MUNI
SCI

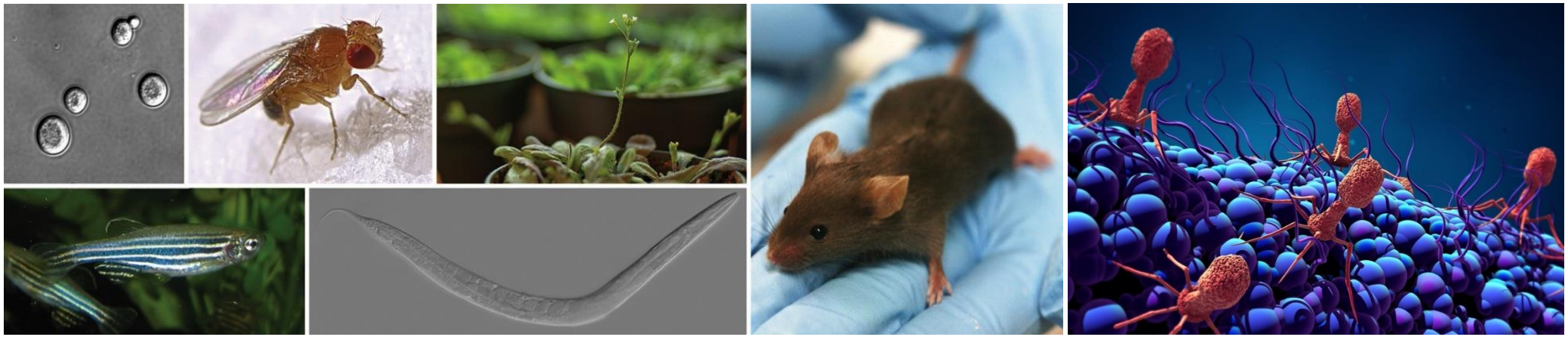
GENE TECHNOLOGIES

Model organisms:

Model organisms used in biotechnology - bacteria (*E. coli*), yeasts (*Pichia, Saccharomyces*) and fungi (*Penicillium*), *Caenorhabditis elegans* (nematode), *Drosophila melanogaster*, *Danio rerio* (Zebra fish), house mouse, animal cell cultures, *Arabidopsis thaliana*, viruses (bacteriophages, retroviruses).

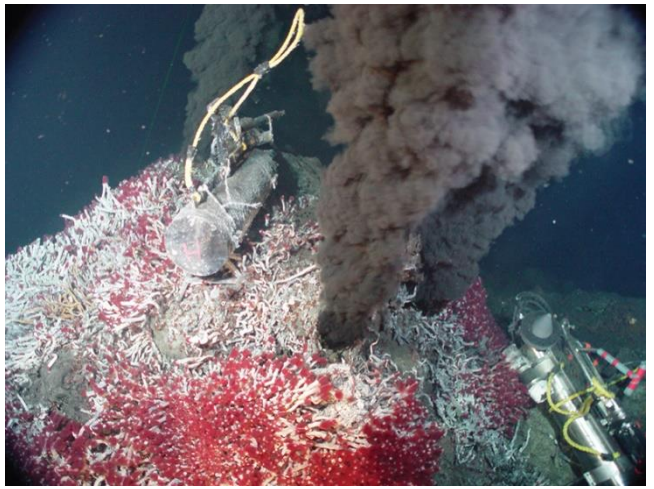
Model Organisms

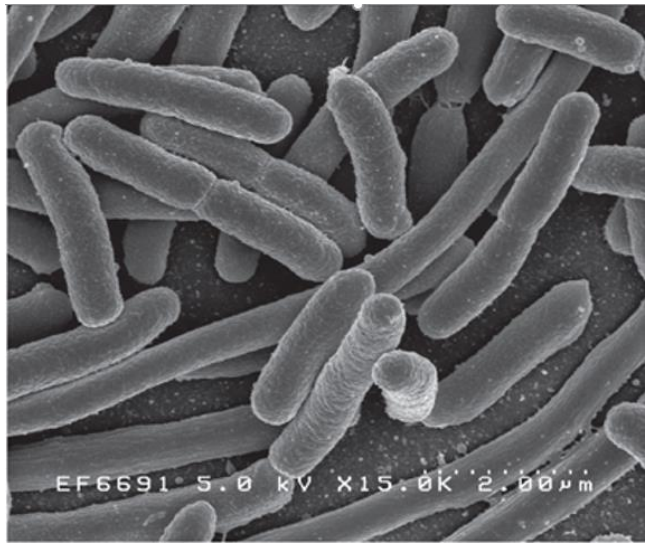
- DNA is found in all living organisms and viruses
- Only a fraction of so-called model organisms are studied in detail
- In model organisms, we now know the complete genome
- We use model organisms:
  - as a model for studying similar organisms
  - in a wide range of biotechnological processes



