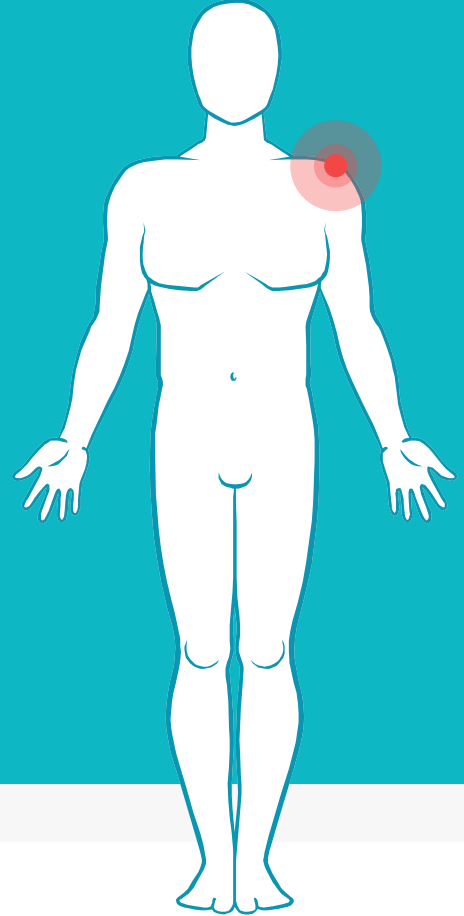


# DNA Sequencing

## Lecture – 3

Nafis Neehal, Lecturer, Department of CSE, DIU



# CONTENTS

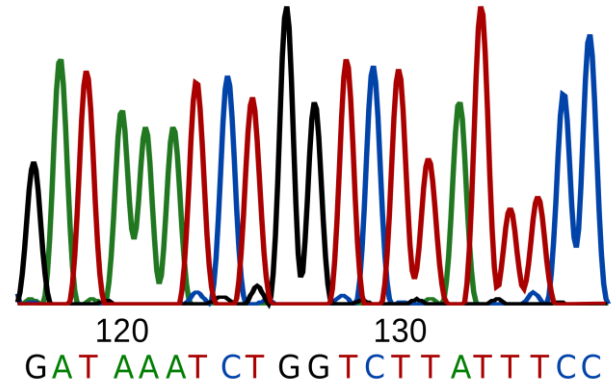
1. DNA Sequencing
2. First Gen Sequencing
  - Sanger Method (1977)
3. Second / Next Gen Sequencing
  - 454/Roche (2005)
  - ABI SOLiD (2006)
  - Illumina/Solexa (2007)
4. Third / Next-Next Gen Sequencing
  - Pacific Biosciences (PacBio)
  - Oxford Nanopore
5. Miscellaneous Terms

# 1. DNA Sequencing

Determining nucleotide sequences

# DNA Sequencing

▷ DNA sequencing is the process of determining the precise order of nucleotides (A, T, G, C) within a DNA molecule.

