พันธุวิศวกรรมและการโคลนยีน
DNA fingerprinting
Gel electrophoresis
DNA sequencing

มรรษภูมิ เดชเจริญ
matsapume.d@psu.ac.th
หลักสูตรชีววิทยา

20230105

Polymerase Chain Reaction (PCR)

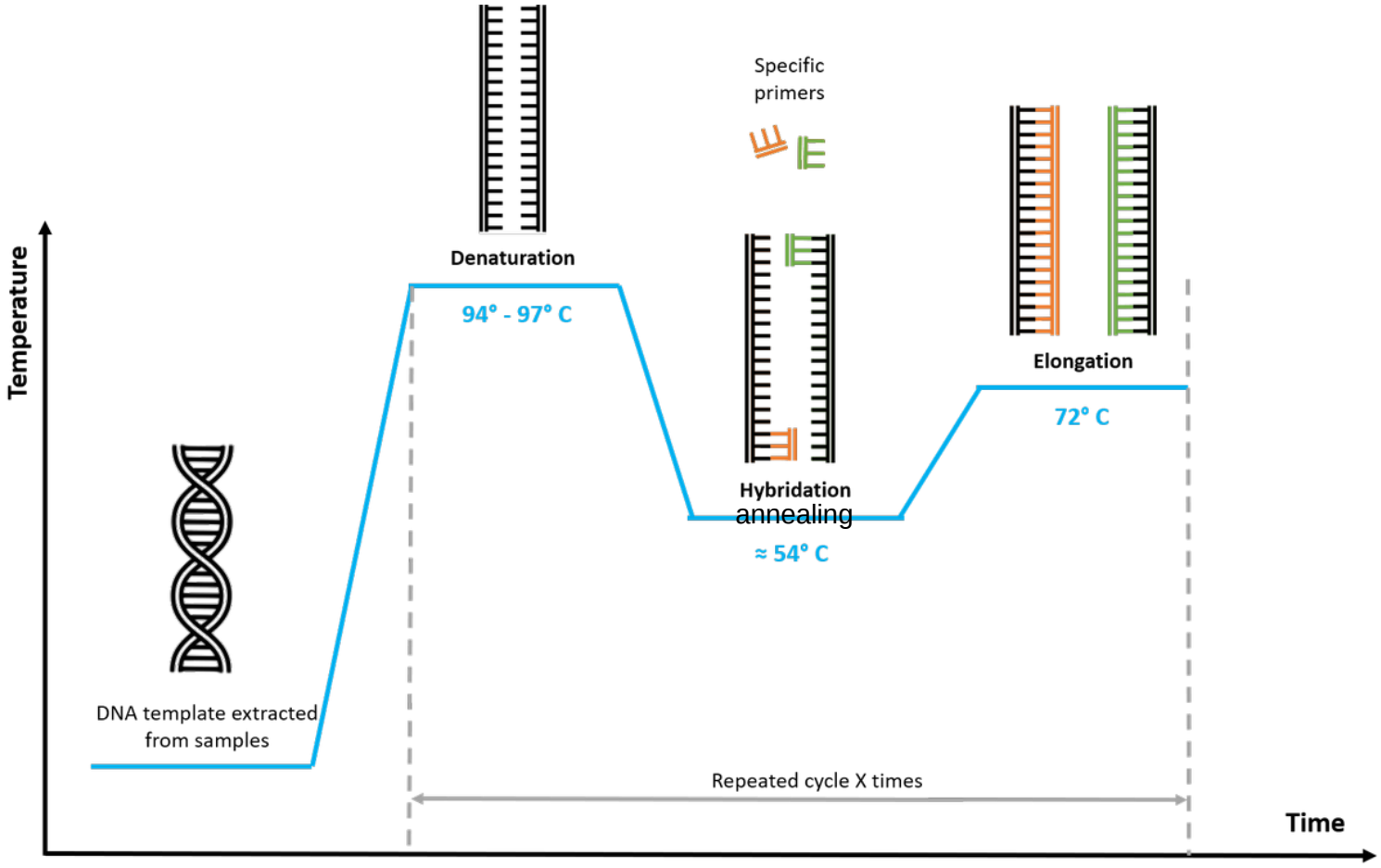
- Template DNA
- DNA polymerase
- Primers (forward and reverse)
- Deoxynucleoside triphosphates (dNTPs: dATP, dCTP, dGTP, and dTTP)
- Required cofactor for activity of DNA polymerases: Mg^{2+}
- Buffer: suitable chemical environment for activity of DNA polymerase

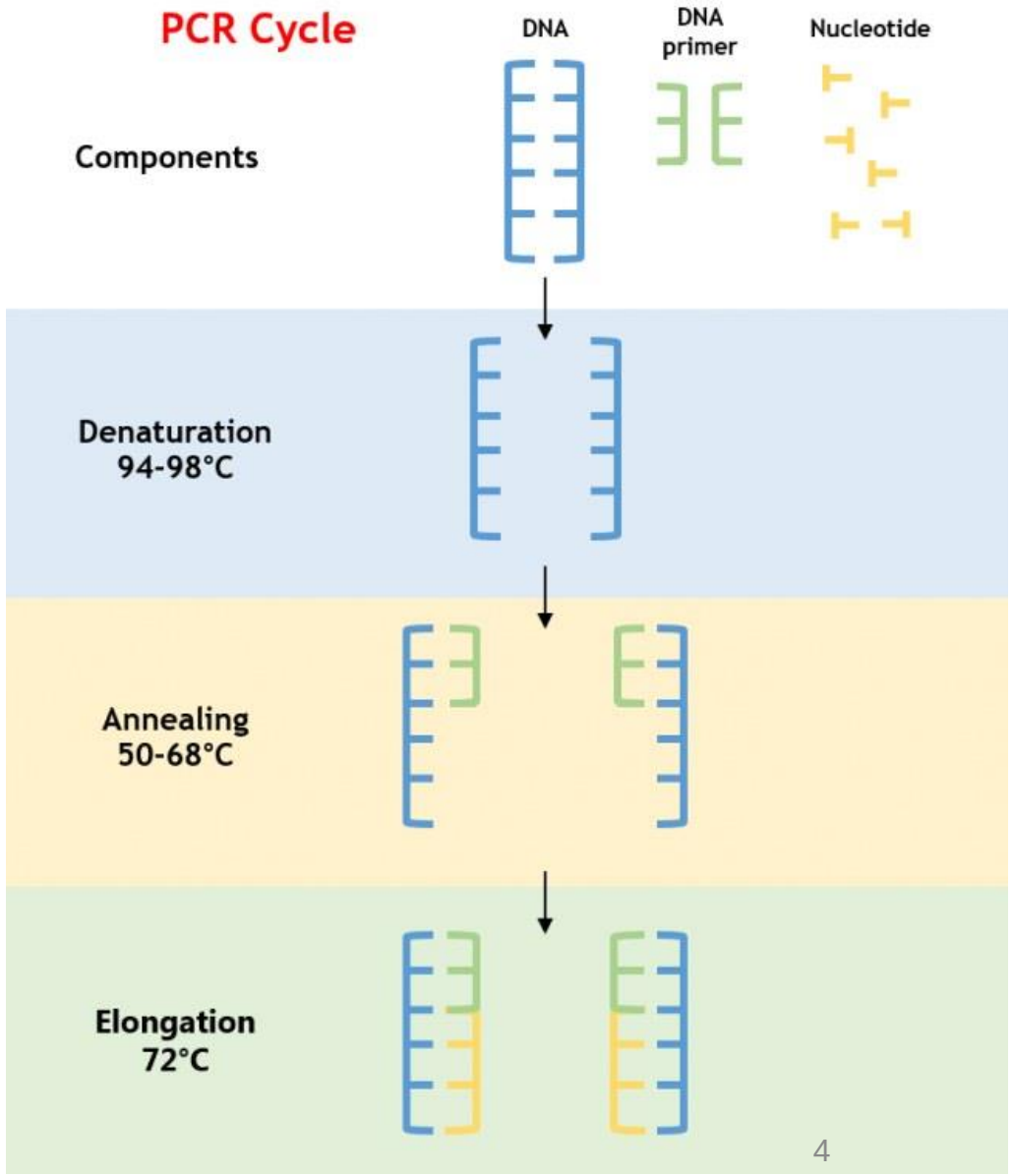
Polymerase Chain Reaction (PCR)

PCR thermocycler

Polymerase Chain Reaction (PCR)

Steps





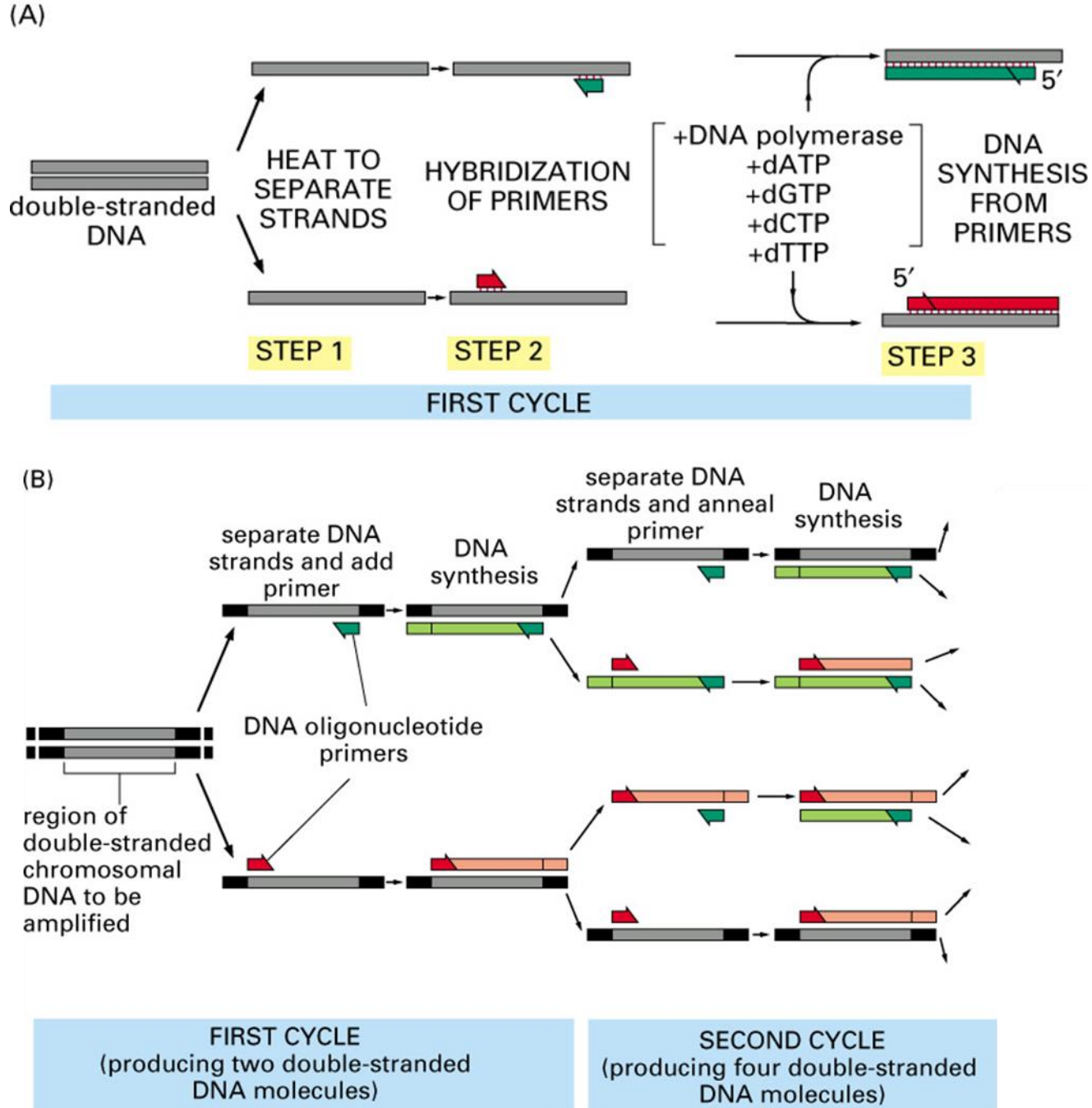
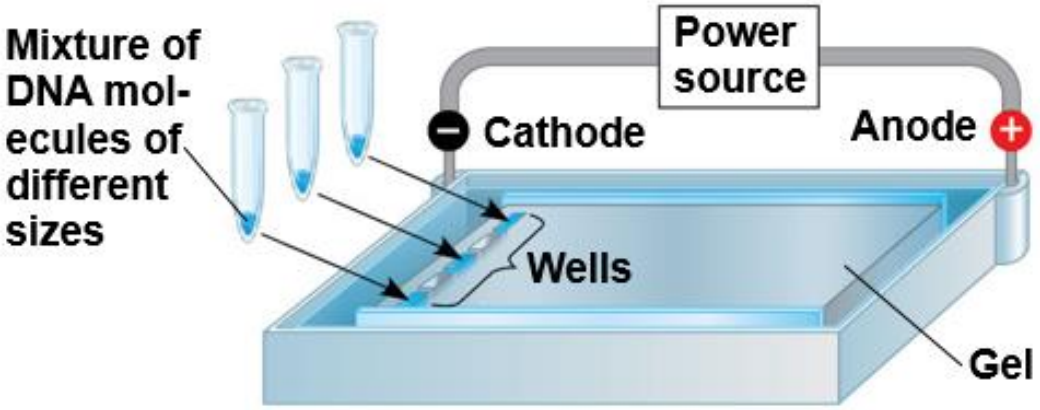
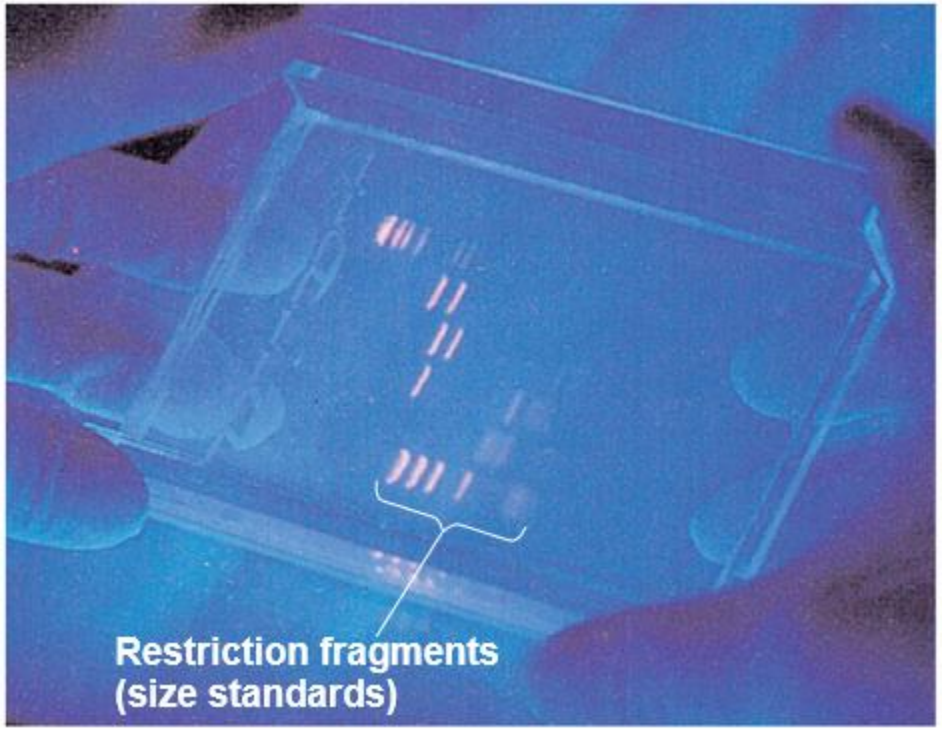


