749).

89. "New Software, Polymerase for Sequel System Boost Throughput and Affordability – PacBio" (https://www.pacb.com/blog/new-software-polymerase-sequel-system-boost-throughput-affordability/). 7 March 2018.

90. "After a Year of Testing, Two Early PacBio Customers Expect More Routine Use of RS Sequencer in 2012" (http://www.genomeweb.com/sequencing/after-year-testing-two-early-pacbio-customers-expect-more-routine-use-rs-sequenc). GenomeWeb. 10 January 2012.(registration required)

91. "Pacific Biosciences Introduces New Chemistry With Longer Read Lengths to Detect Novel Features in DNA Sequence and Advance Genome Studies of Large Organisms" (http://globenewswire.com/news-release/2013/10/03/577891/10051072/en/Pacific-Biosciences-Introduces-New-Chemistry-With-Longer-Read-Lengths-to-Detect-Novel-Features-in-DNA-Sequence-and-Advance-Genome-Studies-of-Large-Organisms.html) (Press release). 2013.

92. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J (2013). "Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data". *Nat. Methods*. **10** (6): 563–69. doi:10.1038/nmeth.2474 (https://doi.org/10.1038%2Fnmeth.2474). PMID 23644548 (https://pubmed.ncbi.nlm.nih.gov/23644548). S2CID 205421576 (https://api.semanticscholar.org/CorpusID:205421576).

93. "De novo bacterial genome assembly: a solved problem?" (http://flxlexblog.wordpress.com/2013/07/05/de-novo-bacterial-genome-assembly-a-solved-problem/). 5 July 2013.

94. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK (25 August 2011). "Origins of the Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168948). *N Engl J Med*. **365** (8): 709–17. doi:10.1056/NEJMoa1106920 (https://doi.org/10.1056%2FNEJMoa1106920). PMC 3168948 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168948). PMID 21793740 (https://pubmed.ncbi.nlm.nih.gov/21793740).

95. Tran B, Brown AM, Bedard PL, Winquist E, Goss GD, Hotte SJ, Welch SA, Hirte HW, Zhang T, Stein LD, Ferretti V, Watt S, Jiao W, Ng K, Ghai S, Shaw P, Petrocelli T, Hudson TJ, Neel BG, Onetto N, Siu LL, McPherson JD, Kamel-Reid S, Dancey JE (1 January 2012). "Feasibility of real time next generation sequencing of cancer genes linked to drug response: Results from a clinical trial" (https://doi.org/10.1002%2Fijc.27817). *Int. J. Cancer*. **132** (7): 1547–55. doi:10.1002/ijc.27817 (https://doi.org/10.1002%2Fijc.27817). PMID 22948899 (https://pubmed.ncbi.nlm.nih.gov/22948899). S2CID 72705 (https://api.semanticscholar.org/CorpusID:72705).(subscription required)

96. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ (2 October 2012). "The methylomes of six bacteria" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526280). *Nucleic Acids Research*. **40** (22): 11450–62. doi:10.1093/nar/gks891 (https://doi.org/10.1093%2Fnar%2Fgks891). PMC 3526280 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3526280). PMID 23034806 (https://pubmed.ncbi.nlm.nih.gov/23034806).

97. "Ion 520 & Ion 530 ExT Kit-Chef – Thermo Fisher Scientific" (https://www.thermofisher.com/order/catalog/product/A30670). *thermofisher.com*.

98. "Raw accuracy" (https://web.archive.org/web/20180330075720/http://129.130.90.13/ion-docs/GUID-C6419130-57D8-4DE2-BCF8-47157CB3C9A2.html). Archived from the original (http://129.130.90.13/ion-docs/GUID-C6419130-57D8-4DE2-BCF8-47157CB3C9A2.html) on 30 March 2018. Retrieved 29 March 2018.

99. van Vliet AH (1 January 2010). "Next generation sequencing of microbial transcriptomes: challenges and opportunities" (https://doi.org/10.1111%2Fj.1574-6968.2009.01767.x). *FEMS Microbiology Letters*. **302** (1): 1–7. doi:10.1111/j.1574-6968.2009.01767.x (https://doi.org/10.1111%2Fj.1574-6968.2009.01767.x). PMID 19735299 (https://pubmed.ncbi.nlm.nih.gov/19735299).

100. "BGI and MGISEQ" (http://en.mgitech.cn/product/30.html). *en.mgitech.cn*. Retrieved 5 July 2018.

101. Huang YF, Chen SC, Chiang YS, Chen TH, Chiu KP (2012). "Palindromic sequence impedes sequencing-by-ligation mechanism" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3521181). *BMC Systems Biology*. **6** (Suppl 2): S10. doi:10.1186/1752-0509-6-S2-S10 (https://doi.org/10.1186%2F1752-0509-6-S2-S10). PMC 3521181 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3521181). PMID 23281822 (https://pubmed.ncbi.nlm.nih.gov/23281822).

