UNIT 2

History and Evolution of Sequencing

Year	Event	Description
1970	Laborious methods based on 2D chromatography	First DNA sequences obtained
1975	Sanger's plus and minus method	Developed for DNA sequencing
1977	Sanger's chain-termination method	Revolutionized DNA sequencing, enabling rapid determination of long DNA sequences
1980s	Development of bioinformatics	Enabled by information technology. Processor computing allowed for the automation of overlapping sequences by similarity using dedicated computer programs, leading to the development of the first genome assemblers.
1990s	Recombinant DNA technology	Enabled isolation of DNA samples from various sources, leading to significant advancements in sequencing technology.
1995	First complete genome sequenced (Haemophilus influenzae)	Published using whole – genome shotgun sequencing
2001	Draft human genome sequence	Produced using shotgun sequencing
2000s	Emergence of NGS platforms	enabled the rapid sequencing of larger genomes Increased throughput, decreased costs
2010s	Third-generation sequencing technologies	Developed for higher quality and more accurate genome assemblies

Generation	Developed	Method/Technology	Applications
First Generation	1970s	Sanger dideoxy sequencing (Chain termination method), Maxam – Gilbert method (Chemical method)	Sequencing first genomes (e.g., human genome)
Second Generation (Next-Generation)	Early 2000s	454 pyrosequencing, Illumina sequencing, SOLiD sequencing	Rapid sequencing of entire genomes, genomics revolution
Third Generation	Late 2000s	Pacific Biosciences SMRT sequencing, Oxford Nanopore's nanopore sequencing	Real-time sequencing, improved genome assembly, structural variation detection

Current Approaches and Challenges

1. **De Novo Genome Assembly -** The process of deducing the complete genome sequence from overlapping shorter fragments, known as reads, remains a significant challenge in bioinformatics.
2. **Reference Genomes -** High-quality reference genomes are crucial for downstream applications, such as gene function understanding and gene manipulation experiments.
3. **Bioinformatics Tools -** Advances in analytical methods and bioinformatics tools have improved the efficiency and accuracy of genome assembly and analysis.

Future Directions

1. **Single-Molecule Sequencing -** This technology has the potential to generate genomes of higher quality and accuracy.
2. **Long-Read Sequencing -** Long-read sequencing methods are being developed to further improve the assembly of large genomes.
3. **Artificial Intelligence and Machine Learning -** Integration of AI and ML techniques is expected to enhance the efficiency and accuracy of genome assembly and analysis.

The history of sequencing in bioinformatics has been marked by significant advancements, from the early laborious methods to the rapid and high-throughput technologies of today. Ongoing research and development are expected to continue improving the efficiency, accuracy, and quality of genome sequencing and assembly.

First Generation Technologies

Method	Developed by	Technique	Impact
Sanger Dideoxy Sequencing	Frederick Sanger (1970s)	Dideoxynucleotides terminate DNA synthesis at specific points	Instrumental in DNA structure discovery and first genome sequencing
Maxam-Gilbert Sequencing	Allan Maxam & Walter Gilbert (1970s)	Chemical reactions cleave DNA at specific points	Useful for sequencing longer DNA regions

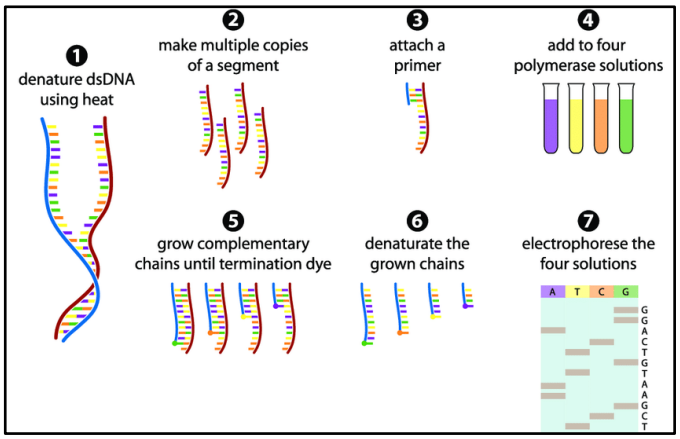
1. Sanger Dideoxy Sequencing

Principle

Sanger dideoxy sequencing is a method of DNA sequencing that involves the incorporation of chain-terminating dideoxynucleotides (ddNTPs) during DNA amplification.

This method is based on the random incorporation of ddNTPs by DNA polymerase, which causes DNA strand elongation to terminate at specific points.

