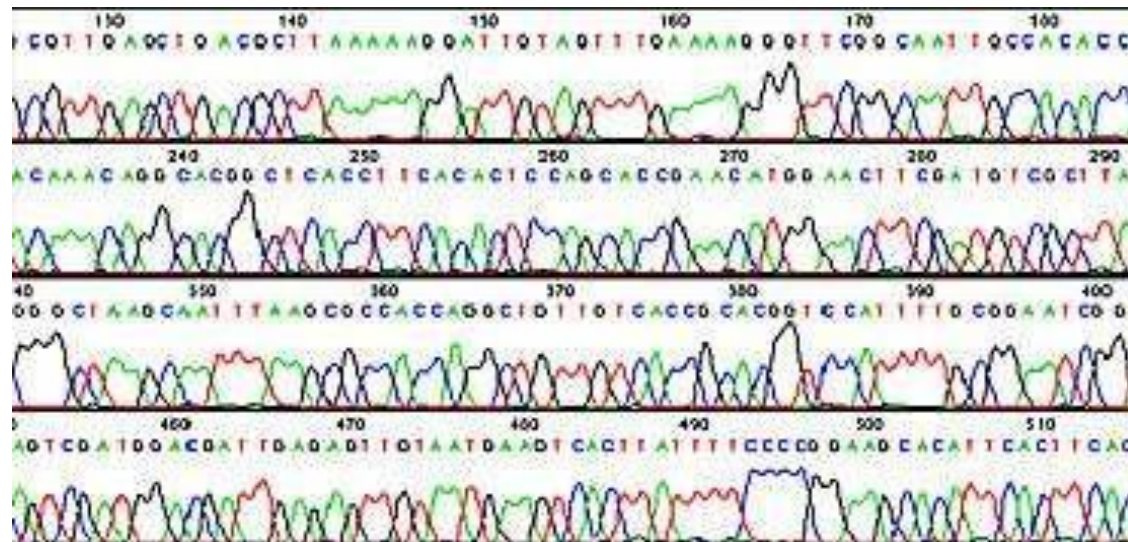
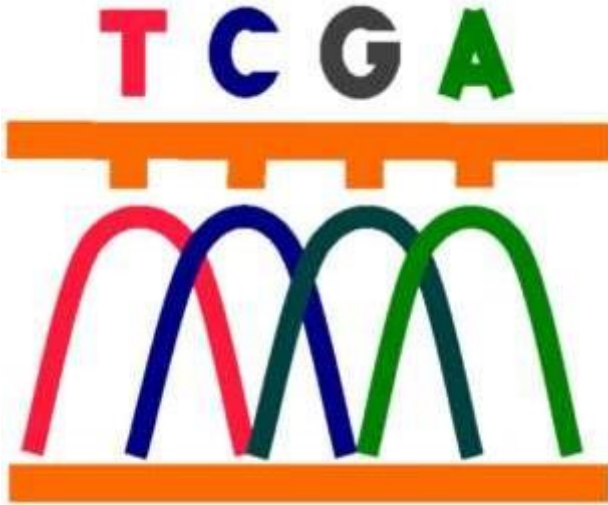


DNA Sequencing and its types

A T G C

Dr. N. Yuvaraj
Assistant Professor
Achariya Arts and Science College
Villianur, Puducherry
yuvaraj.aasc@achariya.org

What is DNA Sequencing ?

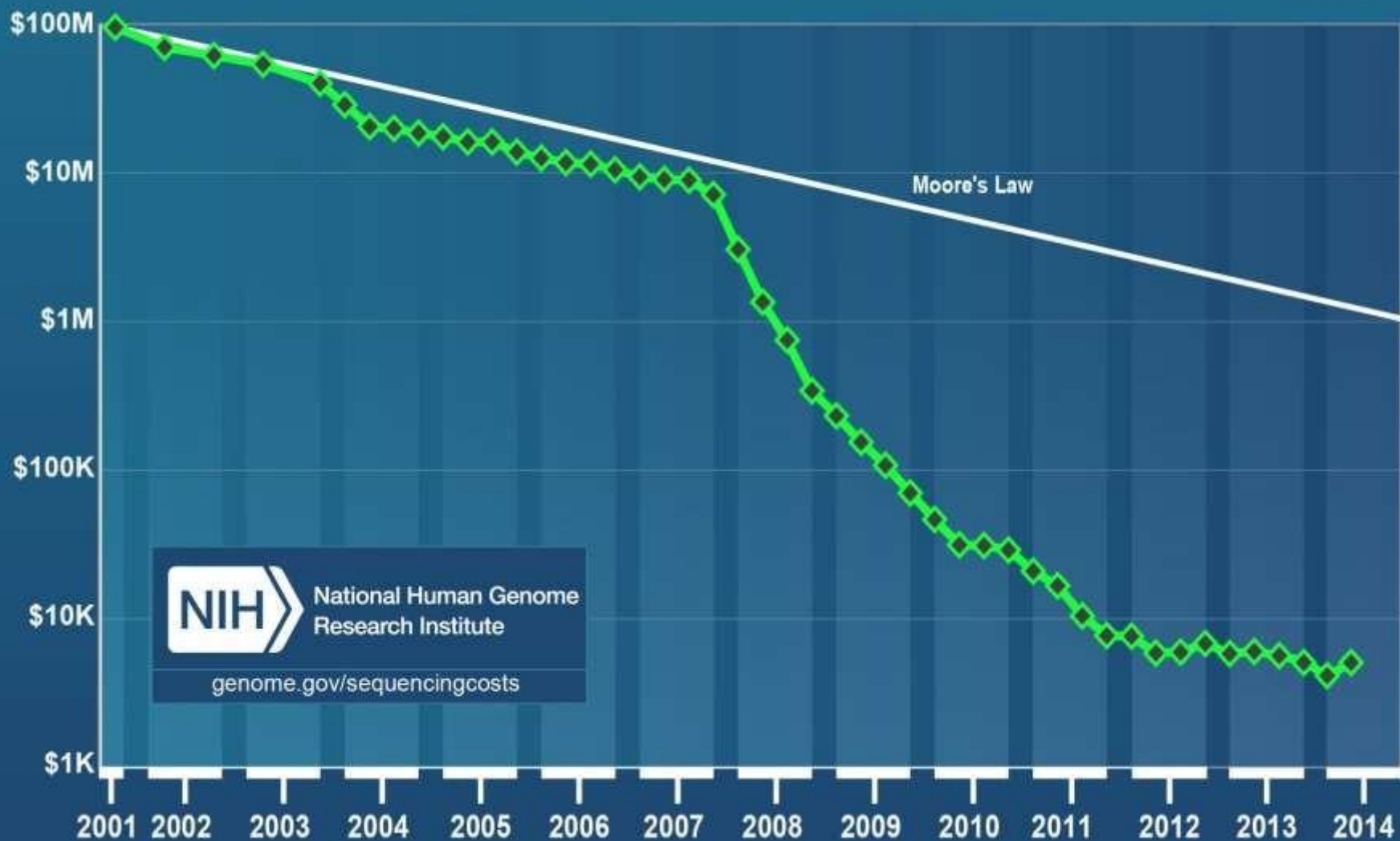


- The process of determining the order of nucleotides adenine (A), thymine (T), cytosine (C), and guanine (G) along a DNA strand.
- We need to know the order of nucleotide bases in a strand of DNA for sequencing.
- All the information required for the growth and development of an organism is encoded in the DNA of its genome.
- So, DNA sequencing is fundamental to genome analysis and understanding the biological processes in general.