102. Loose, Matthew; Rakyan, Vardhman; Holmes, Nadine; Payne, Alexander (3 May 2018). "Whale watching with BulkVis: A graphical viewer for Oxford Nanopore bulk fast5 files". bioRxiv 10.1101/312256 (https://doi.org/10.1101%2F312256).

103. "PacBio Sales Start to Pick Up as Company Delivers on Product Enhancements" (http://www.genomeweb.com/sequencing/pacbio-sales-start-pick-company-delivers-product-enhancements). 12 February 2013.

104. "Bio-IT World" (https://web.archive.org/web/20200729220749/http://www.bio-itworld.com/2015/9/30/pacbio-announces-sequel-sequencing-system.aspx). *bio-itworld.com*. Archived from the original (http://www.bio-itworld.com/2015/9/30/pacbio-announces-sequel-sequencing-system.aspx) on 29 July 2020. Retrieved 16 November 2015.

105. "PacBio Launches Higher-Throughput, Lower-Cost Single-Molecule Sequencing System" (https://www.genomeweb.com/business-news/pacbio-launches-higher-throughput-lower-cost-single-molecule-sequencing-system). October 2015.

106. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (April 2009). "Continuous base identification for single-molecule nanopore DNA sequencing". *Nature Nanotechnology*. **4** (4): 265–70. Bibcode:2009NatNa...4..265C (https://ui.adsabs.harvard.edu/abs/2009NatNa...4..265C). doi:10.1038/nnano.2009.12 (https://doi.org/10.1038%2Fnnano.2009.12). PMID 19350039 (https://pubmed.ncbi.nlm.nih.gov/19350039).

107. dela Torre R, Larkin J, Singer A, Meller A (2012). "Fabrication and characterization of solid-state nanopore arrays for high-throughput DNA sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3557807). *Nanotechnology*. **23** (38): 385308. Bibcode:2012Nanot..23L5308D (https://ui.adsabs.harvard.edu/abs/2012Nanot..23L5308D). doi:10.1088/0957-4484/23/38/385308 (https://doi.org/10.1088%2F0957-4484%2F23%2F38%2F385308). PMC 3557807 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3557807). PMID 22948520 (https://pubmed.ncbi.nlm.nih.gov/22948520).

108. Pathak B, Lofas H, Prasongkit J, Grigoriev A, Ahuja R, Scheicher RH (2012). "Double-functionalized nanopore-embedded gold electrodes for rapid DNA sequencing" (https://zenodo.org/record/890231). *Applied Physics Letters*. **100** (2): 023701. Bibcode:2012ApPhL.100b3701P (https://ui.adsabs.harvard.edu/abs/2012ApPhL.100b3701P). doi:10.1063/1.3673335 (https://doi.org/10.1063%2F1.3673335).

109. Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW (2008). "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2234111). *Proceedings of the National Academy of Sciences*. **105** (4): 1176–81. Bibcode:2008PNAS..105.1176K (https://ui.adsabs.harvard.edu/abs/2008PNAS..105.1176K). doi:10.1073/pnas.0710982105 (https://doi.org/10.1073%2Fpnas.0710982105). PMC 2234111 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2234111). PMID 18216253 (https://pubmed.ncbi.nlm.nih.gov/18216253).

110. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM (9 September 2005). "Accurate multiplex polony sequencing of an evolved bacterial genome" (https://doi.org/10.1126%2Fscience.1117389). *Science*. **309** (5741): 1728–32. Bibcode:2005Sci...309.1728S (https://ui.adsabs.harvard.edu/abs/2005Sci...309.1728S). doi:10.1126/science.1117389 (https://doi.org/10.1126%2Fscience.1117389). PMID 16081699 (https://pubmed.ncbi.nlm.nih.gov/16081699). S2CID 11405973 (https://api.semanticscholar.org/CorpusID:11405973).

111. Bentley DR, Balasubramanian S, et al. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2581791). *Nature*. **456** (7218): 53–59. Bibcode:2008Natur.456...53B (https://ui.adsabs.harvard.edu/abs/2008Natur.456...53B). doi:10.1038/nature07517 (https://doi.org/10.1038%2Fnature07517). PMC 2581791 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2581791). PMID 18987734 (https://pubmed.ncbi.nlm.nih.gov/18987734).

112. Canard B, Sarfati S (13 October 1994), *Novel derivatives usable for the sequencing of nucleic acids* (http://www.google.ge/patents/CA2158975A1), retrieved 9 March 2016

113. Canard B, Sarfati RS (October 1994). "DNA polymerase fluorescent substrates with reversible 3'-tags". *Gene*. **148** (1): 1–6. doi:10.1016/0378-1119(94)90226-7 (https://doi.org/10.1016%2F0378-1119%2894%2990226-7). PMID 7523248 (https://pubmed.ncbi.nlm.nih.gov/7523248).