The resulting DNA fragments are then separated by size using gel electrophoresis, allowing for the determination of the DNA sequence.



Working Mechanism

1. Template Preparation

A single-stranded DNA template is prepared, which serves as the starting point for the sequencing reaction.
A DNA primer is annealed to the template at a specific location.

↓

2. Sequencing Reaction

The sequencing reaction involves the use of a DNA polymerase, normal deoxynucleoside triphosphates (dNTPs), and modified dideoxynucleoside triphosphates (ddNTPs).
The ddNTPs are chemically altered to lack a 3'-OH group, which prevents further nucleotide incorporation.
The DNA polymerase incorporates the ddNTPs into the growing DNA chain, causing strand elongation to terminate at specific points.

↓

3. Separation by Size

The resulting DNA fragments are heat denatured and separated by size using gel electrophoresis.
Each of the four ddNTPs carries a distinct fluorescent label, allowing for the identification of the nucleotide incorporated at each termination point.

↓

4. Sequence Determination

The DNA bands are visualized by autoradiography or UV light.
The DNA sequence is directly read off the X-ray film or gel image by identifying the positions of the fluorescently labeled bands.

2. Maxam Gilbert Sequencing

Principle

Maxam-Gilbert sequencing is a method of DNA sequencing that involves the use of nucleobase-specific partial chemical modification of DNA and subsequent cleavage at sites adjacent to the modified nucleotides. This method allows for the direct sequencing of purified DNA samples without the need for cloning.

Working Mechanism

1. Radioactive Labeling

The DNA fragment to be sequenced is radioactively labeled at one 5' end using a kinase reaction with gamma- 32*P* ATP.

↓

2. Chemical Modification

The labeled DNA is then subjected to chemical treatment to introduce breaks at specific nucleotide bases. This involves four reactions:

- **(a) G Reaction:** Guanine is methylated by dimethyl sulfate, and then the sugar phosphate backbone is cleaved using hot piperidine.
- **(b) A+G Reaction:** Adenine and guanine are depurinated using formic acid.
- **(c) C Reaction:** Cytosine is hydrolyzed using hydrazine.
- **(d) C+T Reaction:** Cytosine and thymine are hydrolyzed using hydrazine in the presence of 1.5 M sodium chloride.

↓

3. Cleavage and Fragmentation

The modified DNA is then cleaved at the sites of modification using hot piperidine. This generates a series of labeled fragments, each ending at a specific nucleotide base.