Historical Timeline

- 1870 – Friedrich Miescher discovers DNA
- 1940 - Avery: Proposes DNA as 'Genetic Material'
- 1953 – Watson & Crick “double helical structure”
- 1970 - Wu: Sequences λ Cohesive End DNA
- 1977 – Sanger: Dideoxy Chain Termination
- 1977 – Gilbert: Chemical Degradation
- 1986 – Partial Automation
- 1990 – Cycle Sequencing, Improved Sequencing Enzymes,
Improved fluorescent detection schemes
- 2002 – NGS: 454 Pyrosequencing

Cost per Genome



SEQUENCING METHODS

To determine the order of the nucleotide bases adenine, guanine, cytosine, and thymine in a molecule of DNA two methods were used

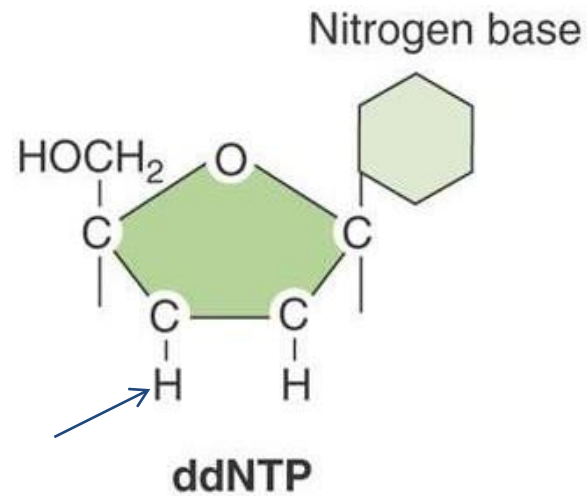
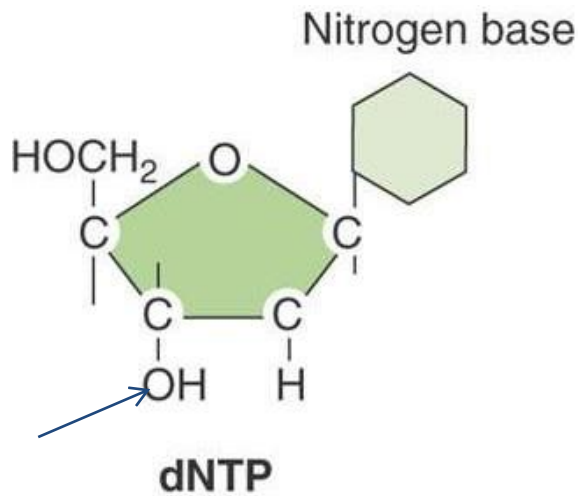
1. Sanger; Chain Termination Sequencing method
2. Maxam and Gilbert; Chemical Sequencing method

These two methods are most popular conventional methods

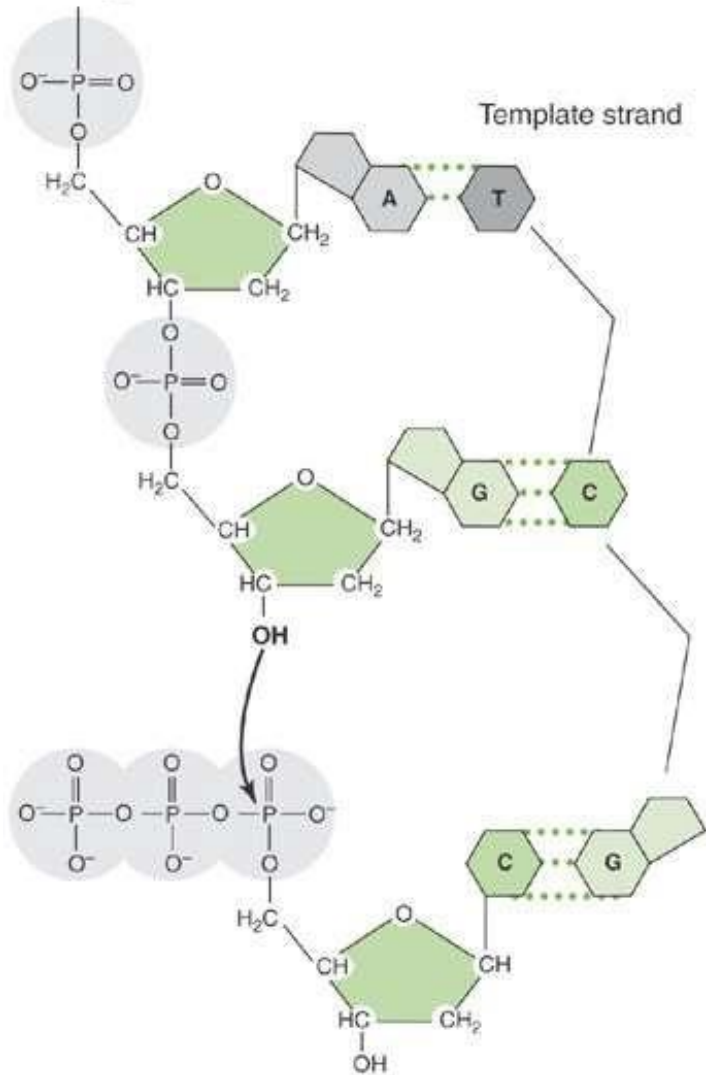
➤ Robotics and automated sequencing are based on these methods

Sanger's- Chain Termination Sequencing

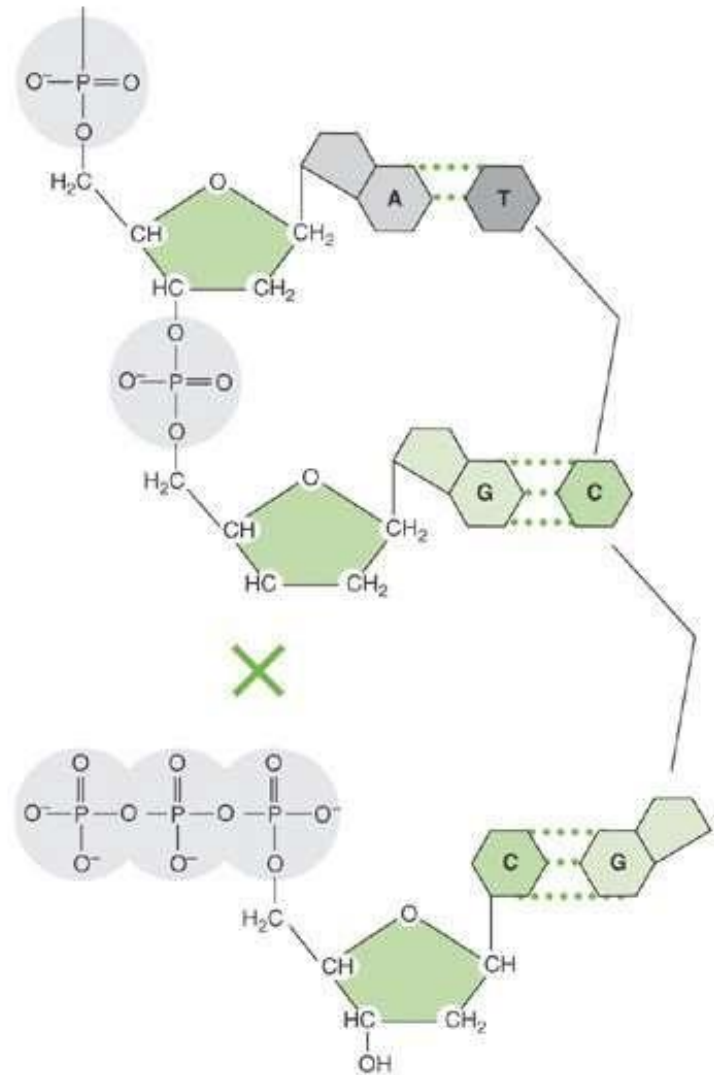
- It is PCRbased method
- A modified DNA replication reaction
- Growing chains are terminated by dideoxynucleotides



Growing strand



Template strand

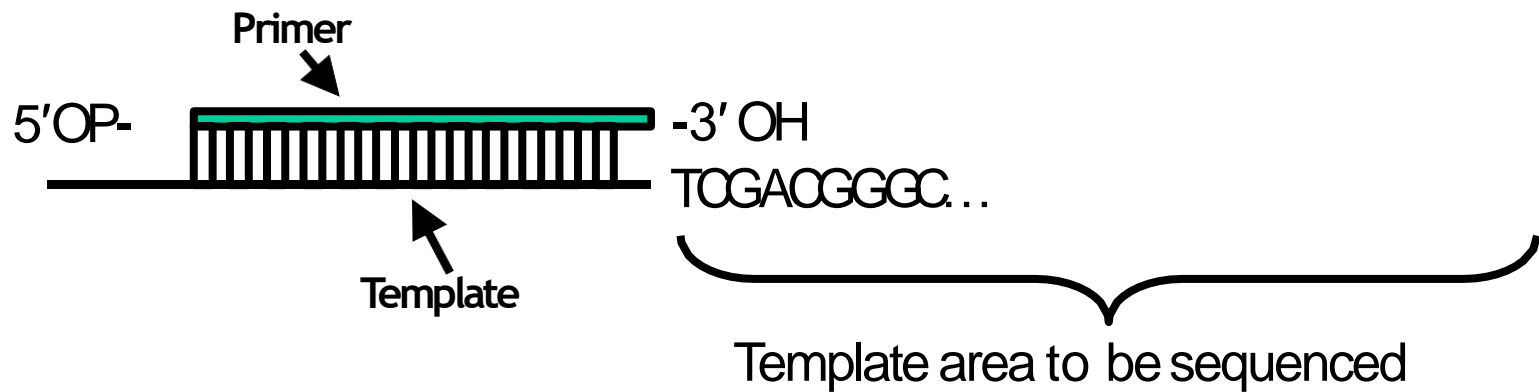


The 3'-OH group necessary for formation of the phosphodiester bond is missing in ddNTPs.