Bacteria

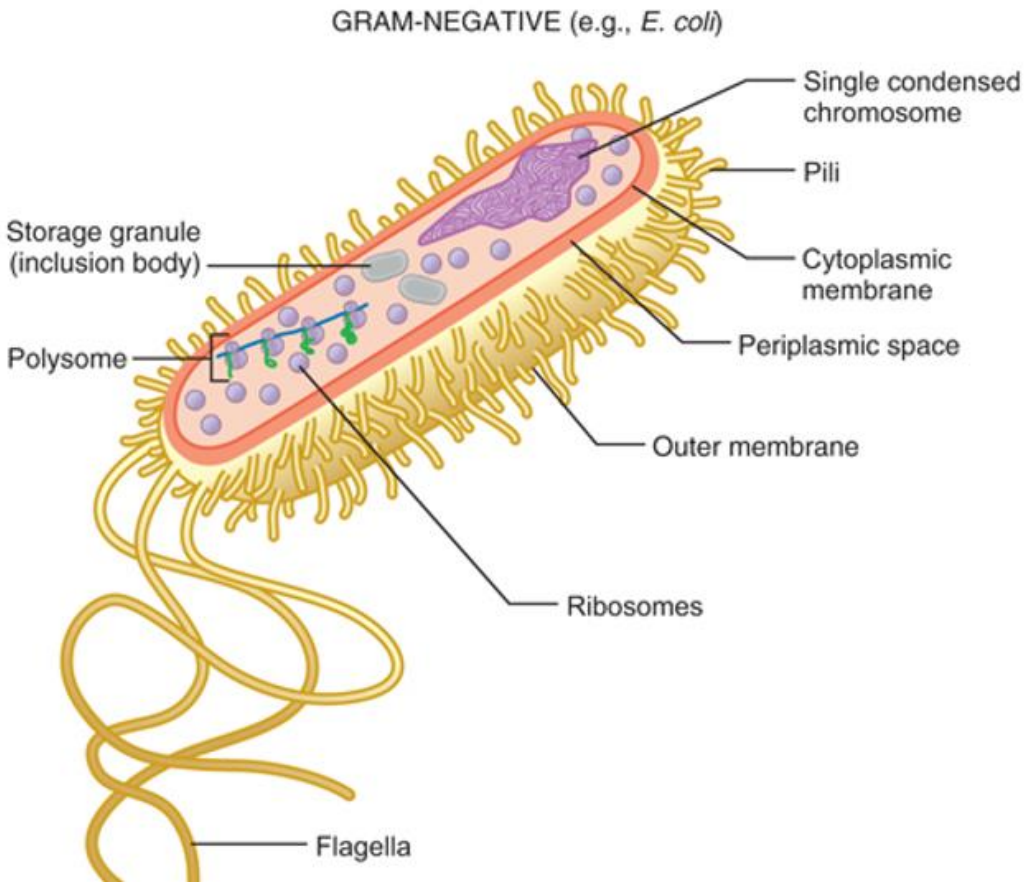
- Master of model organisms
- Makes up approx. 50% of all living organisms (5 x 10^{30})
- Ability to survive in extreme conditions -temperature (*Thermus aquaticus*), pH (*Acidothiobacillus*)
- *Escherichia coli* is the most commonly used:
  - Gram-negative rod
  - has about 10 flagella and thousands of pili on its surface
  - most strains are harmless
  - *E. coli* O157:H7 - two toxins responsible for bloody diarrhea