↓

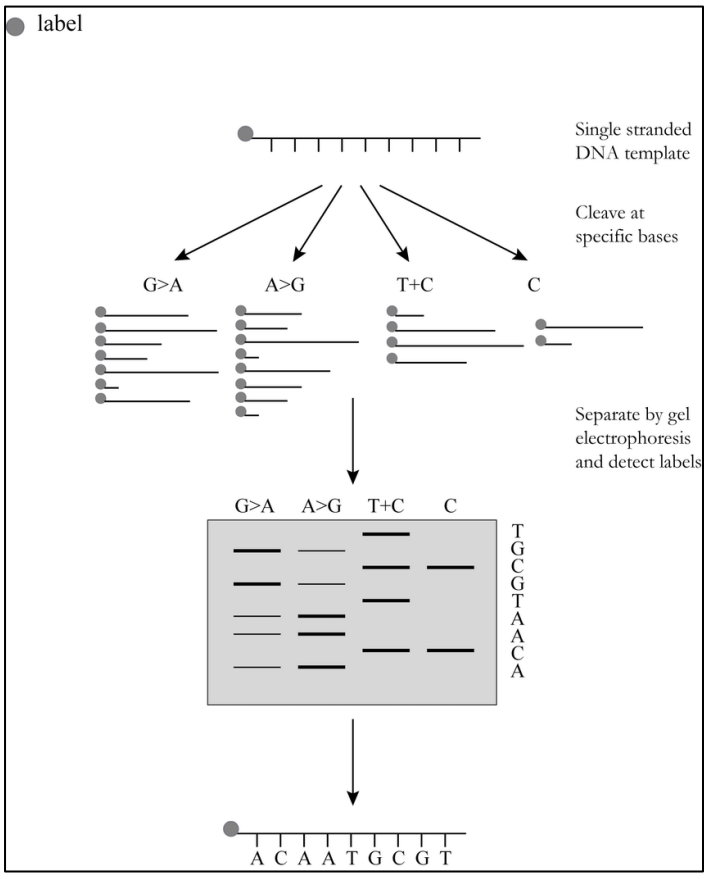
4. Size Separation

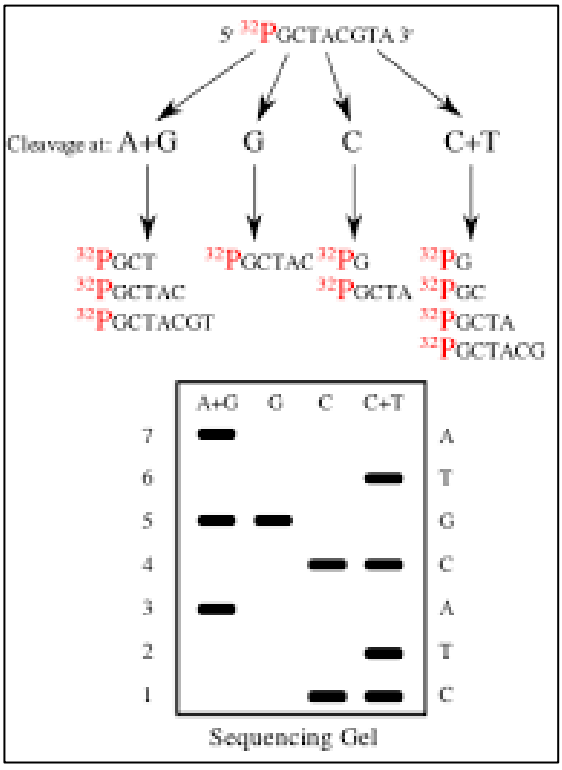
The fragments are separated by size using denaturing acrylamide gel electrophoresis.

↓

5. Autoradiography

The gel is exposed to X-ray film for autoradiography, yielding a series of dark bands corresponding to the radiolabeled DNA fragments.





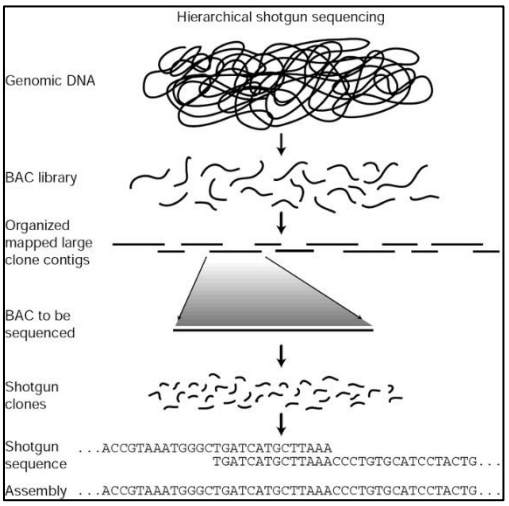
TECHNOLOGIES USED IN HUMAN GENOME PROJECT

The Human Genome Project (HGP) was a 13-year international research effort that aimed to determine the base pairs that make up human DNA and to identify, map, and sequence all of the genes of the human genome from both a physical and functional standpoint. The project was completed in 2003 and included about 92% of the genome. The HGP used several sequencing technologies to achieve this goal:

1. **Shotgun Sequencing:** This method involved breaking the genome into smaller fragments, sequencing each fragment, and then assembling the fragments into a complete genome. The shotgun sequencing approach was used to generate the majority of the human genome sequence.
2. **Capillary Electrophoresis:** used capillary tubes to separate DNA fragments based on their size. It was more efficient and cost-effective than earlier methods and played a key role in the Human Genome Project.
3. **Physical Mapping:** This involved creating a physical map of the human genome by identifying specific DNA sequences called sequence-tagged sites (STSs) that could be used to locate specific regions of the genome.
4. **Bacterial Artificial Chromosome (BAC) Cloning:** This involved cloning large segments of DNA into bacterial artificial chromosomes, which were then sequenced and assembled into a complete genome.
5. **Automated DNA Sequencing:** This involved using automated sequencing machines to rapidly sequence large amounts of DNA.
6. **Pyrosequencing:** This method uses a combination of enzymes and chemicals to determine the order of nucleotides in a DNA sequence. It was used to sequence the human genome and has since become a widely used technology for high-throughput sequencing.

These technologies, along with others, enabled the HGP to sequence the entire human genome and has since led to significant advances in our understanding of human biology and the development of personalized medicine.