Sanger Sequencing: Process

1. Get enough quantity of DNA (Run PCR)
2. Aliquot DNA into four different tubes
3. Prepare PCR reaction mix as below:
 - Primer, Taq polymerase, template (ssDNA), dNTPS (All) and ddNTPs (ddATP, ddGTP, ddCTP & ddTTP respectively)
1. Run PCR
2. Perform Gel Electrophoresis
3. Interpret results

- A sequencing reaction mix includes labeled primer and template.



- Dideoxynucleotides are added separately to each of the four tubes.

- With addition of enzyme (DNA polymerase), the primer is extended until a ddNTP is encountered.
- The chain will end with the incorporation of the ddNTP.
- With the proper dNTP:ddNTP ratio, the chain will terminate throughout the length of the template.
- All terminated chains will end in the ddNTP added to that reaction.

- The collection of fragments is a **sequencing ladder**.
- The resulting terminated chains are resolved by electrophoresis.
- Fragments from each of the four tubes are placed in four separate gel lanes.

Sanger Sequencing: An Example

5'-TACACGATCGA-3'

3'-ATGTGCTAGCT-5'

Denature the sequence

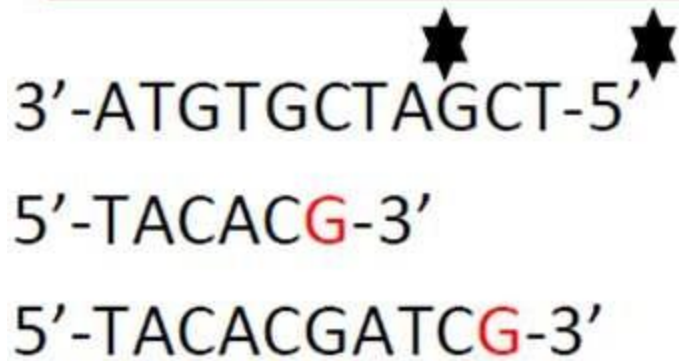
Use only forward primer i.e. use 3'-5' strand of

DNA

Amplification in ddTTP



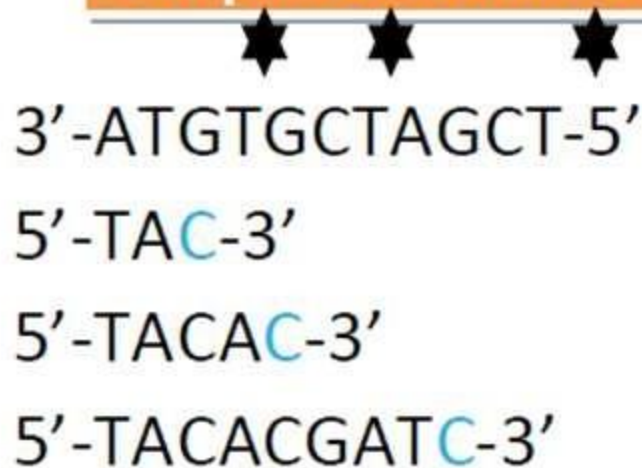
Amplification in ddGTP



Amplification in ddATP



Amplification in ddCTP



Maxam–Gilbert sequencing method

- **Maxam-Gilbert sequencing is a method of DNA sequencing developed by Allan Maxam and Walter Gilbert in 1976–1977.**
- This method is based on nucleobase-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified nucleotides.

Maxam Gilbert Sequencing: Process Summarized

1. Label 5'- end of DNA
2. Aliquot DNA sample in 4 tubes
3. Perform base modification reaction
4. Perform Cleavage reaction
5. Perform Gel Electrophoresis
6. Perform Autoradiography
7. Interpret results

- I. Chemical Modification of DNA; radioactive labeling at one 5' end of the DNA (typically by a kinase reaction using gamma-³²P ATP)
- II. Purification of the DNA fragment to be sequenced
- III. Chemical treatment generates breaks in DNA
- IV. Run on the gel

Chemical Modification and Cleavage

Poly nucleotide Kinase radioactive label at one 5' end of the DNA using gamma-³²P

5' GACGTGCAACGAA 3'

³²P 5' GACGTGCAACGAA 3'

