DNA sequencing in Microbiology

Dr. Nick Coleman

DNA sequencing in microbiology

Learning outcomes. Be able to…

- **Describe in detail how DNA sequencing methods work**
  - **Sanger method vs. 454 pyrosequencing**
- **Discuss advantages / limitations of different seq. methods**
- **Describe and compare strategies for sequencing**
  - **Primer walking / Shotgun / Direct**
- **Explain the significance of sequencing technologies for microbiology**

DNA sequencing technologies

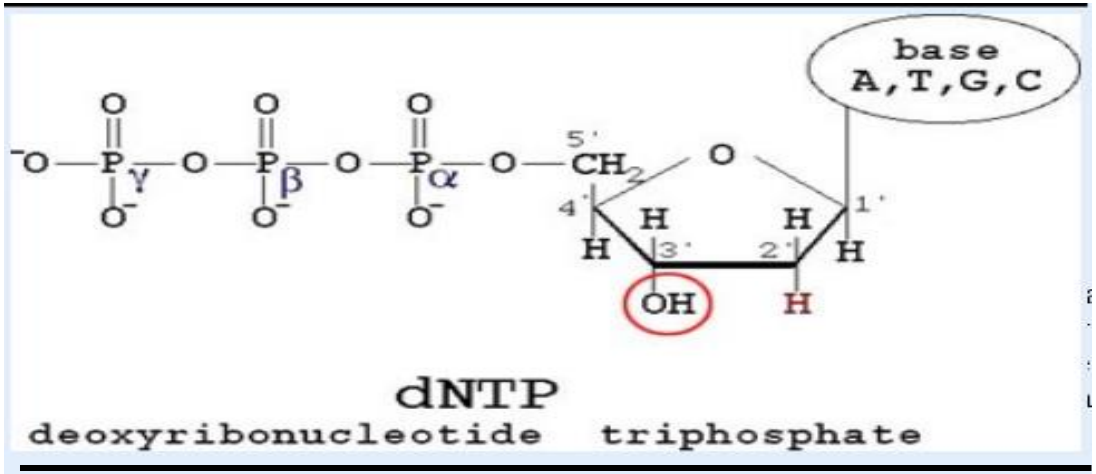
- Traditional
  - Maxam Gilbert sequencing ('chemical method')
  - Sanger sequencing (dideoxy terminators)☆
- Next generation
  - 454 pyrosequencing ☆
  - Solexa sequencing "Illumina"

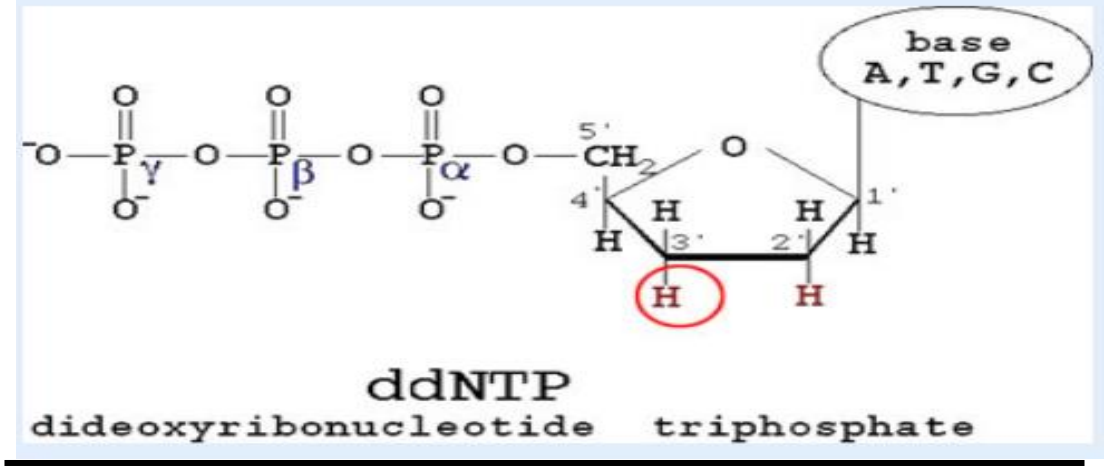
Sequencing strategies

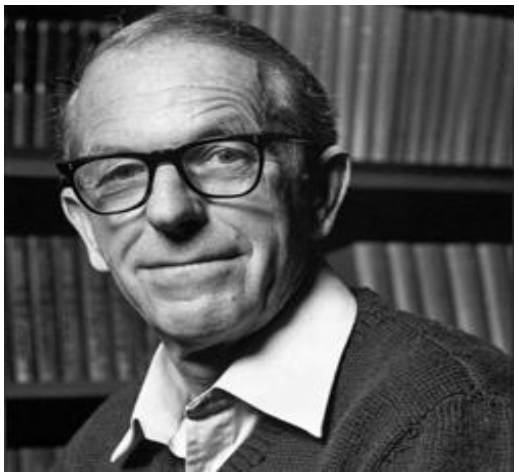
- Large insert clones : Primer walking
- Small insert clones : Shotgun sequencing
- No cloning: direct sequencing from adapters

Sanger sequencing (dideoxy method)

- **dNTPs vs. ddNTPs**
- **The key to the Sanger sequencing method is the use of dideoxynucleotides (ddNTPs) → act as chain-terminators**