SHOTGUN SEQUENCING



Principle

Shotgun sequencing is a method of DNA sequencing that involves randomly breaking up a large DNA sequence into smaller fragments, sequencing each fragment, and then reassembling the fragments to determine the original sequence.

This method is particularly useful for sequencing large genomes, such as the human genome, where traditional methods like Sanger sequencing are too time-consuming and labor-intensive.

Working Mechanism

1. Random Fragmentation:

The DNA sequence is broken up into smaller fragments of varying sizes, typically ranging from 2,000 to 300,000 bases long. This is done using mechanical shearing or other methods.

↓

2. Sequencing:

Each of the fragmented DNA pieces is then sequenced using a sequencing method like Sanger sequencing or next-generation sequencing (NGS) technologies.

↓

3. Assembly:

The sequenced fragments are then assembled into a complete sequence using computer algorithms that look for overlapping regions between the fragments. This process is often referred to as "assembly" or "reconstruction."

↓

4. Error Correction:

The assembled sequence is then checked for errors and corrected using various methods, such as comparing the assembled sequence to a reference genome or using additional sequencing data.

Advantages and disadvantages of 1st generation sequencing methods

Advantages

1. **High Accuracy-** These methods are known for their high accuracy, which is crucial in certain applications where precision is essential.
2. **Helps in Validating Findings of NGS-** First-generation sequencing methods can be used to validate the findings of next-generation sequencing (NGS) methods, ensuring the accuracy of the results.
3. **Low Error Rate-** The error rate in these methods is relatively low, which is important for ensuring the quality of the sequencing data.

Disadvantages

1. **High Cost-** First-generation sequencing methods are generally more expensive than newer methods, which can limit their use in large-scale projects.
2. **Low Throughput-** The throughput of these methods is relatively low compared to newer technologies, which can increase the time and cost required for sequencing.
3. **Long Run Time-** The run time for these methods can be quite long, which can be a significant limitation in many applications.
4. **Limited Read Length-** The read length of first-generation sequencing methods is generally shorter than that of newer methods, which can make assembly and analysis more challenging.
5. **Time-Consuming Sample Preparation-** Sample preparation for these methods can be time-consuming and labor-intensive, which can add to the overall cost and time required for sequencing.

DIFFERENCE BETWEEN FIRST GENERATION SEQUENCING AND SECONG GENERATION SEQUENCING METHODS

First Generation	Feature	Second Generation (NGS)
Short (400-900 bp)	**Read Lengths produced**	Much shorter (50-400 bp)
Low	**Sensitivity**	High
Low (up to a few hundred kb/day)	**Throughput**	High (millions to billions of reads/run), enabling sequencing of entire genomes in a matter of days
More expensive per bp sequenced	**Cost**	Significantly reduced cost per bp, making it more accessible for large – scale genomic projects
Time-consuming cloning due to cloning of DNA fragments into vectors	**Sample Preparation**	Streamlined fragmentation and library preparation
1. Sanger dideoxy sequencing 2. Maxam Gilbert Sequencing	**Methods involved**	
Targeted sequencing of specific genes of interest, validation of NGS results	**Applications**	Whole genome, transcriptome, metagenome sequencing

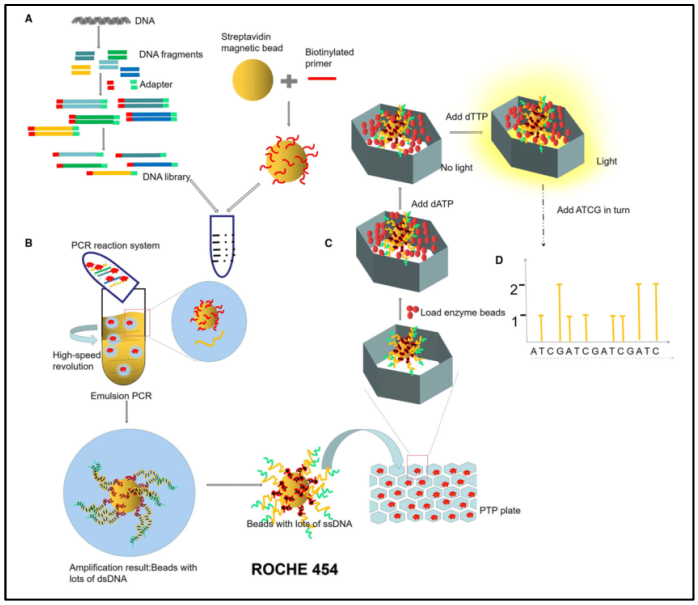
2ND GENERATION – NGS PLATFORMS

ROCHE 454

Principle of Roche 454 Sequencing

Roche 454 sequencing, also known as pyrosequencing, is a method of DNA sequencing that relies on the detection of pyrophosphate release during DNA synthesis.