114. Mardis ER (2008). "Next-generation DNA sequencing methods". *Annu Rev Genom Hum Genet*. **9**: 387–402. doi:10.1146/annurev.genom.9.081307.164359 (https://doi.org/10.1146%2Fannurev.genom.9.081307.164359). PMID 18576944 (https://pubmed.ncbi.nlm.nih.gov/18576944).

115. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. (January 2010). "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays" (https://doi.org/10.1126%2Fscience.1181498). *Science*. **327** (5961): 78–81. Bibcode:2010Sci...327...78D (http://ui.adsabs.harvard.edu/abs/2010Sci...327...78D). doi:10.1126/science.1181498 (https://doi.org/10.1126%2Fscience.1181498). PMID 19892942 (https://pubmed.ncbi.nlm.nih.gov/19892942). S2CID 17309571 (https://api.semanticscholar.org/CorpusID:17309571).

116. brandonvd. "About Us – Complete Genomics" (http://www.completegenomics.com/). *Complete Genomics*. Retrieved 2 July 2018.

117. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, et al. (May 2017). "A reference human genome dataset of the BGISEQ-500 sequencer" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467036). *GigaScience*. **6** (5): 1–9. doi:10.1093/gigascience/gix024 (https://doi.org/10.1093%2Fgigascience%2Fgix024). PMC 5467036 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467036). PMID 28379488 (https://pubmed.ncbi.nlm.nih.gov/28379488).

118. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM (July 2008). "A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2493394). *Genome Res*. **18** (7): 1051–63. doi:10.1101/gr.076463.108 (https://doi.org/10.1101%2Fgr.076463.108). PMC 2493394 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2493394). PMID 18477713 (https://pubmed.ncbi.nlm.nih.gov/18477713).

119. Rusk N (2011). "Torrents of sequence" (https://doi.org/10.1038%2Fnmeth.f.330). *Nat Methods*. **8** (1): 44. doi:10.1038/nmeth.f.330 (https://doi.org/10.1038%2Fnmeth.f.330). S2CID 41040192 (https://api.semanticscholar.org/CorpusID:41040192).

120. Drmanac R, Sparks AB, et al. (2010). "Human Genome Sequencing Using Unchained Base Reads in Self-Assembling DNA Nanoarrays" (https://doi.org/10.1126%2Fscience.1181498). *Science*. **327** (5961): 78–81. Bibcode:2010Sci...327...78D (https://ui.adsabs.harvard.edu/abs/2010Sci...327...78D). doi:10.1126/science.1181498 (https://doi.org/10.1126%2Fscience.1181498). PMID 19892942 (https://pubmed.ncbi.nlm.nih.gov/19892942). S2CID 17309571 (https://api.semanticscholar.org/CorpusID:17309571).

121. Porreca GJ (2010). "Genome Sequencing on Nanoballs". *Nature Biotechnology*. **28** (1): 43–44. doi:10.1038/nbt0110-43 (https://doi.org/10.1038%2Fnbt0110-43). PMID 20062041 (https://pubmed.ncbi.nlm.nih.gov/20062041). S2CID 54557996 (https://api.semanticscholar.org/CorpusID:54557996).

122. "HeliScope Gene Sequencing / Genetic Analyzer System : Helicos BioSciences" (https://web.archive.org/web/2009110204 1828/http://www.helicosbio.com/Products/HelicosregGeneticAna lysisSystem/HeliScopetradeSequencer/tabid/87/Default.aspx). 2 November 2009. Archived from the original (http://www.helicosbi o.com/Products/HelicosregGeneticAnalysisSystem/HeliScopetra deSequencer/tabid/87/Default.aspx) on 2 November 2009.

123. Thompson JF, Steinmann KE (October 2010). "Single molecule sequencing with a HeliScope genetic analysis system" (https://w ww.ncbi.nlm.nih.gov/pmc/articles/PMC2954431). *Current Protocols in Molecular Biology*. Chapter 7: Unit7.10. doi:10.1002/0471142727.mb0710s92 (https://doi.org/10.1002% 2F0471142727.mb0710s92). PMC 2954431 (https://www.ncbi.nl m.nih.gov/pmc/articles/PMC2954431). PMID 20890904 (https:// pubmed.ncbi.nlm.nih.gov/20890904).

124. "tSMS SeqLL Technical Explanation" (https://web.archive.org/we b/20140808055229/http://seqll.com/technical-description/). SeqLL. Archived from the original (http://seqll.com/technical-des cription/) on 8 August 2014. Retrieved 9 August 2015.

125. Heather, James M.; Chain, Benjamin (January 2016). "The sequence of sequencers: The history of sequencing DNA" (http s://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727787). *Genomics*. **107** (1): 1–8. doi:10.1016/j.ygeno.2015.11.003 (http s://doi.org/10.1016%2Fj.ygeno.2015.11.003). PMC 4727787 (htt ps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727787). PMID 26554401 (https://pubmed.ncbi.nlm.nih.gov/26554401).