Chemical Modification and Cleavage

➤ Base Modification using Dimethyl sulphate

– Purine

- Adenine
- Guanine

– Only DMS----- G

– DMS+ Formic acid-----G+A

➤ Cleavage of Sugar Phosphate backbone using

Piperidine

➤ Base modification using Hydrazine

– Pyrimidine

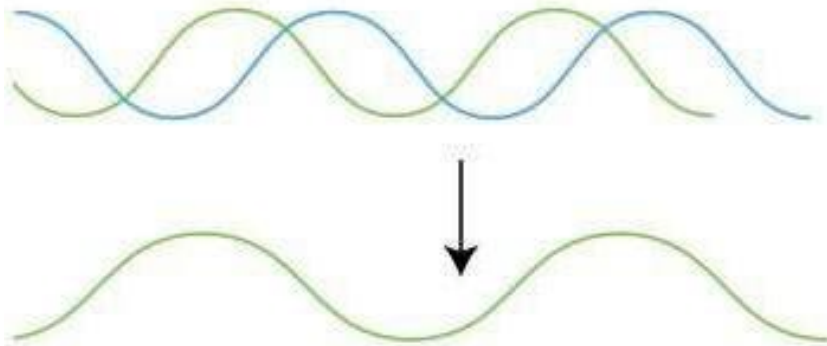
- Cytosine
- Thymine

– Hydrazine----- C+T

– Hydrazine + NaCl-----C

➤ Cleavage of Sugar Phosphate backbone using

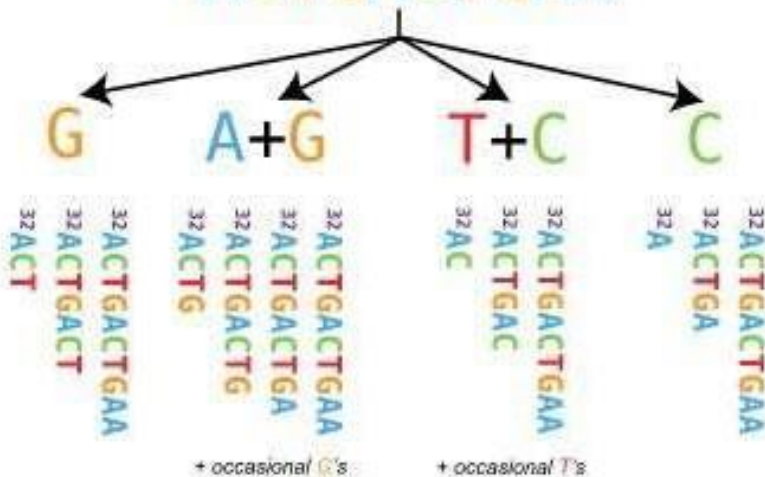
Piperidine



1) Obtain single stranded DNA

ACTGACTGAA

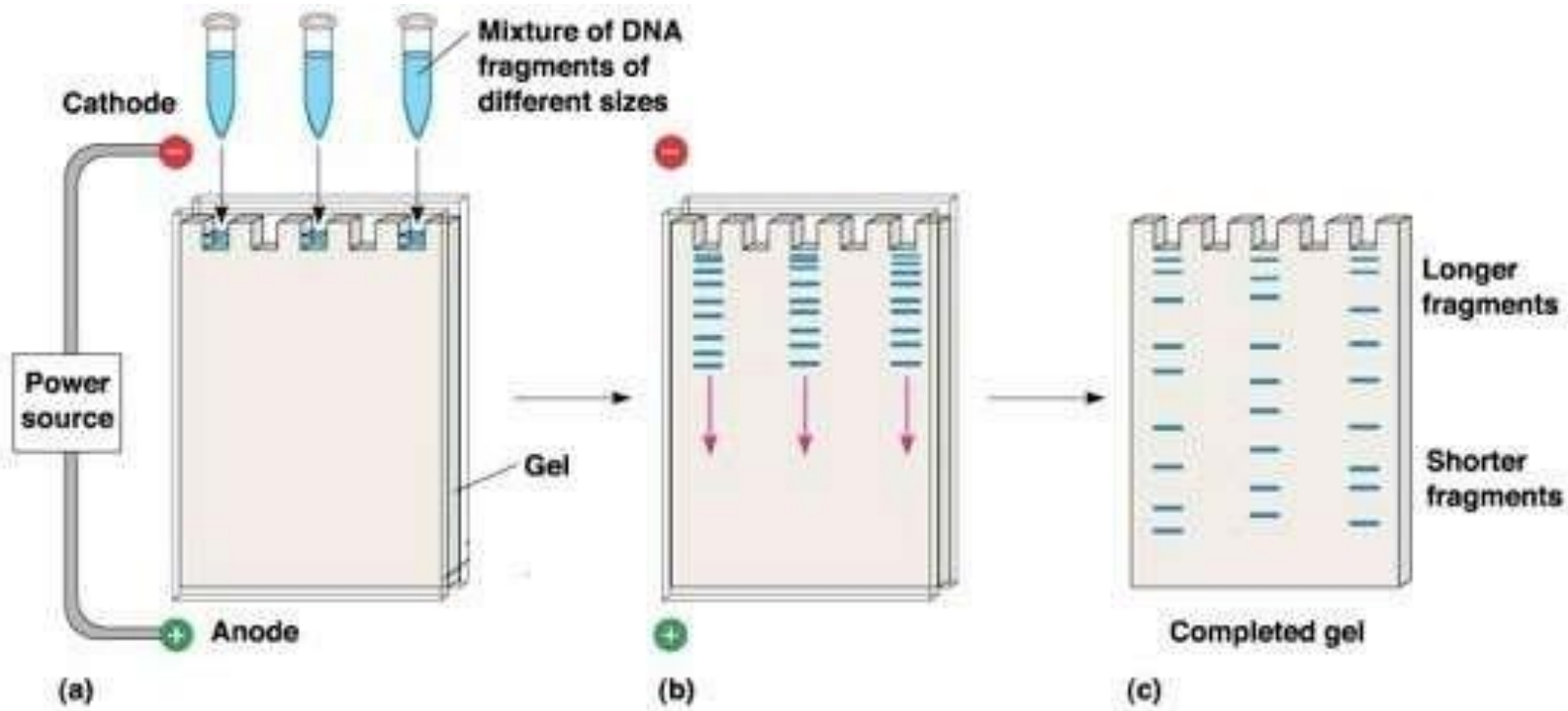
³²ACTGACTGAA



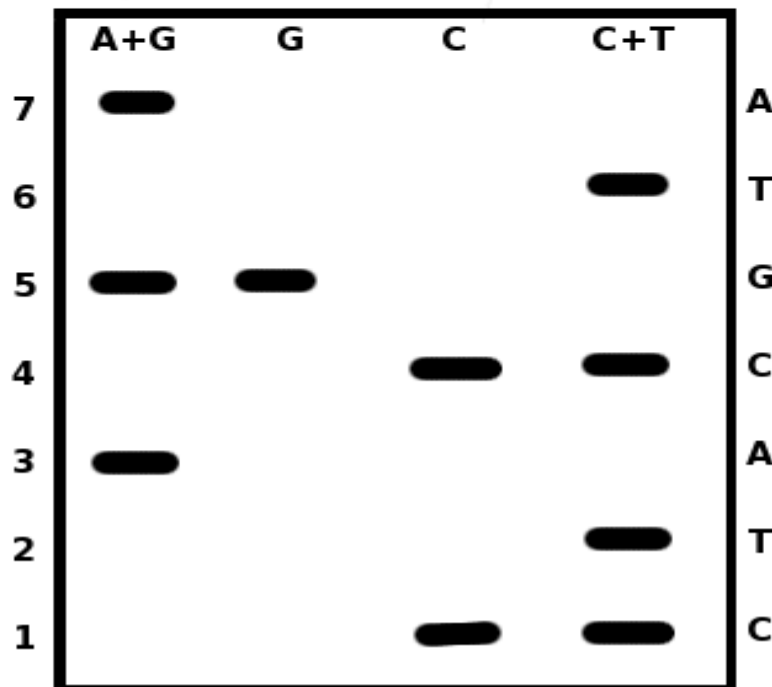
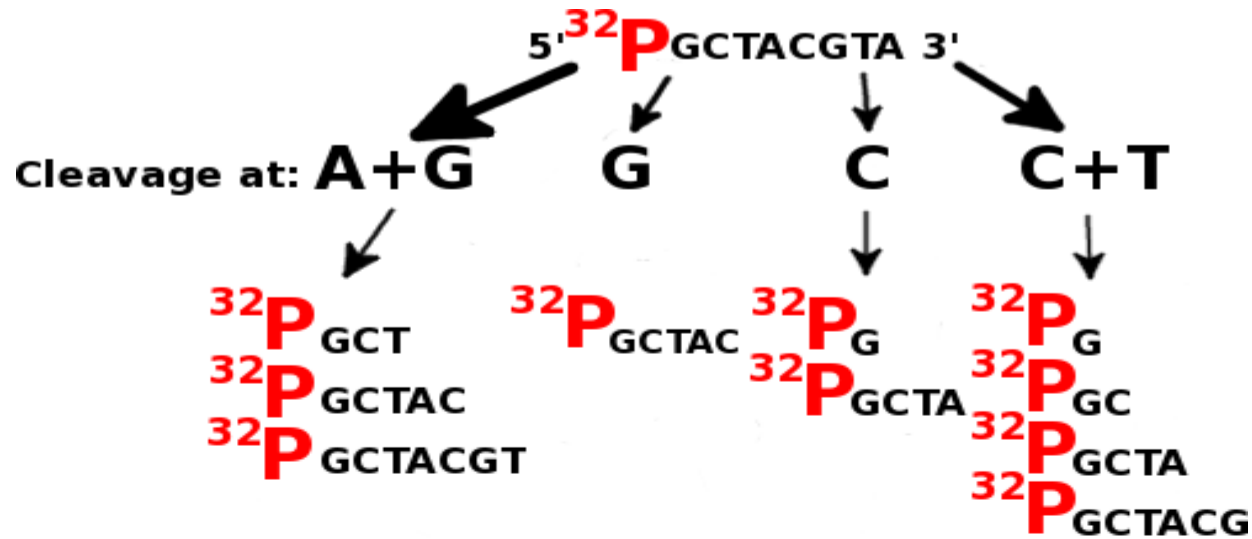
2) Add a ³²P to 5' end

3) Cleave at specific nucleotides

4) Differently sized DNA strands



An Example for Maxam-Gilbert Sequencing



Sequencing Gel

- The principle behind Next Generation Sequencing (NGS) is similar to that of [Sanger sequencing](#), which relies on capillary electrophoresis.
- The genomic strand is fragmented, and the bases in each fragment are identified by emitted signals when the fragments are ligated against a template strand.
- The NGS method uses array-based sequencing which combines the techniques developed in Sanger sequencing to process millions of reactions in parallel, resulting in very high speed and throughput at a reduced cost.