The key principle behind Roche 454 sequencing is the use of pyrosequencing, which detects the release of pyrophosphate (PPi) during DNA synthesis. When a nucleotide is incorporated into the growing DNA strand, the pyrophosphate group is released, and this release is detected and converted into a light signal.



Working Mechanism of Roche 454 Sequencing

1. DNA Library Preparation:

DNA samples are fragmented into smaller pieces, typically 300-800 base pairs (bp) in length.
Adapters are added to the DNA fragments to enable binding to beads and subsequent amplification.

↓

2. Emulsion PCR (emPCR):

The DNA fragments are combined with beads coated with complementary adapters, and the mixture is emulsified in oil.
Each bead becomes encapsulated in a water-in-oil microdroplet, creating an independent microreactor for DNA amplification.
The DNA fragments bound to the beads are amplified using PCR, resulting in millions of copies of each fragment on a single bead.

↓

3. Sequencing:

The beads with amplified DNA fragments are loaded onto a PicoTiterPlate (PTP), a specialized plate with tiny wells that can accommodate a single bead per well.
The sequencing reaction is initiated by the addition of the four DNA nucleotides (A, T, C, and G) in a predetermined order.
When a nucleotide is incorporated into the growing DNA strand, the release of pyrophosphate (PPi) is detected by a series of enzymatic reactions, resulting in the emission of light.
The light signal is detected by a charge-coupled device (CCD) camera, and the sequence of nucleotides is determined based on the pattern of light signals.

↓

4. Data Analysis:

The raw sequence data generated by the Roche 454 system is processed and assembled using bioinformatics tools and algorithms.
The assembled sequences are then compared to reference genomes or databases to identify genes, genetic variations, and other genomic features.

Advantages of Roche 454 –

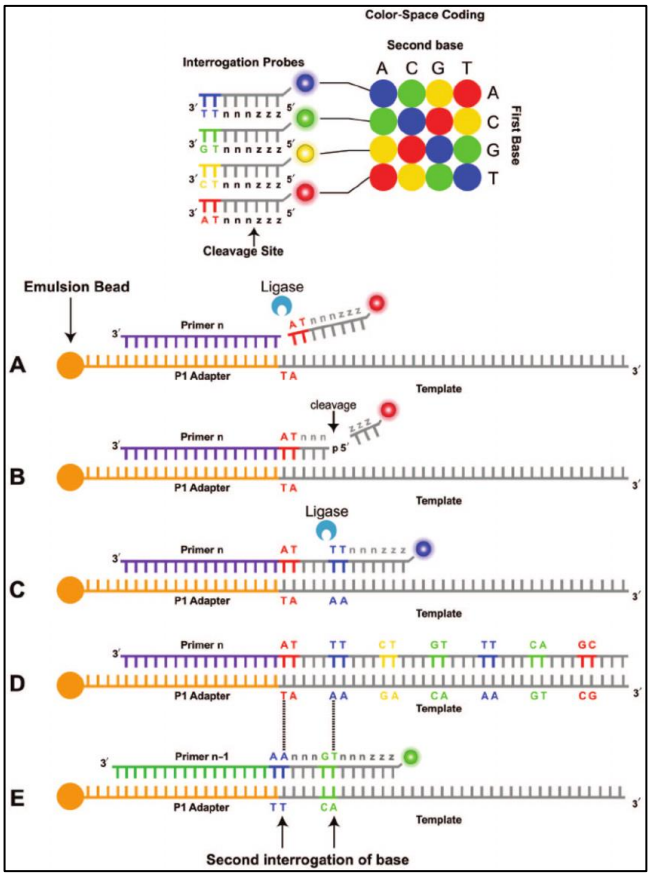
1. ability to generate longer read lengths (up to 1,000 bp) compared to other NGS technologies
2. its relatively high throughput.

Disadvantages

1. difficulty in accurately sequencing homopolymer regions (regions with consecutive identical nucleotides), which can lead to insertion and deletion errors.

