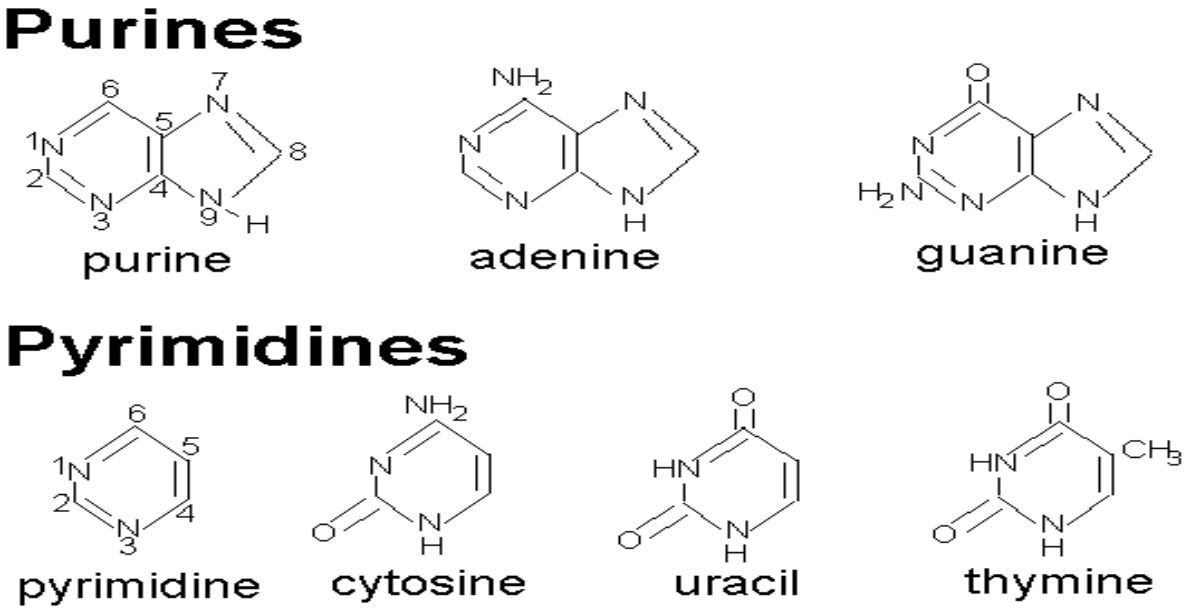
DNA SEQUENCING

DNA SEQUENCING

- Determining the precise order of nucleotides within a DNA molecule.



- Used to determine the sequence of individual genes, larger genetic regions, full chromosomes or entire genomes.
- The resulting sequences may be used by researchers in molecular biology or genetics to further scientific progress.

HISTORY OF DNA SEQUENCING

- 1972 – Earliest nucleotide sequencing – RNA sequencing of Bacteriophage MS2 by WALTER FIESERR
- Early sequencing was performed with tRNA through a technique developed by Richard Holley, who published the first structure of a tRNA in 1964.
- 1977 - DNA sequencing FREDRICK SANGER by Chain termination method
- Chemical degradation method by ALLAN MAXAM and WALTER GILBERT
- 1977 - First DNA genome t be sequenced of Bacteriophage ΦX174
- 1986 - LOREY and SMITH gave Semiautomated sequencing
- 1987 – Applied biosystems marketed Fully automated sequencing machines

- 1995 – CRAIG VENTER, HAMILTON SMITH and collegues published first complete genome sequence of *Haemophilus influenzae*

- 2003 – Human genome project

- 2ND Generation of DNA sequencing

- 3RD Generation of DNA sequencing

Determining the Sequence of DNA

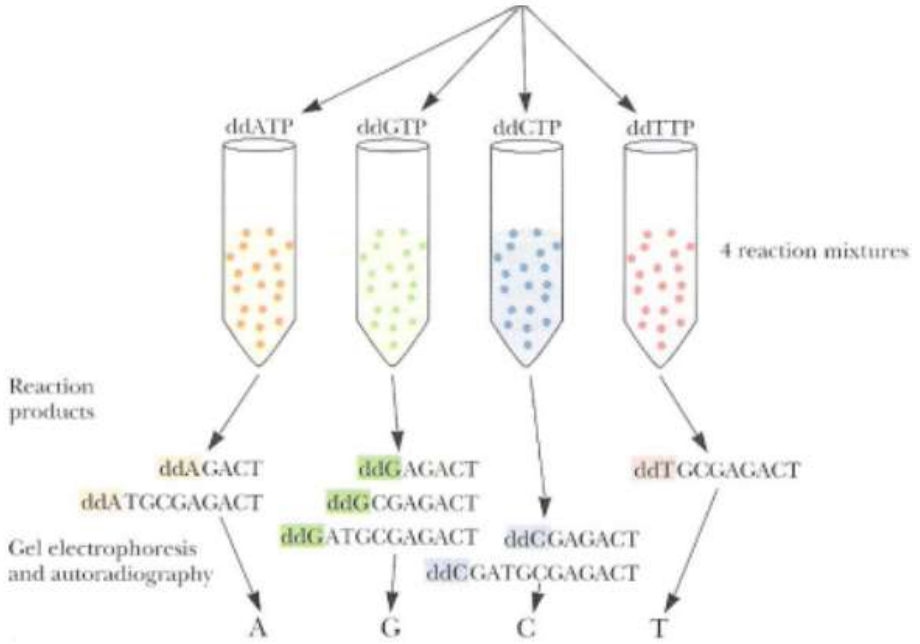
- Methods:

1) Maxam and Gilbert chemical degradation method
2) Chain termination or Dideoxy method
   - Fredrick Sanger
3) Genome sequencing method
   - Shotgun sequencing
   - Clone contig approach
4) 2nd generation sequencing methods
   - Pyrosequencing
   - Nanopore sequencing
   - Illumina sequencing
   - Solid sequencing

SANGER SEQEUNCING

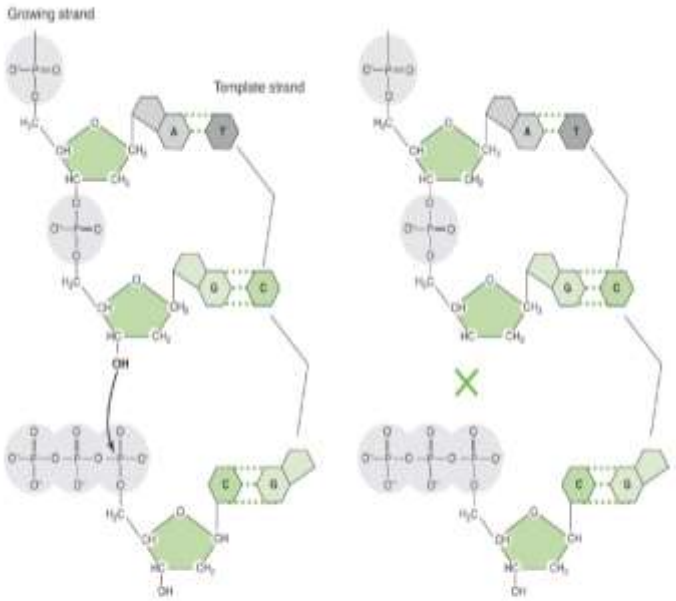
- Chain termination method of DNA sequencing.
- It involves following components:

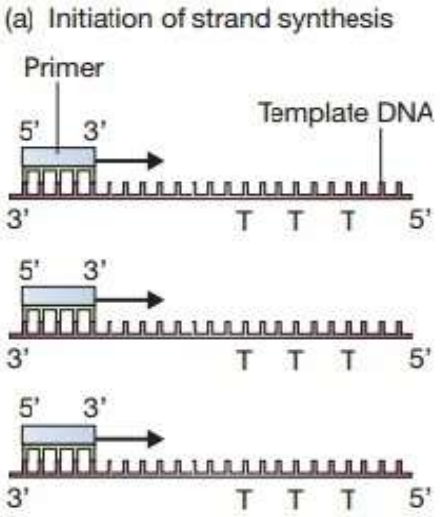
a) 1. Primer
b) 2. DNA template
c) 3. DNA polymerase
d) 4.. dNTPs(A,T,G,C)
e) 5. ddNTPs