Three general steps in NGS

1. **Library preparation:** libraries are created using random fragmentation of DNA, followed by ligation with custom linkers
2. **Amplification:** the library is amplified using clonal amplification methods and [PCR](#)
3. **Sequencing:** DNA is sequenced using one of several different approaches

LIBRARY PREPARATION

- Firstly, DNA is fragmented either enzymatically or by sonication (excitation using ultrasound) to create smaller strands.
- Adaptors (short, double-stranded pieces of synthetic DNA) are then ligated to these fragments with the help of DNA ligase, an enzyme that joins DNA strands.
- The adaptors enable the sequence to become bound to a complementary counterpart.

- Adaptors are synthesized so that one end is 'sticky' whilst the other is 'blunt' (non-cohesive) with the view to joining the blunt end to the blunt ended DNA.
- This could lead to the potential problem of base pairing between molecules and therefore dimer formation.
- To prevent this, the chemical structure of DNA is utilised, since ligation takes place between the 3'-OH and 5'-P ends.
- By removing the phosphate from the sticky end of the adaptor and therefore creating a 5'-OH end instead, the DNA ligase is unable to form a bridge between the two termini.

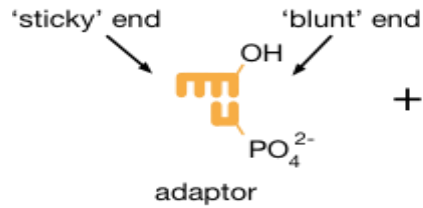


genomic DNA

sonication or
digestive enzyme
→



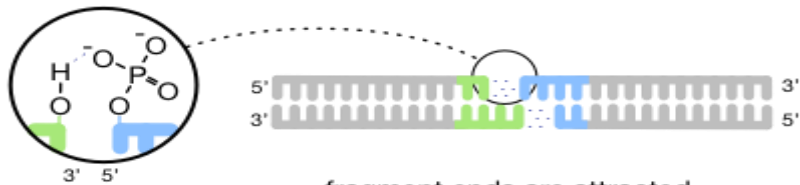
double-stranded DNA fragments



+



Ligase
→



fragment ends are attracted
to one another by H bonding

Ligase
→



ligation of two fragments can occur



modified adaptor with 5'-OH terminus

Ligase
→

dimerization cannot occur

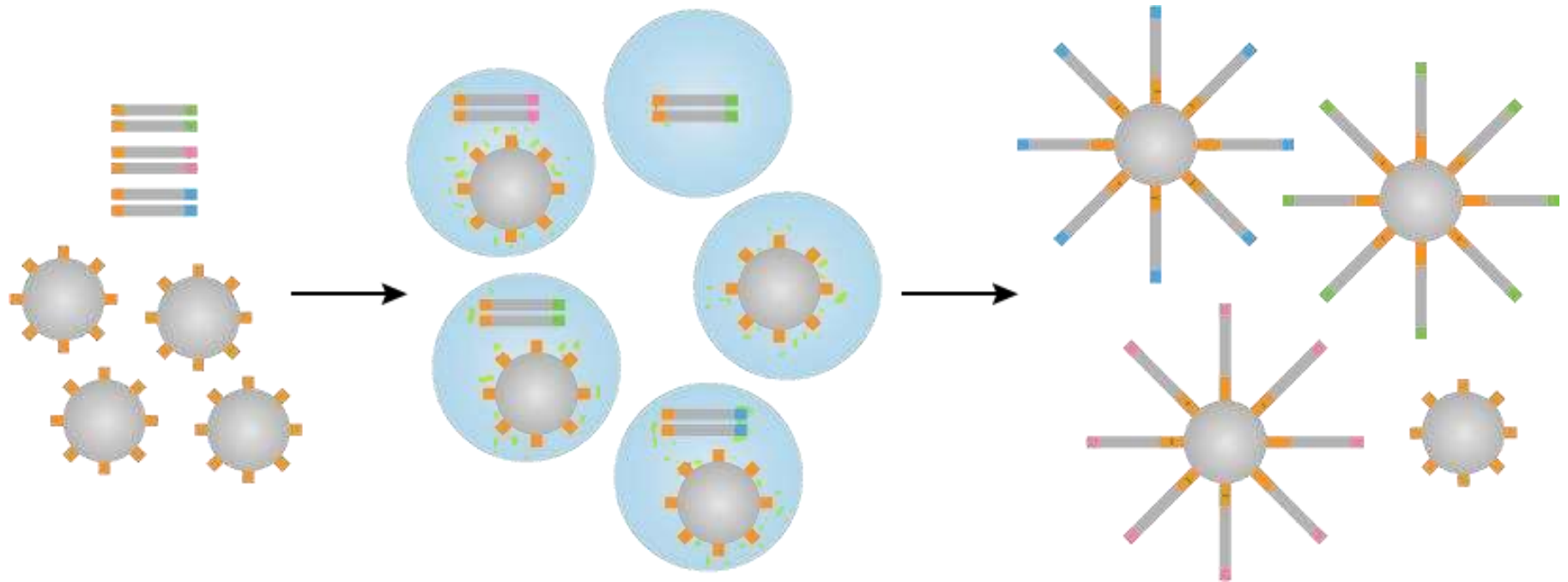
- In order for sequencing to be successful, the library fragments need to be spatially clustered in PCR colonies or 'colonies' as they are conventionally known, which consist of many copies of a particular library fragment.
- Since these colonies are attached in a planar fashion, the features of the array can be manipulated enzymatically in parallel.
- This method of library construction is much faster than the previous labour intensive procedure of colony picking and *E. coli* cloning used to isolate and amplify DNA for Sanger sequencing, however, this is at the expense of read length of the fragments.

AMPLIFICATION

- Library amplification is required so that the received signal from the sequencer is strong enough to be detected accurately.
- With enzymatic amplification, phenomena such as 'biasing' and 'duplication' can occur leading to preferential amplification of certain library fragments.
- Instead, there are several types of amplification process which use PCR to create large numbers of DNA clusters.

Emulsion PCR

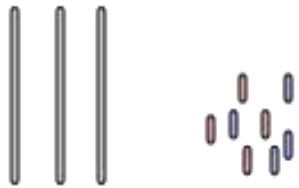
Emulsion oil, beads, PCRmix and the library DNA are mixed to form an emulsion which leads to the formation of micro wells



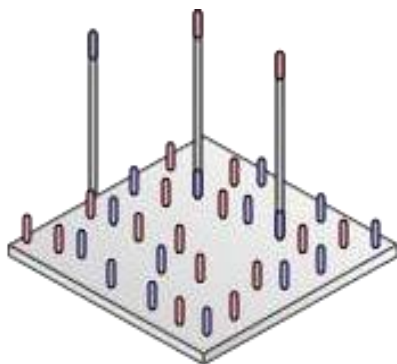
- In order for the sequencing process to be successful, each micro well should contain one bead with one strand of DNA (approximately 15% of micro wells are of this composition).
- The PCR then denatures the library fragment leading two separate strands, one of which (the reverse strand) anneals to the bead.
- The annealed DNA is amplified by polymerase starting from the bead towards the primer site.

- The original reverse strand then denatures and is released from the bead only to re-anneal to the bead to give two separate strands.
- These are both amplified to give two DNA strands attached to the bead.
- The process is then repeated over 30-60 cycles leading to clusters of DNA.

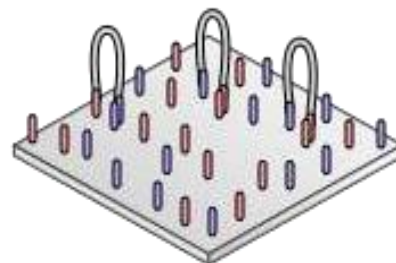
- This technique has been criticized for its time consuming nature, since it requires many steps (forming and breaking the emulsion, PCR amplification, enrichment etc) despite its extensive use in many of the NGS platforms.
- It is also relatively inefficient since only around two thirds of the emulsion micro reactors will actually contain one bead.
- Therefore an extra step is required to separate empty systems leading to more potential inaccuracies.