126. Sara El-Metwally; Osama M. Ouda; Mohamed Helmy (2014). "New Horizons in Next-Generation Sequencing". *Next Generation Sequencing Technologies and Challenges in Sequence Assembly*. SpringerBriefs in Systems Biology. Vol. 7. Next Generation Sequencing Technologies and Challenges in Sequence Assembly, Springer Briefs in Systems Biology Volume 7. pp. 51–59. doi:10.1007/978-1-4939-0715-1_6 (https://doi.org/ 10.1007%2F978-1-4939-0715-1_6). ISBN 978-1-4939-0714-4.

127. Fair RB, Khlystov A, Tailor TD, Ivanov V, Evans RD, Srinivasan V, Pamula VK, Pollack MG, Griffin PB, Zhou J (January 2007). "Chemical and Biological Applications of Digital-Microfluidic Devices". *IEEE Design & Test of Computers*. **24** (1): 10–24. CiteSeerX 10.1.1.559.1440 (https://citeseerx.ist.psu.edu/viewdo c/summary?doi=10.1.1.559.1440). doi:10.1109/MDT.2007.8 (htt ps://doi.org/10.1109%2FMDT.2007.8). hdl:10161/6987 (https://h dl.handle.net/10161%2F6987). S2CID 10122940 (https://api.se manticscholar.org/CorpusID:10122940).

128. Boles DJ, Benton JL, Siew GJ, Levy MH, Thwar PK, Sandahl MA, et al. (November 2011). "Droplet-based pyrosequencing using digital microfluidics" (https://www.ncbi.nlm.nih.gov/pmc/arti cles/PMC3690483). *Analytical Chemistry*. **83** (22): 8439–47. doi:10.1021/ac201416j (https://doi.org/10.1021%2Fac201416j). PMC 3690483 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3 690483). PMID 21932784 (https://pubmed.ncbi.nlm.nih.gov/219 32784).

129. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L (January 2017). "Single-cell barcoding and sequencing using droplet microfluidics". *Nature Protocols*. **12** (1): 44–73. doi:10.1038/nprot.2016.154 (https://doi.org/10.103 8%2Fnprot.2016.154). PMID 27929523 (https://pubmed.ncbi.nl m.nih.gov/27929523). S2CID 767782 (https://api.semanticschol ar.org/CorpusID:767782).

130. "The Harvard Nanopore Group" (https://web.archive.org/web/20 020221002907/http://mcb.harvard.edu/branton/index.htm). Mcb.harvard.edu. Archived from the original (http://mcb.harvard. edu/branton/index.htm) on 21 February 2002. Retrieved 15 November 2009.

131. "Nanopore Sequencing Could Slash DNA Analysis Costs" (htt p://www.physorg.com/news157378086.html).

132. US patent 20060029957 (https://worldwide.espacenet.com/textd oc?DB=EPODOC&IDX=US20060029957), ZS Genetics, "Systems and methods of analyzing nucleic acid polymers and related components", issued 2005-07-14

133. Xu M, Fujita D, Hanagata N (December 2009). "Perspectives and challenges of emerging single-molecule DNA sequencing technologies". *Small*. **5** (23): 2638–49. doi:10.1002/smll.200900976 (https://doi.org/10.1002%2Fsmll.20 0900976). PMID 19904762 (https://pubmed.ncbi.nlm.nih.gov/19 904762).

134. Schadt EE, Turner S, Kasarskis A (2010). "A window into third-generation sequencing" (https://doi.org/10.1093%2Fhmg%2Fdd q416). *Human Molecular Genetics*. **19** (R2): R227–40. doi:10.1093/hmg/ddq416 (https://doi.org/10.1093%2Fhmg%2Fd dq416). PMID 20858600 (https://pubmed.ncbi.nlm.nih.gov/2085 8600).

135. Xu M, Endres RG, Arakawa Y (2007). "The electronic properties of DNA bases" (https://doi.org/10.1002%2Fsmll.200600732). *Small*. **3** (9): 1539–43. doi:10.1002/smll.200600732 (https://doi.o rg/10.1002%2Fsmll.200600732). PMID 17786897 (https://pubm ed.ncbi.nlm.nih.gov/17786897).

136. Di Ventra M (2013). "Fast DNA sequencing by electrical means inches closer" (https://zenodo.org/record/896777). *Nanotechnology*. **24** (34): 342501. Bibcode:2013Nanot..24H2501D (https://ui.adsabs.harvard.edu/a bs/2013Nanot..24H2501D). doi:10.1088/0957-4484/24/34/342501 (https://doi.org/10.1088%2F0957-4484%2F 24%2F34%2F342501). PMID 23899780 (https://pubmed.ncbi.nl m.nih.gov/23899780). S2CID 140101884 (https://api.semanticsc holar.org/CorpusID:140101884).