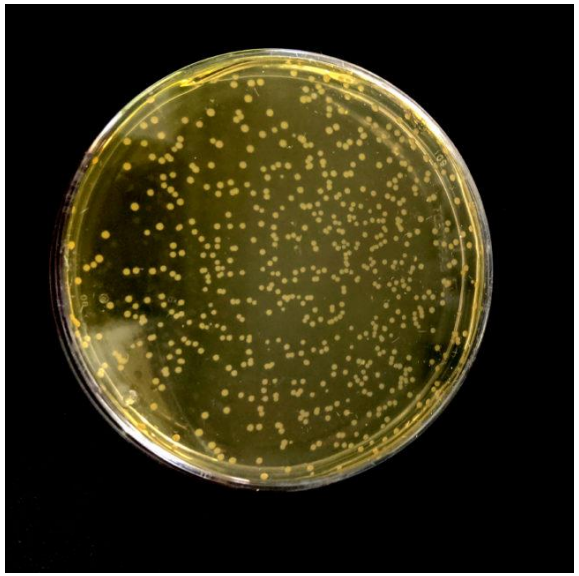


Clark and Pazdernik, 2016

Bacteria

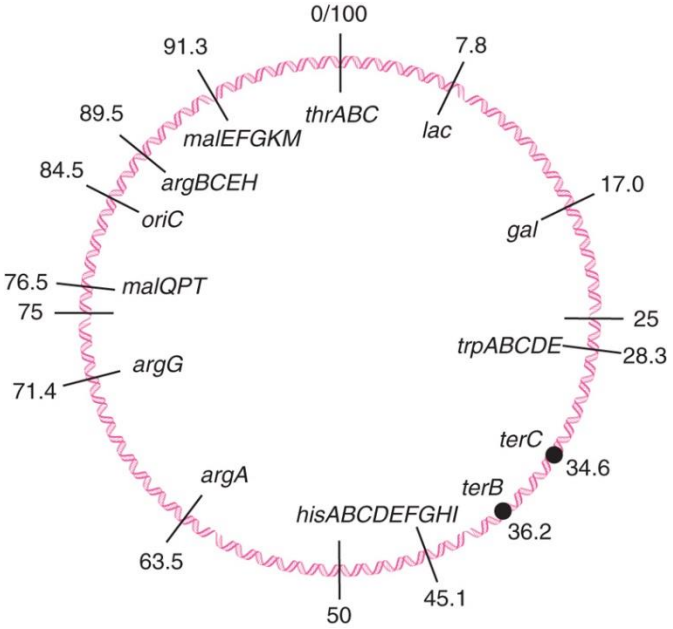


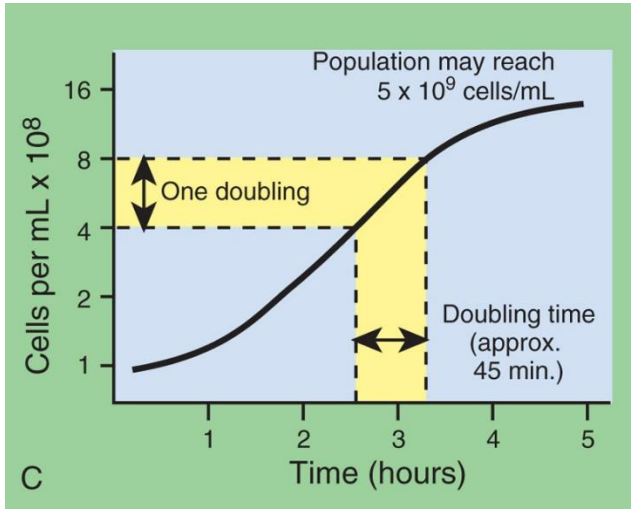
Clark and Pazdernik, 2016



E. coli

- Rapid growth of culture
- Can only grow in a medium containing mineral salts and sugar
- Liquid culture will last for weeks in the refrigerator
- Can be frozen at -70°C for up to 20 years
- Can grow under both aerobic and anaerobic conditions
- Has one circular chromosome containing about 4000 genes