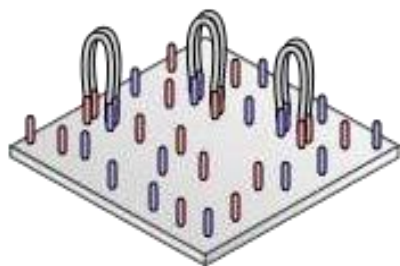
DNA fragments Primers



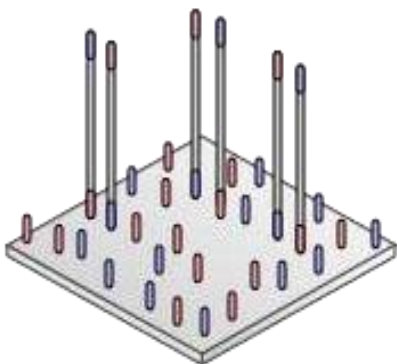
DNA strands are attached to cell surface at one end



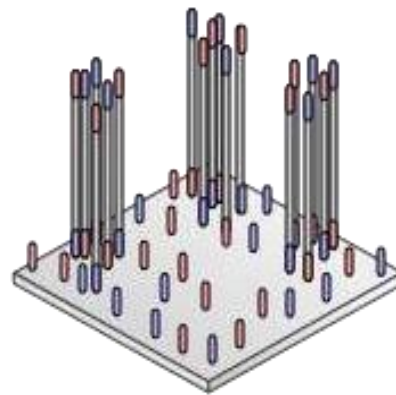
Ends are attached to surface by complimentary primers



Enzymes create double strands



Denaturation forms two separate DNA fragments



Repetition forms clusters of identical strands

Bridge PCR

- The surface of the flow cell is densely coated with primers that are complementary to the primers attached to the DNA library fragments.
- The DNA is then attached to the surface of the cell at random where it is exposed to reagents for polymerase based extension.
- On addition of nucleotides and enzymes, the free ends of the single strands of DNA attach themselves to the surface of the cell via complementary primers, creating bridged structures.

- Enzymes then interact with the bridges to make them double stranded, so that when the denaturation occurs, two single stranded DNA fragments are attached to the surface in close proximity.
- Repetition of this process leads to clonal clusters of localized identical strands.
- In order to optimize cluster density, concentrations of reagents must be monitored very closely to avoid overcrowding.

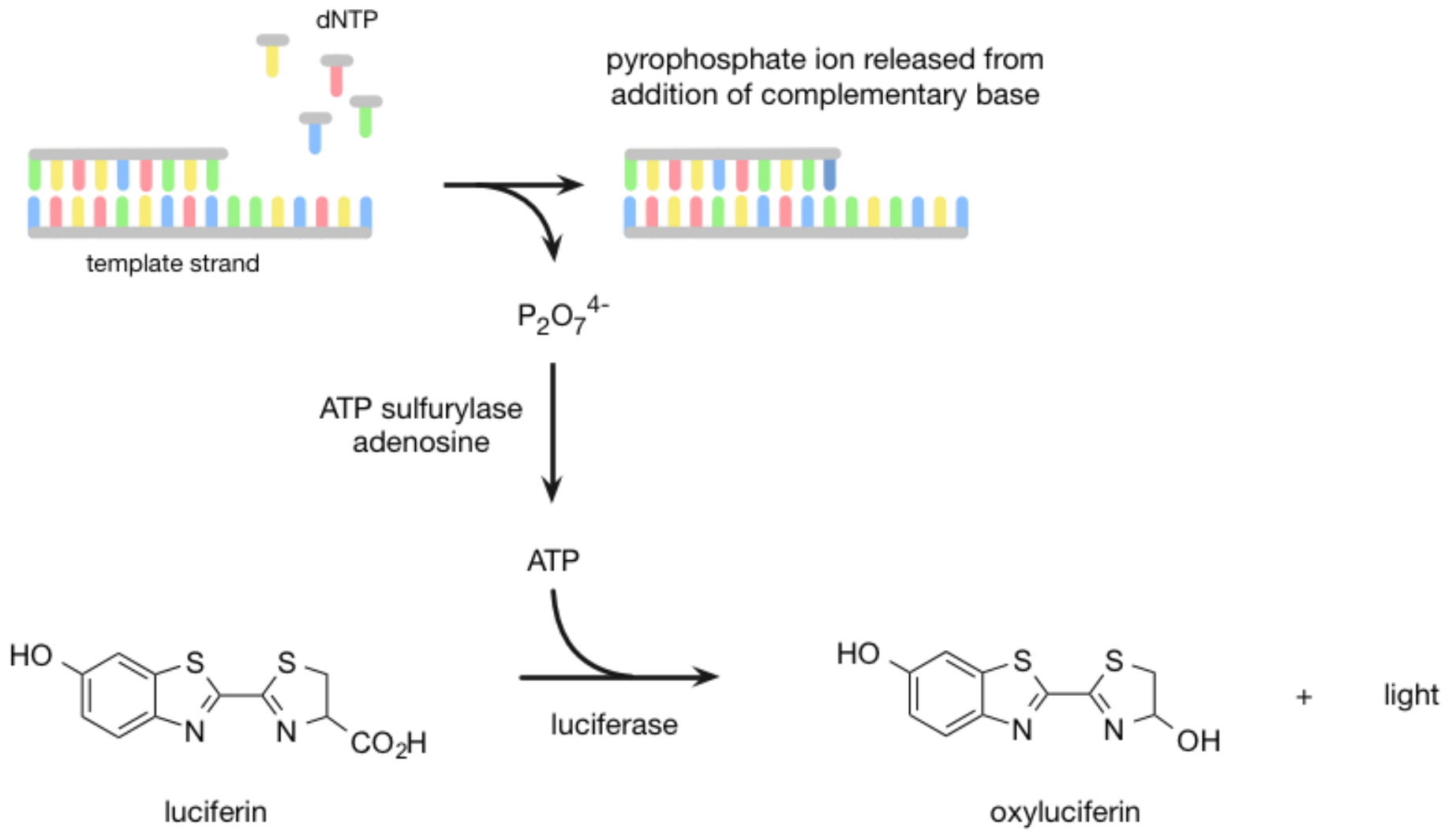
SEQUENCING

Several competing methods of Next Generation Sequencing have been developed by different companies.

1. 454 Pyrosequencing
2. Ion torrent semiconductor sequencing
3. Sequencing by ligation (SOLiD)
4. Reversible terminator sequencing (Illumina)
 - 3'-O-blocked reversible terminators
 - 3'-unblocked reversible terminators

1. 454 Pyrosequencing

- Pyrosequencing is based on the 'sequencing by synthesis' principle, where a complementary strand is synthesized in the presence of polymerase enzyme.
- In contrast to using dideoxynucleotides to terminate chain amplification (as in Sanger sequencing), pyrosequencing instead detects the release of pyrophosphate when nucleotides are added to the DNA chain.



➤ It initially uses the emulsion PCR technique to construct the colonies required for sequencing and removes the complementary strand.

➤ Next, a ssDNA sequencing primer hybridizes to the end of the strand (primer-binding region), then the four different dNTPs are then sequentially made to flow in and out of the wells over the colonies.

➤ When the correct dNTP is enzymatically incorporated into the strand, it causes release of pyrophosphate