Figure 8–39 part 2 of 3. Molecular Biology of the Cell, 4th Edition.

Gel Electrophoresis

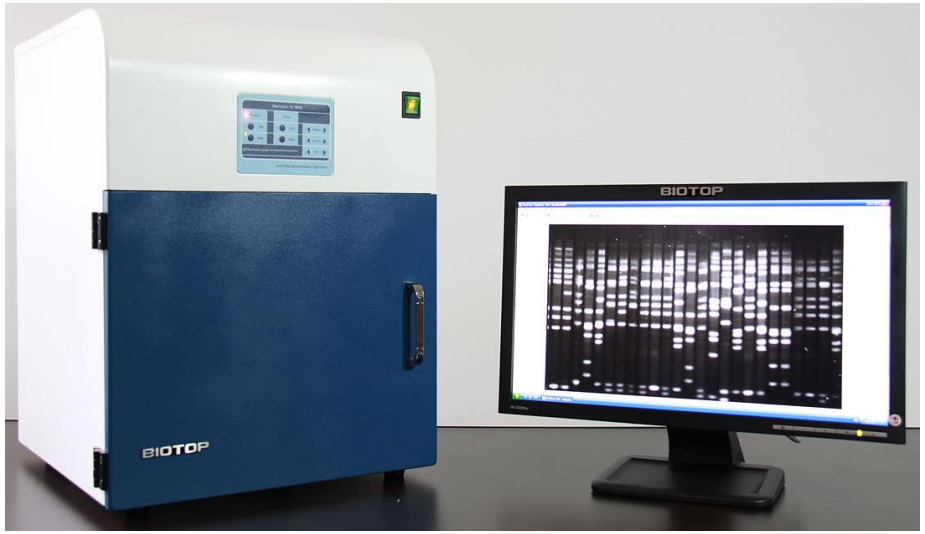


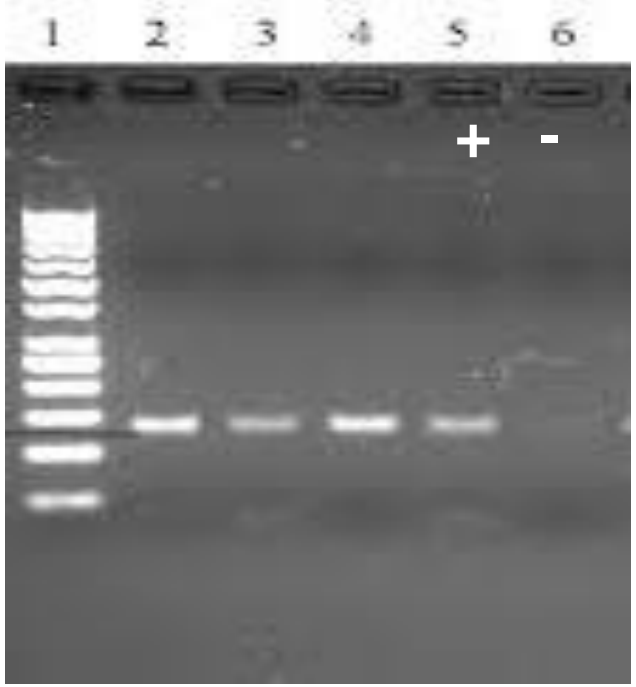
a) Negatively charged DNA molecules will move toward the positive electrode.



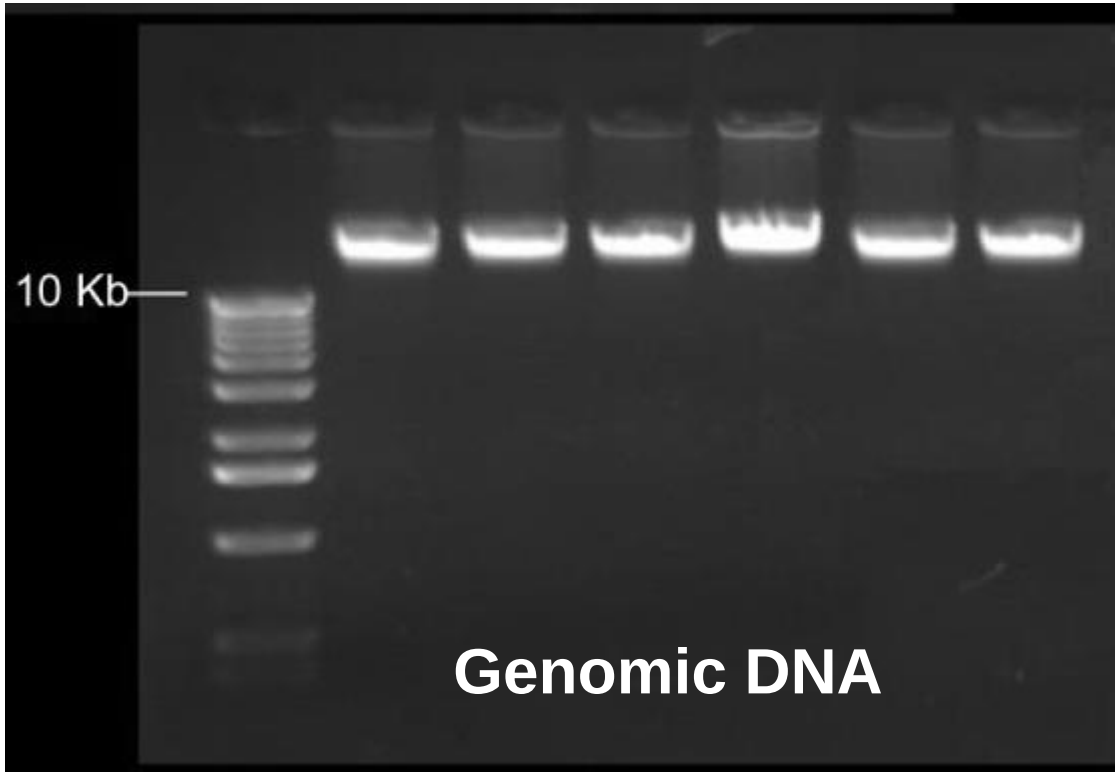
b) Shorter molecules are slowed down less than longer ones, so they move faster through the gel.