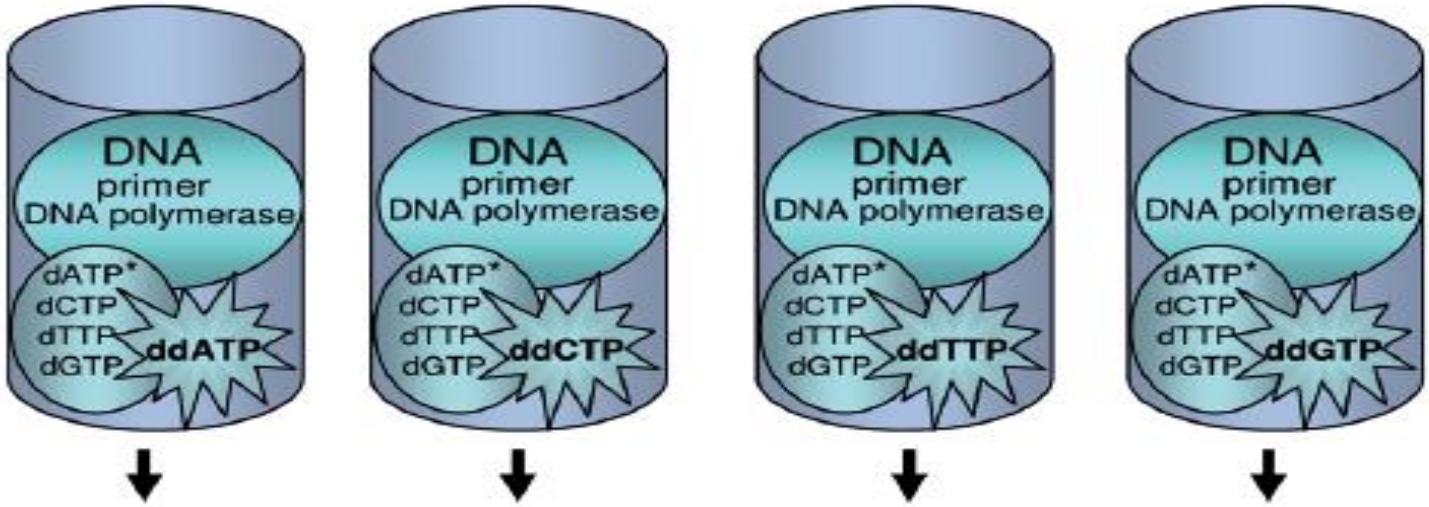
Frederick Sanger

Protein sequence of insulin
- Nobel prize 1958

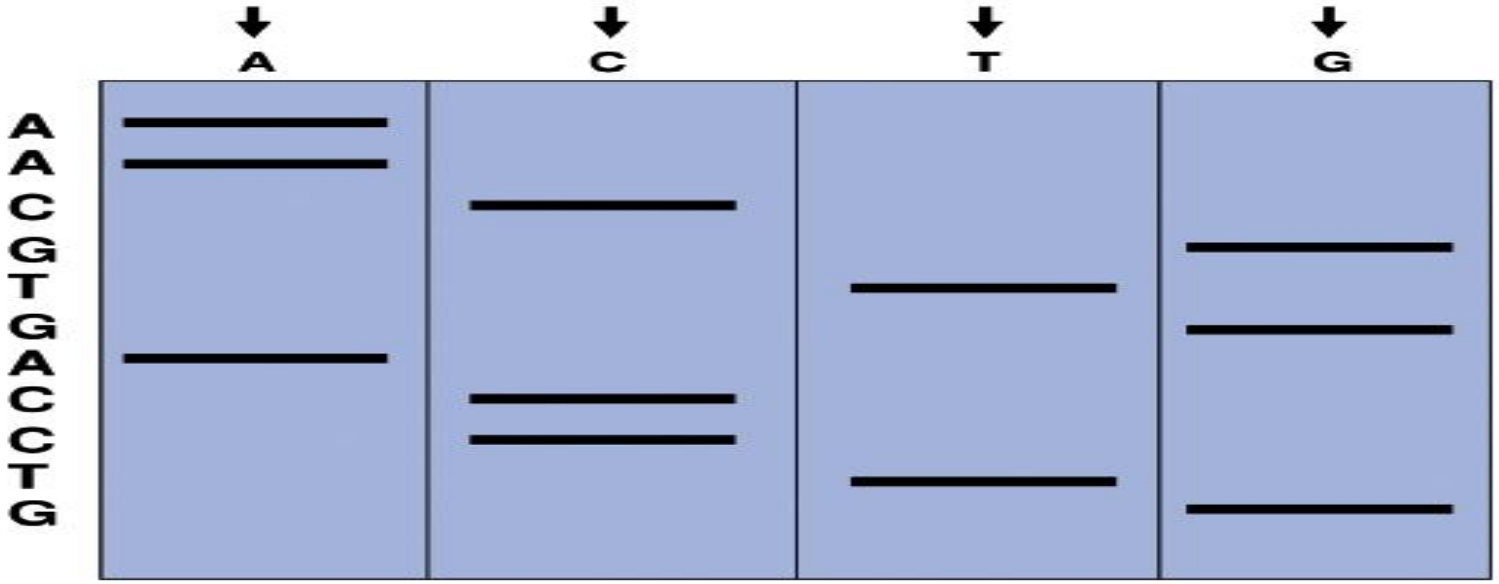
DNA sequencing method
- Nobel prize 1980

Sanger sequencing

- **A polymerase-mediated linear amplification of DNA (similarities / differences to PCR?)**
- **Sequencing mixture: template DNA, primer, DNA pol. I (Klenow), dNTPs (high conc.), *one* ddNTP (low conc.)**
- **Four parallel sequencing reactions are done, each has a different dideoxy nucleotide (ddNTP) added.**



Sanger sequencing



- <u>ddNTPs lack the 3' OH group</u> that is needed to make the phosphodiester bond of DNA backbone.

  → chain elongation terminates at different positions

- By adjusting ratio of normal : dideoxy nucleotides, can get a spread of termination fragments of different sizes

Sanger sequencing – gels.

- **Very high resolution gels (single nucleotide level) are required to separate the products of dideoxy sequencing**

  → Denaturing polyacrylamide gels

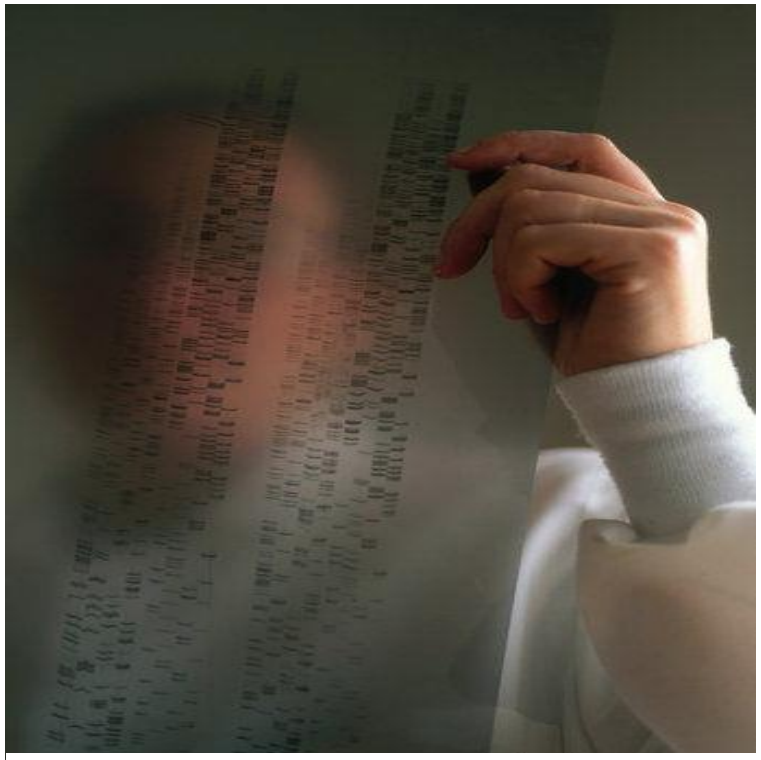
- **Thin vertical gels (radioactive method, manual)**
  or Capillary gels (fluorescent method, automated)

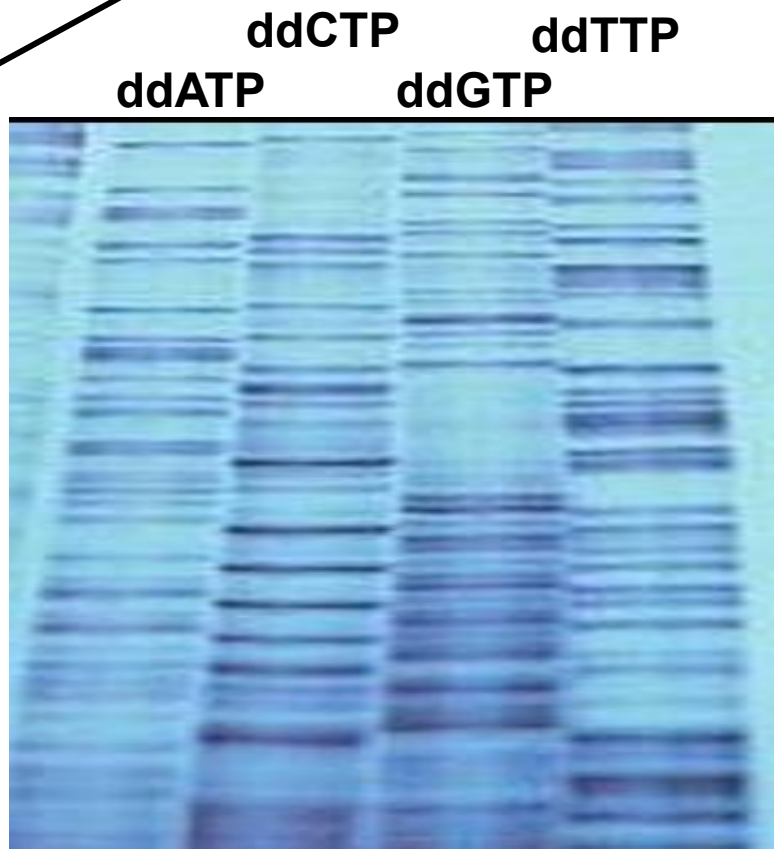
- **Acrylamide gel good for separating small nucleotides (<500 bp); agarose better for larger fragments.**