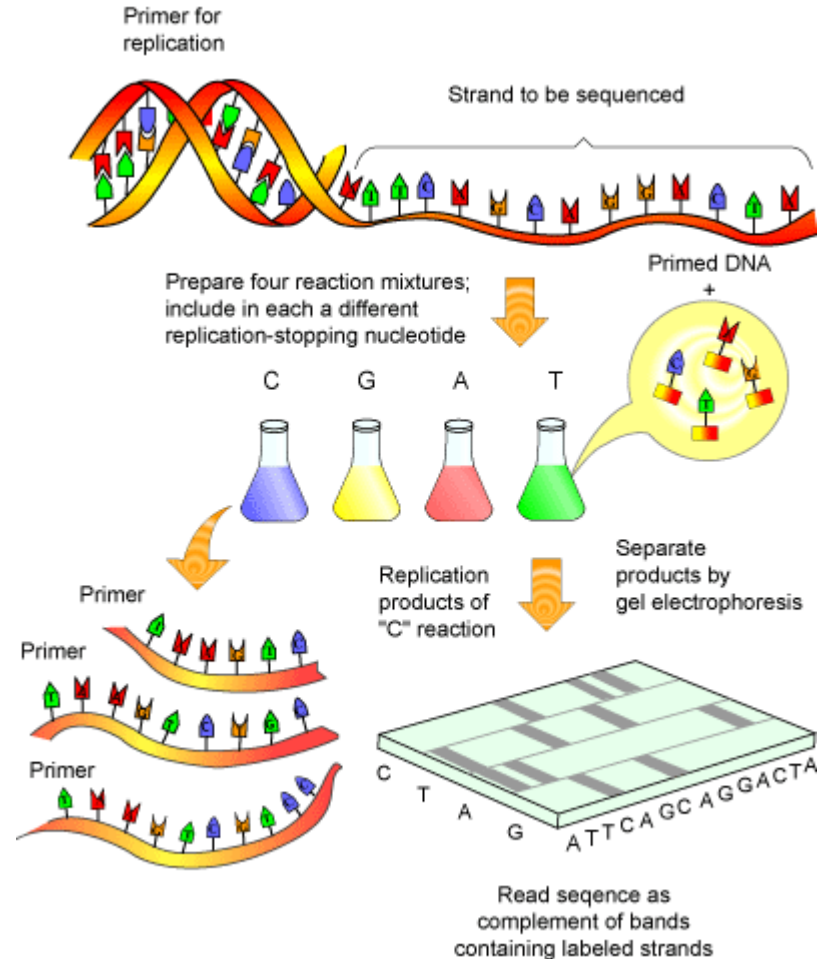


## 2. First Generation Sequencing

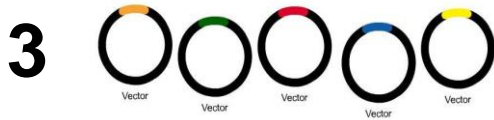
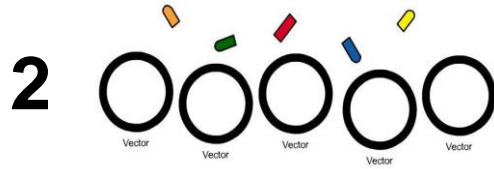
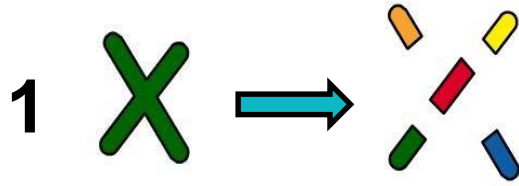
Predominant method for sequencing for decades

# Sanger Method

- ▶ Developed by Frederick Sanger in 1977
- ▶ Most popular and predominant method for DNA Sequencing for decades
- ▶ Can read up to 2000 bps
- ▶ Slow and expensive
- ▶ Labor intensive
- ▶ Human Genome Project was completed using Sanger Sequencing