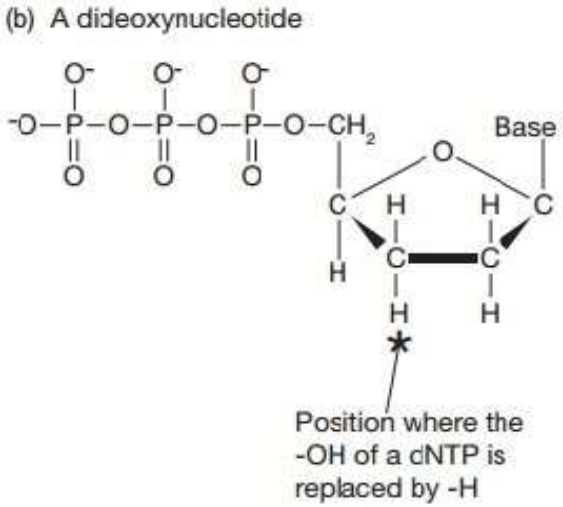


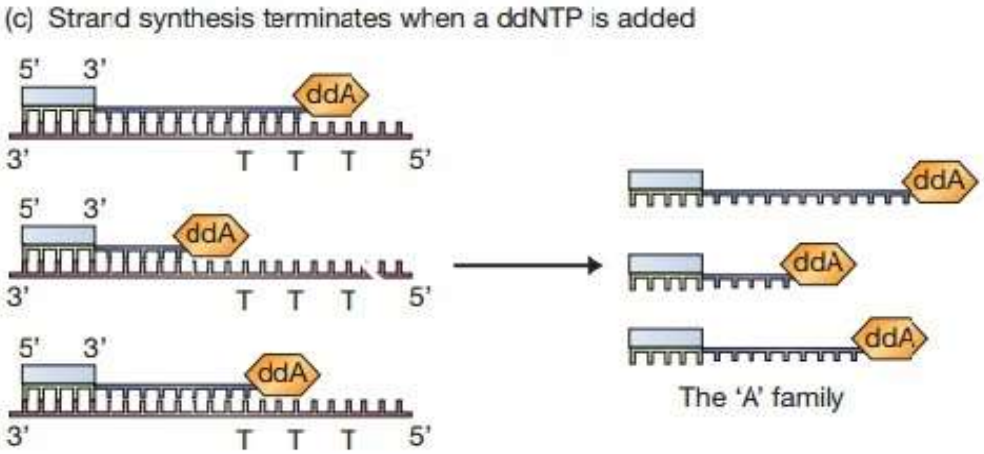
- 4 Steps:
    1. Denaturation
    2. Primer attachment and extension of bases
    3. Termination
    4. Poly acrylamide gel electrophoresis