- **Urea added as denaturant → DNA runs single-stranded**

Sanger sequencing – using radiolabel

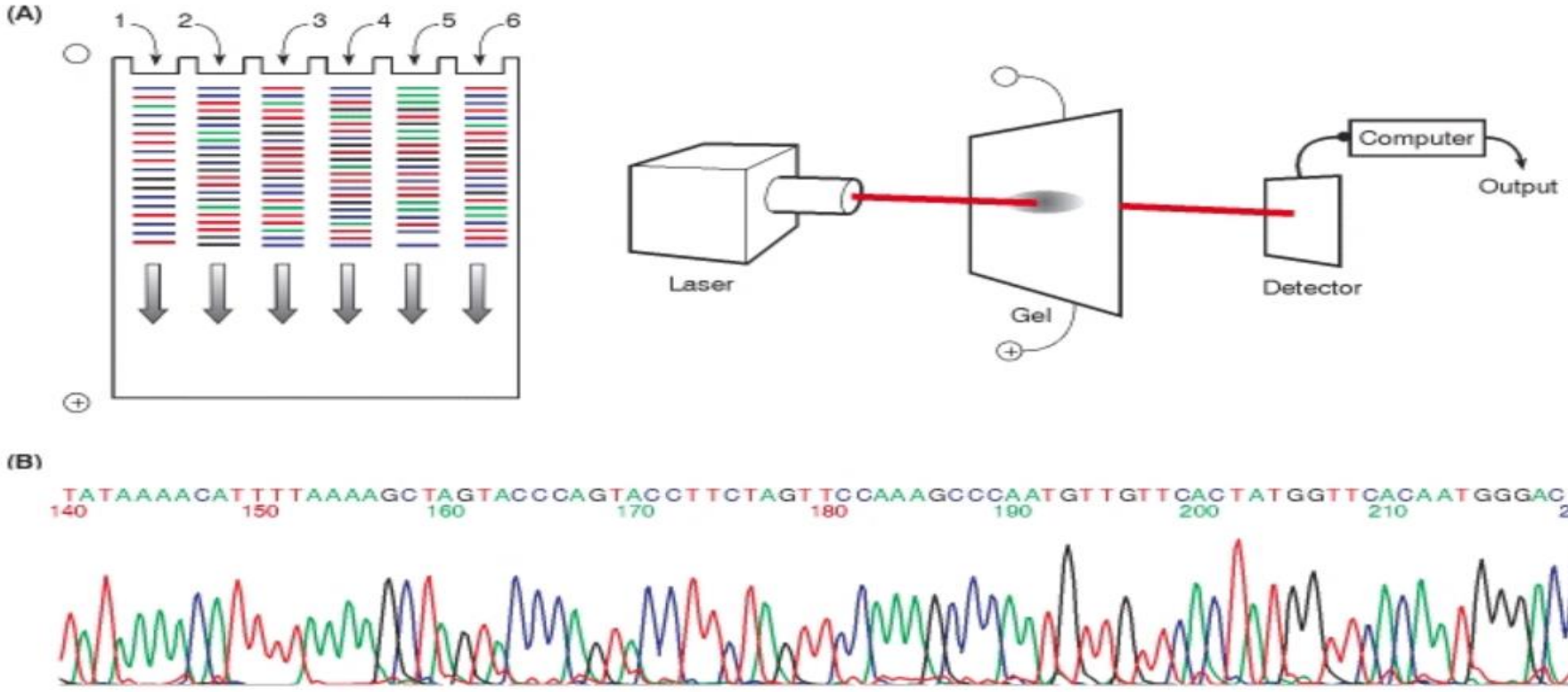
- Early days: Radiolabelled ddNTPs
- Uses 35-S or 33-P isotopes
- Each reaction then run in separate lane on gel
- Thin vertical PAGE gel, run and interpret manually



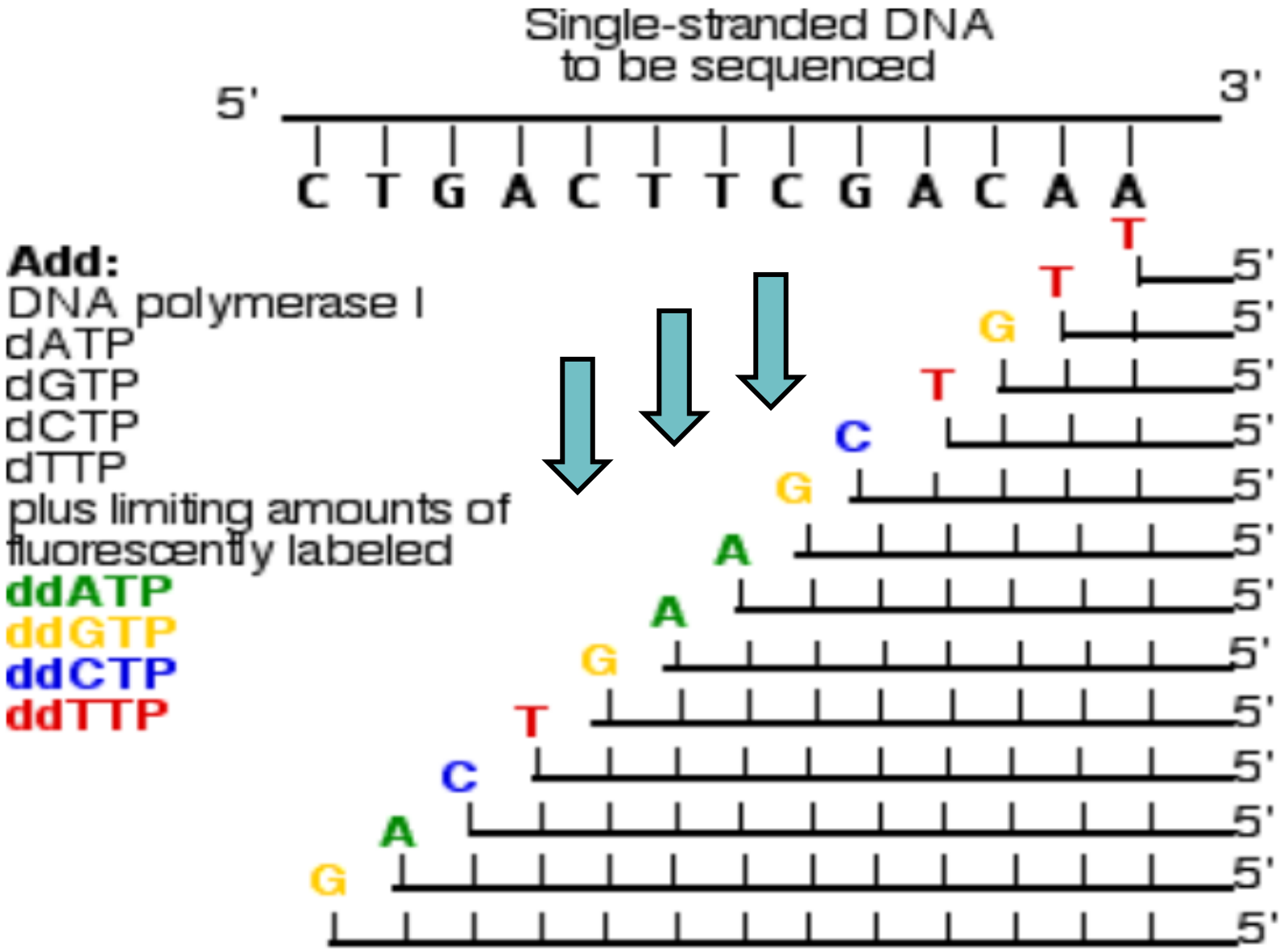


Sanger sequencing – using fluorescence

- Modern sequencing : fluorescent ddNTPs
- Each ddNTP is labelled with a different fluorophore.
- Can mix all 4 sequencing reactions in one lane of gel.
- Could be vertical gel or capillary gel (thin tube)
- Detection and analysis is automated



Sanger sequencing - using fluorescence



Sanger sequencing - overview

- **Advantages:**
  - **Long reads (800-900 nuc), fairly accurate (quoted error rates range from 0.001% to 1%)**
  - **Relatively simple, widely available**
- **Limitations:**
  - **Sequence read limit (800-900 nuc) ! Still a limit.**
  - **Automated sequencers: 384 samples/run, up to 24 runs a day → 7.4 million bases (7.4 Mb) – pretty good.**
  - **But… not "high throughput" enough for genomic applications – ultimately limited by the technology, each sample needs separate gel lane or capillary gel.**