Gel documentation system

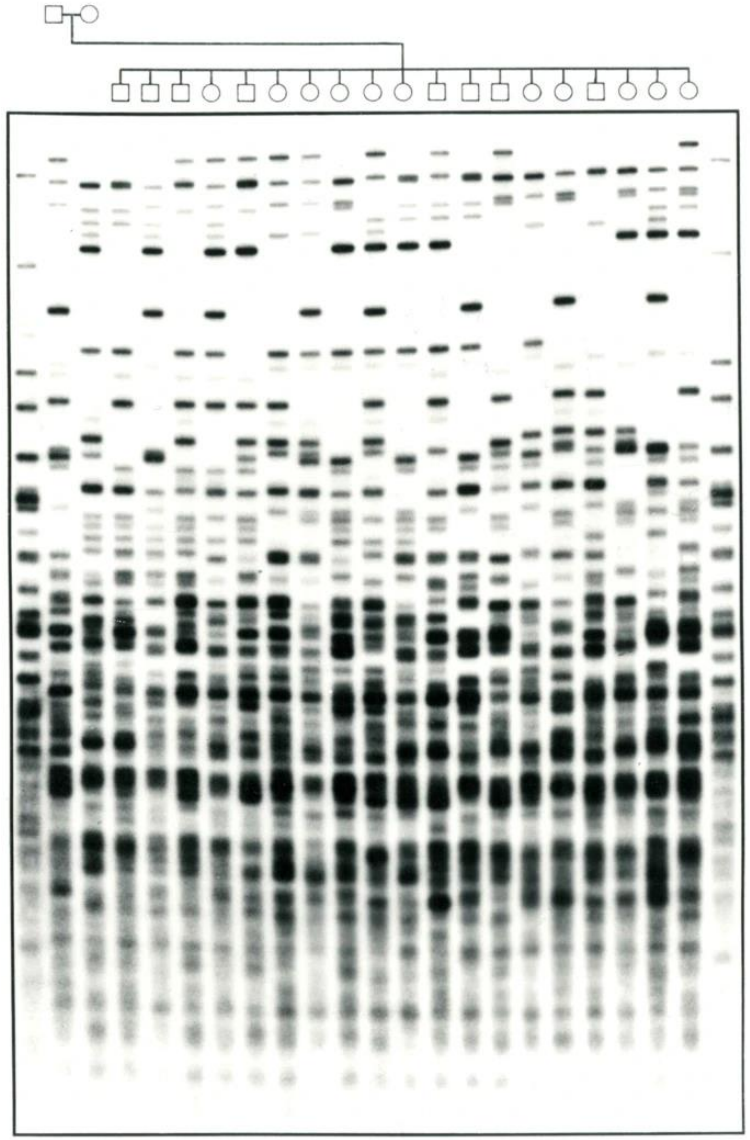


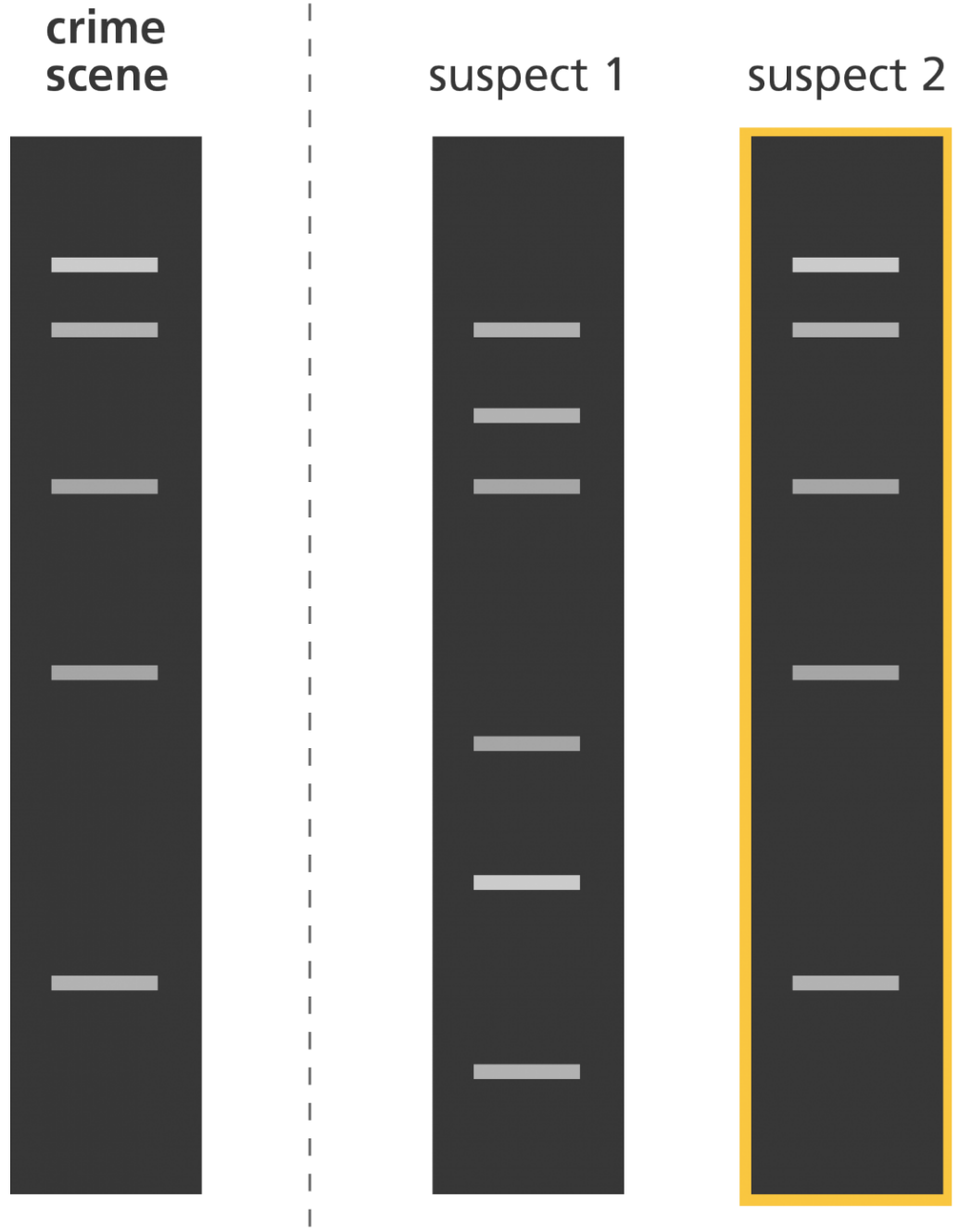


PCR product: Amplicon

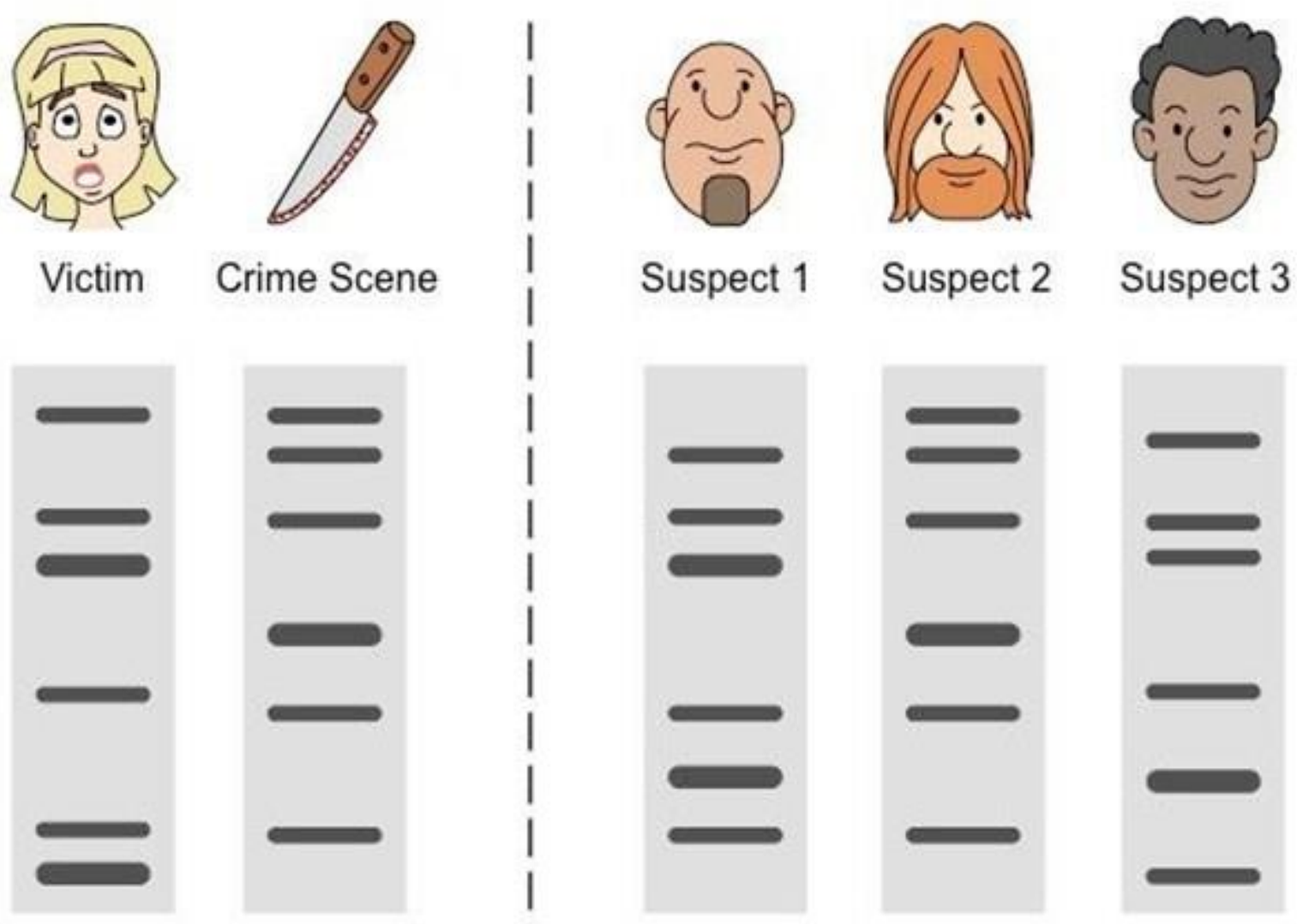


DNA fingerprinting





DNA fingerprinting

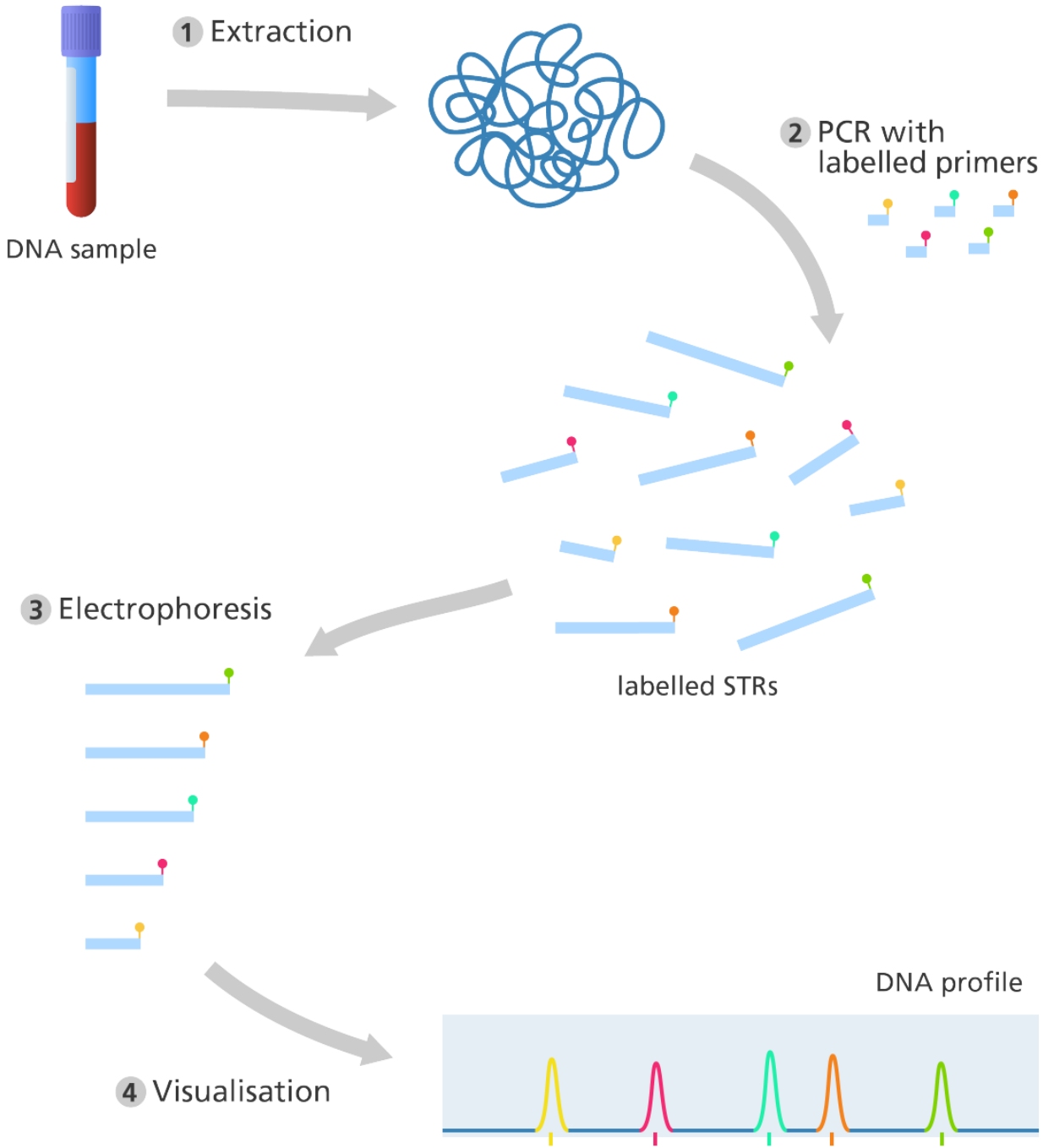


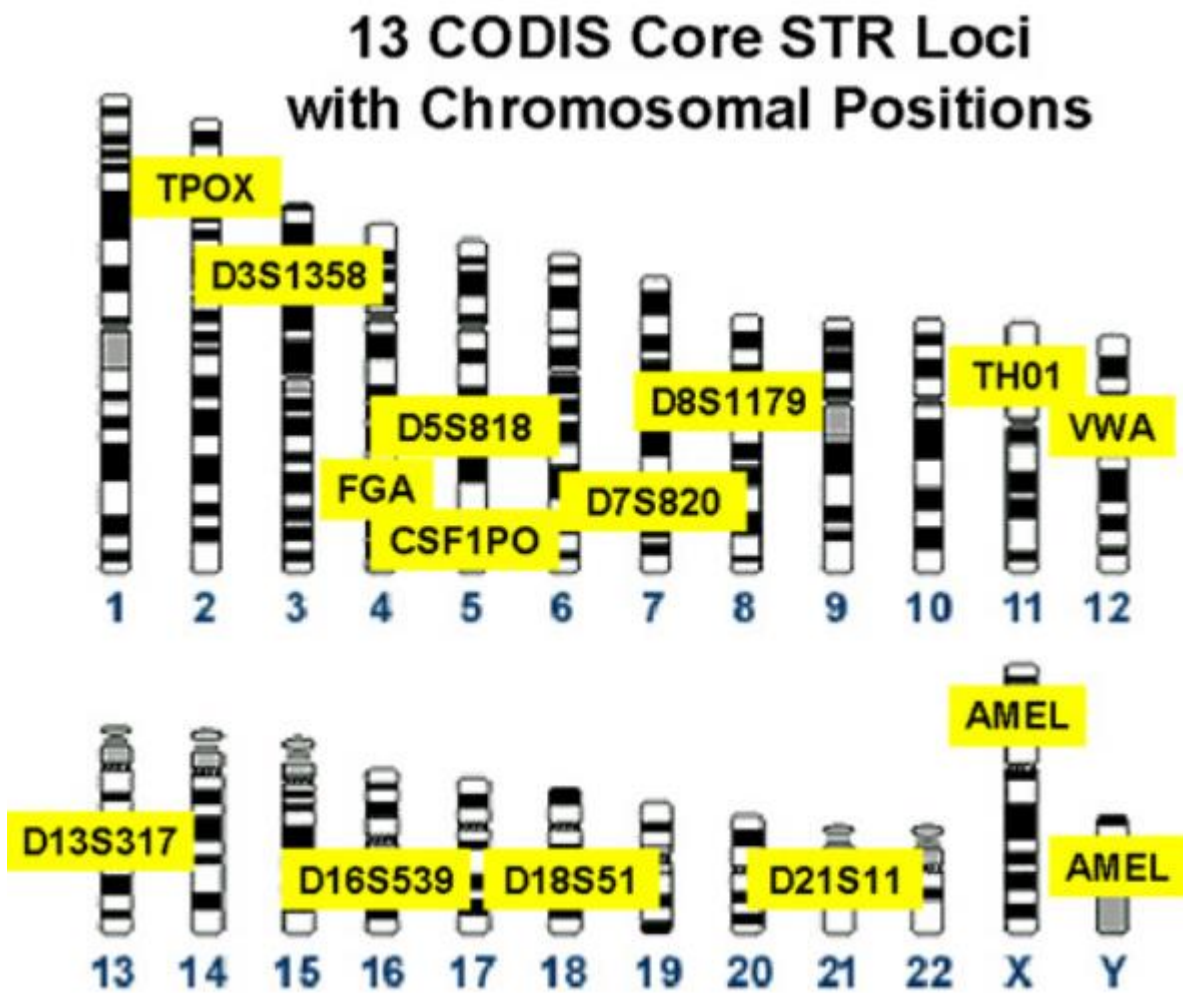
DNA fingerprinting

DNA repeats

- ATATATATATATATATATATATAT, $(AT)_{11}$
- GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA, $(GCA)_{12}$
- TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
- GTCAGTCAGTCAGTCAGCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGT CAGTCA

- Minisatellites
- Microsatellites


1 Extraction
DNA sample