ABI SOLiD

Principle of ABI SOLiD Sequencing



The ABI SOLiD (Sequencing by Oligonucleotide Ligation and Detection) system is a next-generation sequencing technology that uses a sequencing-by-ligation approach. The key principle behind SOLiD sequencing is the use of DNA ligase, an enzyme that can join two DNA strands together, to determine the sequence of nucleotides in a DNA sample.

Working Mechanism of ABI SOLiD Sequencing

1. DNA Library Preparation:

DNA samples are fragmented and adapters are added to the fragments to enable binding to beads.

↓

2. Emulsion PCR (emPCR):

The DNA fragments are combined with beads coated with complementary adapters, and the mixture is emulsified in oil. Each bead becomes encapsulated in a water-in-oil microdroplet, allowing for the clonal amplification of individual DNA fragments.

↓
↓
↓

3. Bead Deposition:

The beads with amplified DNA fragments are deposited onto a glass slide or flow cell, creating a high-density array of template beads.

↓

4. Sequencing by Ligation:

A sequencing primer is hybridized to the adapter region of the DNA fragments on the beads.
A library of fluorescently labeled 8-base probes is introduced, and DNA ligase is used to join the complementary probes to the sequencing primer.
The fluorescent signal from the incorporated probe is detected, and the probe is then cleaved to remove the fluorescent dye.
This process is repeated multiple times, using shorter sequencing primers (N-1, N-2, etc.) to sequence the DNA fragments from multiple positions.

↓

5. Data Analysis:

The raw sequence data generated by the SOLiD system is processed and assembled using bioinformatics tools and algorithms.
The assembled sequences are then compared to reference genomes or databases to identify genes, genetic variations, and other genomic features.

Advantages

1. high accuracy (up to 99.999% with sufficient coverage)
2. ability to detect small genetic variations, such as single nucleotide polymorphisms (SNPs)

Disadvantages

1. It has relatively shorter read lengths compared to other next-generation sequencing technologies.

ION TORRENT

Ion Torrent sequencing is a next-generation sequencing (NGS) technology developed by Life Technologies. It uses a "sequencing by synthesis" approach, where a new DNA strand is synthesized one base at a time.

Principle of Ion Torrent Sequencing

The principle behind Ion Torrent sequencing is the detection of hydrogen ions released during DNA polymerization. When a nucleotide is incorporated into the growing DNA strand, a hydrogen ion is released, which changes the pH of the solution. This change in pH is detected by a semiconductor chip, allowing for the determination of the DNA sequence.

Working Mechanism of Ion Torrent Sequencing

1. Library Preparation:

DNA samples are fragmented and adapters are added to the fragments to enable binding to beads.

↓

2. Emulsion PCR (emPCR):

The DNA fragments are combined with beads coated with complementary adapters, and the mixture is emulsified in oil. Each bead becomes encapsulated in a water-in-oil microdroplet, allowing for the clonal amplification of individual DNA fragments.

↓

3. Loading onto Ion Chip:

The beads with amplified DNA fragments are loaded onto an Ion chip, which is a semiconductor chip that detects the pH changes caused by the release of hydrogen ions during DNA polymerization.

↓

4. Sequencing:

A sequencing primer is hybridized to the adapter region of the DNA fragments on the beads.
A library of nucleotides is introduced, and DNA polymerase is used to incorporate the nucleotides into the growing DNA strand.
The release of hydrogen ions during DNA polymerization causes a change in pH, which is detected by the Ion chip.
The pH change is proportional to the number of hydrogen ions released, which is in turn proportional to the number of nucleotides incorporated.

↓

5. Signal Processing:

The raw sequence data generated by the Ion Torrent system is processed and assembled using bioinformatics tools and algorithms.
The assembled sequences are then compared to reference genomes or databases to identify genes, genetic variations, and other genomic features.

Advantages

1. high accuracy,
2. speed
3. lower cost compared to other NGS technologies

Disadvantages

1. It has relatively short read lengths
2. Can be challenging to enumerate the number of identical bases added consecutively
3. Can make it difficult to decode repetitive sequences.

Advantages and disadvantages of 2nd generation sequencing methods

Advantages:

1. **High Throughput:** Second-generation sequencing technologies can generate millions to billions of short reads per run, making them much faster than first-generation sequencing methods.
2. **Low Cost:** The cost of sequencing has decreased significantly with the advent of 2ndGS technologies, making them more accessible to researchers and clinicians.
3. **Clinical Applications:** 2ndGS technologies have been used in various clinical applications, including the diagnosis of genetic diseases and the monitoring of cancer treatment.
4. **Short Run Time:** The run time for 2ndGS technologies is relatively short compared to first-generation sequencing methods, allowing for faster turnaround times.