454 pyrosequencing

- "Sequencing-by-synthesis"
- Detects incorporation of nucleotides via enzymatic conversion of released pyrophosphate into light
- One DNA Fragment → One Bead → One Read

http://454.com

454 is an automated, high throughput method

- approx. 500 base read length
- 1,000,000 reads per run

⇩

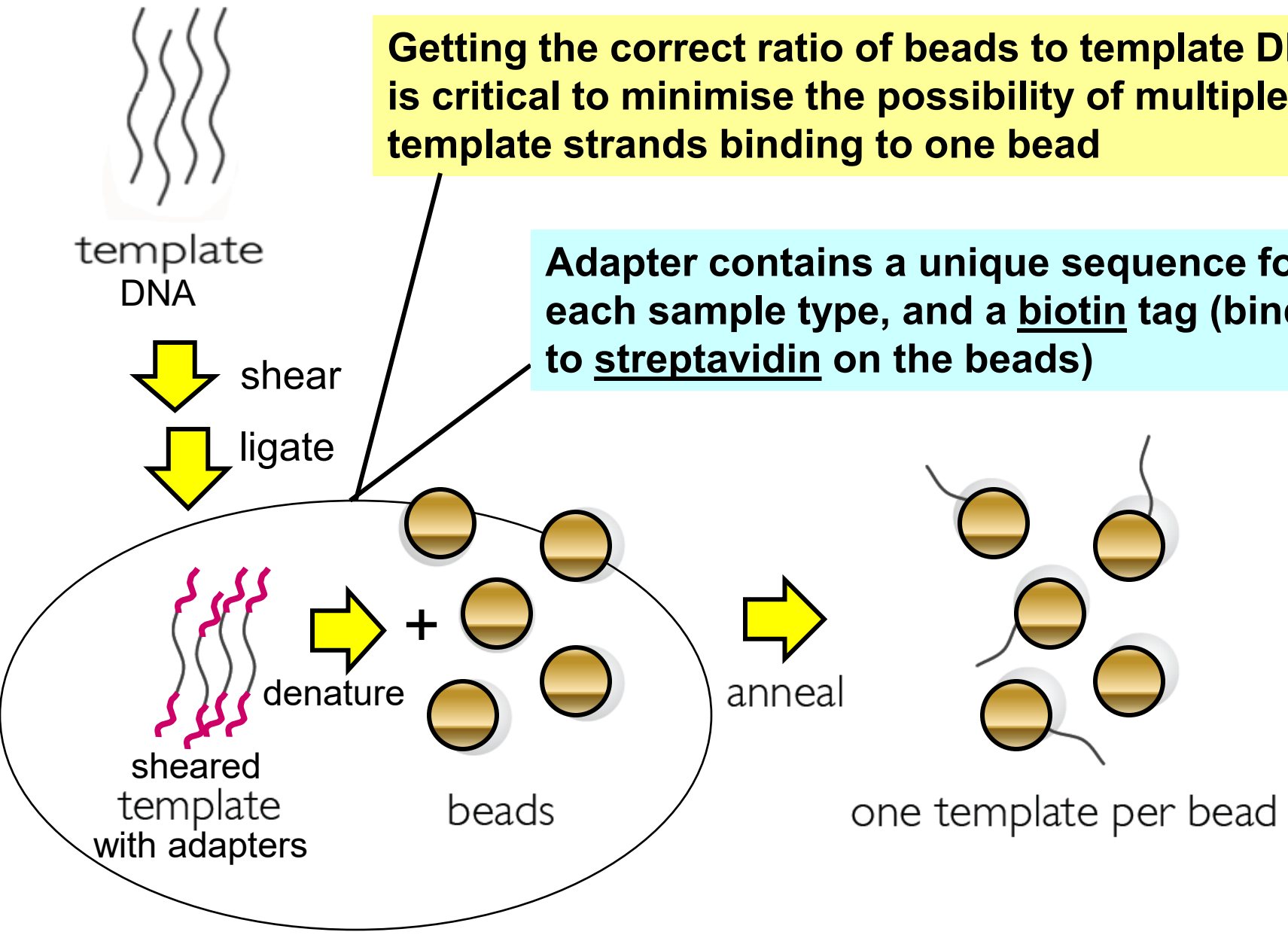
- 500,000,000 bases per run
- = 500 Mb of DNA

454 pyrosequencing

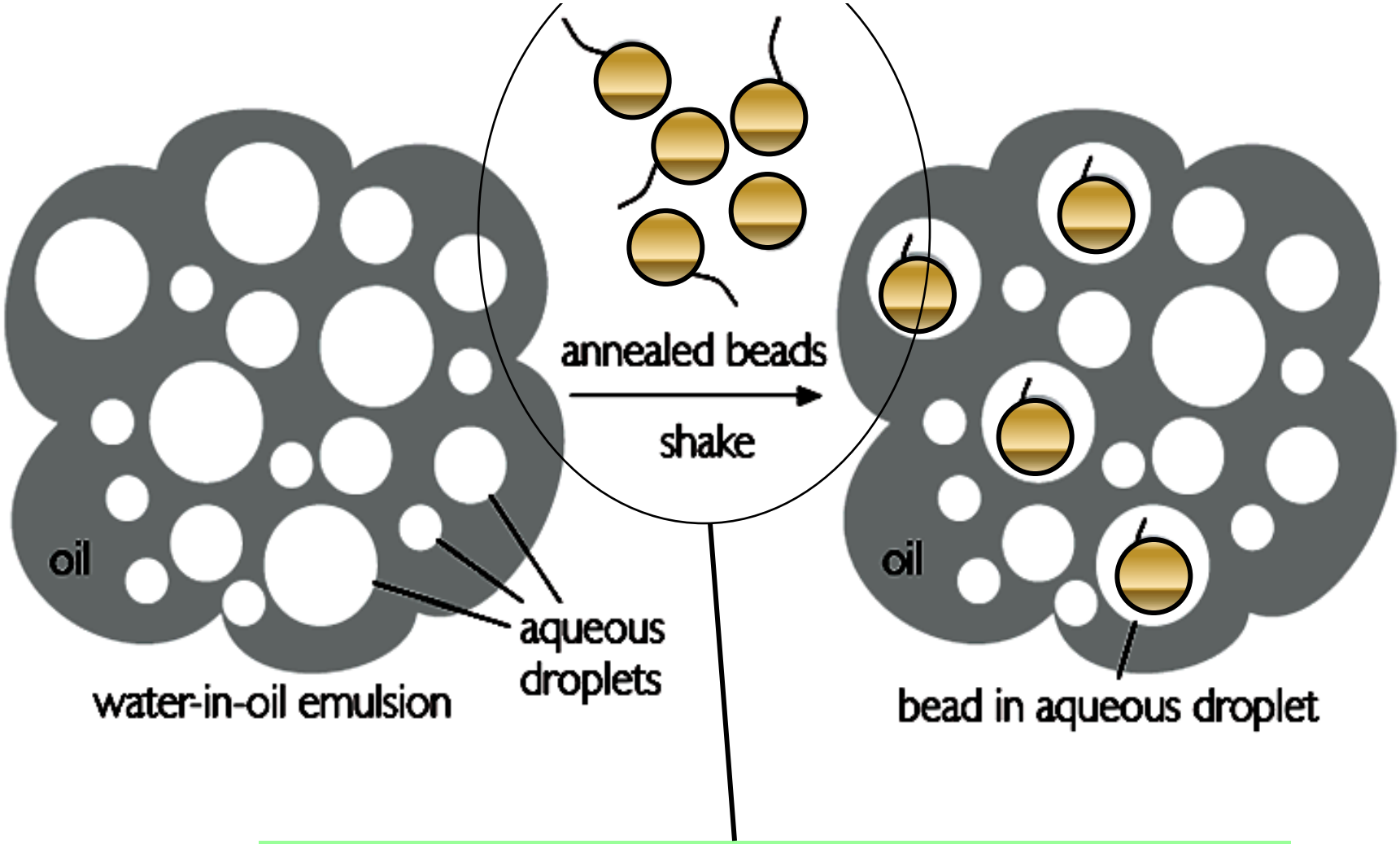
- **Steps:**

1. **Fragment template DNA (→500-800 bp chunks)**
2. **Ligate template to adapters.**
3. **Denature template → ss DNA**
4. **Ligate template to beads**
5. **Add beads to oil/water emulsion**
6. **Amplify by emulsion PCR**
7. **Load beads on picotitre plate**
8. **Read sequence by pyrosequencing**