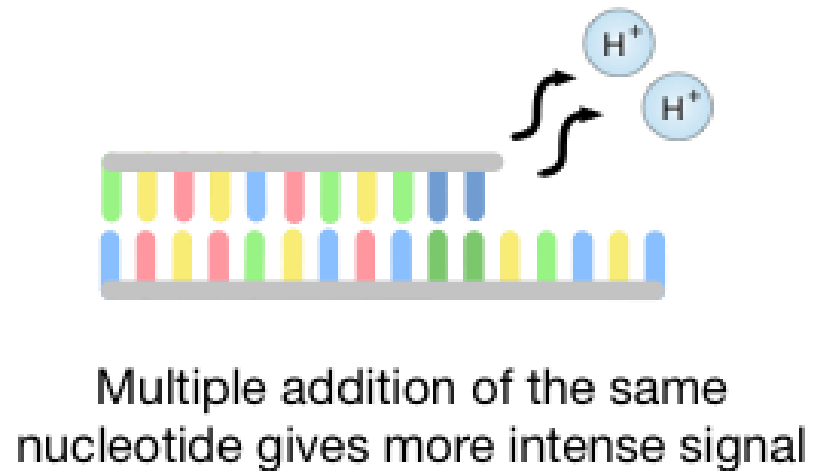
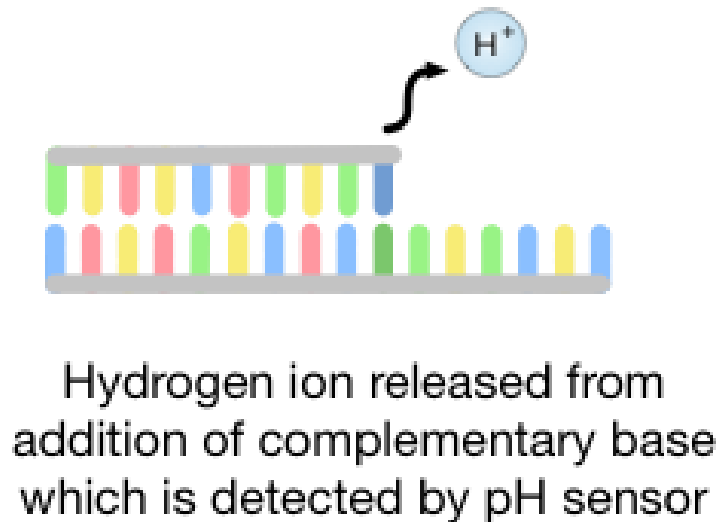
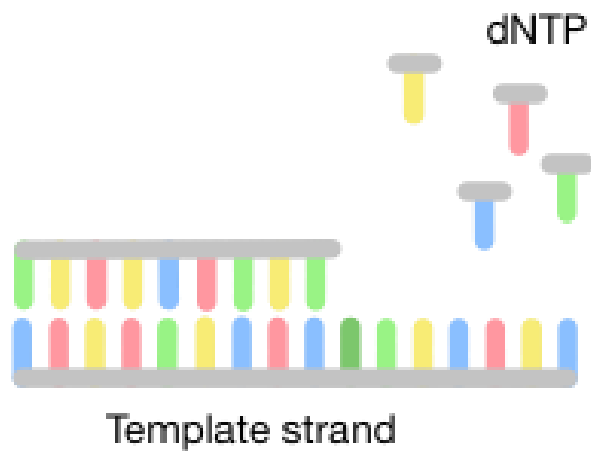
- In the presence of ATP sulfurylase and adenosine, the pyrophosphate is converted into ATP.
- This ATP molecule is used for luciferase-catalysed conversion of luciferin to oxyluciferin, which produces light that can be detected with a camera.
- The relative intensity of light is proportional to the amount of base added (i.e. a peak of twice the intensity indicates two identical bases have been added in succession).

➤ Pyrosequencing, developed by 454 Life Sciences, was one of the early successes of Next-generation sequencing; indeed, 454 Life Sciences produced the first commercially available Next-generation sequencer.

➤ However, the method was eclipsed by other technologies and, in 2013, new owners Roche announced the closure of 454 Life Sciences and the discontinuation of the 454 pyrosequencing platform.

2. Ion torrent semiconductor sequencing

- Ion torrent sequencing uses a "sequencing by synthesis" approach, in which a new DNA strand, complementary to the target strand, is synthesized one base at a time.
- A semiconductor chip detects the hydrogen ions produced during DNA polymerization



- Following colony formation using emulsion PCR, the DNA library fragment is flooded sequentially with each nucleoside triphosphate (dNTP), as in pyrosequencing.
- The dNTP is then incorporated into the new strand if complementary to the nucleotide on the target strand.
- Each time a nucleotide is successfully added, a hydrogen ion is released, and it detected by the sequencer's pH sensor.
- As in the pyrosequencing method, if more than one of the same nucleotide is added, the change in pH/signal intensity is correspondingly larger.

- Ion torrent sequencing is the first commercial technique not to use fluorescence and camera scanning.
- It is therefore faster and cheaper than many of the other methods.
- Unfortunately, it can be difficult to enumerate the number of identical bases added consecutively.
- For example, it may be difficult to differentiate the pH change for a homo repeat of length 9 to one of length 10, making it difficult to decode repetitive sequences.

3. Sequencing by ligation (SOLiD)

- SOLiD is an enzymatic method of sequencing that uses DNA ligase, an enzyme used widely in biotechnology for its ability to ligate double-stranded DNA strands .
- Emulsion PCR is used to immobilize/amplify a ssDNA primer-binding region (known as an adapter) which has been conjugated to the target sequence (i.e. the sequence that is to be sequenced) on a bead.
- These beads are then deposited onto a glass surface – a high density of beads can be achieved which in turn, increases the throughput of the technique.

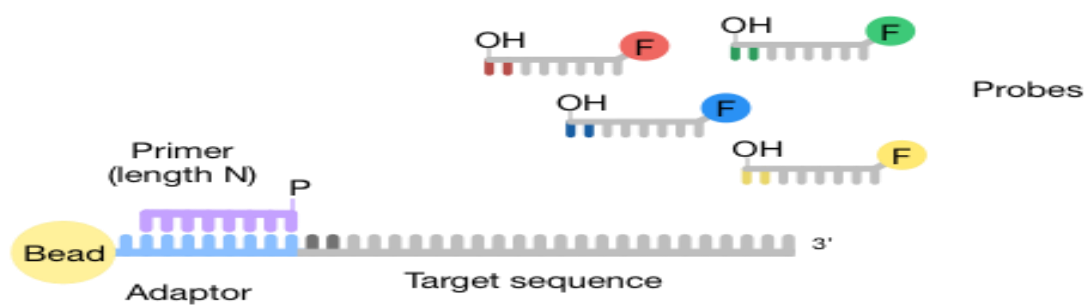
- Once bead deposition has occurred, a primer of length N is hybridized to the adapter, then the beads are exposed to a library of 8-mer probes which have different fluorescent dye at the 5' end and a hydroxyl group at the 3' end.
- Bases 1 and 2 are complementary to the nucleotides to be sequenced whilst bases 3-5 are degenerate and bases 6-8 are inosine bases.
- Only a complementary probe will hybridize to the target sequence, adjacent to the primer.

- DNA ligase is then used to join the 8-mer probe to the primer.
- A phosphorothioate linkage between bases 5 and 6 allows the fluorescent dye to be cleaved from the fragment using silver ions.
- This cleavage allows fluorescence to be measured (four different fluorescent dyes are used, all of which have different emission spectra) and also generates a 5'-phosphate group which can undergo further ligation.

➤ Once the first round of sequencing is completed, the extension product is melted off and then a second round of sequencing is performed with a primer of length $N-1$.

➤ Many rounds of sequencing using shorter primers each time (i.e. $N-2$, $N-3$ etc) and measuring the fluorescence ensures that the target is sequenced.

- Due to the two-base sequencing method (since each base is effectively sequenced twice), the SOLiD technique is highly accurate (at 99.999% with a sixth primer, it is the most accurate of the second generation platforms) and also inexpensive.
- It can complete a single run in 7 days and in that time can produce 30 Gb of data.
- Unfortunately, its main disadvantage is that read lengths are short, making it unsuitable for many applications.