Disadvantages:

1. **Short Read Length:** Second-generation sequencing technologies typically generate short reads (50-400 bp), which can make it difficult to assemble and analyze the data.
2. **Difficult Sample Preparation:** The sample preparation process for 2ndGS technologies can be challenging, requiring specialized equipment and expertise.
3. **PCR Amplification:** The use of PCR amplification in 2ndGS technologies can introduce errors and bias into the sequencing data.
4. **Long Run Time:** While the run time for 2ndGS technologies is relatively short, the overall time required to complete a sequencing project can still be lengthy due to the need for sample preparation and data analysis.

3RD GENERATION SEQUENCING

PACBIO / SMRT

PacBio sequencing, also known as Single Molecule Real-Time (SMRT) sequencing, is a third-generation sequencing technology developed by Pacific Biosciences (PacBio). It uses a unique approach to sequence DNA molecules directly, without the need for amplification or fragmentation.

Principle of PacBio Sequencing

The principle behind PacBio sequencing is the use of DNA polymerase to synthesize DNA molecules in real-time, while simultaneously detecting the incorporation of nucleotides. This is achieved through the use of a specialized sequencing chip called the Zero-Mode Waveguide (ZMW) array.

Working Mechanism of PacBio Sequencing

1. Library Preparation:

High-molecular-weight DNA (HMW DNA) is extracted, quantified, and fragmented.
DNA damage is repaired.

↓

2. Sequencing:

The DNA fragments are bound to the ZMW array, where DNA polymerase is used to synthesize new DNA strands.
The incorporation of nucleotides is detected by the ZMW array, which emits a fluorescent signal.
The fluorescent signal is collected by a charge-coupled device (CCD) camera.

↓

3. Data Analysis:

The raw sequence data is processed and assembled using bioinformatics tools and algorithms.
The assembled sequences are then compared to reference genomes or databases to identify genes, genetic variations, and other genomic features.

Advantages of PacBio Sequencing

1. **Long Read Lengths:** PacBio sequencing can generate long reads (up to tens of thousands of base pairs) that are highly accurate and contiguous.
2. **High Accuracy:** The technology is highly accurate, with error rates as low as 0.1%.
3. **Real-Time Sequencing:** PacBio sequencing is a real-time sequencing technology, allowing for rapid sequencing and analysis.
4. **No Amplification Required:** PacBio sequencing does not require DNA amplification, which can introduce errors and bias into the sequencing data.

Applications of PacBio Sequencing

1. **Whole Genome Sequencing:** PacBio sequencing is used for whole genome sequencing, allowing for the rapid and accurate sequencing of entire genomes.
2. **Variant Detection:** PacBio sequencing is used for variant detection, enabling the identification of genetic variations and mutations.
3. **RNA Sequencing:** PacBio sequencing is used for RNA sequencing, allowing for the analysis of transcriptomes and the detection of alternative splicing events.
4. **Epigenetics:** PacBio sequencing is used for epigenetic analysis, enabling the study of DNA methylation and histone modifications.

Overall, PacBio sequencing is a powerful tool for genomics research, offering high accuracy, long read lengths, and real-time sequencing capabilities.

OXFORD NANOPORE

Oxford Nanopore Technologies (ONT) is a company that develops and manufactures nanopore sequencing technologies. The company's flagship product is the MinION, a portable, real-time sequencing device that uses a biological nanopore to detect and sequence nucleic acid molecules.

Principle of Oxford Nanopore Sequencing

The principle behind Oxford Nanopore sequencing is the use of a biological nanopore, such as the α-hemolysin protein, to detect and sequence nucleic acid molecules. The nanopore is embedded in a membrane and is surrounded by electrolyte solution. When a nucleic acid molecule passes through the nanopore, it causes a disruption in the ionic current flowing through the pore, which is detected by electrodes on either side of the membrane.

Working Mechanism of Oxford Nanopore Sequencing

1. Library Preparation:

DNA or RNA samples are prepared for sequencing by fragmenting them into smaller pieces and attaching adapters to the ends.

↓

2. Sequencing:

The prepared samples are loaded onto the MinION device, which consists of a flow cell and a nanopore array.
The flow cell contains a membrane with nanopores, and the nanopore array contains the biological nanopores.
The sample is then passed through the flow cell, and the nanopores detect the nucleic acid molecules as they pass through.