454 pyrosequencing – ligating DNA to beads

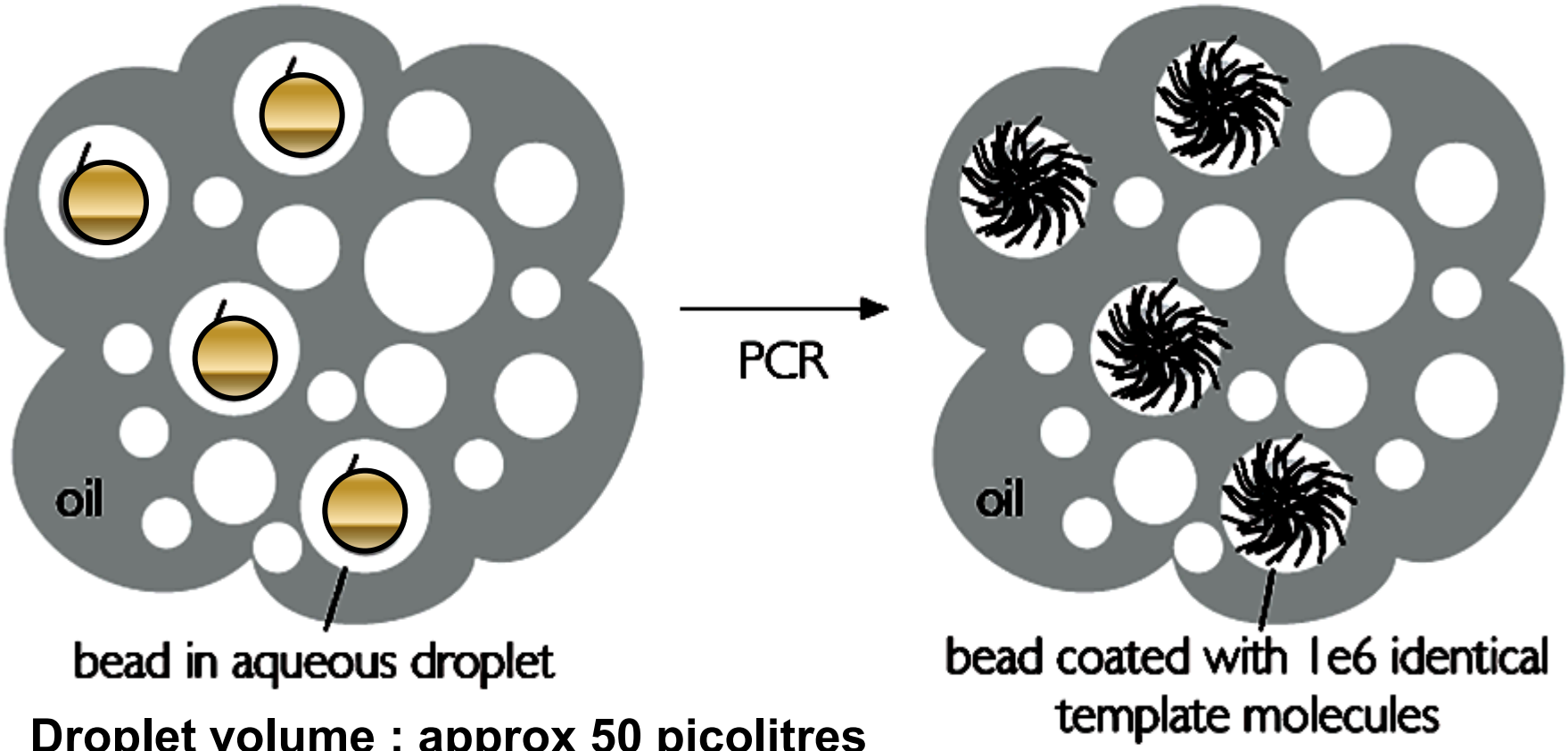


454 pyrosequencing – adding beads to emulsion



Getting the correct ratio of beads to aqueous droplets is critical to minimise the possibility of multiple beads amplifying in the same droplet