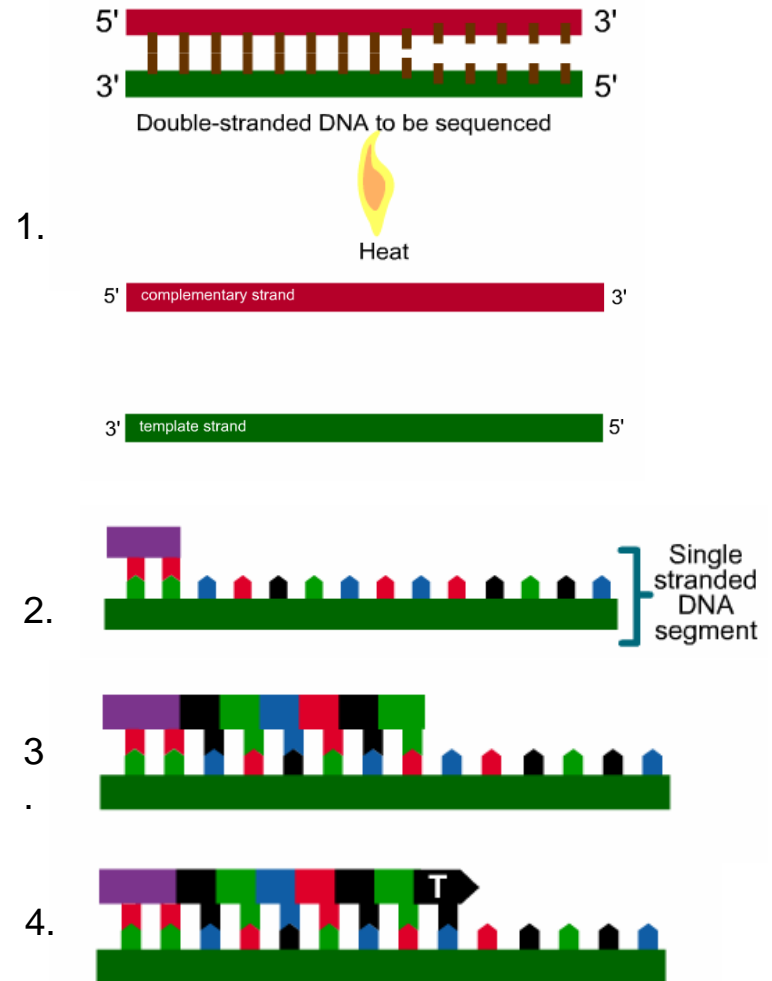
## Step 1 - DNA Preparation



- Cut DNA into a smaller piece for sequencing
- Insert into Plasmid
- Insert Plasmid inside Bacteria Cell and let it multiply
- Extract all the necessary Plasmids and from Plasmid, isolate the DNA for sequencing

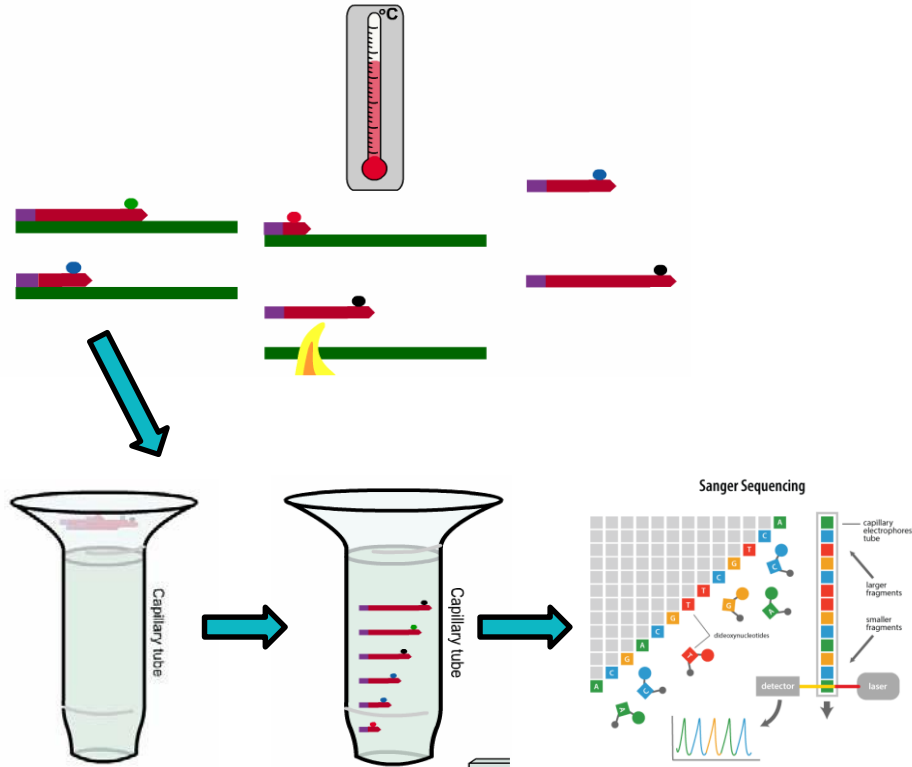
## Step 2 – Sequencing Reaction

- ▶ Strand Separation
  - Heat DNA in 96° C (denaturation)
- ▶ Primer Annealing
  - Lower temperature to 50° C (annealing)
  - Primer binds to DNA
- ▶ Primer Extension
  - Increase temperature to 60° C
  - DNA Polymerase binds to Primer
  - Add complimentary bases (dNTP) after Primer until terminator base is added (ddNTP)
- ▶ Termination
  - Terminate chain after ddNTP is added
  - ddNTP is fluorescently labelled (different colors for A, T, G, C)