2 PCR with labelled primers
labelled STRs
3 Electrophoresis
4 Visualisation
DNA profile


13 CODIS Core STR Loci with Chromosomal Positions
TPOX
D3S1358
D5S818
FGA
CSF1PO
D7S820
D8S1179
TH01
VWA
D13S317
D16S539
D18S51
D21S11
AMEL
AMEL
1 2 3 4 5 6 7 8 9 10 11 12
13 14 15 16 17 18 19 20 21 22 X Y

TPOX

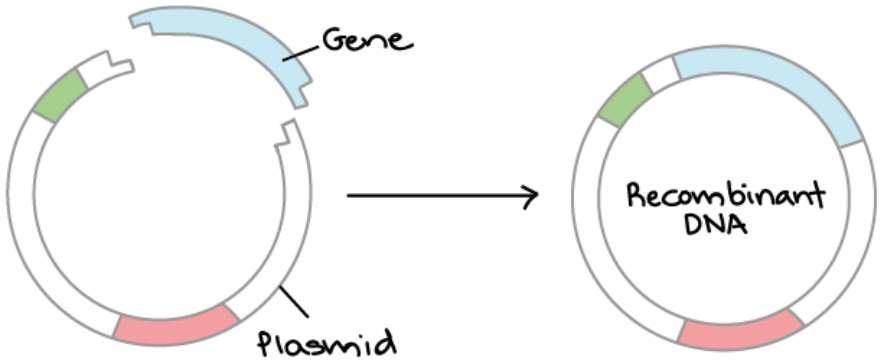
Other Names	Chromosomal Location
hTPO, TPO UniSTS: 240638	**2p25.3**; intron 10 of human thyroid peroxidase gene Chr 2; 1.472 Mb (May 2004, NCBI build 35)

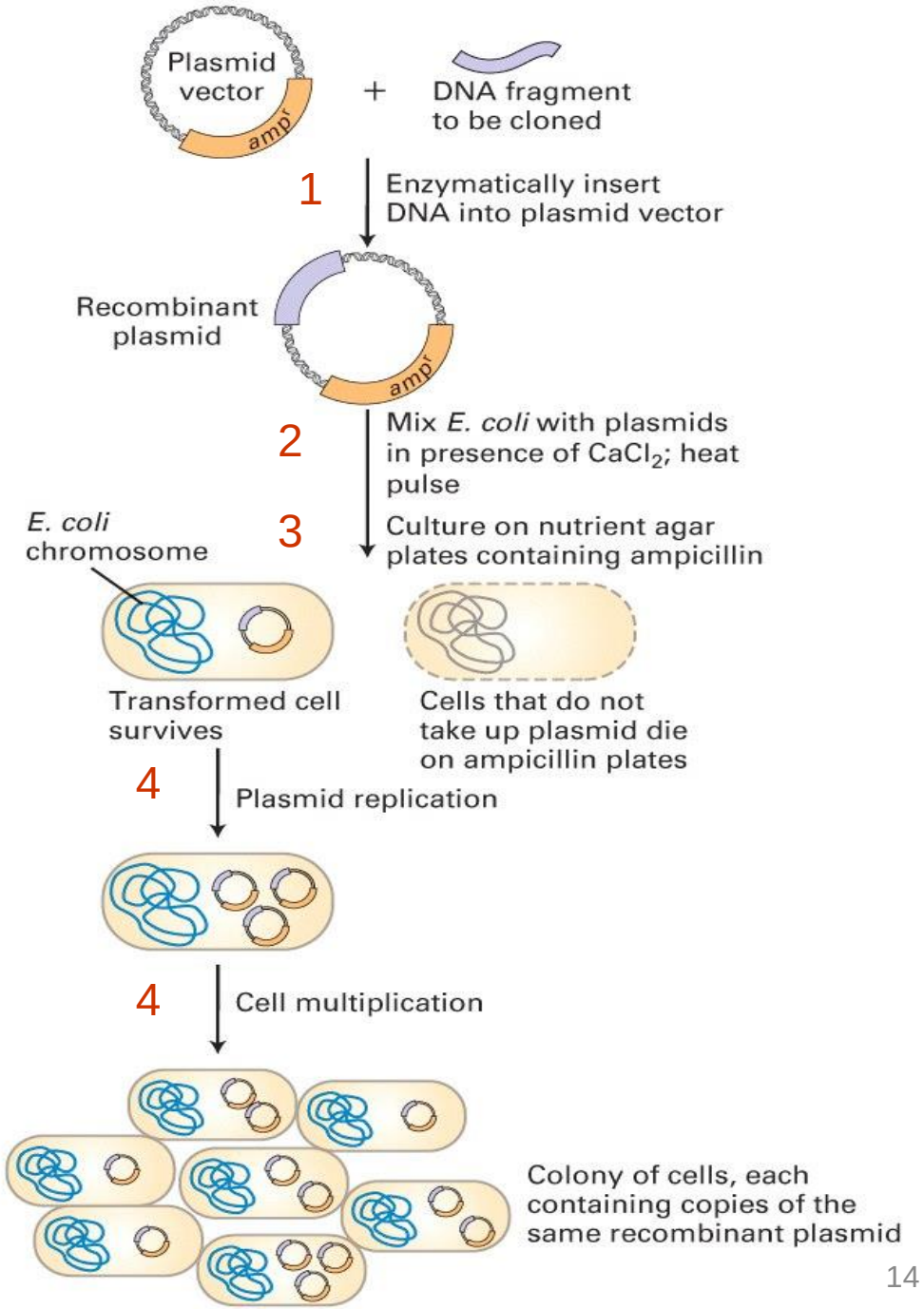
FBI CODIS Core STR Loci:

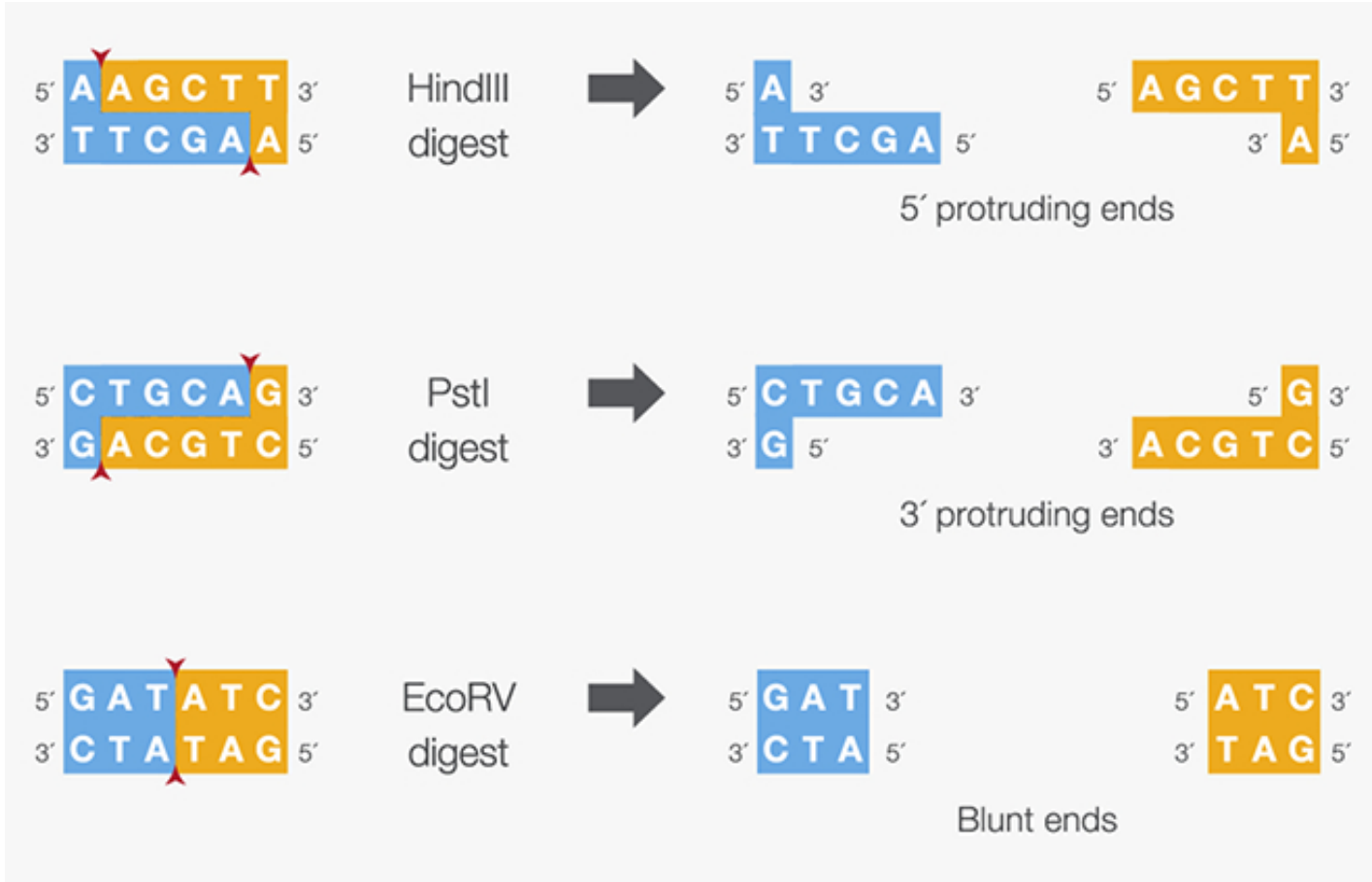
ENFSI
EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES

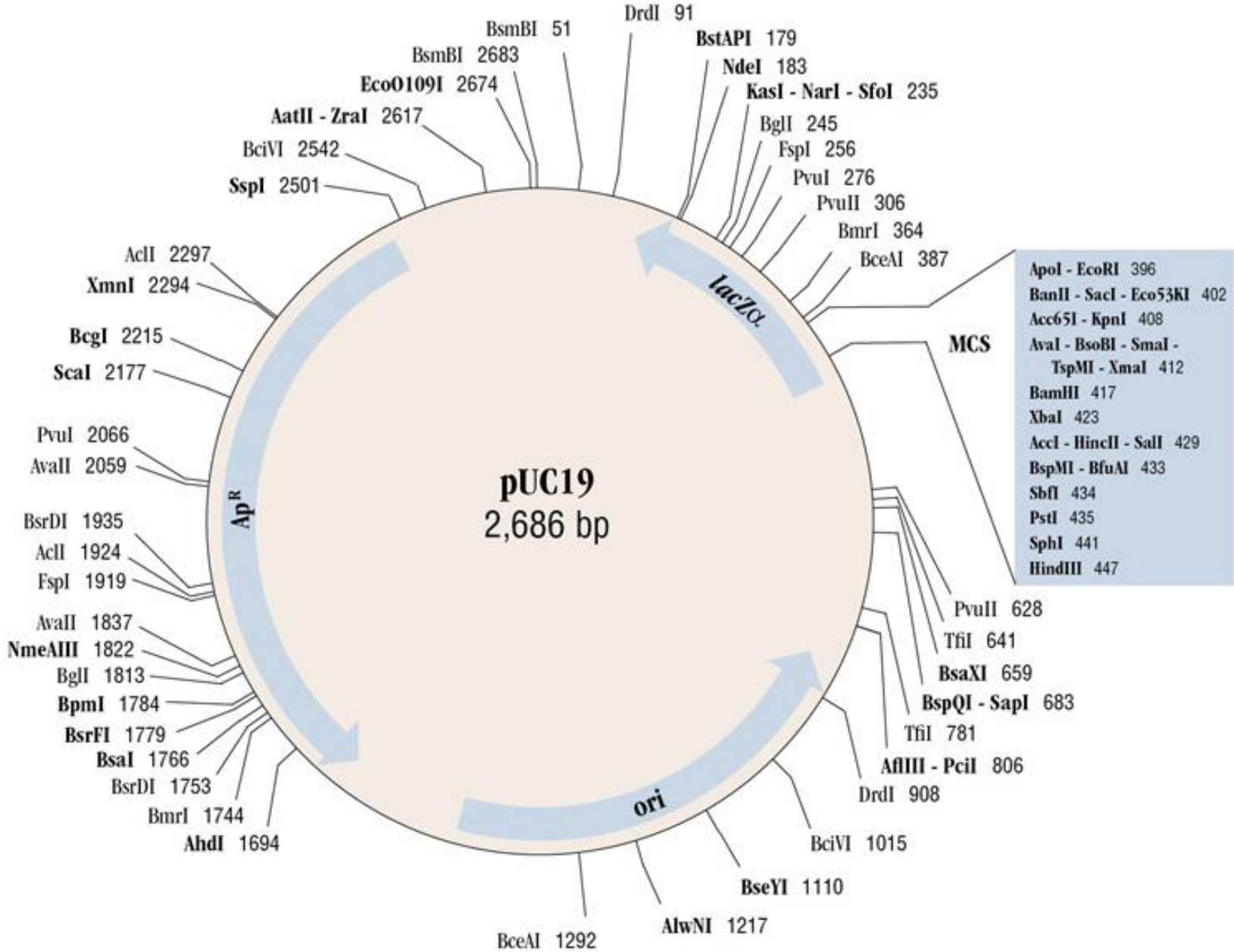
98 bp
102 bp
106 bp
110 bp
113 bp
114 bp
118 bp
122 bp
123 bp
125 bp
126 bp
130 bp
134 bp
135 bp
138 bp
142 bp
146 bp

DNA cloning

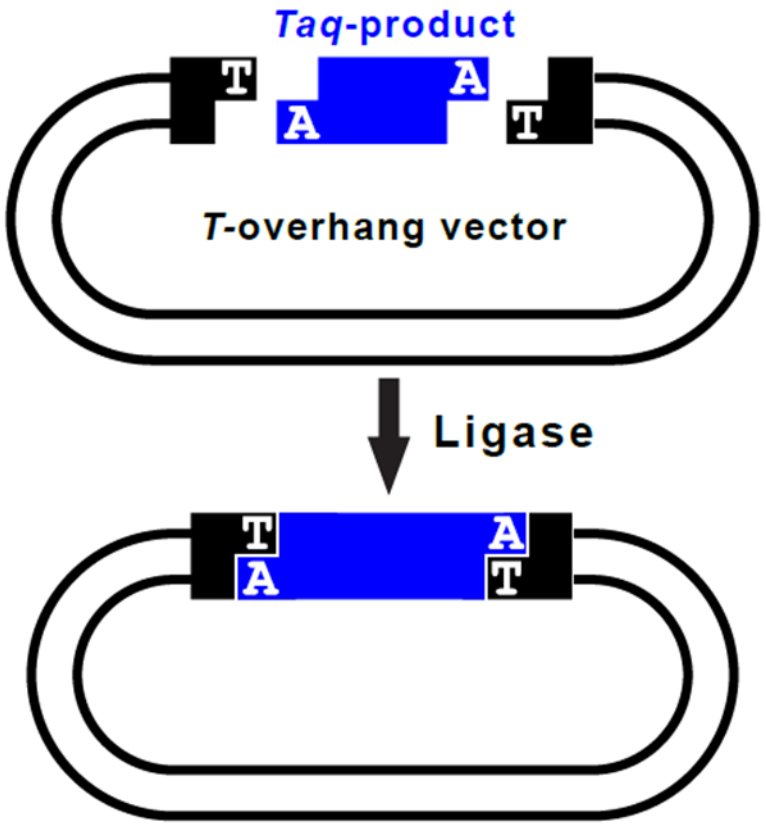

Gene
Plasmid
Recombinant DNA

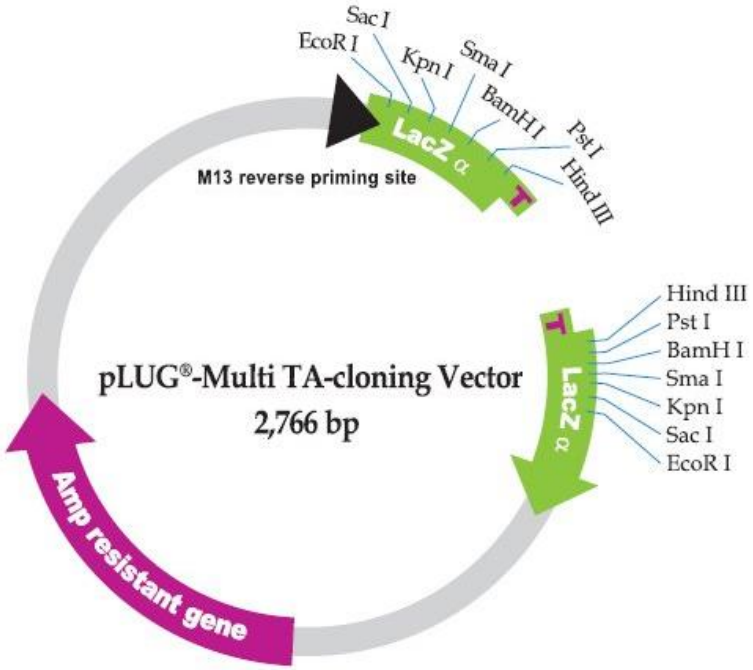

Plasmid vector
ampr
+
DNA fragment to be cloned
1
Enzymatically insert DNA into plasmid vector
Recombinant plasmid
ampr
2
Mix E. coli with plasmids in presence of CaCl2; heat pulse
3
Culture on nutrient agar plates containing ampicillin
E. coli chromosome
Transformed cell survives
Cells that do not take up plasmid die on ampicillin plates
4
Plasmid replication
4
Cell multiplication
Colony of cells, each containing copies of the same recombinant plasmid

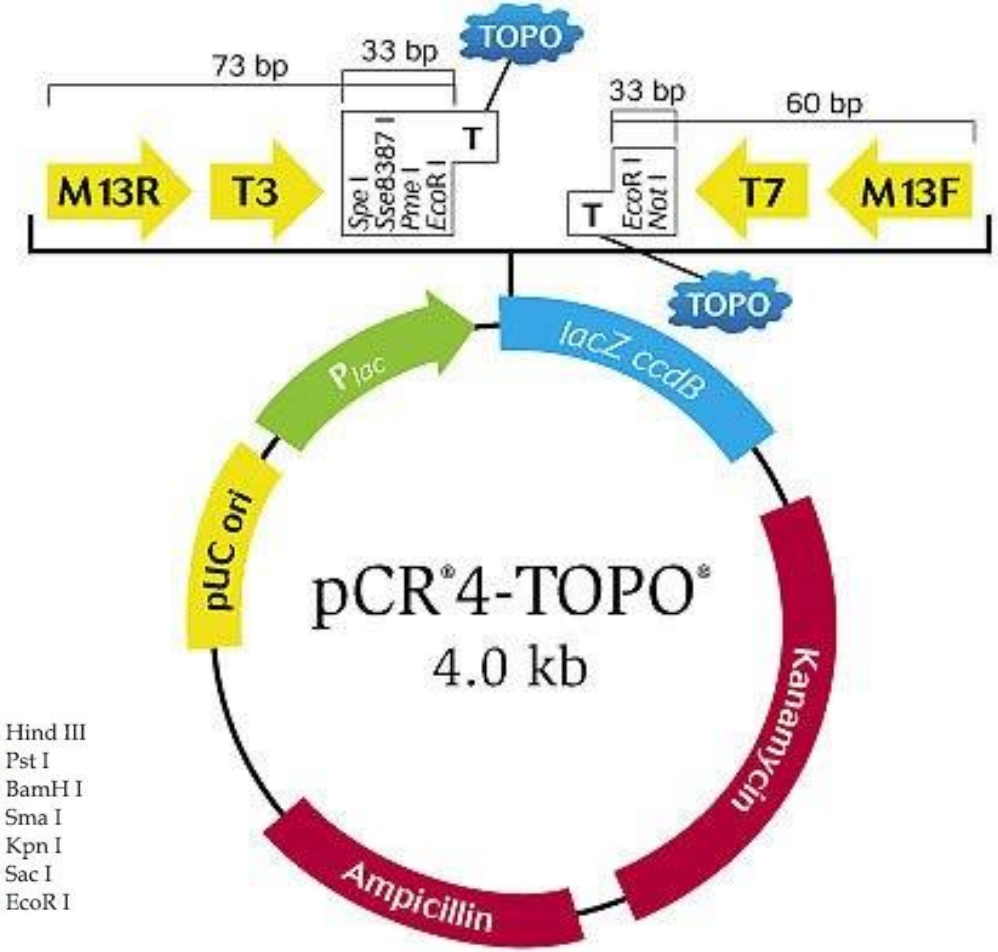

5′ A A G C T T 3′
3′ T T C G A A 5′
HindIII digest
5′ A 3′
3′ T T C G A 5′
5′ A G C T T 3′
3′ A 5′
5′ protruding ends
5′ C T G C A G 3′
3′ G A C G T C 5′
PstI digest
5′ C T G C A 3′
3′ G 5′
5′ G 3′
3′ A C G T C 5′
3′ protruding ends
5′ G A T A T C 3′
3′ C T A T A G 5′
EcoRV digest
5′ G A T 3′
3′ C T A 5′
5′ A T C 3′
3′ T A G 5′
Blunt ends