454 pyrosequencing – emulsion PCR

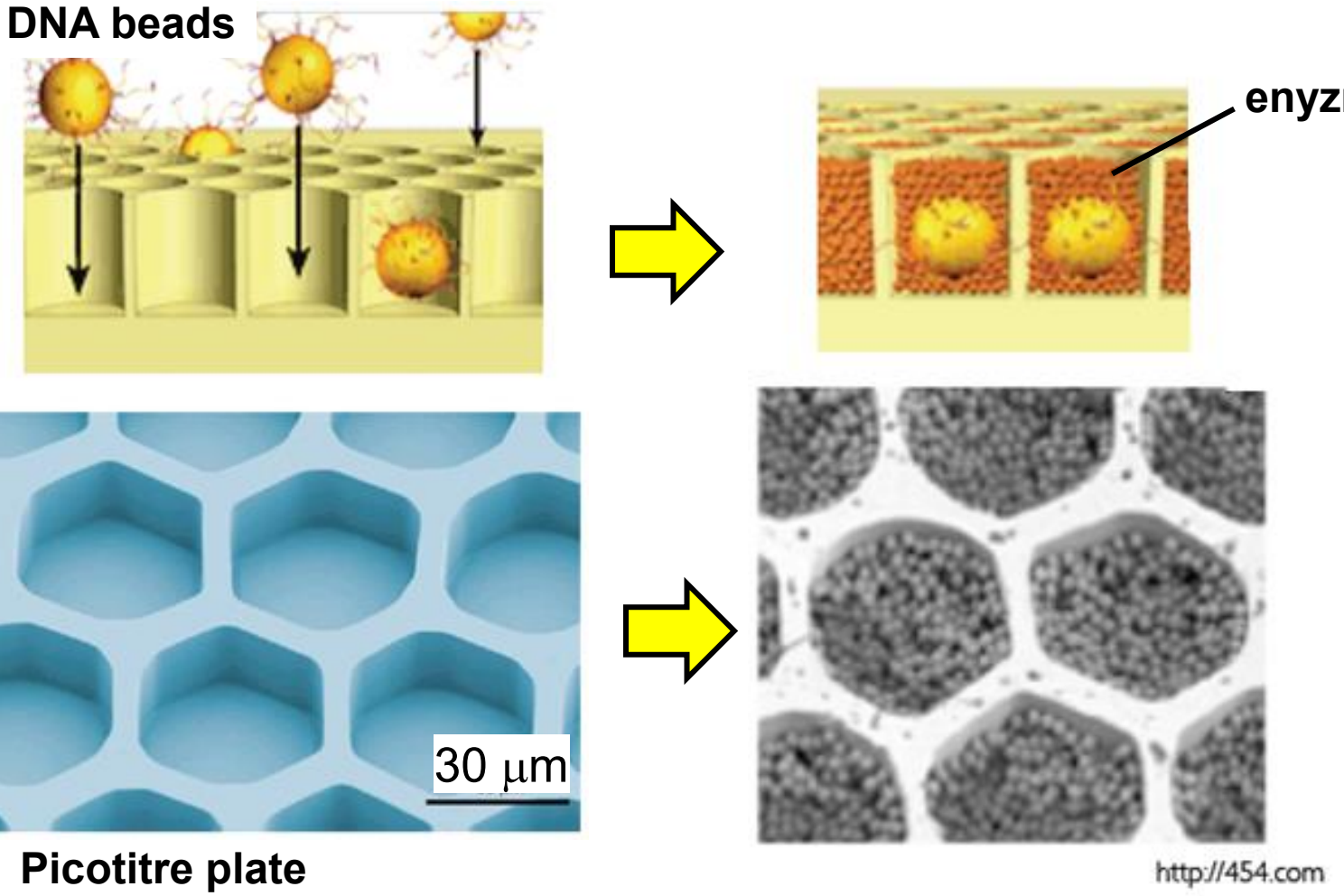


Droplet volume : approx 50 picolitres

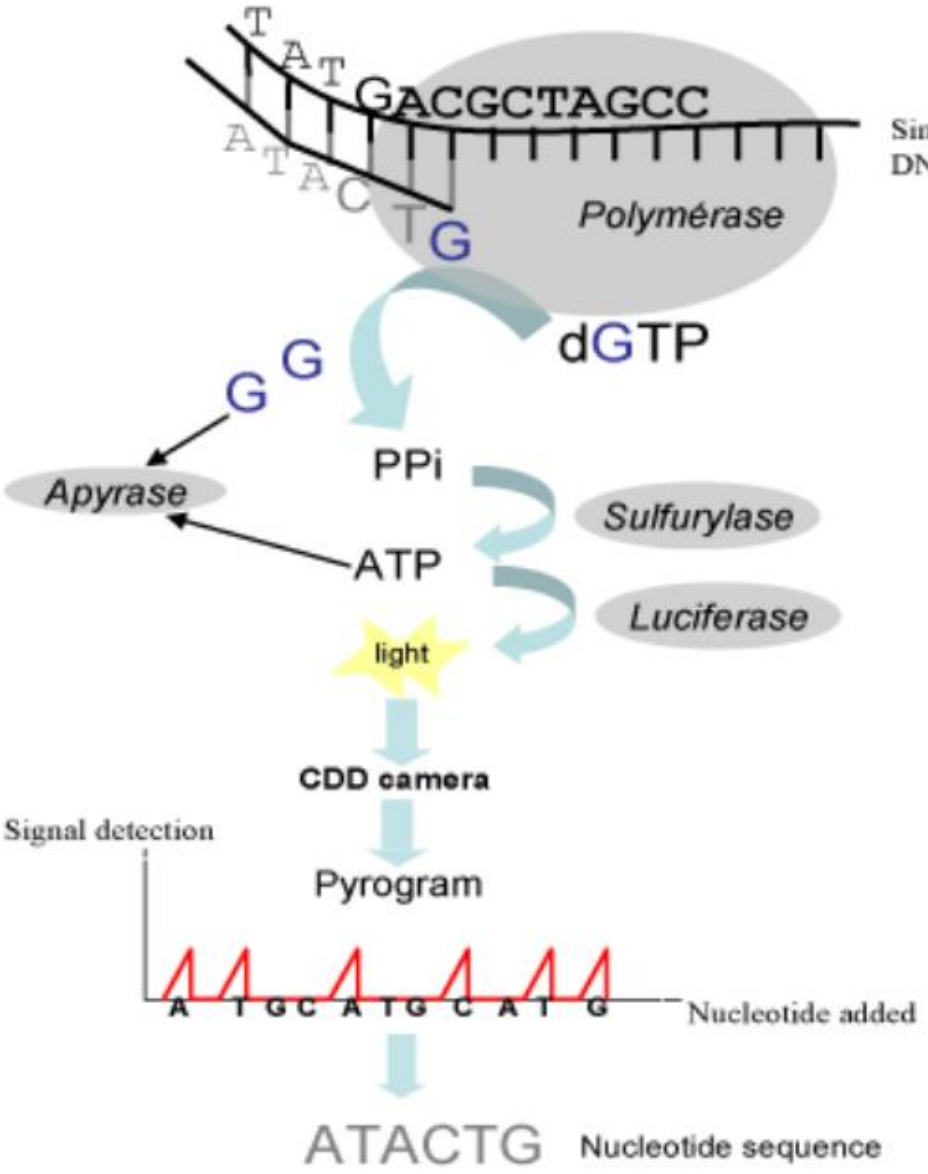
Each different PCR product is contained within its own droplet.
PCR products bind to the beads due to the biotin-labelled primers

454 pyrosequencing – picotitre plate

- DNA-coated beads added to picotitre plate. One bead per well. Pyroseq. reagents (polymerase, sulfurylase, luciferase) added on "enzyme beads"



454 pyrosequencing – the sequencing chemistry



- One nucleotide at a time is supplied to the picotitre plate
- If polymerase incorporates this nucl., pyrophosphate is made (PPi)
- Sulfurylase: PPi → ATP
- Luciferase: ATP→ light
- Light from each well captured by camera
- Then add next nucleotide…

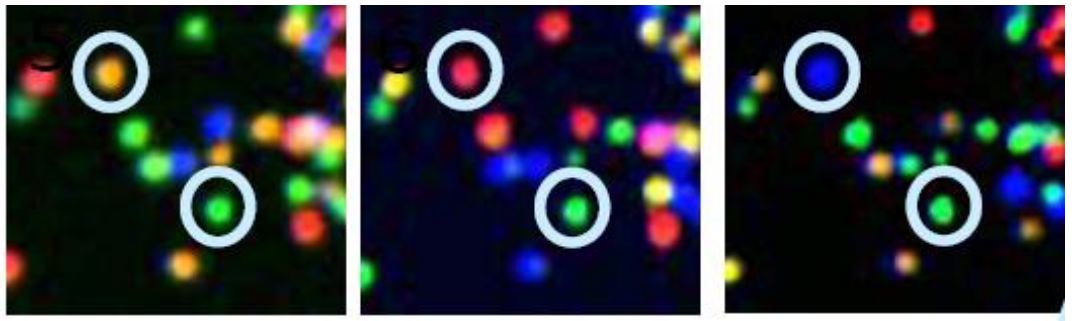
454 pyrosequencing

- **Advantages:**
  - **High throughput (500 Mb per run) compared to traditional Sanger method (7.4 Mb per run)**
- **Limitations:**
  - **Sequence read limit (500 nuc) lower than Sanger**
  - **Error rate is fairly high (1% = 1 mistake / 100 bp)**
  - **Homopolymers cause problems (>5 of the same nucleotides in a row are difficult to count)**
  - **Machine more expensive than Sanger sequencer, technology / expertise not as widely available**

Newer sequencing methods

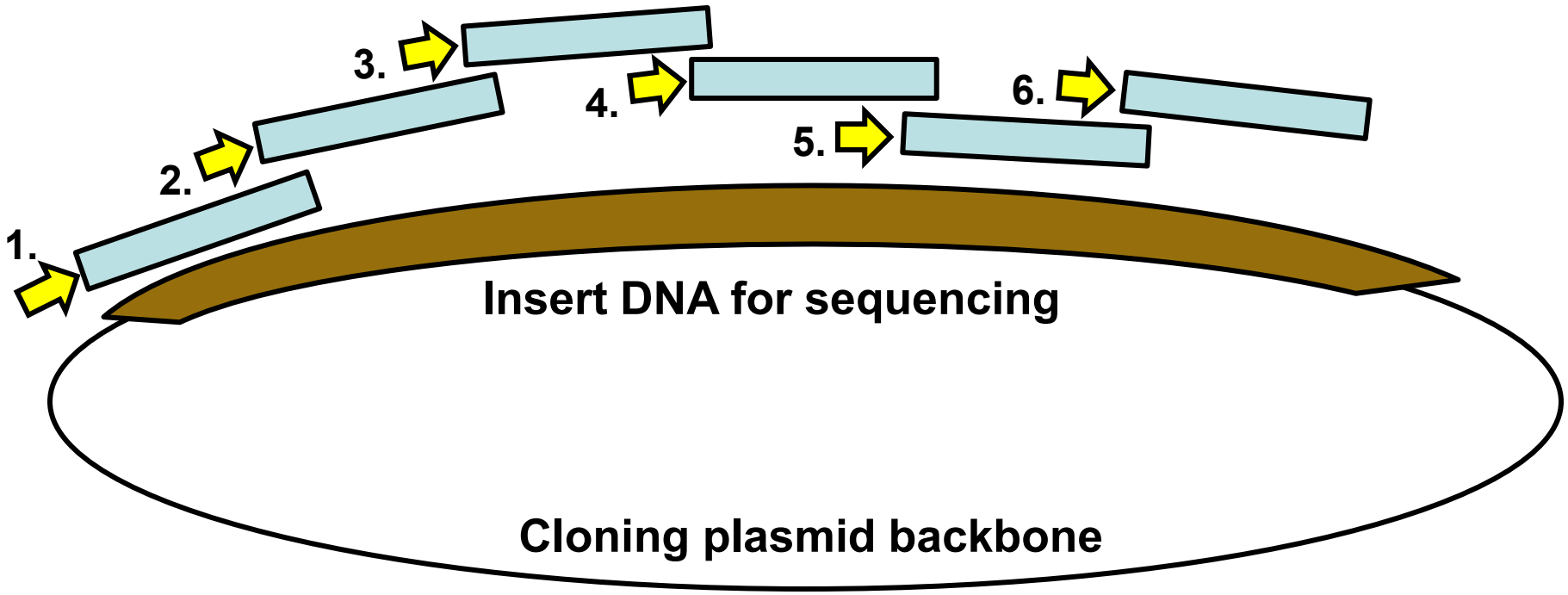
- **Shorter read lengths, but MUCH higher throughput**

- **Illumina method**
  - **'bridge-amplification' in microscopic flow cell**
  - **reversible fluorescent ddNTP terminator**
  - **approx. 30 bp read length**
  - **8 Gb per run**