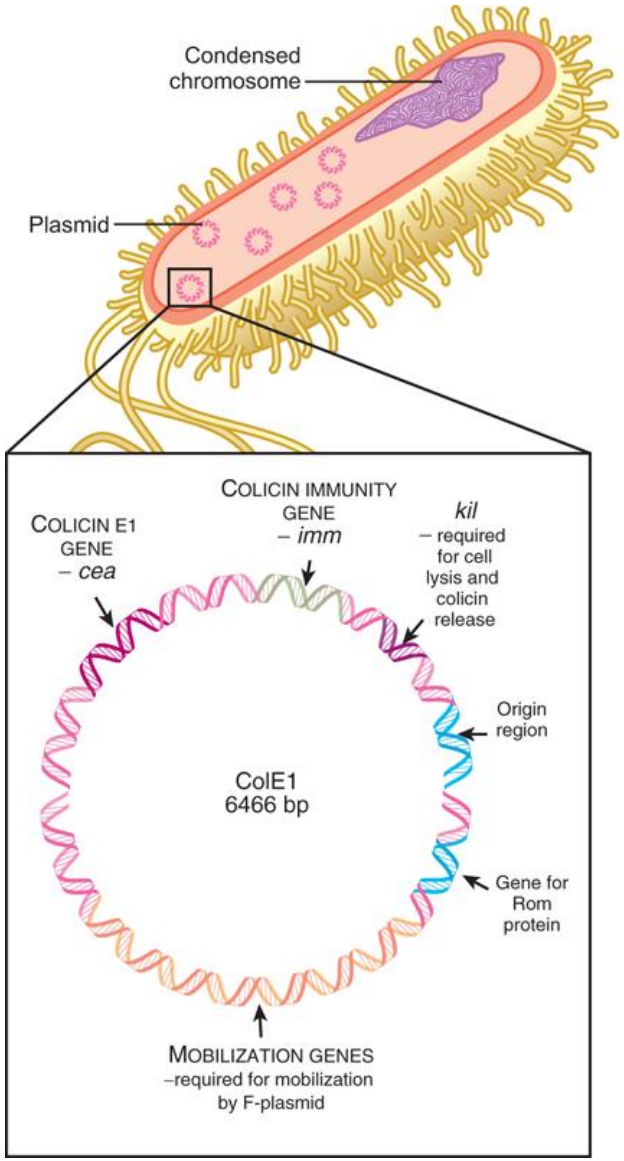




Clark and Pazdernik, 2016

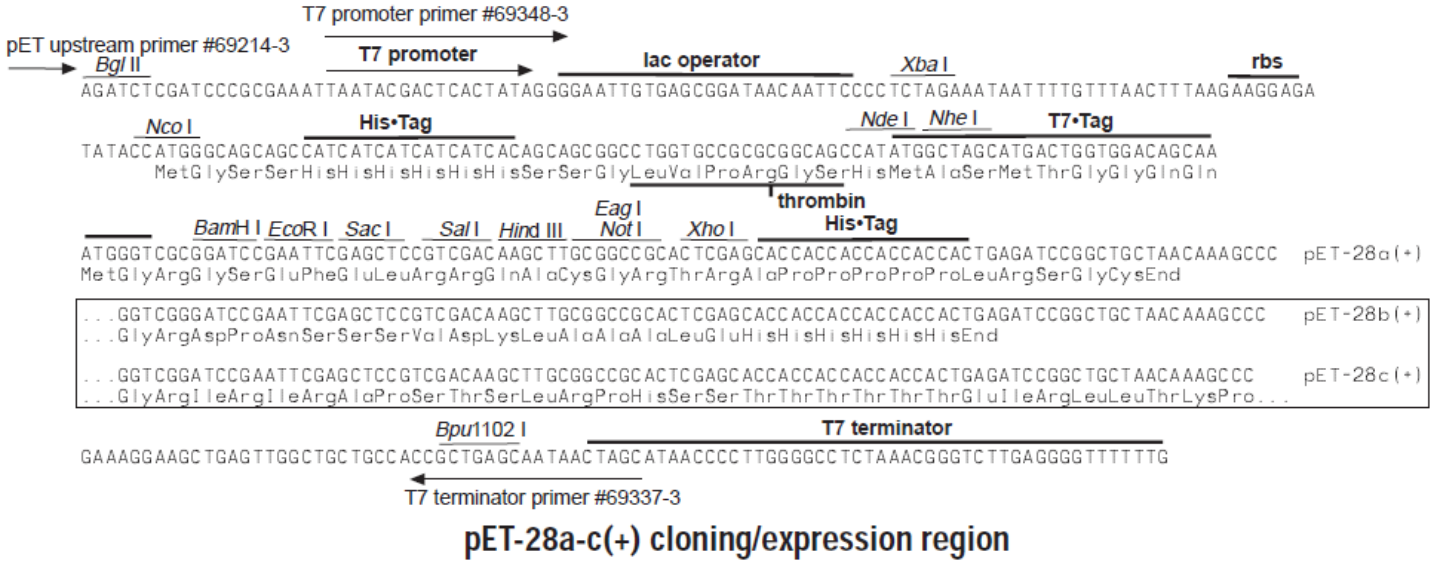
Plasmids

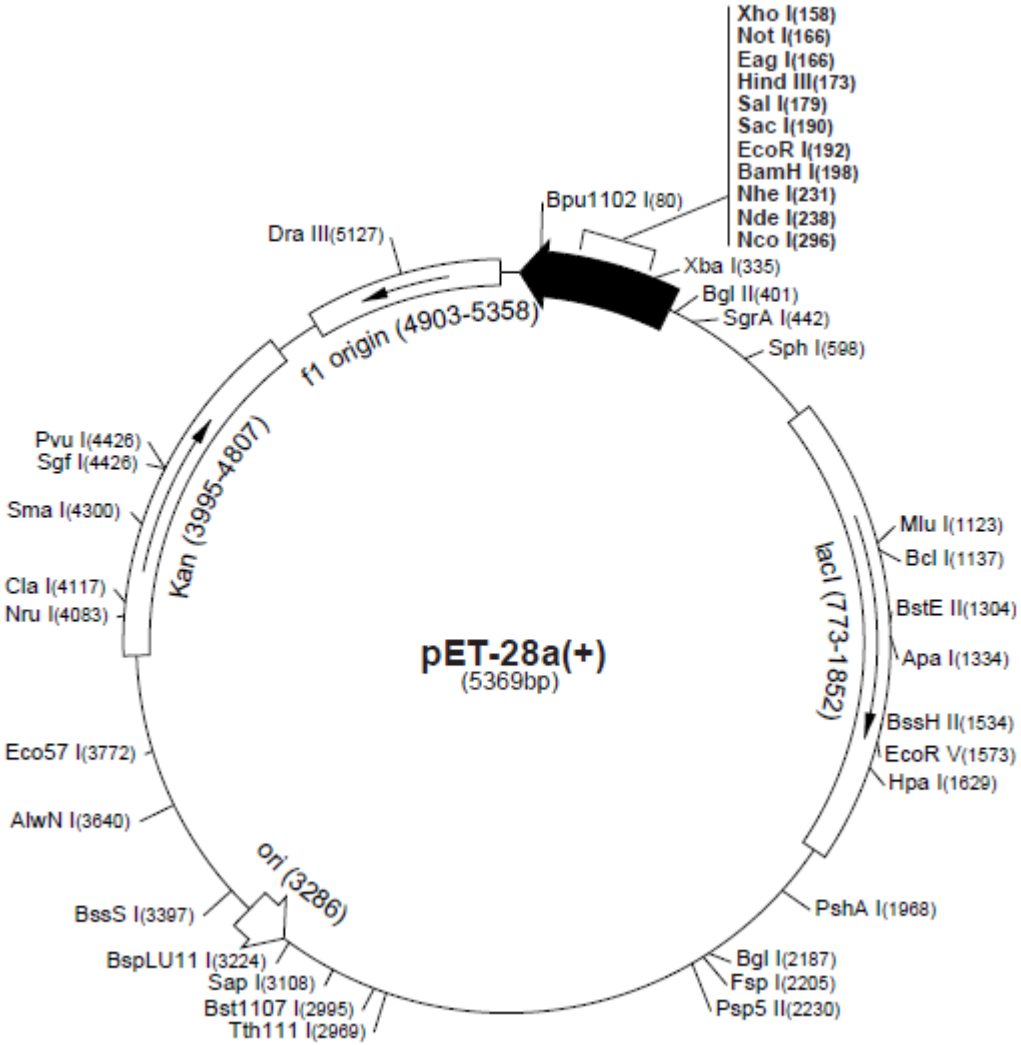
- Survival strategy requires cooperation with other organisms
- A number of bacteria secrete toxins called bacteriocins
- *E. coli* produces so-called colicins (E1, M) - perforation of the plasma membrane, DNA/RNA degradation
- The bacteria's immune proteins neutralise the effect of the toxins
- The ability to produce colicins is due to the presence of plasmids (ori site)
- These plasmids have been modified for biotechnological purposes



