L3

▼ RE

- for doing the sequence analysis : we use RE and DNA ligase
- RE cuts DNA, DNA ligase bonds pieces
- Sequencing ka process: identify ROI → isolate → copy kro for further use

▼ How bacterias prevent themselves from bactriophages

- they chop up foreign DNA using RE

▼ Why the RE doesnt attack its host

A bacterium that makes a particular restriction endonuclease,
also synthesizes a companion DNA methyltransferase

methylates the DNA target sequence for that
restriction enzyme, thereby protecting it from cleavage

but attack krne wale ke paas ye methyl grps nhi to wo to gya

▼ DNA fragmentation

Blunt end: simple double stranded straight cut form blunt ends

sticky end: when a cut causes short single stranded overhanging ends [aka cohesive end]

- Length of recognition sequence dictates how frequently the enzyme will cut a DNA sequence
- REs with same restriction sites == isoschizomers
- most recognition sequences are palindromes (kyuki both strands chahiye maybe)

▼ Cloning

process of production of multiple copies of DNA fragments inside the organism

Cloning vector : jo krta hai ye sab by carrying a foreign DNA and mulitplying that inside the host cell

▼ Types of Vectors

▼ DNA sequencing

- determine the precise sequence of nucleotides in a sample of DNA – the order of A, T, G, C
- sequencing ke liye u need enuf starting template jo ki Polymerised Chain Reaction se milta hai
- Primer : short single stranded nucleotides which serve as starting point for DNA synthesis
- Cycling reactions:
  - Denaturation
  - Annealing
  - Extension
- then after formation of product we verify if it is right or not

Steps in PCR Sequencing

I The sequencing reaction

- **Denaturation at 94°C**
- **Annealing at 50°C**
- **Extension at 60°C** ← **instead of 72°C**

II Separation of the fragments

III Detection on an automated sequencer

IV Assembling the sequenced parts

aka Sanger Method aka dideoxy method aka chain termination method

▼ Genome Sequences

Whole Genome Shotgun Method

Adding to the challenge is the sheer computational complexity of the task.

Size of H. genome = 3 x 10^9 bp. Given length of read ~500 bps, for desired coverage of 10x, No. of reads required is:

$$\text{RequiredReads} = \text{GenomeLength} * \text{DesiredCoverage} / \text{ReadLength}$$
$$= 6 * 10^7$$

With 60M reads to assemble, we need algorithms that run in near linear time ($O(nlogn)$)