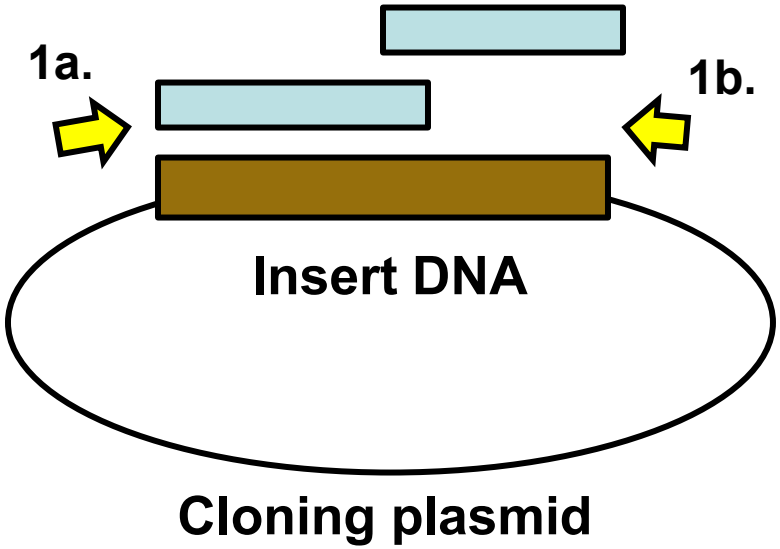
Sequencing strategies – primer walking

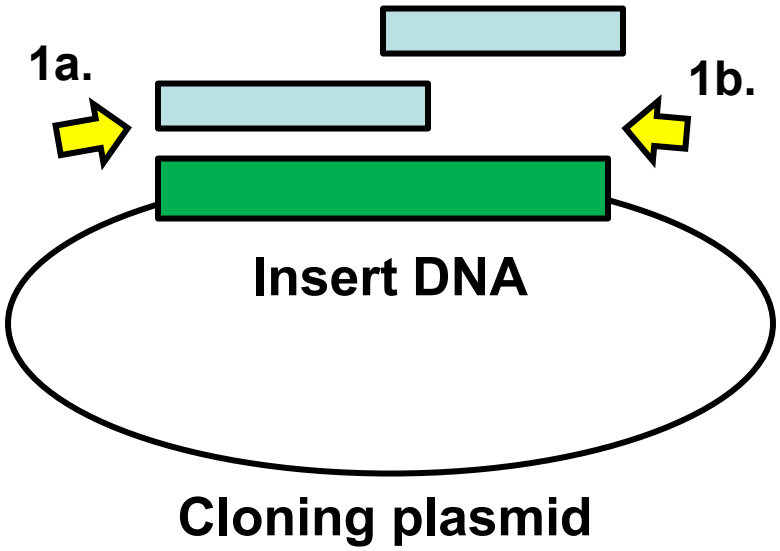
- Appropriate for large pieces of DNA cloned in plasmids
- Done via Sanger sequencing method
- Assembly is easy (new seq. starts within old sequence)
- Method is slow. Not parallel.
- Expensive in primers – need new primer for every read



Sequencing strategies – shotgun sequencing

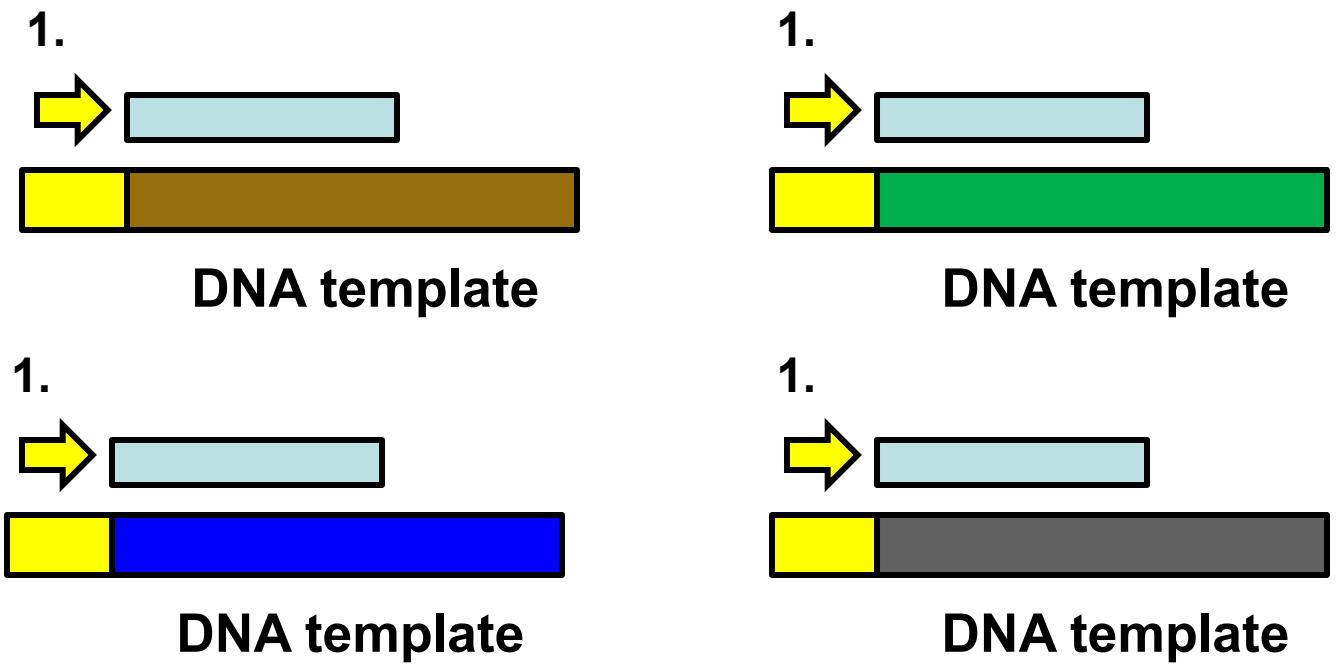
- Appropriate for small pieces of DNA cloned in plasmids
- Done via Sanger or other sequencing methods
- Assembly more difficult (each seq. independent)
- More rapid than primer walking (parallel approach)
- Cheap in primers – same primers for all reads (these bind to the vector, next to the cloning site)





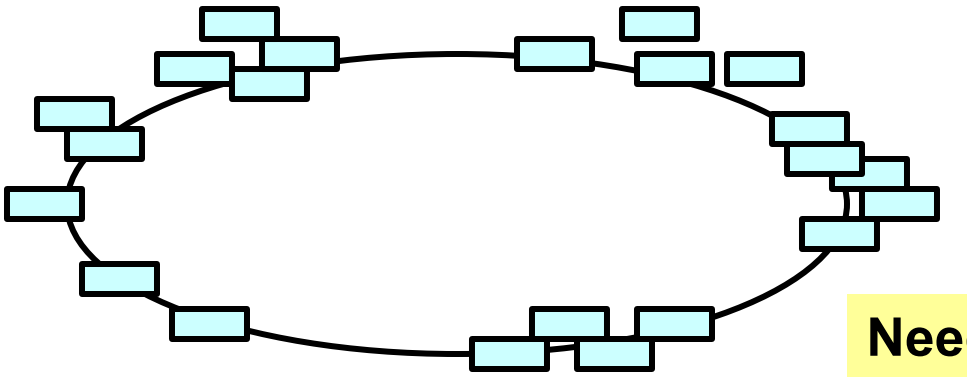
Sequencing strategies – direct

- Use with small pieces of DNA not cloned in plasmids
- DNA must be ligated to adapter seqs first
- Applicable with many sequencing methods
- Assembly difficult (each seq. independent)
- More rapid than primer walking (parallel approach)
- Cheap in primers – same primers for all reads



Sequence assembly

- For shotgun and direct sequencing methods, assembly of the original seq. from fragments not straightforward.
- Need multiple-fold coverage of the original sequence to minimise the number of gaps (eg. 20 x coverage)
- Need for multiple coverage reflects the random nature of the cloning / sequencing process.
- eg. for a 5 Mb bacterial genome, would actually need approx 100 Mb of raw sequence to ensure assembly



Sequence reads will not be uniformly distributed around a target genome sequence
→ WHY?