Clark and Pazdernik, 2016

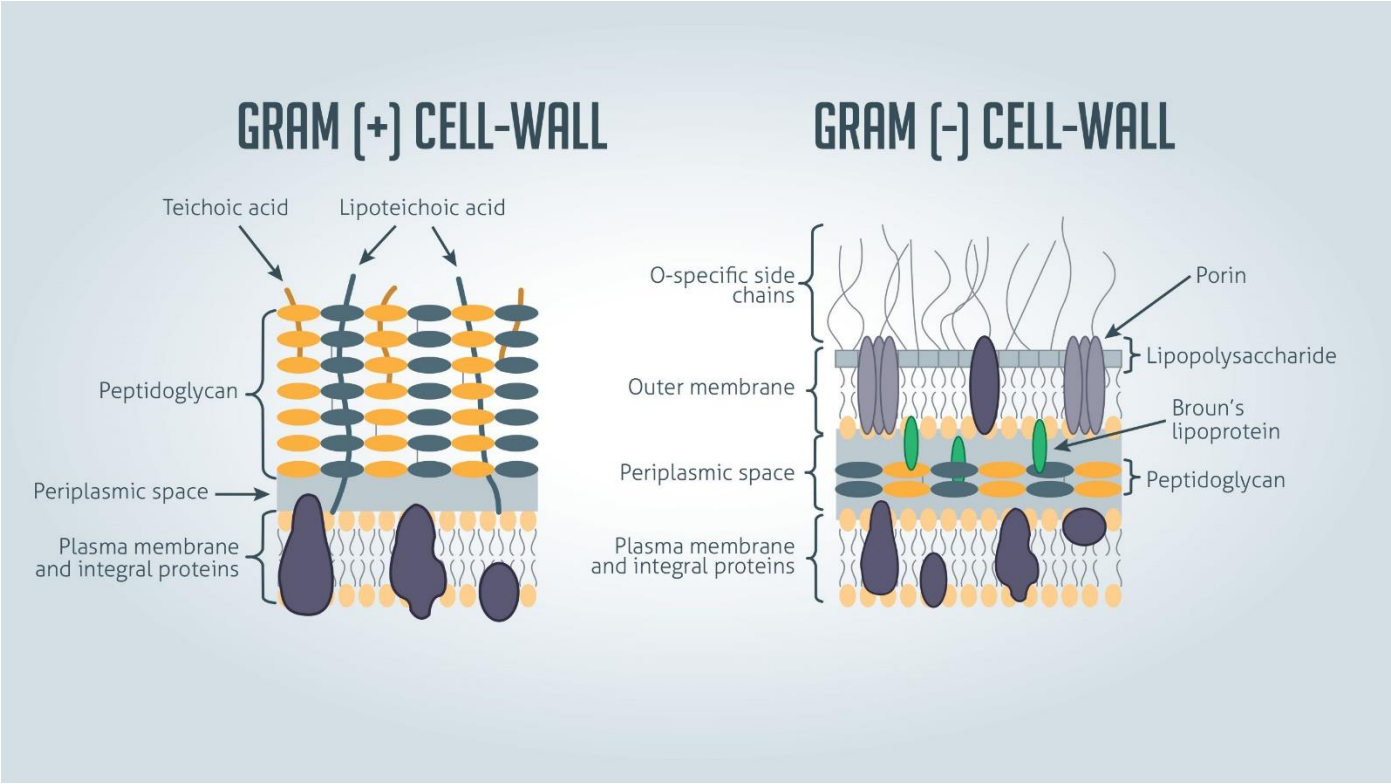
Plasmid pET28





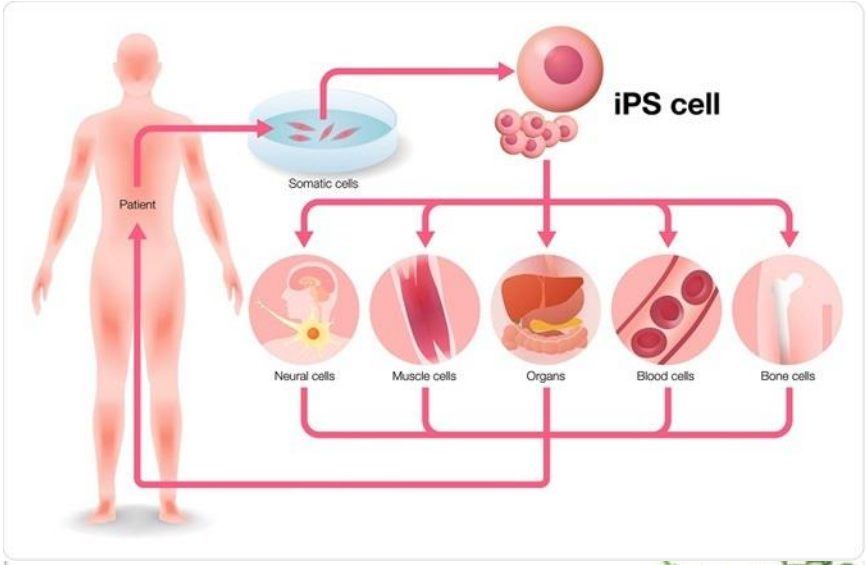
Bacteria in Biotechnology

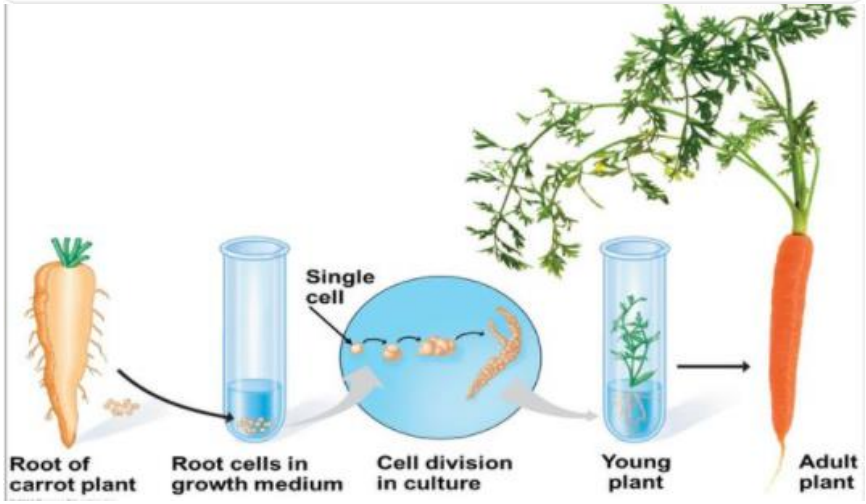
- *Bacillus subtillis* - production of proteases and amylases
- *Pseudomonas putida* – the ability to degrade a range of aromatic compounds
- *Streptomyces coelicolor* - degrades cellulose and chitin, production of a range of antibiotics (Clorobiocin, Undecylprodigiosin, Actinorhodin)
- *Corynebacterium glutamicum* - production of L-glutamate and L-lysine
- *Streptococcus zooepidemicus* - production of hyaluronic acid