SANGER'S METHOD



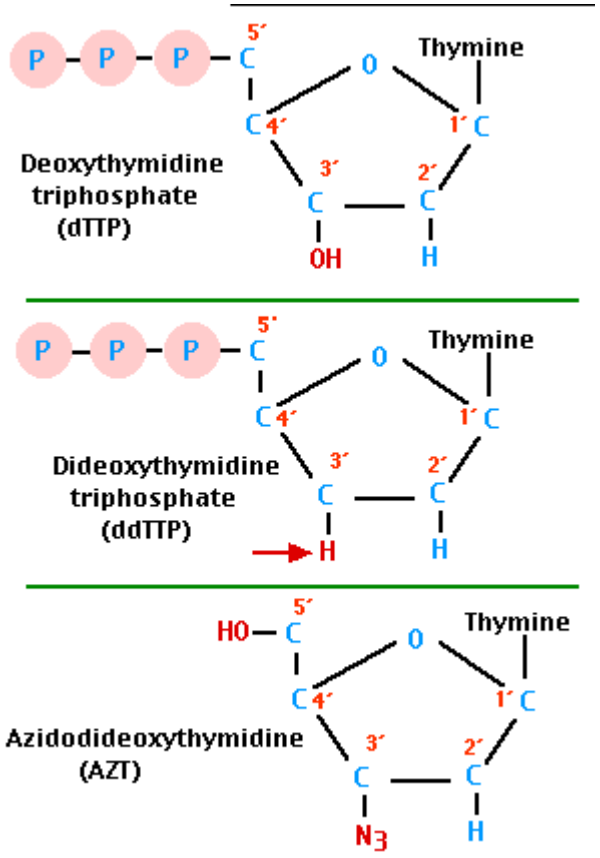




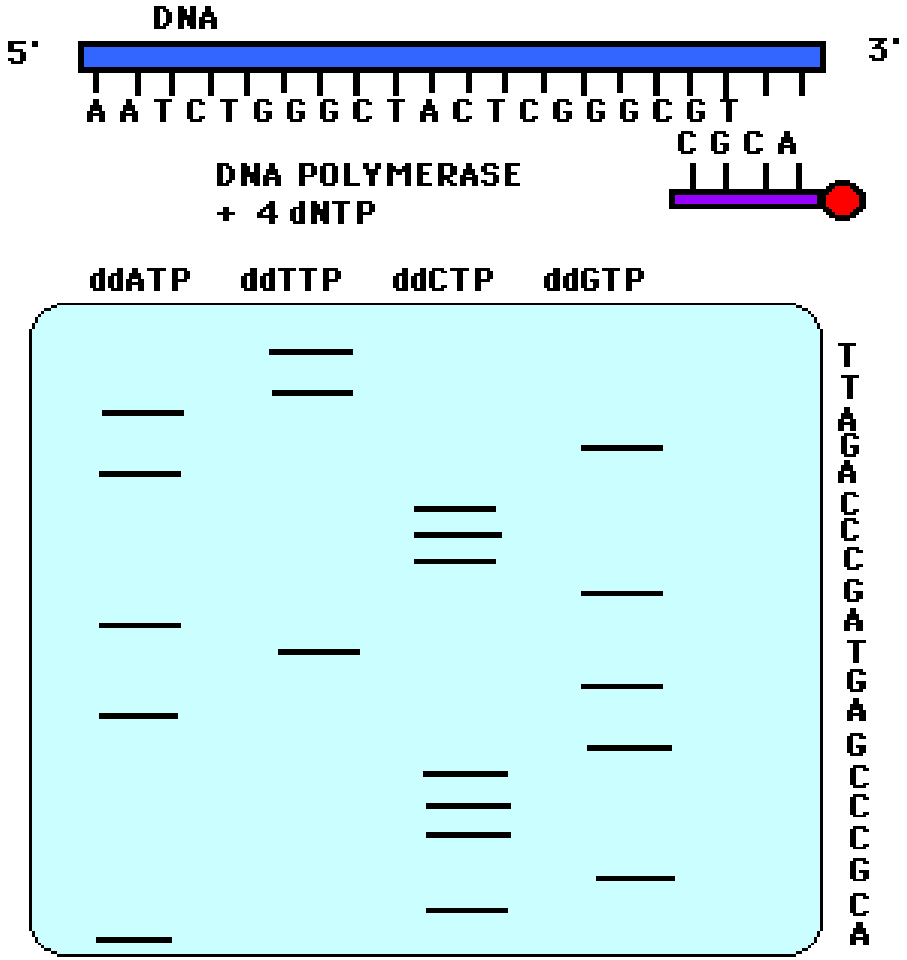


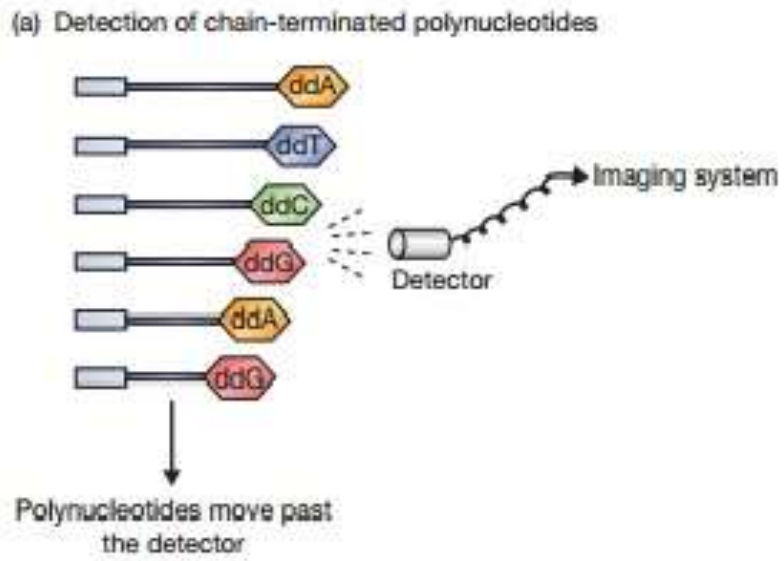
Chain Termination (Sanger) Sequencing

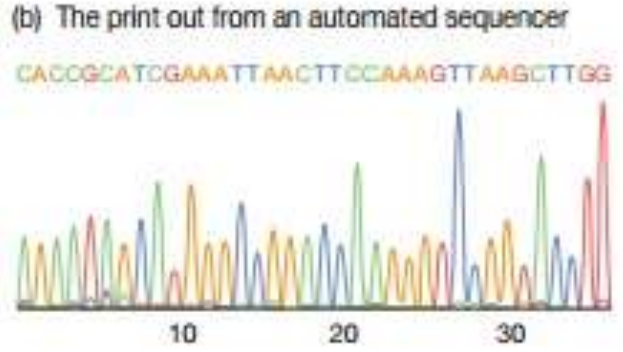
A	ddATP + four dNTPs	ddA dAdGdCdTdGdCdCdCdG
C	ddCTP + four dNTPs	dAdGddC dAdGdCdTdGddC dAdGdCdTdGdCddC dAdGdCdTdGdCdCddC
G	ddGTP + four dNTPs	dAddG dAdGdCdTddG dAdGdCdTdGdCdCdCddG
T	ddTTP + four dNTPs	dAdGdCddT dAdGdCdTdGdCdCdCdG



Determination of nucleotide sequence







SANGER'S METHOD

- Not all polymerases can be used as they have mixed activity of polymerizing and degrading.
- Both exonuclease activities are detrimental.
- Klenow fragment was used in orignal method but it has low processivity.
- So Sequenase from bacteriophage T7 was uesd with high processivity and no exonuclease added.
- Method requires ss DNA. So it is obtained by
  - Denaturation with alkali or boiling
  - DNA can be cloned in phagemid containg M13 ori and can take up DNA fragments of 10kb

PYROSEQUENCING

- Pyrosequencing is the second important type of DNA sequencing methodology in use today.

- The addition of a DNTP is accompanied by release of a molecule of pyrophosphate.

- Reaction mixture contains

❖ DNA sample to be sequenced
❖ Primers
❖ Deoxynucleotides
❖ DNA polymerase
❖ Sulfurylase

- The release of pyrophosphate is converted by the enzyme sulfurylase into a flash of chemiluminescence which is easily automated.

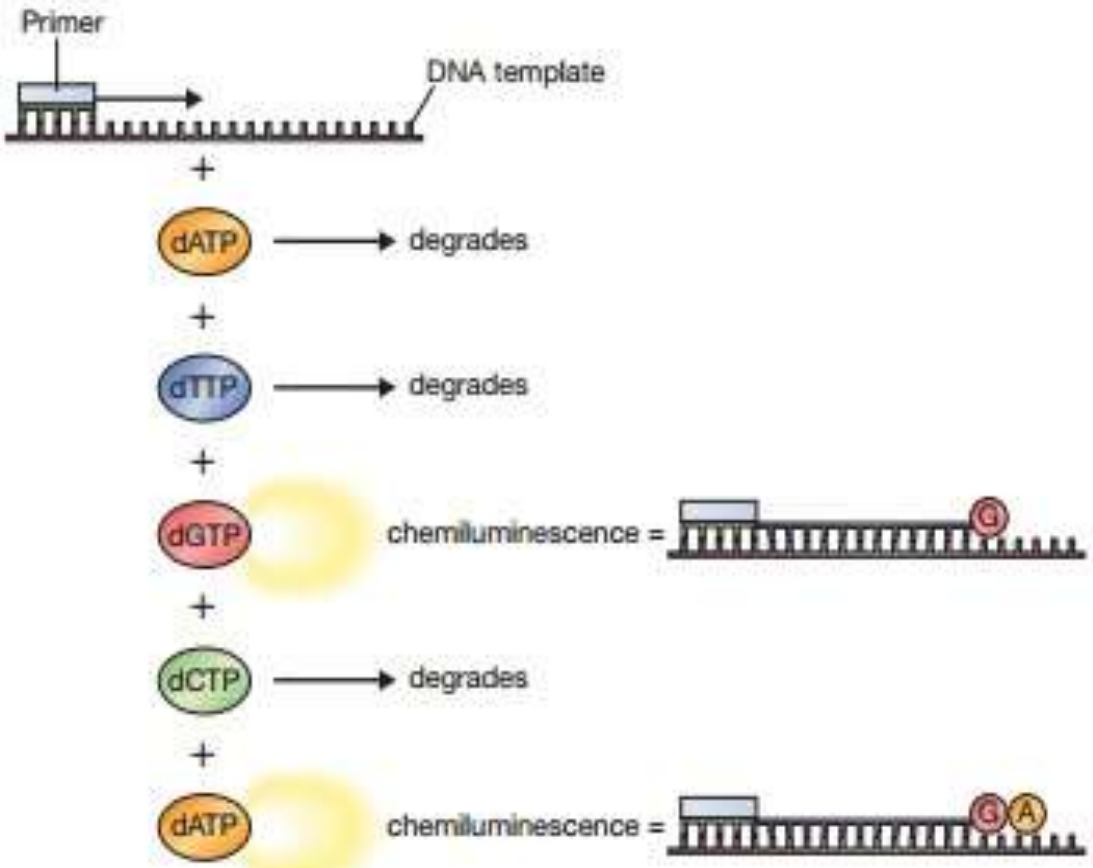
PYROSEQUENCING

❑**Advantages:**

- ❑Accurate
- ❑Parallel processing
- ❑Easily automated
- ❑Eliminates the need for labeled primers and nucleotides
- ❑No need for gel electrophoresis