↓

3. Data Analysis:

The raw sequence data is processed and assembled using bioinformatics tools and algorithms.
The assembled sequences are then compared to reference genomes or databases to identify genes, genetic variations, and other genomic features.

Advantages of Oxford Nanopore Sequencing

1. **Portability:** The MinION is a portable device that can be used in a variety of settings, including the field.
2. **Real-Time Sequencing:** The MinION can sequence nucleic acid molecules in real-time, allowing for rapid analysis and decision-making.
3. **Long Read Lengths:** The MinION can generate long read lengths, which are useful for assembling and analyzing genomes.
4. **Low Cost:** The MinION is relatively low-cost compared to other sequencing technologies.

Applications of Oxford Nanopore Sequencing

1. **Genome Assembly:** The MinION can be used to assemble genomes, including those of bacteria, viruses, and eukaryotes.
2. **Transcriptomics:** The MinION can be used to analyze transcriptomes, including the detection of gene expression and alternative splicing.
3. **Epigenomics:** The MinION can be used to analyze epigenomes, including the detection of DNA methylation and histone modifications.
4. **Clinical Diagnostics:** The MinION can be used for clinical diagnostics, including the detection of genetic diseases and the monitoring of cancer treatment.

Overall, Oxford Nanopore sequencing is a powerful tool for genomics research, offering portability, real-time sequencing, long read lengths, and low cost.

ADVANTAGES AND DISADVANTAGES OF 3RD GENERATION SEQUENCING

Advantages:

1. **Longer Read Lengths:** 3rdGS technologies can generate longer reads (1500-100,000 base pairs) compared to second-generation sequencing (2ndGS) technologies.
2. **Faster Sequencing:** 3rdGS technologies are faster than 2ndGS technologies, allowing for rapid sequencing and analysis.
3. **Portability:** 3rdGS technologies are more portable than 2ndGS technologies, making them suitable for use in various settings.
4. **Low Cost:** 3rdGS technologies are relatively low-cost compared to 2ndGS technologies.
5. **No PCR Amplification:** 3rdGS technologies do not require PCR amplification, which can introduce errors and bias into the sequencing data.
6. **Direct Epigenetic Detection:** 3rdGS technologies can detect epigenetic markers directly, simplifying the detection process.
7. **Improved Accuracy:** 3rdGS technologies have improved accuracy compared to 2ndGS technologies.
8. **Reduced Sample Preparation:** 3rdGS technologies require minimal sample preparation, making them more user-friendly.

Disadvantages:

1. **High Sequencing Error Rate:** 3rdGS technologies have a higher sequencing error rate compared to 2ndGS technologies.
2. **Variable Accuracy:** The accuracy of 3rdGS technologies can vary depending on the specific technology and application.
3. **Longer Run Time:** 3rdGS technologies can have longer run times compared to 2ndGS technologies.
4. **Limited Bioinformatics Tools:** There are limited bioinformatics tools and algorithms available for analyzing 3rdGS data.
5. **Fresh DNA Required:** 3rdGS technologies require fresh DNA samples to ensure the quality of ultralong reads.
6. **Higher Cost:** While 3rdGS technologies are relatively low-cost compared to 2ndGS technologies, they can still be more expensive than other sequencing methods.
7. **Limited Clinical Applications:** 3rdGS technologies have limited clinical applications compared to 2ndGS technologies.

Method	Output	Accuracy	Types of Errors	Advantages	Disadvantages
First Generation Sequencing (Sanger Sequencing)	Short reads (typically 400-900 base pairs)	High accuracy, with error rates as low as 0.1%	Substitution errors, insertion/deletion errors, and homopolymer errors	High accuracy, long read lengths, ability to sequence specific regions	Low throughput, high cost, and labor-intensive process
Second Generation Sequencing (Next Generation Sequencing)	Short reads (typically 50-400 base pairs)	Lower accuracy compared to first generation, error rates 0.5-5%	Substitution errors, insertion/deletion errors, and homopolymer errors	High throughput, lower cost, ability to sequence entire genomes	Lower accuracy, short read lengths, need for assembly and error correction
Third Generation Sequencing (Long-Read Sequencing)	Long reads (typically 1500-100,000 base pairs)	Higher accuracy compared to second generation, error rates 0.1-1%	Substitution errors, insertion/deletion errors, and homopolymer errors	High accuracy, long read lengths, ability to detect epigenetic markers directly	Higher cost, lower throughput, need for specialized equipment and software