Eukaryots

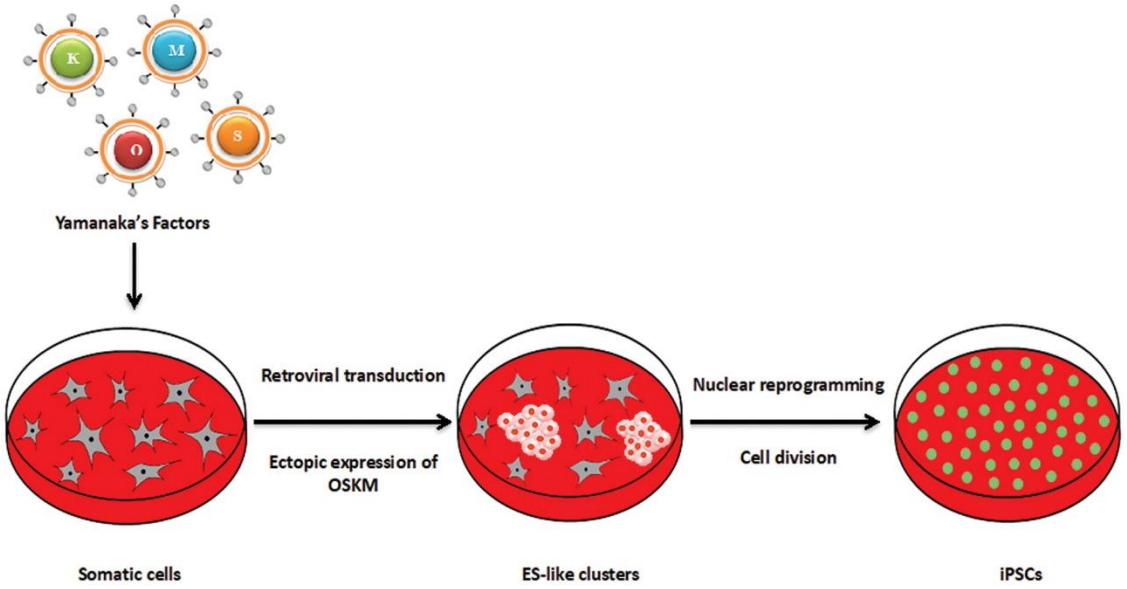
- The entire line of eukaryotes is diploid (two copies of each chromosome)
- In contrast, a whole range of plants are polyploid (wheat = hexaploid, tomato = tetraploid)
- In animals, there is a difference in germ and somatic cells
  - diploid germ lines give rise to haploid gametes (eggs and sperm)
  - somatic cells are diploid
  - somatic mutations are transmitted within the organism
  - somatic mutations are not transmitted to offspring
- In most plants, cells are totipotent
- In animals, only stem cells carry this property