## Step 3 – Electrophoresis in Capillary



- ▶ Sort the newly synthesized DNA strands by length
- ▶ Strands are loaded inside a capillary tube
- ▶ An electrical negative charge pulls positively charged DNA strands through the capillary
- ▶ Emerged strands pass through a laser beam that excites the ddNTP fluorescent dye at the end of each strand
- ▶ Beam causes dye to glow in a specific wavelength/color which is captured by photocell and stored in a computer
- ▶ Computer then maps each color to each nucleotide sequentially and generates final sequence output

# 3. Second / Next Gen Sequencing

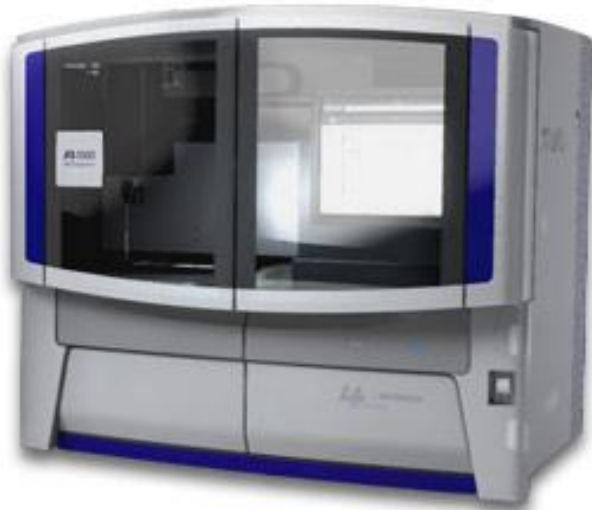
Less Costly methods, mostly Short Read Sequences, High number of reads

## 454/Roche (2005)

- Pyrosequencing technique
- Long Read Sequencing (length up to 700 bps)
- Accuracy 99.9%
- Can sequence up to 1 Million reads/run
- Fast (around 24 hours/run)
- Expensive (costs around \$10 per 1 million base)



## ABI SOLiD (2006)



- SOLiD (Sequence by Ligation)
- Short Read Sequencing (length up to 100 bps)
- Accuracy 99.9%
- Can sequence up to 1.4 Billion reads/run
- Time around 1-2 weeks, Slower than other sequencers
- Cheap (costs around \$0.13 per 1 million base)

# Illumina / Solexa (2007)

- Sequencing by Synthesis
- Short Read Sequencing (length up to 300 bps)
- Accuracy 99.9%
- Can sequence up to 3 Billion reads/run
- Moderately Slow (around 1-11 days/run)
- Expensive Equipment, run cost is low (costs around \$0.05-\$0.15)



# 4. Third / Next-Next Gen Sequencing

Long reads, Higher error rate

# Pacific Biosciences (PacBio)



- Single Molecule Real Time Sequencing
- Long Read Sequencing (length up to 40,000 bps)
- Accuracy 87%
- Can sequence up to 500-1000 Mega reads/run
- Time around 30 minutes – 4 hours, Faster
- Expensive Equipment, run cost is low (costs around \$0.13-\$0.60)

# Oxford Nanopore

- Nanopore sequencing
- Very Long Read Sequencing (length up 500 kb), Portable
- Accuracy 92-97%
- Depends on read length selected by user
- Time around 1 minutes – 48 hours, Faster
- Expensive Equipment, run cost is low (costs around \$500-\$999 per flow cell)