137. Ohshiro T, Matsubara K, Tsutsui M, Furuhashi M, Taniguchi M, Kawai T (2012). "Single-molecule electrical random resequencing of DNA and RNA" (https://www.ncbi.nlm.nih.gov/p mc/articles/PMC3392642). *Sci Rep*. **2**: 501. Bibcode:2012NatSR...2..501O (https://ui.adsabs.harvard.edu/ab s/2012NatSR...2..501O). doi:10.1038/srep00501 (https://doi.org/ 10.1038%2Fsrep00501). PMC 3392642 (https://www.ncbi.nlm.ni h.gov/pmc/articles/PMC3392642). PMID 22787559 (https://pub med.ncbi.nlm.nih.gov/22787559).

138. Hanna GJ, Johnson VA, Kuritzkes DR, Richman DD, Martinez-Picado J, Sutton L, Hazelwood JD, D'Aquila RT (1 July 2000). "Comparison of Sequencing by Hybridization and Cycle Sequencing for Genotyping of Human Immunodeficiency Virus Type 1 Reverse Transcriptase" (https://www.ncbi.nlm.nih.gov/p mc/articles/PMC87006). *J. Clin. Microbiol*. **38** (7): 2715–21. doi:10.1128/JCM.38.7.2715-2721.2000 (https://doi.org/10.112 8%2FJCM.38.7.2715-2721.2000). PMC 87006 (https://www.ncb i.nlm.nih.gov/pmc/articles/PMC87006). PMID 10878069 (https:// pubmed.ncbi.nlm.nih.gov/10878069).

139. Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA (2013). "A glimpse into past, present, and future DNA sequencing". *Molecular Genetics and Metabolism*. **110** (1–2): 3–24. doi:10.1016/j.ymgme.2013.04.024 (https://doi. org/10.1016%2Fj.ymgme.2013.04.024). PMID 23742747 (http s://pubmed.ncbi.nlm.nih.gov/23742747).

140. Qin Y, Schneider TM, Brenner MP (2012). Gibas C (ed.). "Sequencing by Hybridization of Long Targets" (https://www.ncb i.nlm.nih.gov/pmc/articles/PMC3344849). *PLOS ONE*. **7** (5): e35819. Bibcode:2012PLoSO...735819Q (https://ui.adsabs.harv ard.edu/abs/2012PLoSO...735819Q). doi:10.1371/journal.pone.0035819 (https://doi.org/10.1371%2Fj ournal.pone.0035819). PMC 3344849 (https://www.ncbi.nlm.nih. gov/pmc/articles/PMC3344849). PMID 22574124 (https://pubme d.ncbi.nlm.nih.gov/22574124).

141. Edwards JR, Ruparel H, Ju J (2005). "Mass-spectrometry DNA sequencing". *Mutation Research*. **573** (1–2): 3–12. Bibcode:2005MRFMM.573....3E (https://ui.adsabs.harvard.edu/ abs/2005MRFMM.573....3E). doi:10.1016/j.mrfmmm.2004.07.021 (https://doi.org/10.1016%2F j.mrfmmm.2004.07.021). PMID 15829234 (https://pubmed.ncbi. nlm.nih.gov/15829234).

142. Hall TA, Budowle B, Jiang Y, Blyn L, Eshoo M, Sannes-Lowery KA, Sampath R, Drader JJ, Hannis JC, Harrell P, Samant V, White N, Ecker DJ, Hofstadler SA (2005). "Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: A novel tool for the identification and differentiation of humans". *Analytical Biochemistry*. **344** (1): 53–69. doi:10.1016/j.ab.2005.05.028 (https://doi.org/10.1016%2 Fj.ab.2005.05.028). PMID 16054106 (https://pubmed.ncbi.nlm.ni h.gov/16054106).

143. Howard R, Encheva V, Thomson J, Bache K, Chan YT, Cowen S, Debenham P, Dixon A, Krause JU, Krishan E, Moore D, Moore V, Ojo M, Rodrigues S, Stokes P, Walker J, Zimmermann W, Barallon R (15 June 2011). "Comparative analysis of human mitochondrial DNA from World War I bone samples by DNA sequencing and ESI-TOF mass spectrometry" (https://doi.org/10.1016%2Fj.fsigen.2011.05.009). *Forensic Science International: Genetics*. **7** (1): 1–9. doi:10.1016/j.fsigen.2011.05.009 (https://doi.org/10.1016%2Fj.fsigen.2011.05.009). PMID 21683667 (https://pubmed.ncbi.nlm.nih.gov/21683667).