pUC19
2,686 bp
lacZα
ori
Ap^R
MCS
ApoI - EcoRI 396
BanII - SacI - Eco53KI 402
Acc65I - KpnI 408
AvaI - BsoBI - SmaI - TspMI - XmaI 412
BamHI 417
XbaI 423
AccI - HincII - SalI 429
BspMI - BfuAI 433
SbfI 434
PstI 435
SphI 441
HindIII 447
DrdI 91
BstAPI 179
NdeI 183
KasI - NarI - SfoI 235
BglI 245
FspI 256
PvuI 276
PvuII 306
BmrI 364
BceAI 387
PvuII 628
TfiI 641
BsaXI 659
BspQI - SapI 683
TfiI 781
AflIII - PciI 806
DrdI 908
BciVI 1015
BseYI 1110
AlwNI 1217
BceAI 1292
AhdI 1694
BmrI 1744
BsrDI 1753
BsaI 1766
BsrFI 1779
BpmI 1784
BglI 1813
NmeAIII 1822
AvaII 1837
FspI 1919
AclI 1924
BsrDI 1935
AvaII 2059
PvuI 2066
ScaI 2177
BcgI 2215
XmnI 2294
AclI 2297
SspI 2501
BciVI 2542
AatII - ZraI 2617
EcoO109I 2674
BsmBI 2683
BsmBI 51

DNA cloning


Taq-product
T
A
A
T
T-overhang vector
Ligase
T
A
A
T


Sac I
EcoR I
Kpn I
Sma I
BamH I
Pst I
Hind III
LacZ α
M13 reverse priming site
pLUG®-Multi TA-cloning Vector
2,766 bp
Amp resistant gene
Hind III
Pst I
BamH I
Sma I
Kpn I
Sac I
EcoR I


73 bp
33 bp
TOPO
33 bp
60 bp
M13R
T3
Spe I
Sse8387 I
Pme I
EcoR I
T
T
EcoR I
Not I
T7
M13F
TOPO
lacZ ccdB
P lac
pUC ori
pCR®4-TOPO®
4.0 kb
Kanamycin
Ampicillin

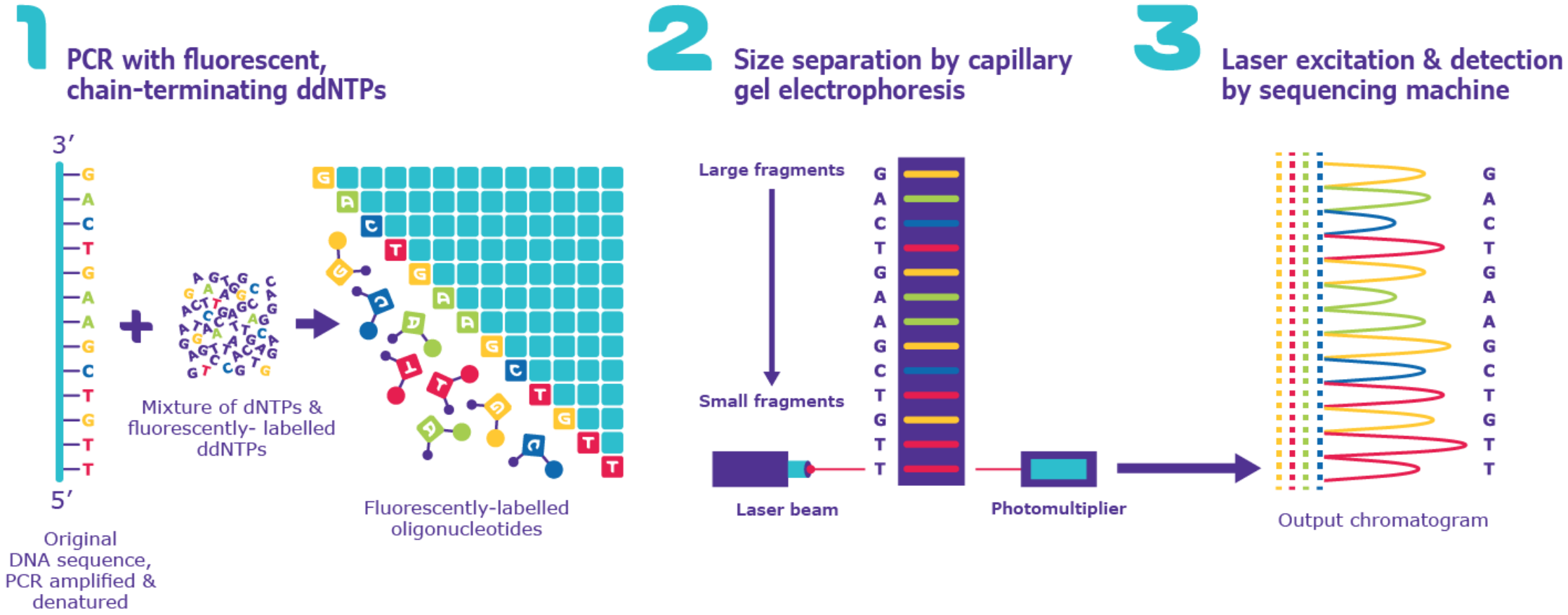
How do you sequence DNA?

- **Sanger** (**dideoxy**, enzymatic) – developed by Frederick Sanger and is still used today with little change to the basic method.

High-throughput DNA sequencing

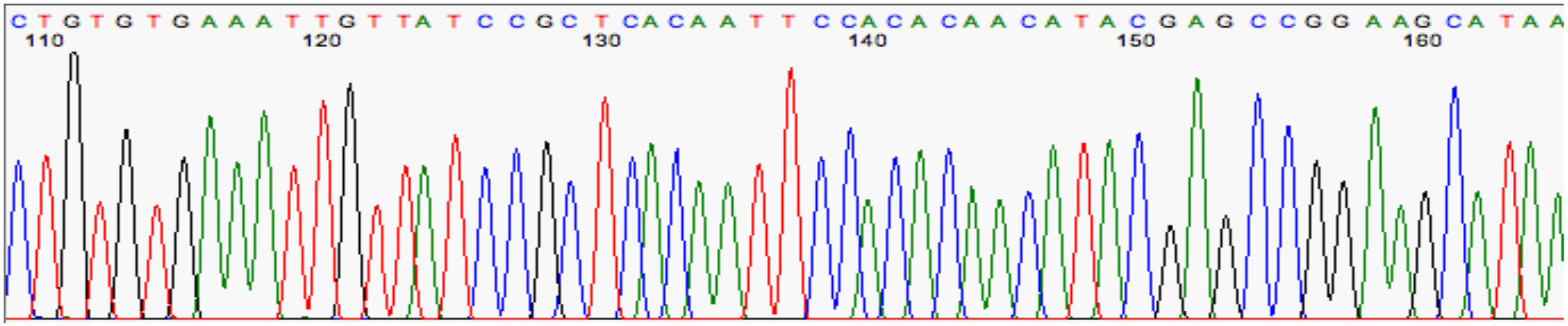
- **Next-generation sequencing**

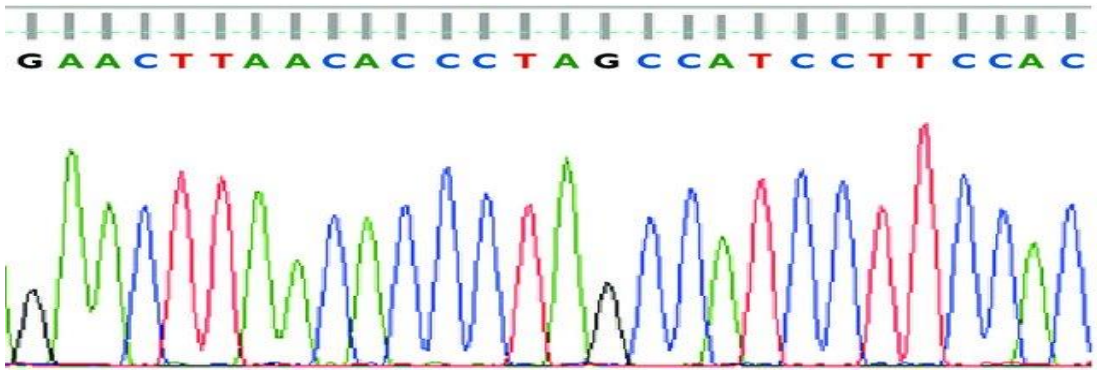
1. All 4 fluorescently-labeled ddNTPs are used in 1 reaction, each a different “color”
2. Fragments are separated in matrix-filled capillary tubes, 1 capillary per reaction
3. Laser detects fluorescence automatically as each fragment exits capillary
4. Computer software “calls bases” and processes sequence files



https://www.sigmaaldrich.com/technical-documents/articles/biology/sanger-sequencing.html

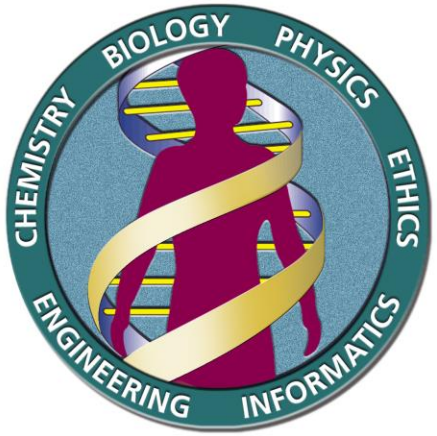
Sequencing chromatogram





A quick history of sequencing

1869 – Discovery of DNA
1909 – Chemical characterization
1953 – Structure of DNA solved
1977 – Sanger sequencing invented
– First genome sequenced – bacteriophage Phi X 174 (5 kb)
1986 – First automated sequencing machine
1990 – Human Genome Project started
1992 – First “sequencing factory” at TIGR



A quick history of sequencing

1995 – First bacterial genome – *H. influenzae* (1.8 Mb)
1998 – First animal genome – *C. elegans* (97 Mb)
2003 – Completion of Human Genome Project (3 Gb). 13 years, $2.7 bn
2005 – First “next-generation” sequencing instrument
2021 – >60,000 genome sequences in NCBI database

Published: 15 February 2001

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium

Nature **409**, 860–921(2001) | Cite this article

121k Accesses | **14976** Citations | **1081** Altmetric | Metrics

Genomics Revolution? *Completion* of the Human Genome

Published: 05 November 2008

My genome. So what?