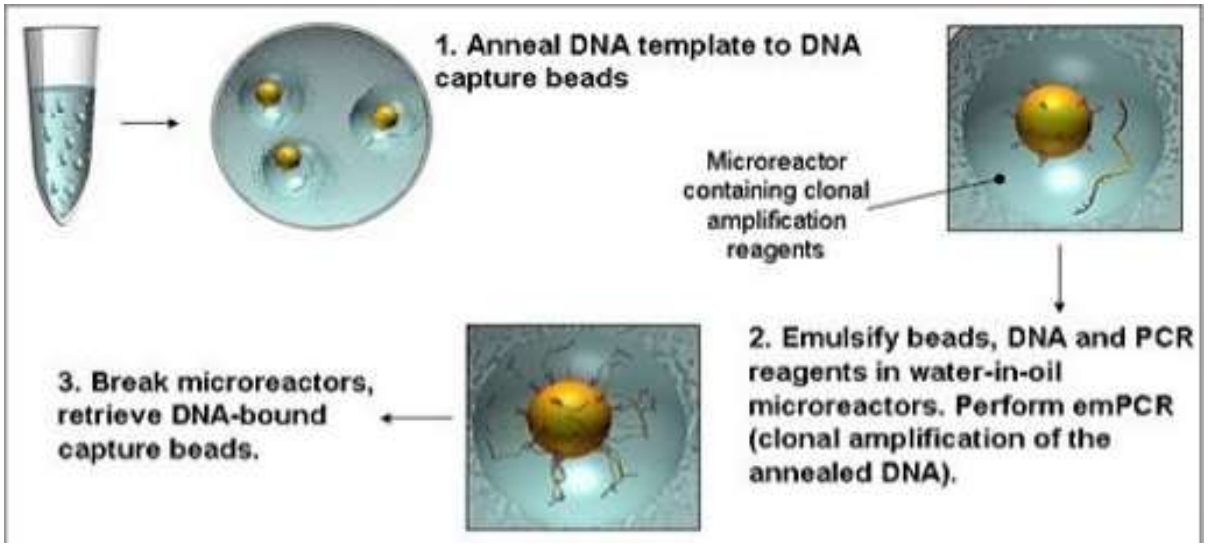
❖**DISADVANTAGES**

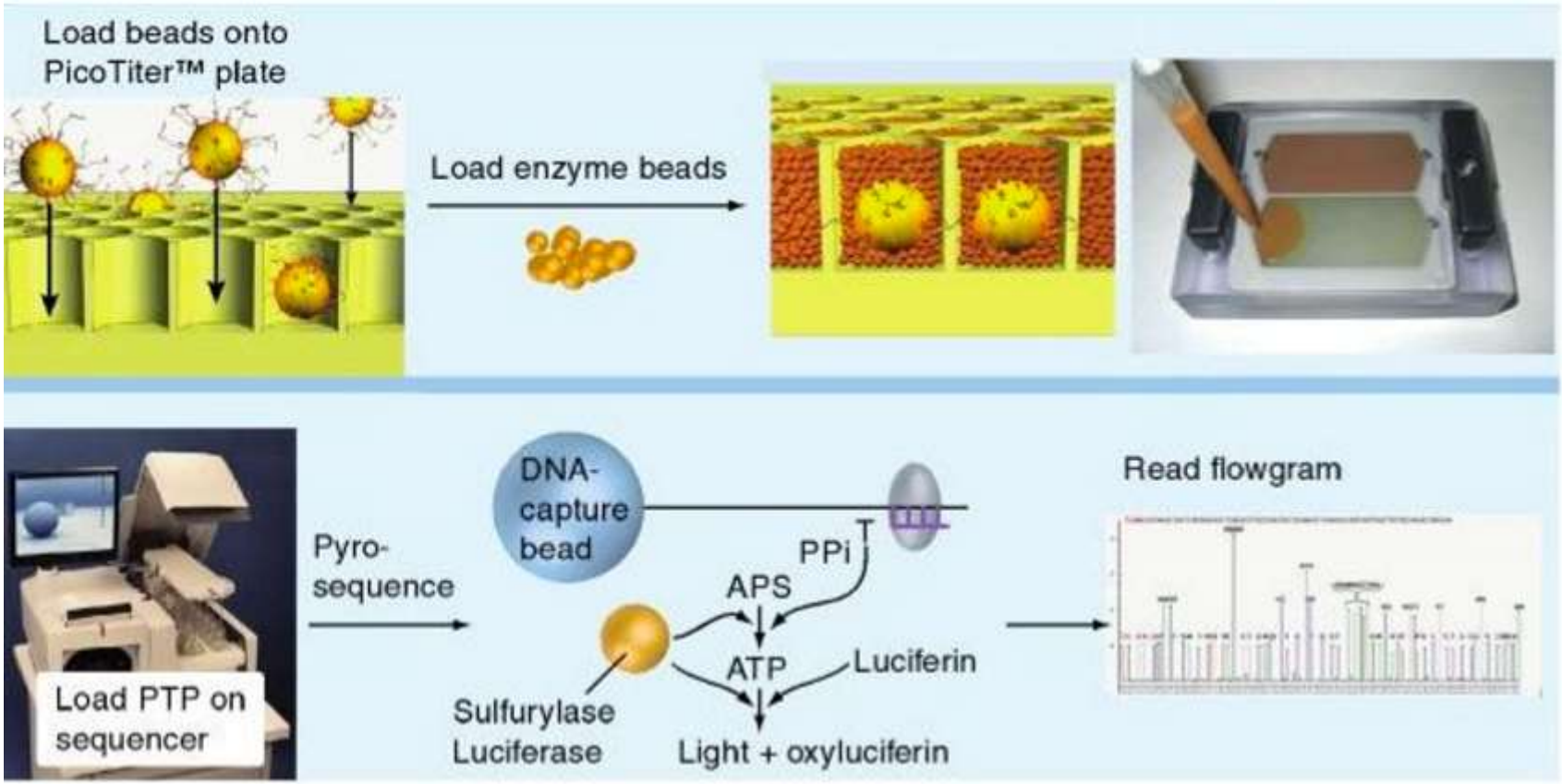
- ❖Smaller sequences
- ❖Nonlinear light response after more than 5-6 identical nucleotides



MASSIVELY PARALLEL PYROSEQUENCING

- The DNA is broken down into fragments between 300 to 500bp
- Each fragment is ligated with a pair of adaptor
    - To attach to the beads
    - Provide annealing sites for the primers for performing PCR
- Adaptors are attached to beads by biotin-streptavidin linkage
- Just one fragment becomes attached to one bead
- Each DNA fragment is now amplified using
- PCR is carried out in a oil emulsion, each bead residing within own droplet in the emulsion
- Each droplet contains all the reagents for PCR and is physically seprated from all the other droplets by the barrier provided by the oil components in the emulsion.
- After PCR, the droplets are transferred on wells on plastic strip and pyrosequencing reactions are carried out


1. Anneal DNA template to DNA capture beads
Microreactor containing clonal amplification reagents
2. Emulsify beads, DNA and PCR reagents in water-in-oil microreactors. Perform emPCR (clonal amplification of the annealed DNA).
3. Break microreactors, retrieve DNA-bound capture beads.


Load beads onto PicoTiter™ plate
Load enzyme beads
Load PTP on sequencer
Pyro-sequence
DNA-capture bead
PPi
APS
ATP
Luciferin
Sulfurylase Luciferase
Light + oxyluciferin
Read flowgram

SHOTGUN SEQUENCING

• **Shotgun sequencing**, also known as **shotgun cloning**, is a method used for sequencing long DNA strands or the whole genome.

•In shotgun sequencing, DNA is broken up randomly into numerous small segments and overlapping regions are identified between all the individual sequences that are generated.

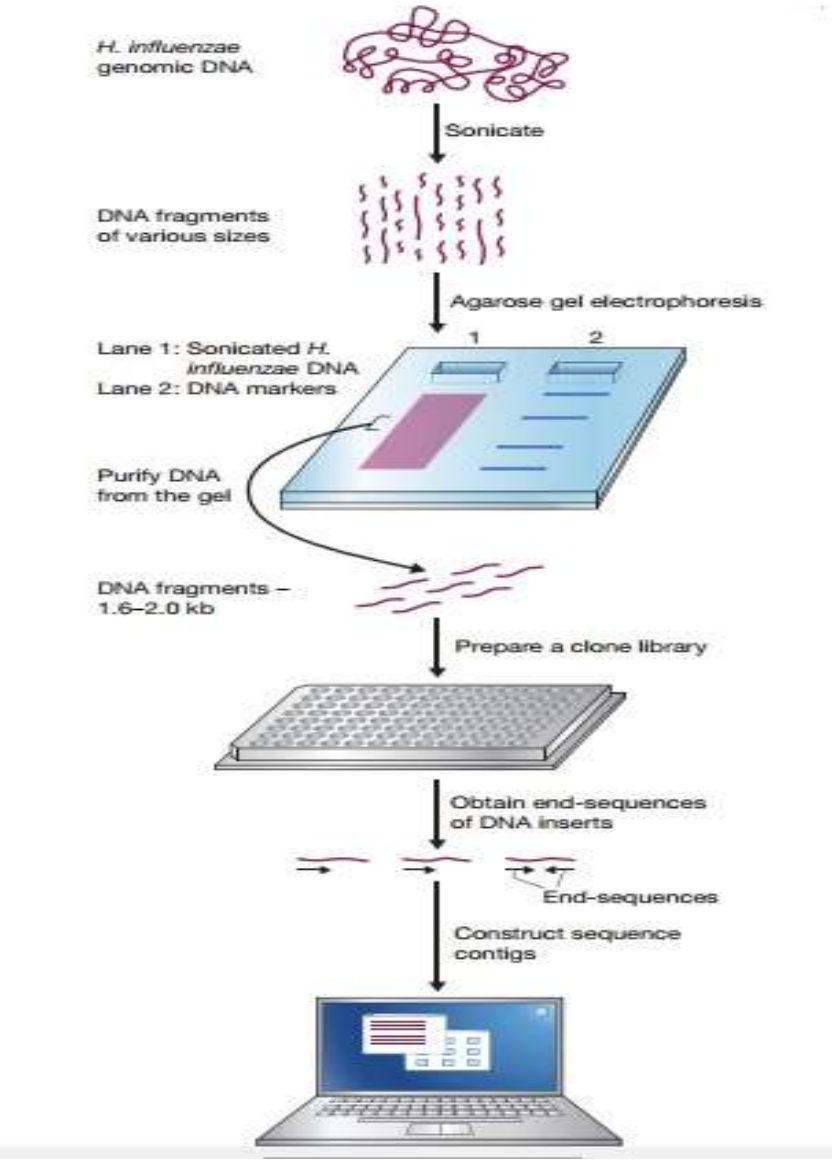
• Multiple overlapping reads for the target DNA are obtained by performing several rounds of this fragmentation and sequencing.