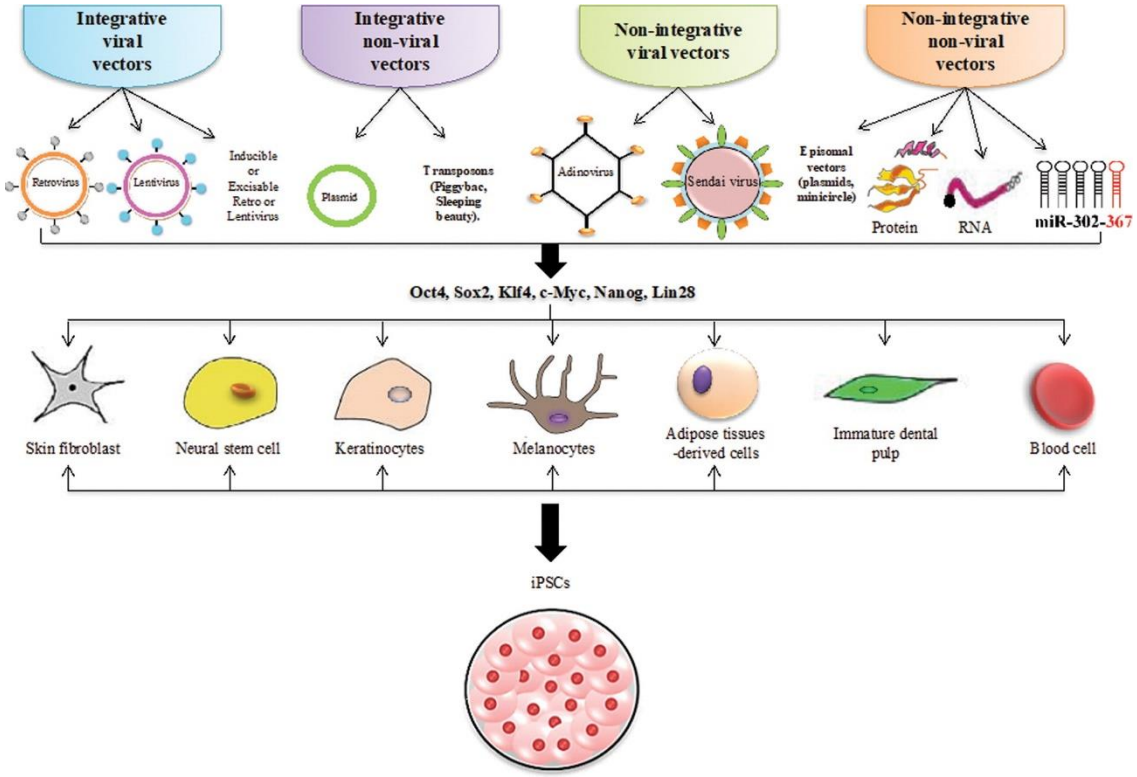


iPSC (induced Pluripotent Stem Cell)