Need to CLOSE THE GAPS ! eg. PCR

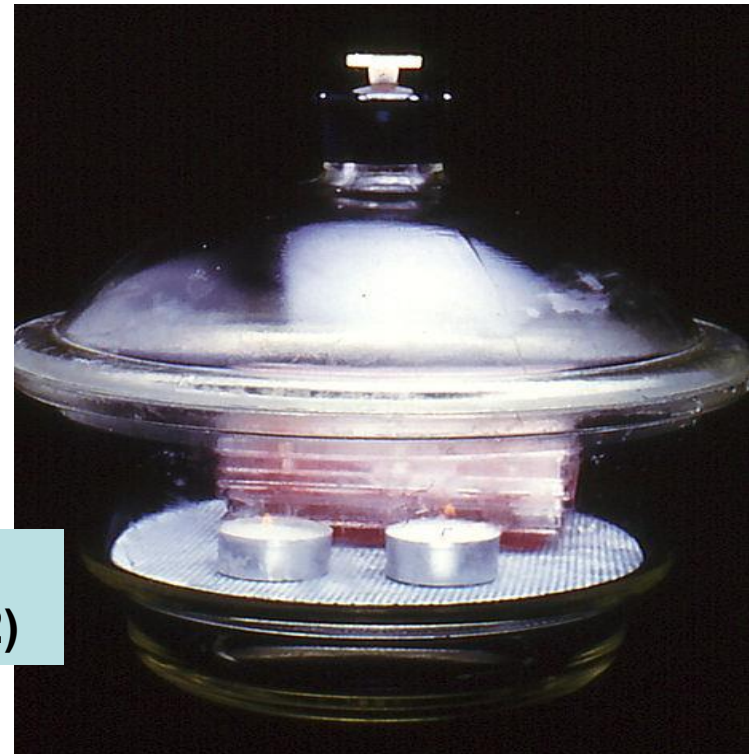
Significance of sequencing for microbiology

- **Identification of unknown microorganisms**

Modern microbiology depends on PCR for identifying unknowns, typing strains, diagnosis of disease

Sequencing of PCR products eg. 16S rDNA is a rapid and reliable way of identifying unknowns.

PCR/sequencing allow detection and identification of unculturable or difficult-to-grow microbes eg. anaerobes, viruses, protists



Plasmodium **(malaria) cultures growing in blood medium in a candle jar (elevated CO2)**

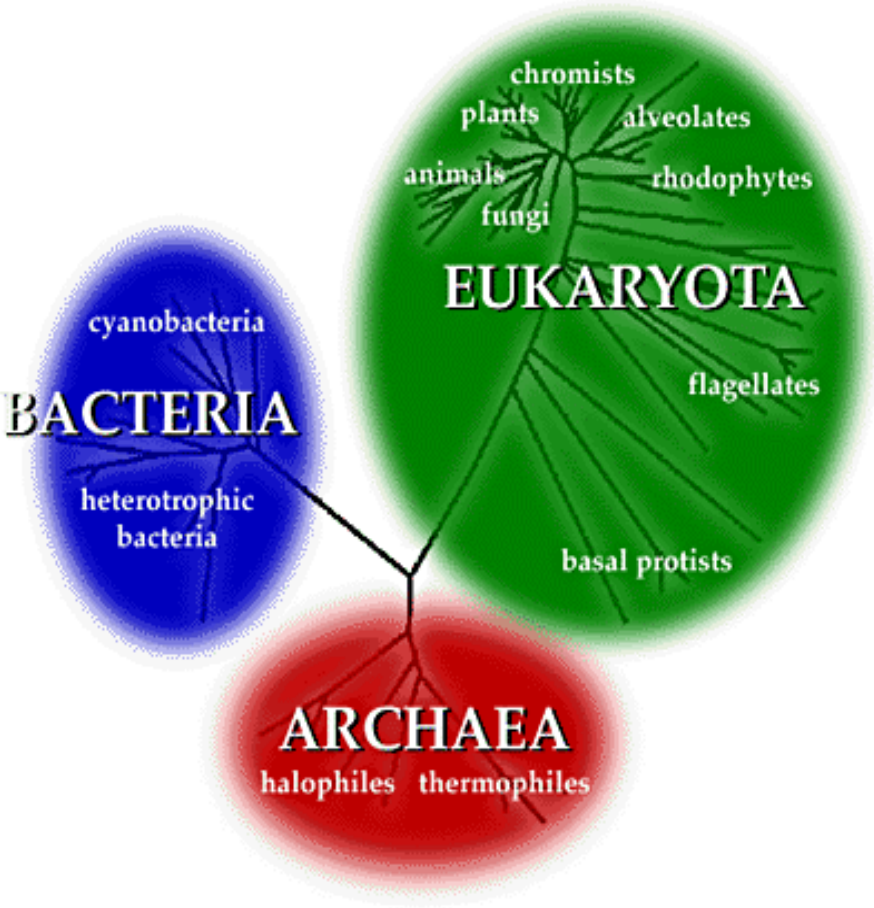
Significance of sequencing for microbiology

- **Classification of microorganisms**

Collection of sequences in Genbank allows reliable classifications, construction of phylogenetic trees.

Modern view of biology (Three Domains) is a direct outcome of sequencing, → distinctiveness of Archaea and Bacteria.

Three Domains view is 'microbe-centric', highlights the importance of microbes



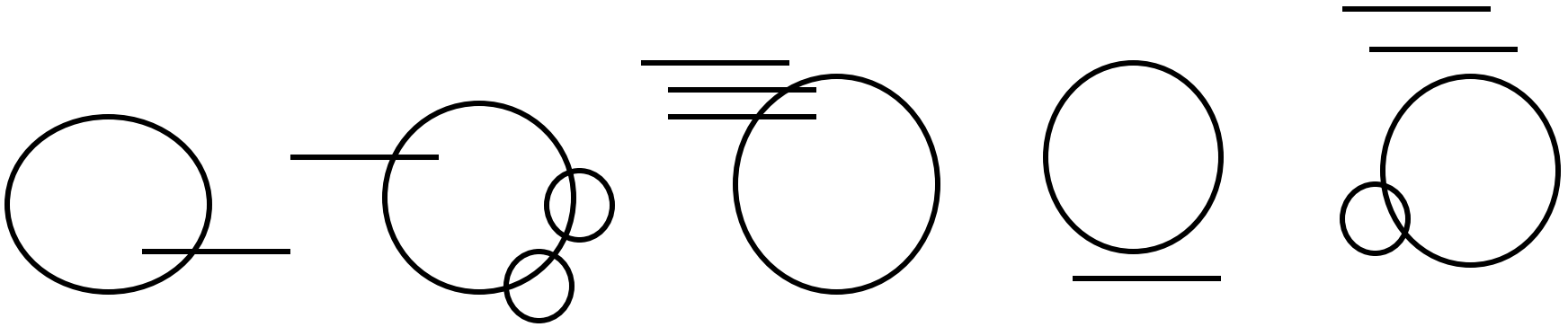
Significance of sequencing for microbiology

- **Understanding microbes from genomes / metagenomes**

Next-generation sequencing methods have greatly accelerated microbial genomics and metagenomics

These depend on simultaneous advances in computation power and software to enable analysis of Gb of data !

Genome-level study of bacteria is now very affordable, gives a large amount of information about the organism, and enables functional predictions.



Genomes of *Mycobacterium* strains NBB3, NBB4, JS60, JS617, JS623

Video links

- **Sanger method:**
  http://www.youtube.com/watch?v=oYpllbl0qF8

- **Illumina method:**
  http://www.youtube.com/watch?v=77r5p8lBwJk

- **454 method:**
  http://www.youtube.com/watch?v=kYAGFrbGl6E

Picture credits

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/D/DNAsequencing.html

http://www.csus.edu/indiv/r/rogersa/Bio181/SeqNotes.pdf

http://www.mun.ca/biology/scarr/How_it_works.htm

http://www.sars.no/facilities/sequencing.php

http://www.ocf.berkeley.edu/~edy/genome/sequencing.html

http://www.wiley-vch.de/books/sample/3527320903_c01.pdf

http://www.sciencephoto.com/media/138431/enlarge

http://genomicmedicineinitiative.org/pdfs/DrArnaout_lecture.pdf

http://research.stowers-institute.org/microscopy/

http://www.dac.neu.edu/biology/k.bergman/WebsiteBarney/Sema/sema.htm