144. Monforte JA, Becker CH (1 March 1997). "High-throughput DNA analysis by time-of-flight mass spectrometry". *Nature Medicine*. **3** (3): 360–62. doi:10.1038/nm0397-360 (https://doi.org/10.1038%2Fnm0397-360). PMID 9055869 (https://pubmed.ncbi.nlm.nih.gov/9055869). S2CID 28386145 (https://api.semanticscholar.org/CorpusID:28386145).

145. Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD, Feldgarden M, Birren BW, Fofanov Y, Boos J, Wheaton WD, Honisch C, Musser JM (8 February 2010). "Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2840111). *Proceedings of the National Academy of Sciences*. **107** (9): 4371–76. Bibcode:2010PNAS..107.4371B (https://ui.adsabs.harvard.edu/abs/2010PNAS..107.4371B). doi:10.1073/pnas.0911295107 (https://doi.org/10.1073%2Fpnas.0911295107). PMC 2840111 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2840111). PMID 20142485 (https://pubmed.ncbi.nlm.nih.gov/20142485).

146. Kan CW, Fredlake CP, Doherty EA, Barron AE (1 November 2004). "DNA sequencing and genotyping in miniaturized electrophoresis systems". *Electrophoresis*. **25** (21–22): 3564–88. doi:10.1002/elps.200406161 (https://doi.org/10.1002%2Felps.200406161). PMID 15565709 (https://pubmed.ncbi.nlm.nih.gov/15565709). S2CID 4851728 (https://api.semanticscholar.org/CorpusID:4851728).

147. Chen YJ, Roller EE, Huang X (2010). "DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2881221). *Lab on a Chip*. **10** (9): 1153–59. doi:10.1039/b921417h (https://doi.org/10.1039%2Fb921417h). PMC 2881221 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2881221). PMID 20390134 (https://pubmed.ncbi.nlm.nih.gov/20390134).

148. Bell DC, Thomas WK, Murtagh KM, Dionne CA, Graham AC, Anderson JE, Glover WR (9 October 2012). "DNA Base Identification by Electron Microscopy". *Microscopy and Microanalysis*. **18** (5): 1049–53. Bibcode:2012MiMic..18.1049B (https://ui.adsabs.harvard.edu/abs/2012MiMic..18.1049B). doi:10.1017/S1431927612012615 (https://doi.org/10.1017%2FS1431927612012615). PMID 23046798 (https://pubmed.ncbi.nlm.nih.gov/23046798). S2CID 25713635 (https://api.semanticscholar.org/CorpusID:25713635).

149. Pareek CS, Smoczynski R, Tretyn A (November 2011). "Sequencing technologies and genome sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189340). *Journal of Applied Genetics*. **52** (4): 413–35. doi:10.1007/s13353-011-0057-x (https://doi.org/10.1007%2Fs13353-011-0057-x). PMC 3189340 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189340). PMID 21698376 (https://pubmed.ncbi.nlm.nih.gov/21698376).

150. Pareek CS, Smoczynski R, Tretyn A (2011). "Sequencing technologies and genome sequencing" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189340). *Journal of Applied Genetics*. **52** (4): 413–35. doi:10.1007/s13353-011-0057-x (https://doi.org/10.1007%2Fs13353-011-0057-x). PMC 3189340 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3189340). PMID 21698376 (https://pubmed.ncbi.nlm.nih.gov/21698376).

151. Fujimori S, Hirai N, Ohashi H, Masuoka K, Nishikimi A, Fukui Y, Washio T, Oshikubo T, Yamashita T, Miyamoto-Sato E (2012). "Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466446). *Scientific Reports*. **2**: 691. Bibcode:2012NatSR...2..691F (https://ui.adsabs.harvard.edu/abs/2012NatSR...2..691F). doi:10.1038/srep00691 (https://doi.org/10.1038%2Fsrep00691). PMC 3466446 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3466446). PMID 23056904 (https://pubmed.ncbi.nlm.nih.gov/23056904).

152. "2022 Sequencing Market Share – Same as It Ever Was (For Now)" (https://sandiegomics.com/2022-sequencing-market-share-same-as-it-ever-was-for-now/). 25 June 2023.

153. Harbers M (2008). "The Current Status of cDNA Cloning" (https://doi.org/10.1016%2Fj.ygeno.2007.11.004). *Genomics*. **91** (3): 232–42. doi:10.1016/j.ygeno.2007.11.004 (https://doi.org/10.1016%2Fj.ygeno.2007.11.004). PMID 18222633 (https://pubmed.ncbi.nlm.nih.gov/18222633).

154. Alberti A, Belser C, Engelen S, Bertrand L, Orvain C, Brinas L, Cruaud C, et al. (2014). "Comparison of Library Preparation Methods Reveals Their Impact on Interpretation of Metatranscriptomic Data" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213505). *BMC Genomics*. **15** (1): 912–12. doi:10.1186/1471-2164-15-912 (https://doi.org/10.1186%2F1471-2164-15-912). PMC 4213505 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4213505). PMID 25331572 (https://pubmed.ncbi.nlm.nih.gov/25331572).