•Computer programs then use the overlapping ends of different reads to assemble them into a continuous sequence.

•The shotgun approach was first used successfully with the bacterium *Haemophilus influenzae.*

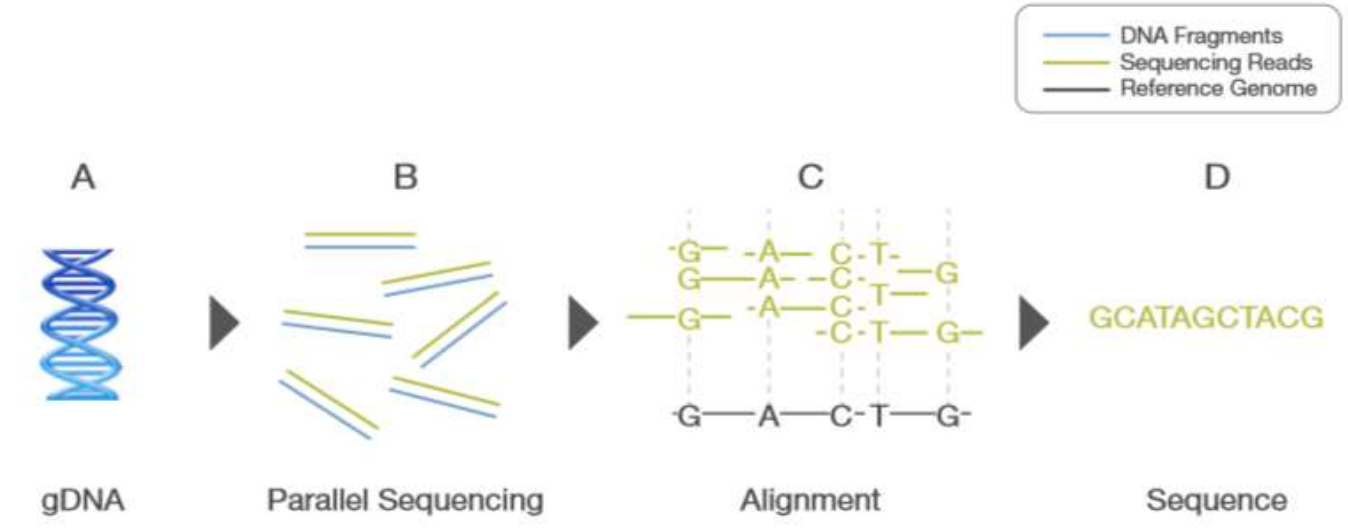
•Craig venter used this method to map the Human genome project in 2001.

Shotgun sequencing



NEXT GENERATION SEQUENCING

- The concept behind NGS – the bases of small fragments of DNAare sequentially identifed as signals emitted as eachfragment is resynthesized from a dna template strand
- NGS extends this process across millions of reactions in a massively parallel fashion rather than being limited to a single or a few dna fragments



A. Extracted gDNA.
B. gDNA is fragmented into a library of small segments that are each sequenced in parallel.
C. Individual sequence reads are reassembled by aligning to a reference genome.
D. The whole-genome sequence is derived from the consensus of aligned reads.

Next Generation Sequencing

Different platforms

Andy Vierstraete,
Department of Biology,
Ghent University. June 2012

- 454 Sequencing / Roche
  - GS Junior System
  - GS FLX+ System
- Illumina (Solexa)
  - HiSeq System
  - Genome analyzer IIx
  - MySeq
- Applied Biosystems - Life Technologies
  - SOLiD 5500 System
  - SOLiD 5500xl System
- Ion Torrent - Life Technologies
  - Personal Genome Machine (PGM)
  - Proton
- Helicos
  - Helicos Genetic Analysis System
- Pacific Biosciences
  - PacBio RS
- Oxford Nanopore Technologies
  - GridION System
  - MinION

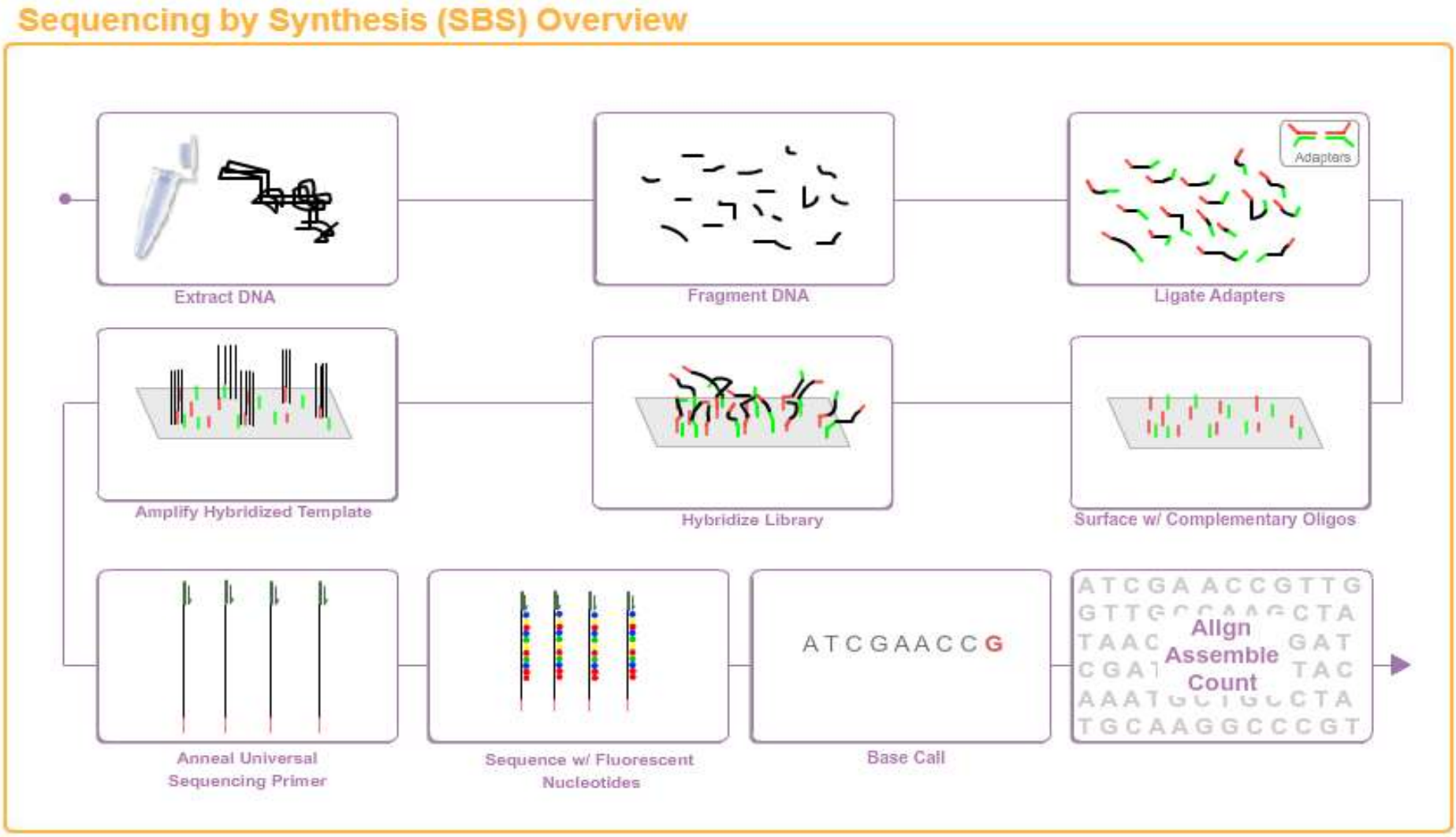
Next Generation Sequencing
Amplified Single Molecule Sequencing

Third Generation Sequencing,
Next Next Generation Sequencing,
Single Molecule Sequencing