Ligase

Complementary probe hybridizes to target sequence

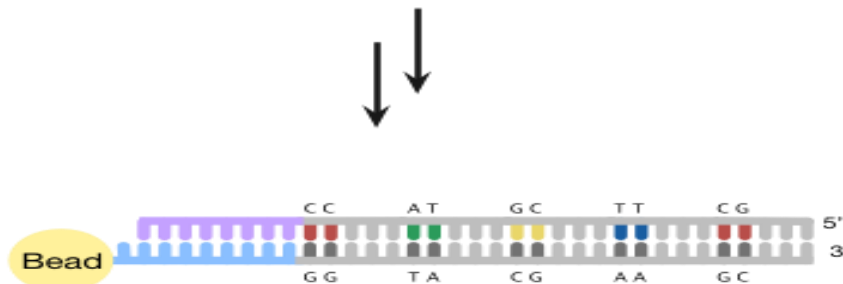


Ligation cycle repeated

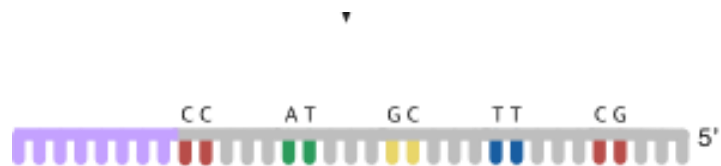
F

Fluorescent dye cleaved and detected by four-color imaging

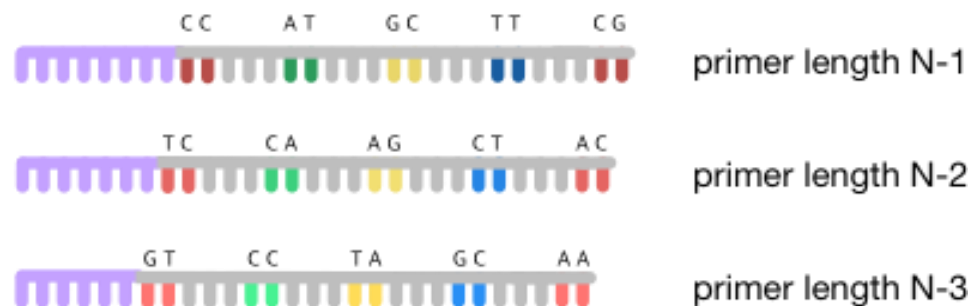
Probes + Ligase



Extension product melted off



Sequencing is repeated,
each time with a primer
which is shorter by one base



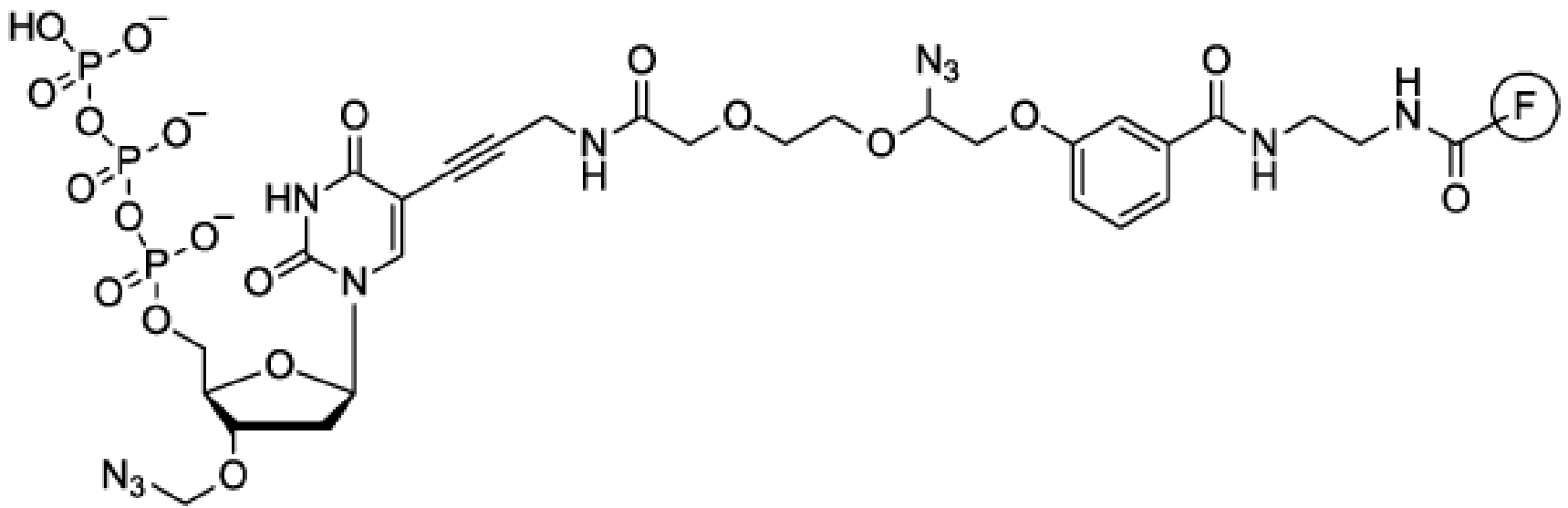
The whole target is sequenced by
observing different emission spectra
from several rounds

4. Reversible terminator sequencing (Illumina)

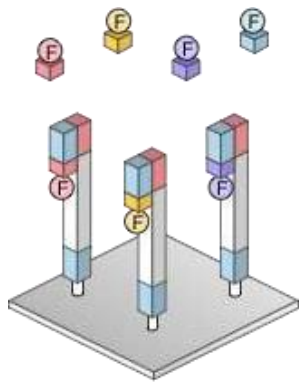
- Reversible terminator sequencing differs from the traditional Sanger method in that, instead of terminating the primer extension irreversibly using dideoxynucleotide, modified nucleotides are used in reversible termination.
- Whilst many other techniques use emulsion PCR to amplify the DNA library fragments, reversible termination uses bridge PCR, improving the efficiency of this stage of the process.

a. 3'-O-blocked reversible terminators

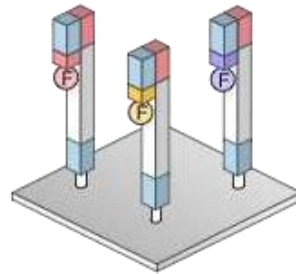
- The mechanism uses a sequencing by synthesis approach, elongating the primer in a stepwisemanner.
- Firstly, the sequencing primers and templates are fixed to a solid support.
- The support is exposed to each of the four DNA bases, which have a different fluorophore attached (to the nitrogenous base) in addition to a 3'-O-azidomethylgroup



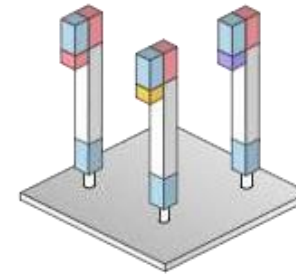
- Only the correct base anneals to the target and is subsequently ligated to the primer.
- The solid support is then imaged and nucleotides that have not been incorporated are washed away and the fluorescent branch is cleaved using TCEP (tris(2-carboxyethyl)phosphine).
- TCEP also removes the 3'-O-azidomethyl group, regenerating 3'-OH, and the cycle can be repeated.



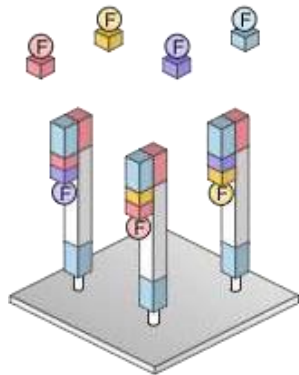
Primers are extended
by one base



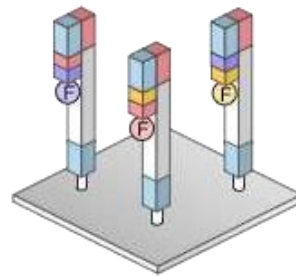
Excess nucleotides
are washed away



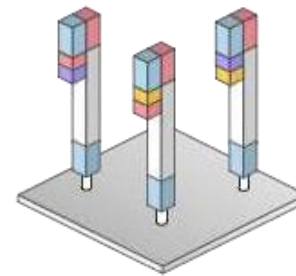
Fluorophore branches
are removed



More nucleotides
are added



Excess nucleotides
are washed away



Fluorophore branches
are removed

A C Fluorophore

G T

b. 3'-unblocked reversible terminators

- The reversible termination group of 3'-unblocked reversible terminators is linked to both the base and the fluorescence group, which now acts as part of the termination group as well as a reporter.
- This method differs from the 3'-O-blocked reversible terminators method in three ways: firstly, the 3'-position is not blocked (i.e. the base has free 3'-OH); the fluorophore is the same for all four bases; and each modified base is flowed in sequentially rather than at the same time.

➤ The main disadvantage of these techniques lies with their poor read length, which can be caused by one of two phenomena.

➤ In order to prevent incorporation of two nucleotides in a single step, a block is put in place, however in the event of no block addition due to a poor synthesis, strands can become out of phase creating noise which limits readlength.

➤ Noise can also be created if the fluorophore is unsuccessfully attached or removed.

- These problems are prevalent in other sequencing methods and are the main limiting factors to read length.
- This technique was pioneered by Illumina, with their HiSeq and MiSeq platforms.
- HiSeq is the cheapest of the second generation sequencers with a cost of \$0.02 per million bases.
- It also has a high data output of 600 Gb per run which takes around 8 days to complete.

Thank you