155. "Scalable Nucleic Acid Quality Assessments for Illumina Next-Generation Sequencing Library Prep" (https://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/library-qc-fragment-analyzer-application-note-770-2017-002.pdf) (PDF). Retrieved 27 December 2017.

156. "Archon Genomics XPRIZE" (https://web.archive.org/web/20130617175241/http://genomics.xprize.org/). *Archon Genomics XPRIZE*. Archived from the original (http://genomics.xprize.org/) on 17 June 2013. Retrieved 9 August 2007.

157. "Grant Information" (http://www.genome.gov/10000004). *National Human Genome Research Institute (NHGRI)*.

158. Severin J, Lizio M, Harshbarger J, Kawaji H, Daub CO, Hayashizaki Y, Bertin N, Forrest AR (2014). "Interactive visualization and analysis of large-scale sequencing datasets using ZENBU". *Nat. Biotechnol*. **32** (3): 217–19. doi:10.1038/nbt.2840 (https://doi.org/10.1038%2Fnbt.2840). PMID 24727769 (https://pubmed.ncbi.nlm.nih.gov/24727769). S2CID 26575621 (https://api.semanticscholar.org/CorpusID:26575621).

159. Shmilovici A, Ben-Gal I (2007). "Using a VOM model for reconstructing potential coding regions in EST sequences" (https://web.archive.org/web/20200531014400/http://www.eng.tau.ac.il/~bengal/VOM_EST.pdf) (PDF). *Computational Statistics*. **22** (1): 49–69. doi:10.1007/s00180-007-0021-8 (https://doi.org/10.1007%2Fs00180-007-0021-8). S2CID 2737235 (https://api.semanticscholar.org/CorpusID:2737235). Archived from the original (http://www.eng.tau.ac.il/~bengal/VOM_EST.pdf) (PDF) on 31 May 2020. Retrieved 10 January 2014.

160. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM (2013). "An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3871669). *PLOS ONE*. **8** (12): e85024. Bibcode:2013PLoSO...885024D (https://ui.adsabs.harvard.edu/abs/2013PLoSO...885024D). doi:10.1371/journal.pone.0085024 (https://doi.org/10.1371%2Fjournal.pone.0085024). PMC 3871669 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3871669). PMID 24376861 (https://pubmed.ncbi.nlm.nih.gov/24376861).

161. Martin, Marcel (2 May 2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads" (https://doi.org/10.14806%2Fej.17.1.200). *EMBnet.journal*. **17** (1): 10. doi:10.14806/ej.17.1.200 (https://doi.org/10.14806%2Fej.17.1.200).

162. Smeds L, Künstner A (19 October 2011). "ConDeTri--a content dependent read trimmer for Illumina data" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198461). *PLOS ONE*. **6** (10): e26314. Bibcode:2011PLoSO...626314S (https://ui.adsabs.harvard.edu/abs/2011PLoSO...626314S). doi:10.1371/journal.pone.0026314 (https://doi.org/10.1371%2Fjournal.pone.0026314). PMC 3198461 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198461). PMID 22039460 (https://pubmed.ncbi.nlm.nih.gov/22039460).

163. Prezza N, Del Fabbro C, Vezzi F, De Paoli E, Policriti A (2012). "Erne-Bs5". *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. Vol. 12. pp. 12–19. doi:10.1145/2382936.2382938 (https://doi.org/10.1145%2F2382936.2382938). ISBN 9781450316705. S2CID 5673753 (https://api.semanticscholar.org/CorpusID:5673753).

164. Schmieder R, Edwards R (March 2011). "Quality control and preprocessing of metagenomic datasets" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3051327). *Bioinformatics*. **27** (6): 863–4. doi:10.1093/bioinformatics/btr026 (https://doi.org/10.1093%2Fbioinformatics%2Fbtr026). PMC 3051327 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3051327). PMID 21278185 (https://pubmed.ncbi.nlm.nih.gov/21278185).

165. Bolger AM, Lohse M, Usadel B (August 2014). "Trimmomatic: a flexible trimmer for Illumina sequence data" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4103590). *Bioinformatics*. **30** (15): 2114–20. doi:10.1093/bioinformatics/btu170 (https://doi.org/10.1093%2Fbioinformatics%2Fbtu170). PMC 4103590 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4103590). PMID 24695404 (https://pubmed.ncbi.nlm.nih.gov/24695404).

166. Cox MP, Peterson DA, Biggs PJ (September 2010). "SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2956736). *BMC Bioinformatics*. **11** (1): 485. doi:10.1186/1471-2105-11-485 (https://doi.org/10.1186%2F1471-2105-11-485). PMC 2956736 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2956736). PMID 20875133 (https://pubmed.ncbi.nlm.nih.gov/20875133).