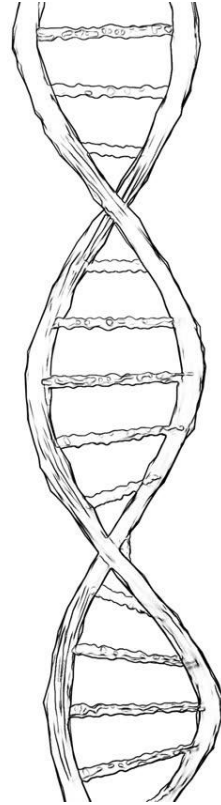


# 5. Miscellaneous Terms

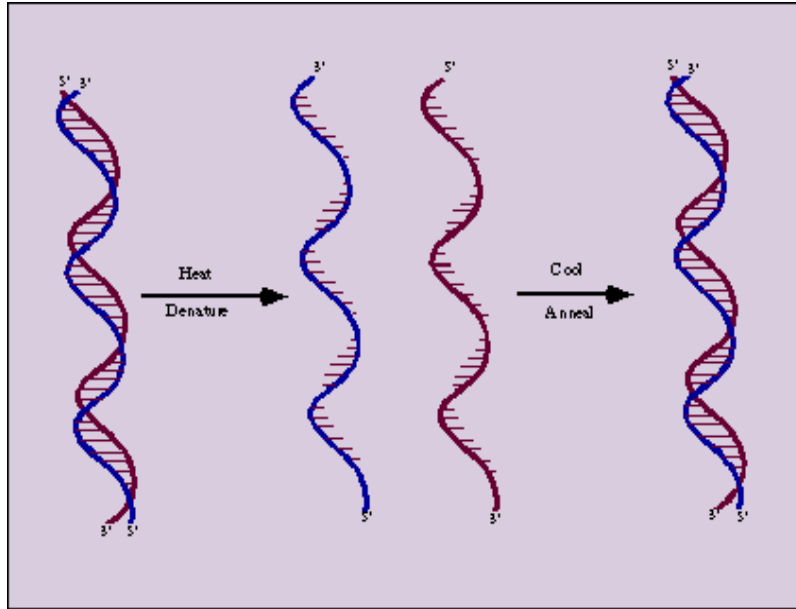
Some comparisons, terms etc.

# Oligonucleotide

- ▷ Short sequences of DNA or RNA
- ▷ Typically less than 20bp
- ▷ Oligonucleotide of 'k' bases length is called k-mer.



# Denaturation and Annealing



## Denaturation

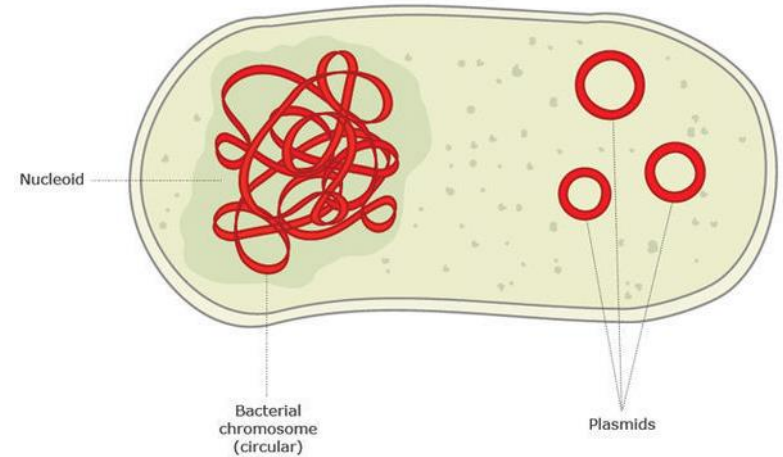
- Energy of heat pull apart two DNA strands
- Happens at a critical temperature denoted  $T_m$

## Annealing

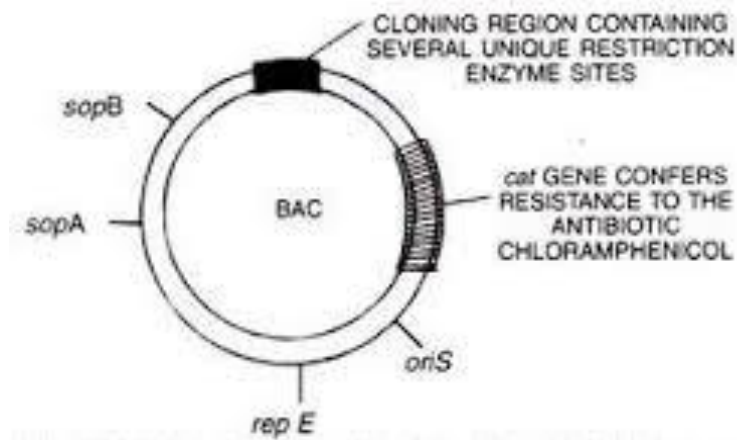
- Decrease temperature, and strands are joined back together
- Only complementary bases will bond