Open Access | Published: 06 November 2008

The diploid genome sequence of an Asian individual

Jun Wang ✉, Wei Wang, [...] Jian Wang ✉

Nature **456**, 60–65(2008) | Cite this article

3520 Accesses | **673** Citations | **75** Altmetric | Metrics

Han Chinese

Open Access | Published: 06 November 2008

Accurate whole human genome sequencing using reversible terminator chemistry

David R. Bentley ✉, Shankar Balasubramanian, [...] Anthony J. Smith

Nature **456**, 53–59(2008) | Cite this article

20k Accesses | **2171** Citations | **80** Altmetric | Metrics

Illumina sequencing (Solexa)

Published: 01 September 2005

Initial sequence of the chimpanzee genome and comparison with the human genome

The Chimpanzee Sequencing and Analysis Consortium

Nature **437**, 69–87(2005) | Cite this article

29k Accesses | **1454** Citations | **384** Altmetric | Metrics

Some examples of sequenced genomes:

Species	**Genome size (*C*)**
Mycoplasma genitalium (bacteria)	580 Kb
Haemophilis influenzae (bacteria)	1.8 Mb
Escherichia coli (bacteria)	4.7 Mb
Saccharomyces cerevisea (yeast)	12.5 Mb
Caenorhabditis elegans (worm)	97 Mb
Arabidopsis thaliana (mustard weed)	125 Mb
Drosophila melanogaster (fruit fly)	180 Mb
Fugu rubripes (puffer fish)	400 Mb
Oryza sativa (rice)	400 Mb
Homo sapiens (human)	3.2 Gb

Next Generation DNA Sequencing

- Next-generation sequencing (NGS or high-throughput sequencing)
- The catch-all term used to describe several different modern sequencing technologies.
- Allow for sequencing of DNA or RNA more quickly and cheaply than the Sanger sequencing

Advantages of NGS

- No *priori* knowledge of the genome required
- Offers single-nucleotide resolution, possible to detect related genes (or features), alternatively spliced transcripts, allelic gene variants and single nucleotide polymorphisms
- Higher dynamic range of signal
- Requires less DNA/RNA as input (nanograms of materials are sufficient)
- Higher reproducibility

NGS technologies

- **Illumina**: works by simultaneously identifying DNA bases, as each base emits a unique fluorescent signal, and adding them to a nucleic acid chain

Illumina sequencing

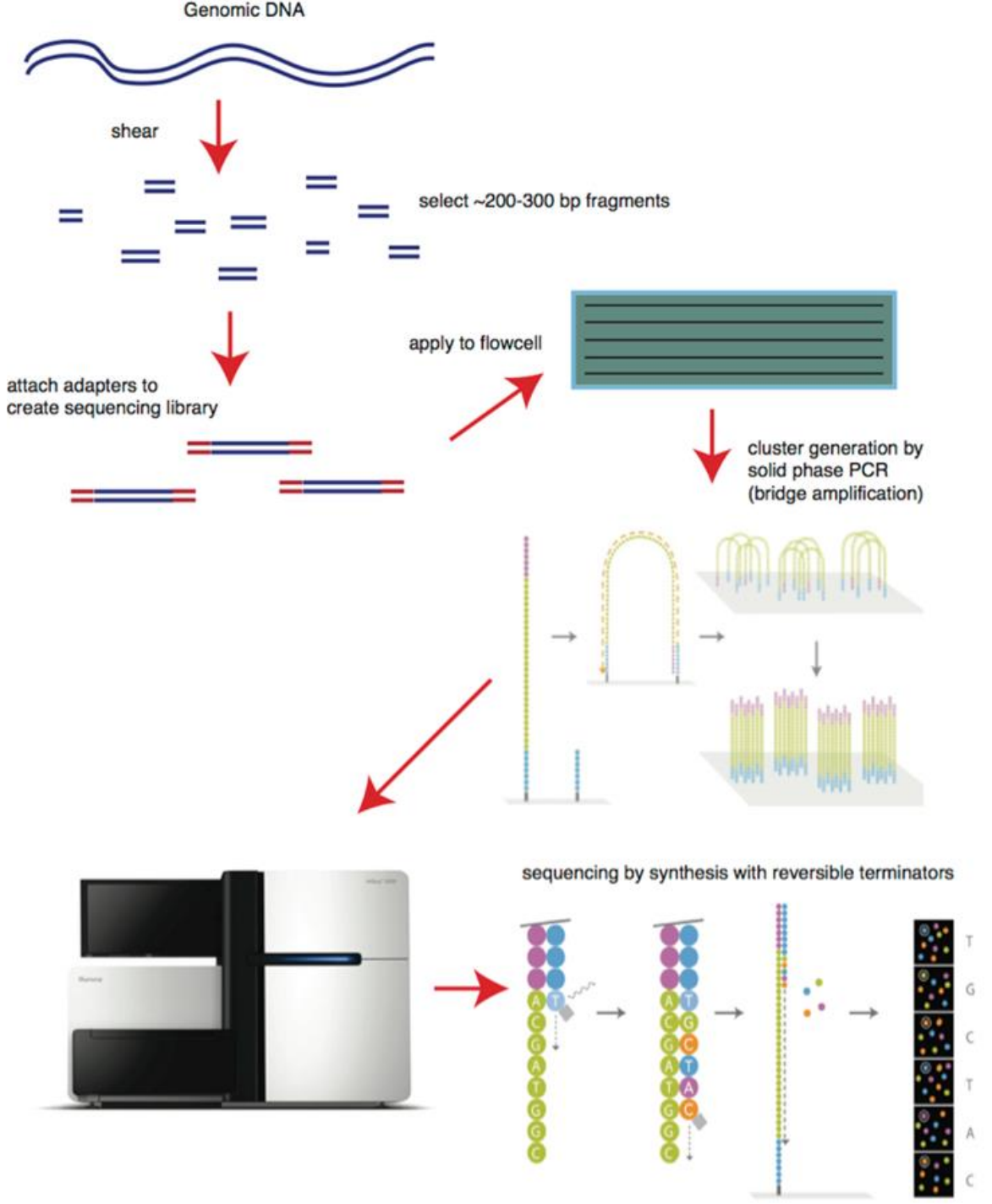
Sequencing-by-synthesis

Vast numbers of short reads are sequenced in a single stroke.

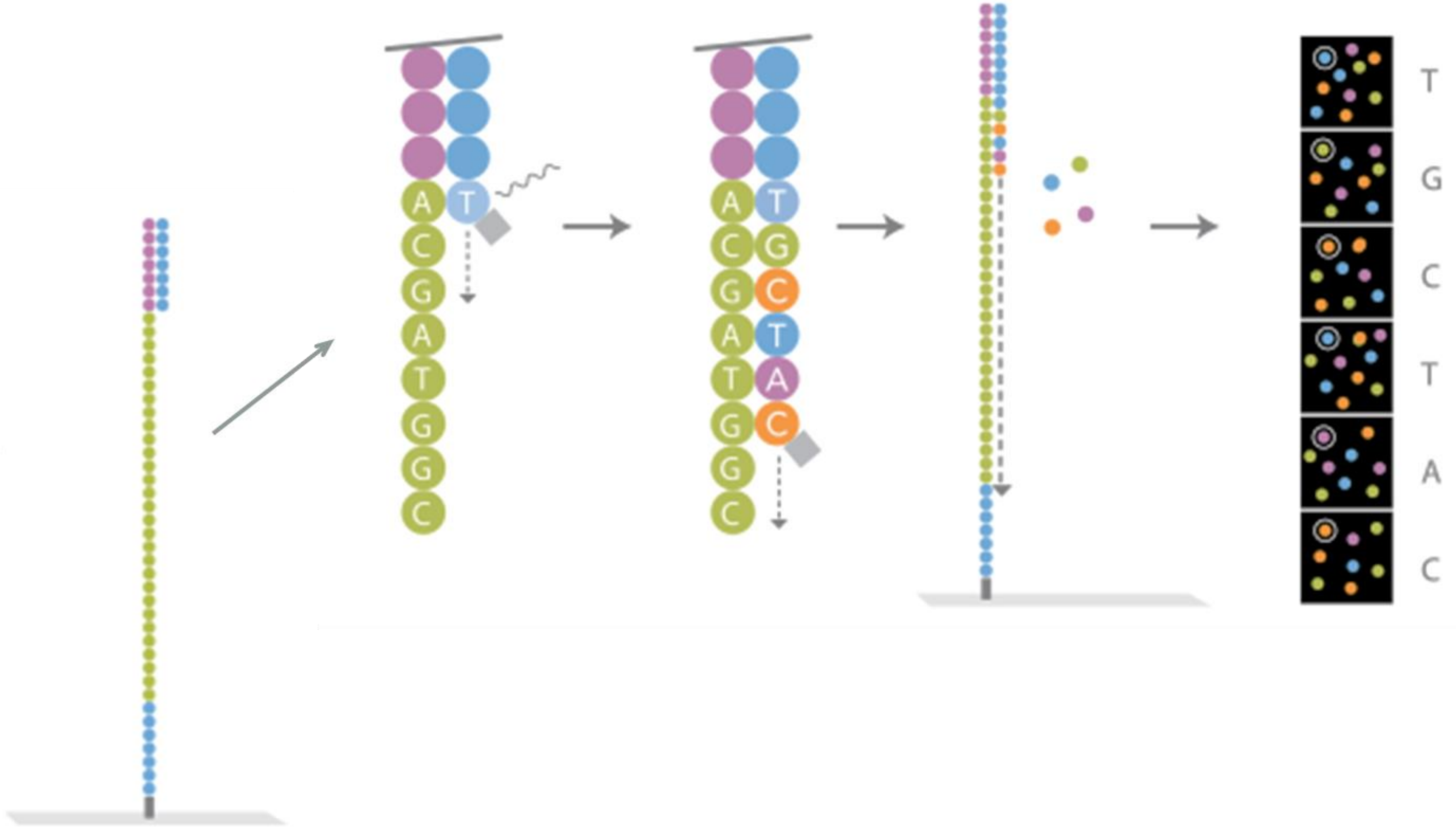
The input sample must be cleaved into short sections (100-300bp)

Illumina sequencing

- Library construction
  - Fragment, attach adapter DNA
- Cluster generation
  - Add to flow cell
  - Bridge amplification
- Sequencing
  - Single base at a time, imaging
- Data analysis
  - Images transformed into base calls and ‘reads’



Illumina sequencing



4 main advantages of NGS over Sanger sequencing

- **Sample size**
  NGS needs significantly less DNA and is more accurate and reliable than Sanger sequencing.
- **Speed**
  NGS is quicker than Sanger sequencing
- **Cost**
  The reduced time, manpower and reagents in NGS mean that the costs are much lower.
- **Accuracy**
  Repeats are intrinsic to NGS, as each read is amplified before sequencing, and because it relies on many short overlapping reads, so each section of DNA or RNA is sequenced multiple times.

Third generation sequencing: Long reads sequencing

SMRT Sequencing: PacBio

Article | Open Access | Published: 07 October 2020

Assembly of the durian chloroplast genome using long PacBio reads

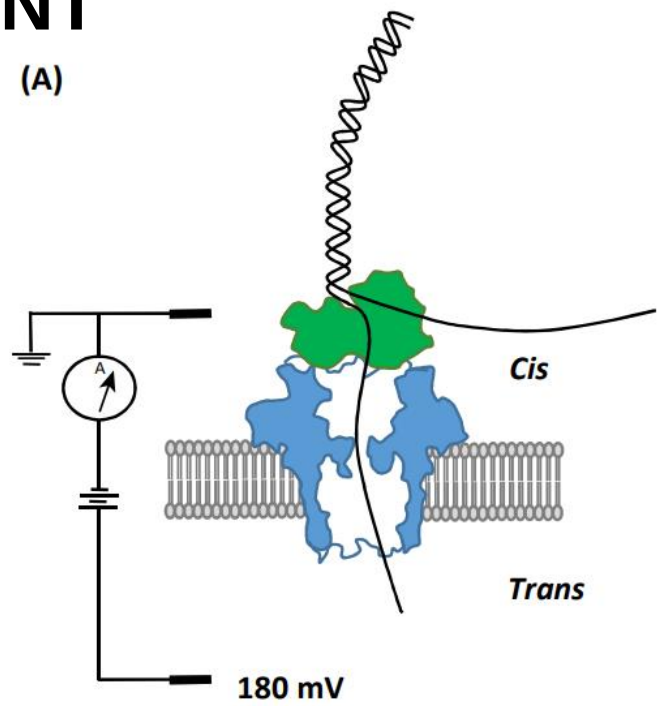
Jeremy R. Shearman, Chutima Sonthirod, Chaiwat Naktang, Duangjai Sangsrakru, Thippawan Yoocha, Ratchanee Chatbanyong, Siriporn Vorakuldumrongchai, Orwintinee Chusri, Sithichoke Tangphatsornruang & Wirulda Pootakham

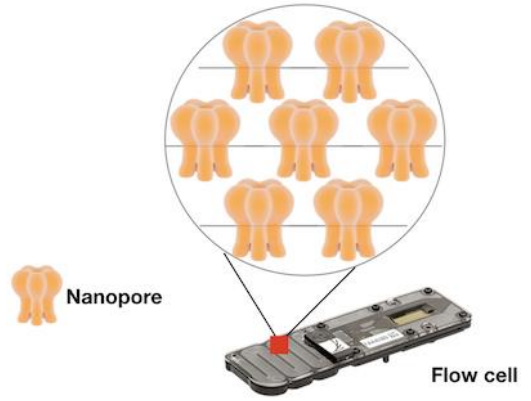
Third generation sequencing: Long reads sequencing

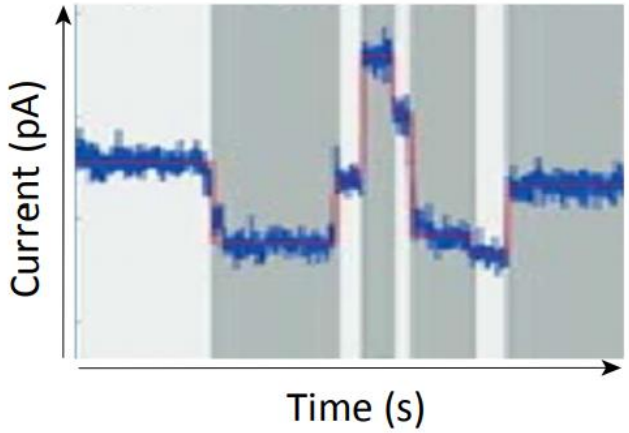
Nanopore Sequencing: ONT

Only one strand DNA passes through the pore.