Differentiating Next Gen technologies

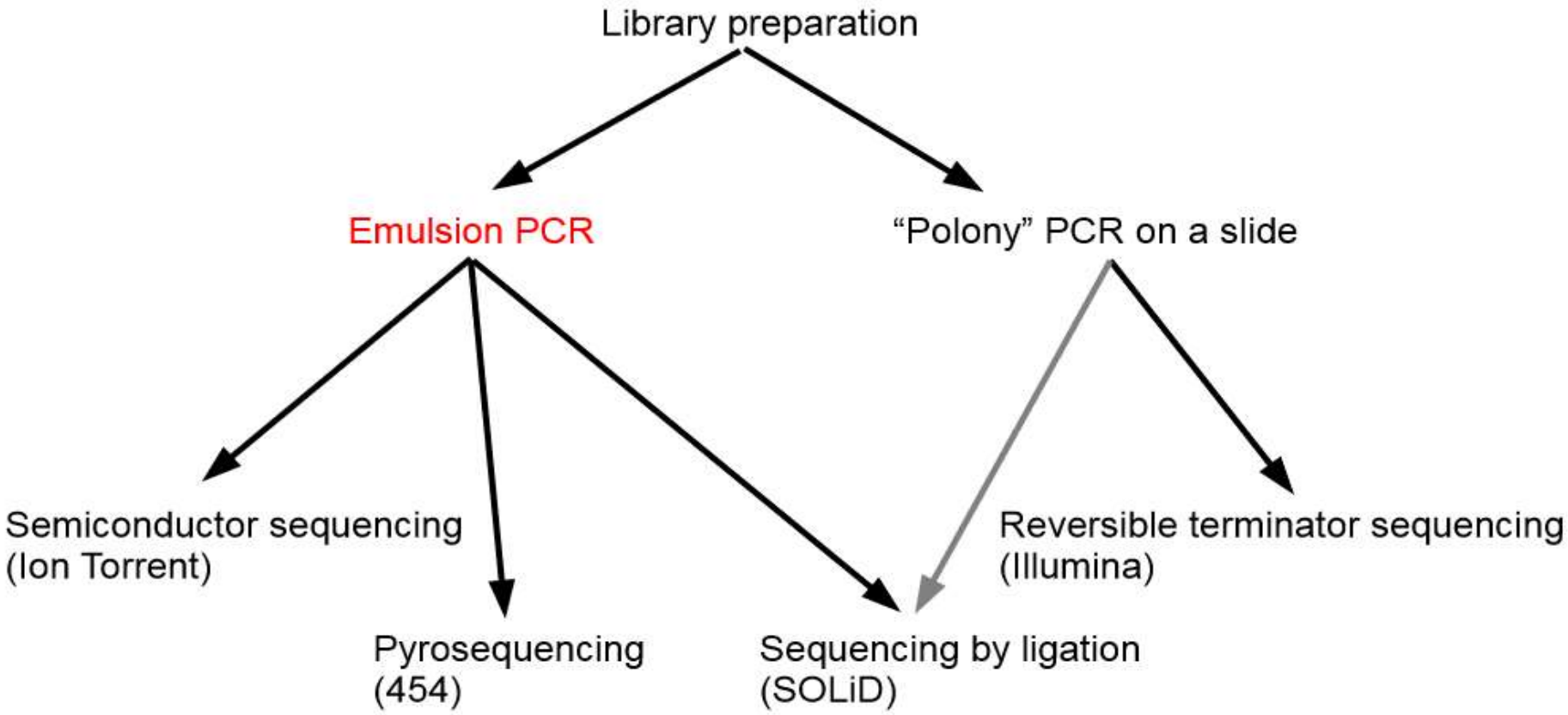
Illumina	Library Construction	Clonal amplification via bridge amplification	Massively parallel sequencing-by-synthesis of DNA clusters
454	Library Construction	Clonal amplification with emulsionPCR and enrichment	Massively parallel pyrosequencing of bead bound DNA templates
SOLiD	Library Construction	Clonal amplification with emulsionPCR and enrichment	Massively parallel ligation-based sequencing of bead bound DNA templates

Illumina sequencing



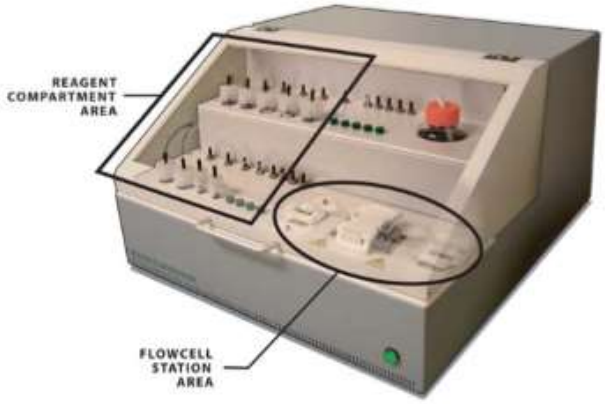
Andy Vierstraete,
Department of Biology,
Ghent University. June 2012

Next Generation Sequencing : Amplified Single Molecule Sequencing



Illumina Genome Analyzer

cBot



Cluster station

Paired end module

The flow cell - a core component

EVERYTHING EXCEPT SAMPLE PREPARATION IS COMPLETED ON THE FLOW CELL

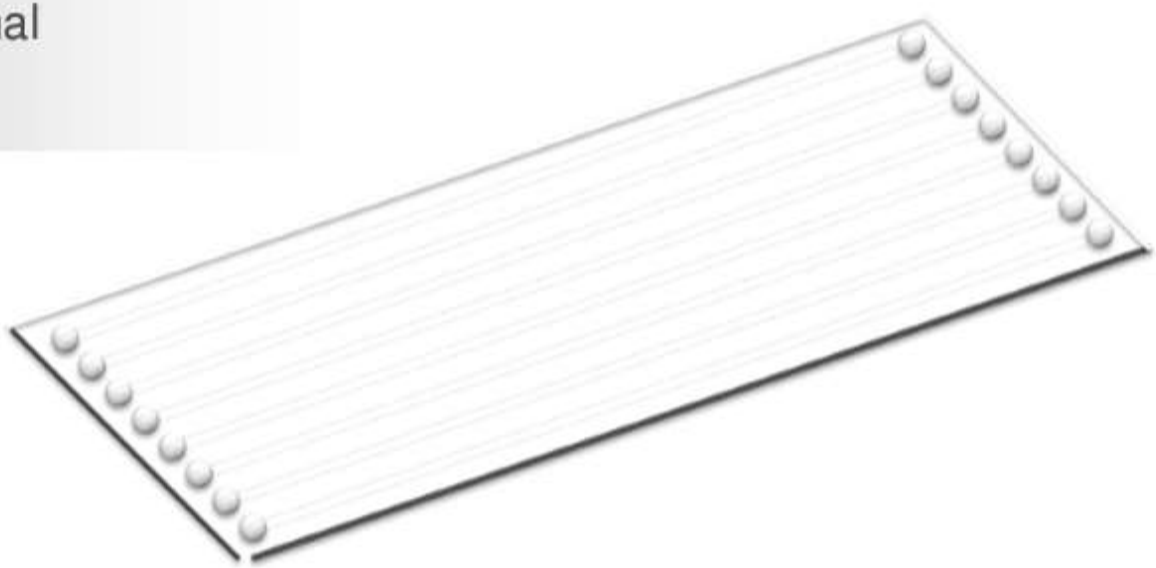
template annealing (1 - 96 samples)