# Plasmid

- ▶ Small, circular piece of DNA often found in bacteria.
- ▶ Sizes of 2.5-20 kb
- ▶ Plasmid using method -
  - \* Isolate them in large quantities
  - \* Cut and splice them, adding whatever DNA needed
  - \* Put them back into bacteria, where they'll replicate along with the bacteria's own DNA
  - \* Isolate them again - getting billions of copies of whatever DNA was inserted into the plasmid



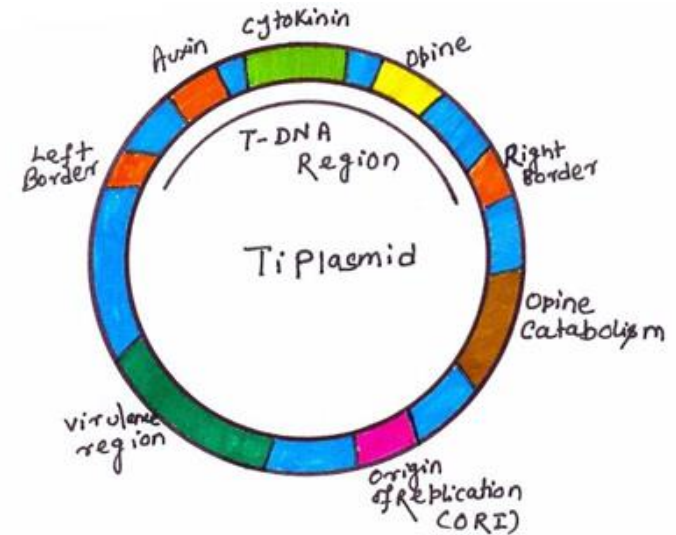
# Bacterial Artificial Chromosome (BAC)



- ▶ Used like a plasmid
- ▶ BACs carry DNA from humans or mice or any other living being, and is inserted into a host bacterium for replication
- ▶ BAC is artificially constructed, unlike Plasmid

# Cloning Vector

- ▶ A cloning vector is a small piece of DNA, taken from a virus, a plasmid, or the cell of a higher organism, that can be stably maintained in an organism, and into which a foreign DNA fragment can be inserted for cloning purposes.



8%

Of Human DNA is made of Ancient Viruses

700 Terabytes

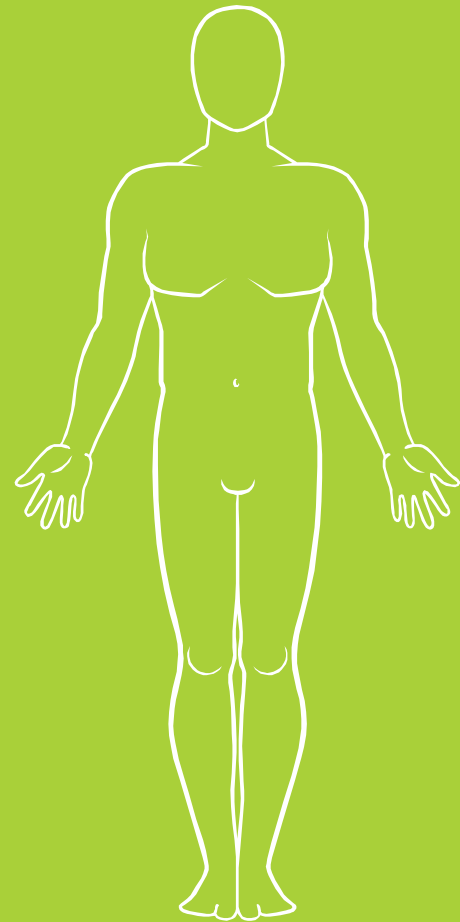
Data can be stored in 1gm DNA

50 Years Time

Type entire human genome at a speed of 60  
wpm

99.9%

Human DNA is identical, 0.01% creates  
human diversity



TO BE CONTINUED

# Impressed?



# Youtube Links

▶ Sanger Sequencing - <https://www.youtube.com/watch?v=ONGdehkB8jU>