167. Murray TH (January 1991). "Ethical issues in human genome research" (https://doi.org/10.1096%2Ffasebj.5.1.1825074). *FASEB Journal*. **5** (1): 55–60. doi:10.1096/fasebj.5.1.1825074 (https://doi.org/10.1096%2Ffasebj.5.1.1825074). PMID 1825074 (https://pubmed.ncbi.nlm.nih.gov/1825074). S2CID 20009748 (https://api.semanticscholar.org/CorpusID:20009748).

168. Robertson JA (August 2003). "The $1000 genome: ethical and legal issues in whole genome sequencing of individuals". *The American Journal of Bioethics*. **3** (3): W–IF1. doi:10.1162/152651603322874762 (https://doi.org/10.1162%2F152651603322874762). PMID 14735880 (https://pubmed.ncbi.nlm.nih.gov/14735880). S2CID 15357657 (https://api.semanticscholar.org/CorpusID:15357657).

169. Henderson, Mark (9 September 2013). "Human genome sequencing: the real ethical dilemmas" (https://www.theguardian.com/science/2013/sep/09/genetics-ethics-human-gene-sequencing). *The Guardian*. Retrieved 20 May 2015.

170. Harmon, Amy (24 February 2008). "Insurance Fears Lead Many to Shun DNA Tests" (https://www.nytimes.com/2008/02/24/health/24dna.html?pagewanted=all&_r=0). *The New York Times*. Retrieved 20 May 2015.

171. Statement of Administration policy (http://www.genome.gov/Pages/PolicyEthics/GeneticDiscrimination/SAPonHR493.pdf), Executive Office of the President, Office of Management and Budget, 27 April 2007

172. National Human Genome Research Institute (21 May 2008). "President Bush Signs the Genetic Information Nondiscrimination Act of 2008" (http://www.genome.gov/27026050). Retrieved 17 February 2014.

173. Baker, Monya (11 October 2012). "US ethics panel reports on DNA sequencing and privacy" (https://blogs.nature.com/news/2012/10/us-ethics-panel-reports-on-dna-sequencing-and-privacy.html). *Nature News Blog*.

174. "Privacy and Progress in Whole Genome Sequencing" (https://web.archive.org/web/20150612001900/http://bioethics.gov/sites/default/files/PrivacyProgress508_1.pdf) (PDF). Presidential Commission for the Study of Bioethical Issues. Archived from the original (http://bioethics.gov/sites/default/files/PrivacyProgress508_1.pdf) (PDF) on 12 June 2015. Retrieved 20 May 2015.

175. Hartnett, Kevin (12 May 2013). "The DNA in your garbage: up for grabs" (https://www.bostonglobe.com/ideas/2013/05/11/the-dna-your-garbage-for-grabs/sU12MtVLkoypL1qu2iF6IL/story.html). *The Boston Globe*. Retrieved 2 January 2023.

176. Goldenberg AJ, Sharp RR (February 2012). "The ethical hazards and programmatic challenges of genomic newborn screening" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3868436). *JAMA*. **307** (5): 461–2. doi:10.1001/jama.2012.68 (https://doi.org/10.1001%2Fjama.2012.68). PMC 3868436 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3868436). PMID 22298675 (https://pubmed.ncbi.nlm.nih.gov/22298675).

177. Hughes, Virginia (7 January 2013). "It's Time To Stop Obsessing About the Dangers of Genetic Information" (http://www.slate.com/articles/health_and_science/medical_examiner/2013/01/ethics_of_genetic_information_whole_genome_sequencing_is_here_and_we_need.html). *Slate Magazine*. Retrieved 22 May 2015.

178. Bloss CS, Schork NJ, Topol EJ (February 2011). "Effect of direct-to-consumer genomewide profiling to assess disease risk" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3786730). *The New England Journal of Medicine*. **364** (6): 524–34. doi:10.1056/NEJMoa1011893 (https://doi.org/10.1056%2FNEJMoa1011893). PMC 3786730 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3786730). PMID 21226570 (https://pubmed.ncbi.nlm.nih.gov/21226570).

179. Rochman, Bonnie (25 October 2012). "What Your Doctor Isn't Telling You About Your DNA" (https://healthland.time.com/2012/10/25/what-your-doctor-isnt-telling-you-about-your-dna/). *Time*. Retrieved 22 May 2015.

180. Sajeer P, Muhammad (4 May 2023). "Disruptive technology: Exploring the ethical, legal, political, and societal implications of nanopore sequencing technology" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10157308). *EMBO Reports*. **24** (5): e56619. doi:10.15252/embr.202256619 (https://doi.org/10.15252%2Fembr.202256619). PMC 10157308 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10157308). PMID 36988424 (https://pubmed.ncbi.nlm.nih.gov/36988424). S2CID 257803254 (https://api.semanticscholar.org/CorpusID:257803254).

External links

- A wikibook on next generation sequencing