template amplification

sequencing primer hybridization

Sequencing-by-synthesis reaction

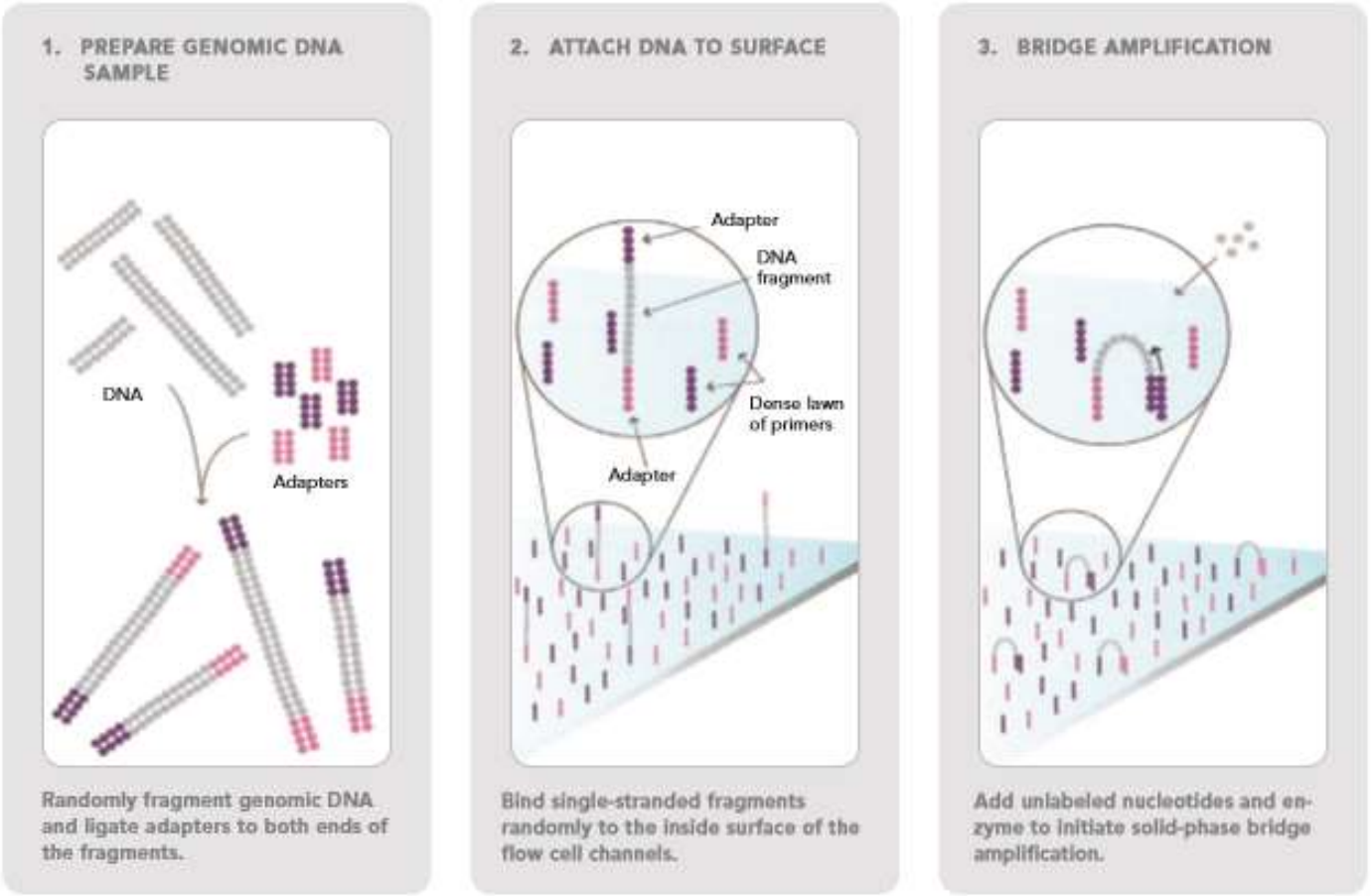
generation of fluorescent signal



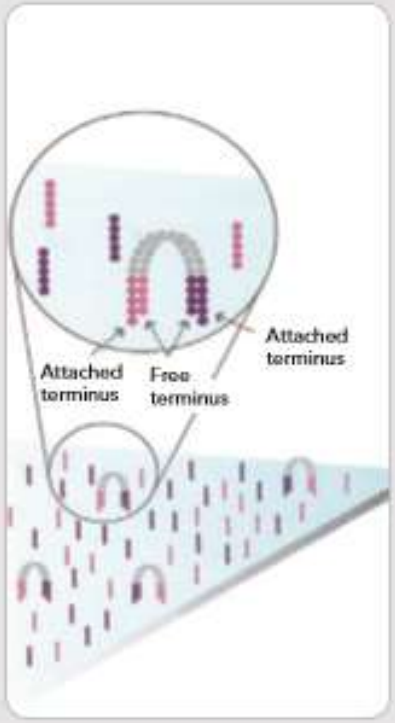
Illumina sequencing

With an experimental design determined, the ten steps of the Illumina Sequencing Workflow begin:

1. Extract Nucleic Acids
2. Quality Assessment
3. Ligate Adapters
4. Quantification of Libraries (optional step)
5. Hybridize to the Surface
6. Amplify Template
7. Linearization
8. Anneal Sequencing Primer
9. Sequence
10. Base Call & Analyze Results

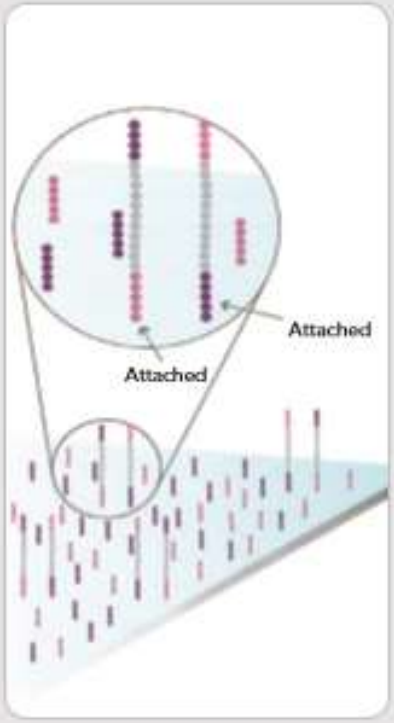


4. FRAGMENTS BECOME DOUBLE-STRANDED



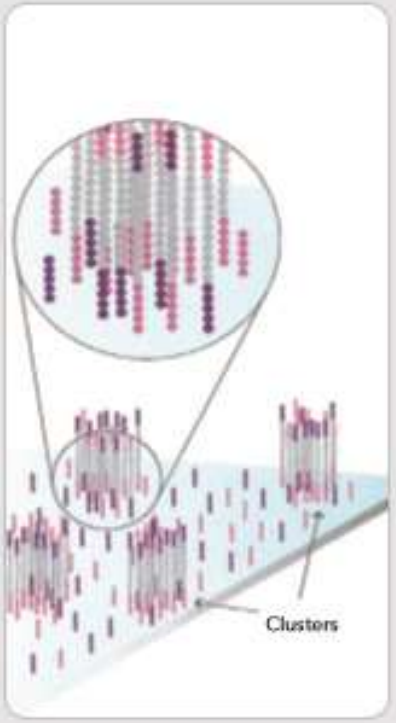
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION

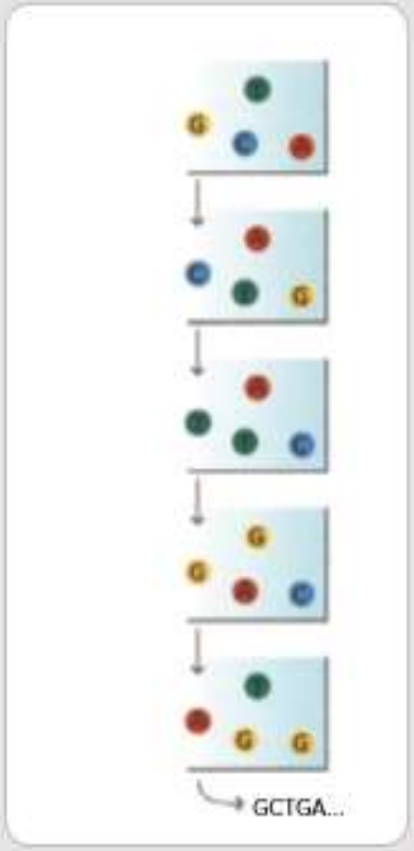


Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE

After laser excitation, the image is captured as before, and the identity of the second base is recorded.

11. SEQUENCING OVER MULTIPLE CHEMISTRY CYCLES



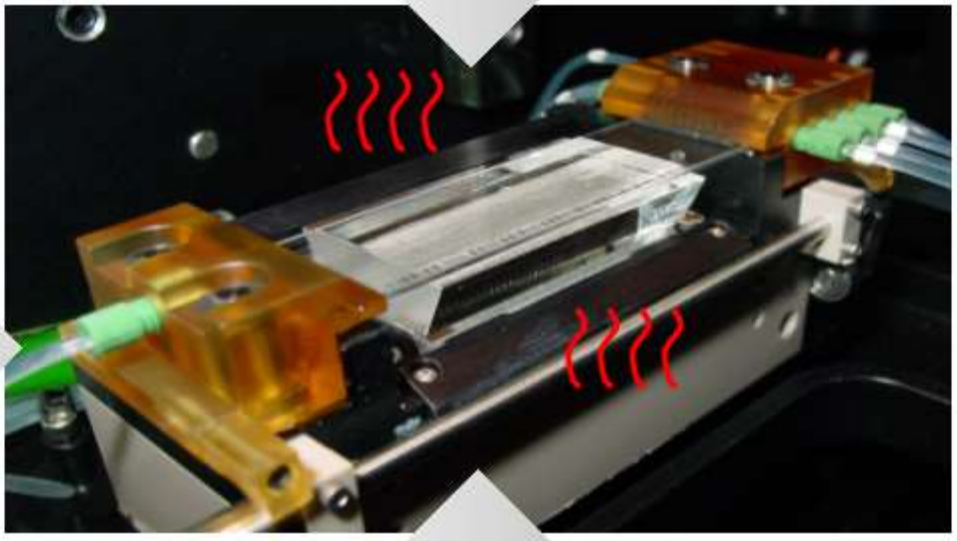
The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

12. ALIGN DATA

The data are aligned and compared to a reference, and sequencing differences are identified.