As the DNA or RNA translocates through the pore, current shifts are recorded in real time







doi.org/10.1016/j.tig.2018.05.008

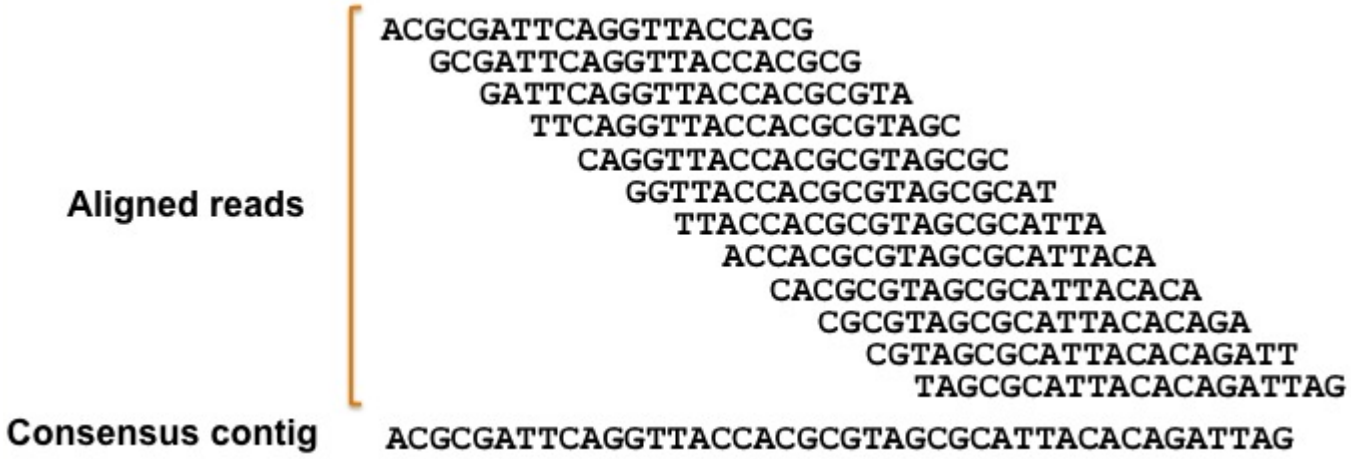
Third generation sequencing: Long reads sequencing

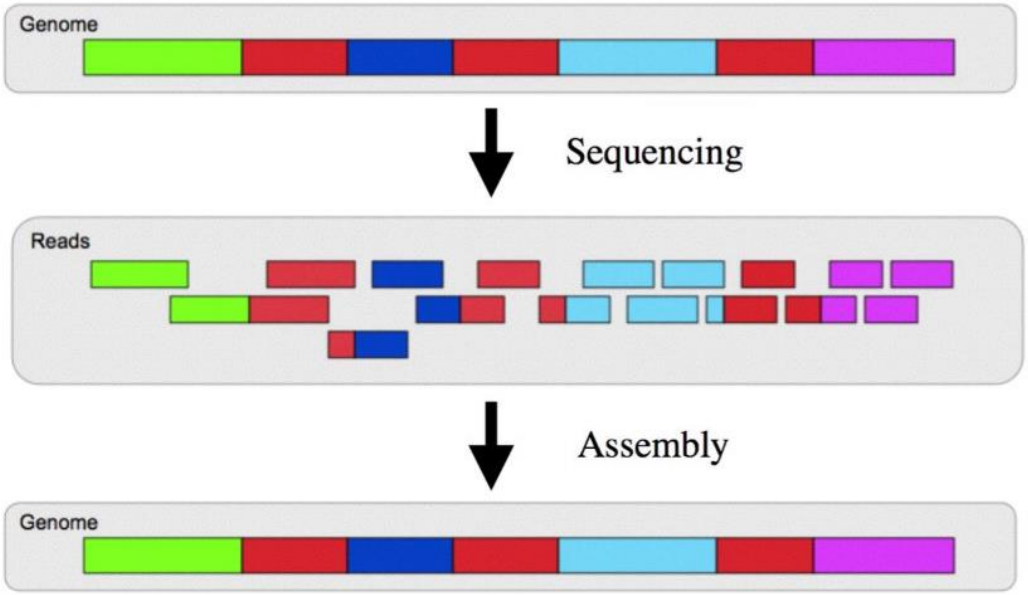
Nanopore Sequencing: ONT

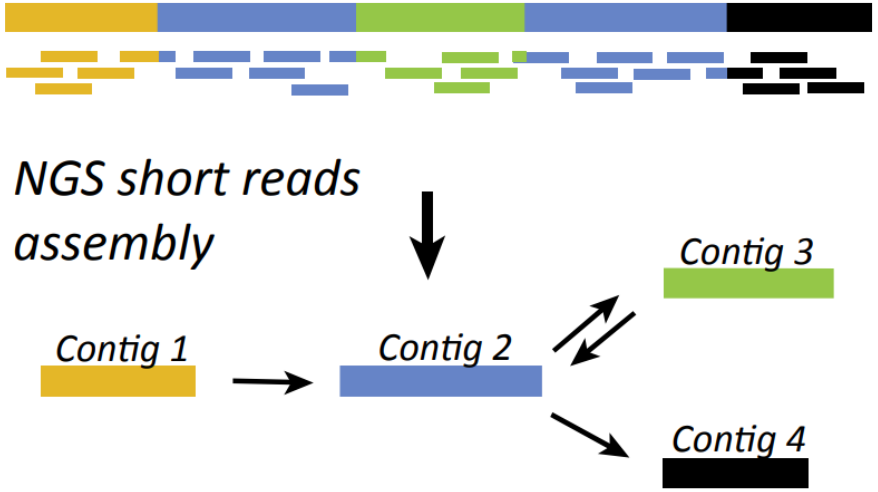
Only one strand DNA passes through the pore.

As the DNA or RNA translocates through the pore, current shifts are recorded in real time

Reads assembly

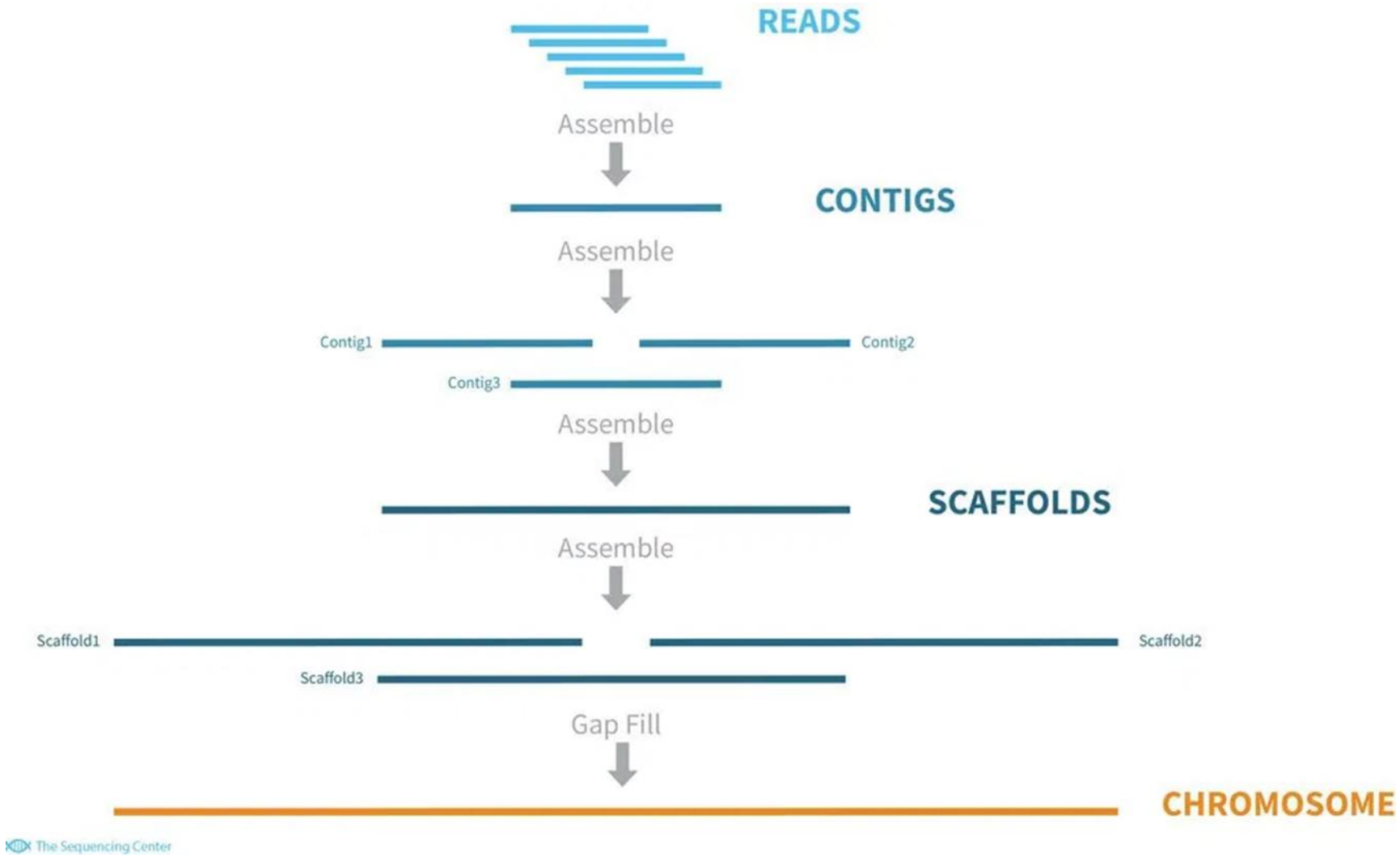






doi.org/10.1016/j.tig.2018.05.008
https://www.youtube.com/watch?v=5wvGapmA5zM

Reads assembly



https://thesequencingcenter.com/knowledge-base/de-novo-assembly/

The pathogenic fungus *Cryphonectria parasitica* infecting American chestnuts

The pathogenic fungus *Cryphonectria parasitica* infecting American chestnuts

They inserted into the tree's genome a wheat gene that codes for an enzyme called oxalate oxidase, or OxO.
It breaks down the oxalic acid the pathogen releases, which is what kills the trees.

EXCLUSIVE: Chinese scientists are creating CRISPR babies

A daring effort is under way to create the first children whose DNA has been tailored using gene editing.

They planned to eliminate a gene called *CCR5* in hopes of rendering the offspring resistant to HIV, smallpox, and cholera.

He claims to have disabled a gene called *CCR5*, which encodes a protein that allows HIV to enter cells.

He responded that his trial was not just for these few patients, but for the millions of children suffering from HIV all over the world.

“I feel proud, actually,” said He.

พระราชบัญญัติ

สัตว์เพื่องานทางวิทยาศาสตร์

พ.ศ. ๒๕๕๘

ประกาศกระทรวงสาธารณสุข

เรื่อง หลักเกณฑ์ วิธีการ และเงื่อนไขที่หน่วยงานตามมาตรา ๒๘ ต้องปฏิบัติ
และการจัดให้มีคณะกรรมการควบคุมความปลอดภัยทางชีวภาพ

พ.ศ. ๒๕๖๐