The flow cell is mounted on the sequencer

CCD camera collects laser-excited fluorescence

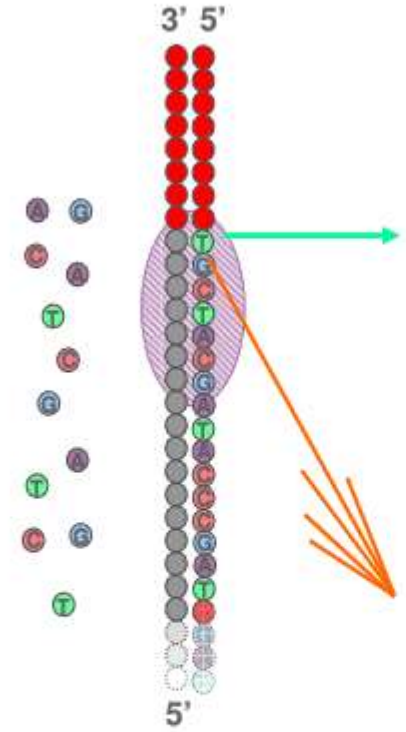


sequencing reagents pass through the 8 lanes inside the flow cell

sequencing reaction is temperature controlled

illumina

A picture is taken every time a new base is added



Sequencing

36bp – 100bp

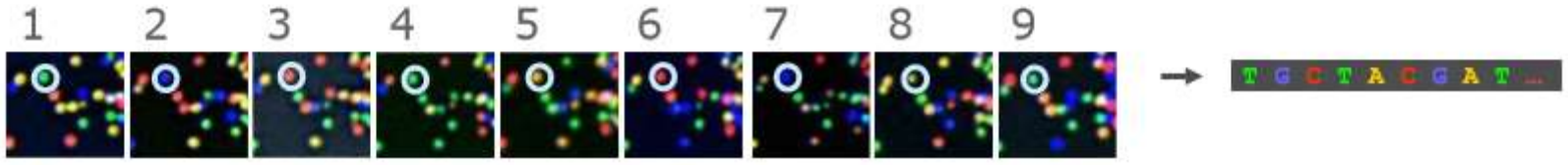


Image acquisition

Base calling

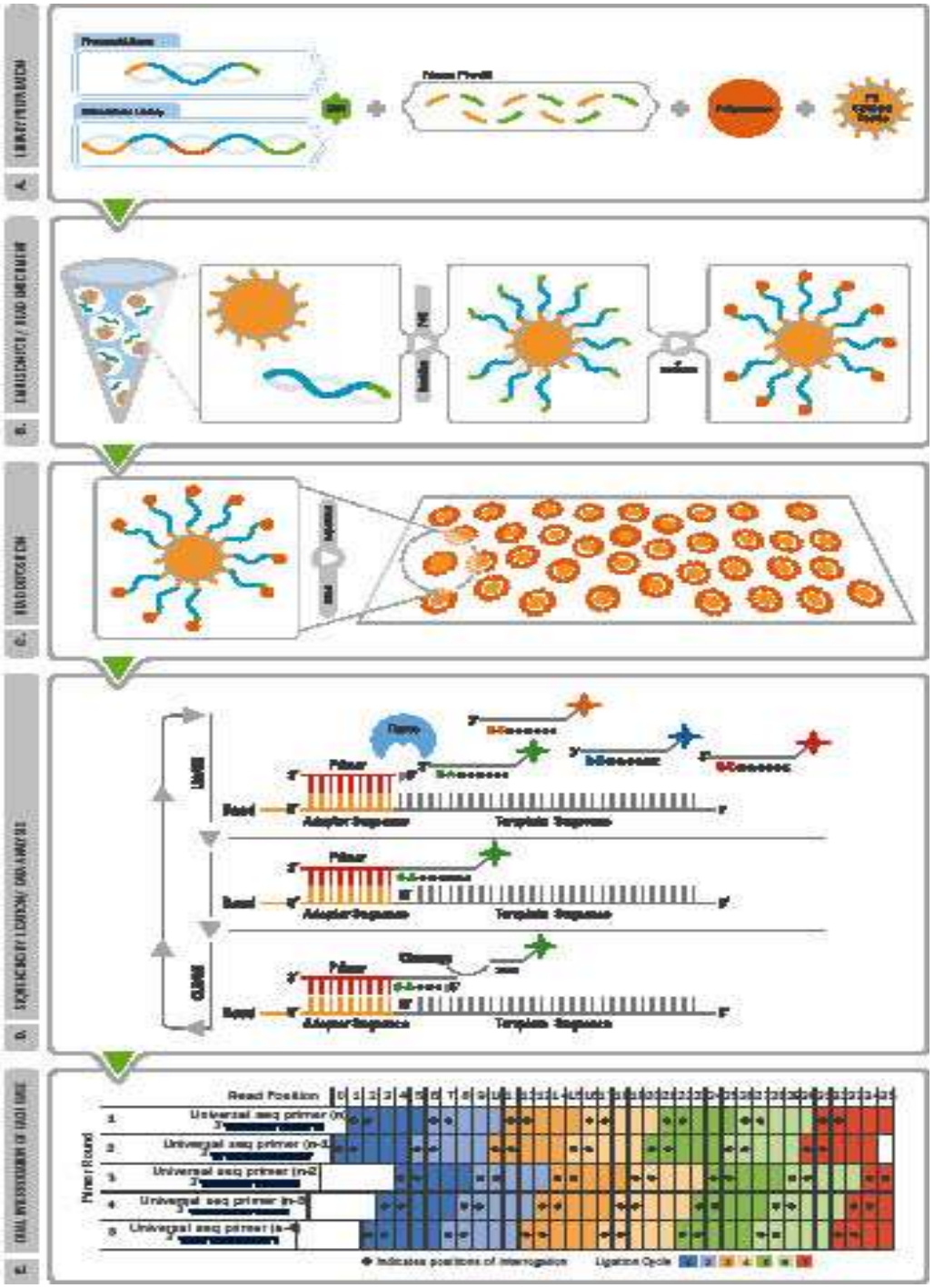
Life Technologies SOLiD

- Sequencing by Oligo Ligation and Detection
- Libraries are bound to beads which are covalently attached to a glass support surface after emulsion PCR
- Uses fluorescently labelled oligomers
- Dibase encoding
- Read lengths up to 2 by 50bp
- Up to 8 samples/slide

Solid sequencing

SOLiD SEQUENCING

- The SOLiD instrument utilizes a series of ligation and detection rounds to sequence millions of fragments simultaneously.
- There are five primer cycles performed on the instrument with each cycle staggered by a single base and including a series of seven or ten ligations for either a 35 or 50 base pair sequencing run.
- Each ligation decodes two bases and is recorded through fluorescent imaging.
- By compiling the fluorescent reads in color space for each fragment, an accurate sequence can be generated.
- Two types of libraries are available for sequencing **Fragment and Mate Pairs.**


A.
B.
C.
D.
E.
Read Position
Primer Round
Universal seq primer (n)
Universal seq primer (n-1)
Universal seq primer (n-2)
Universal seq primer (n-3)
Universal seq primer (n-4)
Indicates positions of interrogation
Ligation Cycle

SOLiD: Advantages

- High throughput –more complete coverage of transcriptome
- Wide dynamic range–simultaneous detection of high and low expressors
- Sensitivity–ability to detect the lowest expressed genes
- Alternative transcriptional start sites(TSS) determination