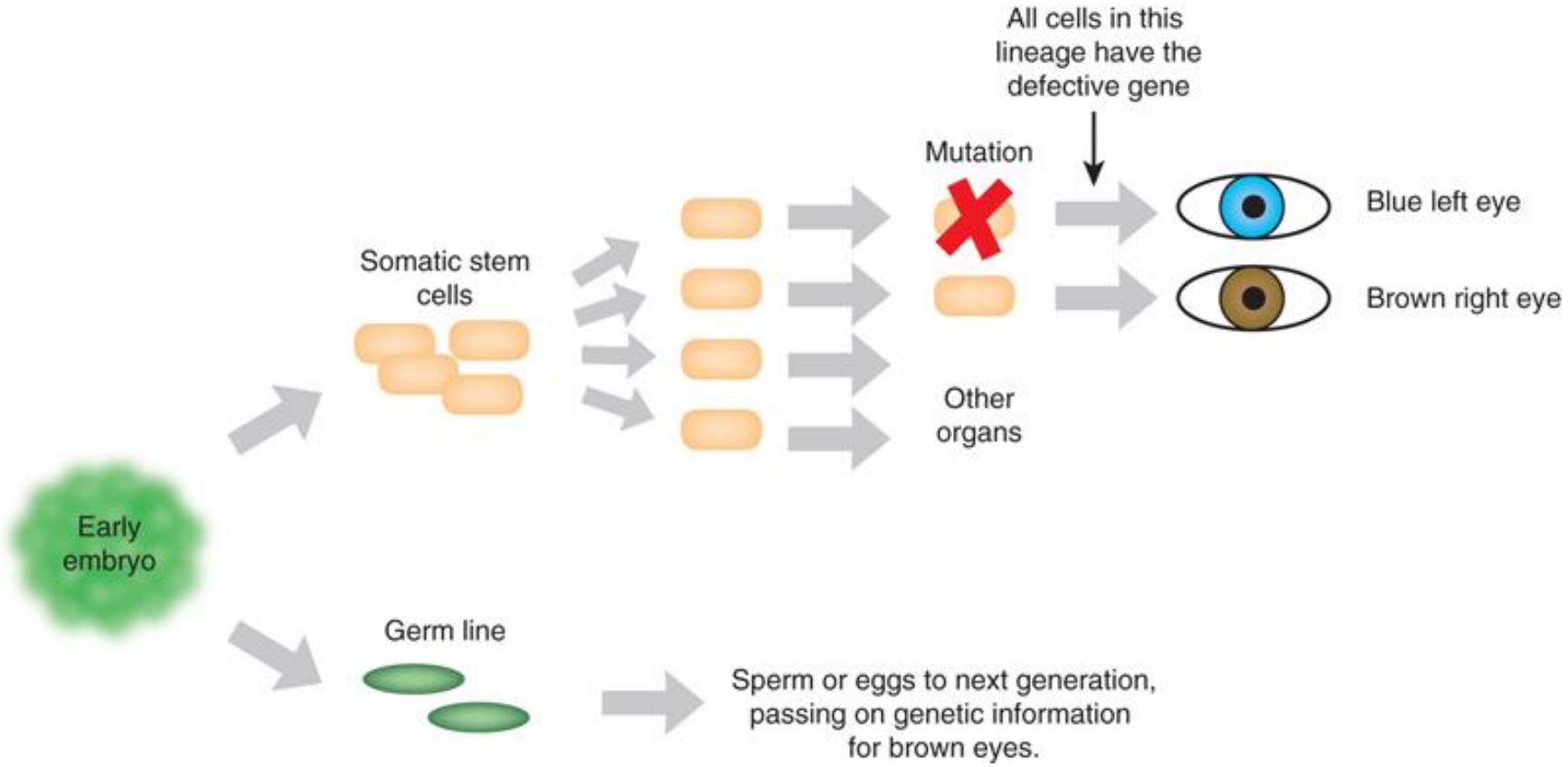
- Method first described in Takahashi and Yamanaka (2006) for induction of iPSCs from fibroblasts
- Requires the expression of 4 transcription factors - octamer-binding transcription factor 3/4 (Oct3/4), SRY (sex determining region Y)-box 2 (Sox2), Krüppel-like factor 4 (Klf4) and cellular-Myelocytomatosis (c-Myc) (OSKM).



Abbar et al., 2020



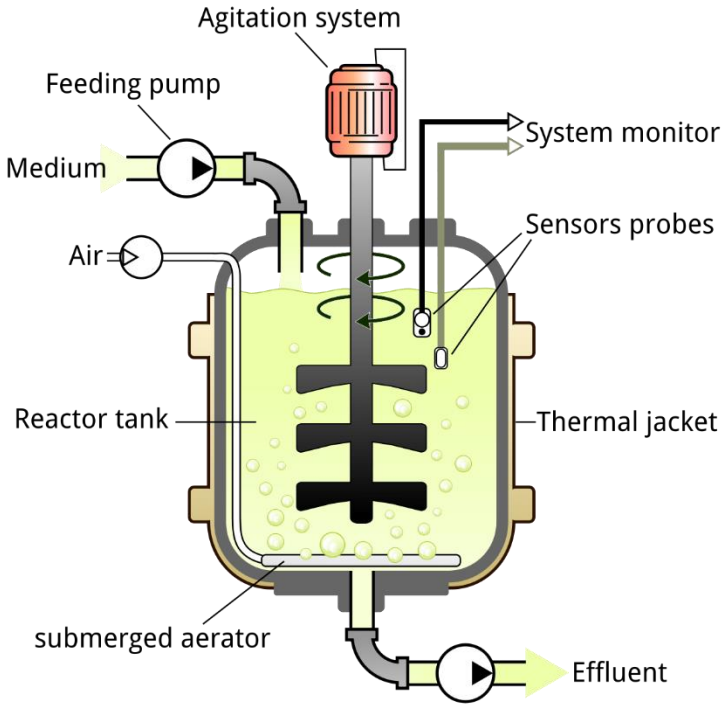
Somatic mutations



Clark and Pazdernik, 2016

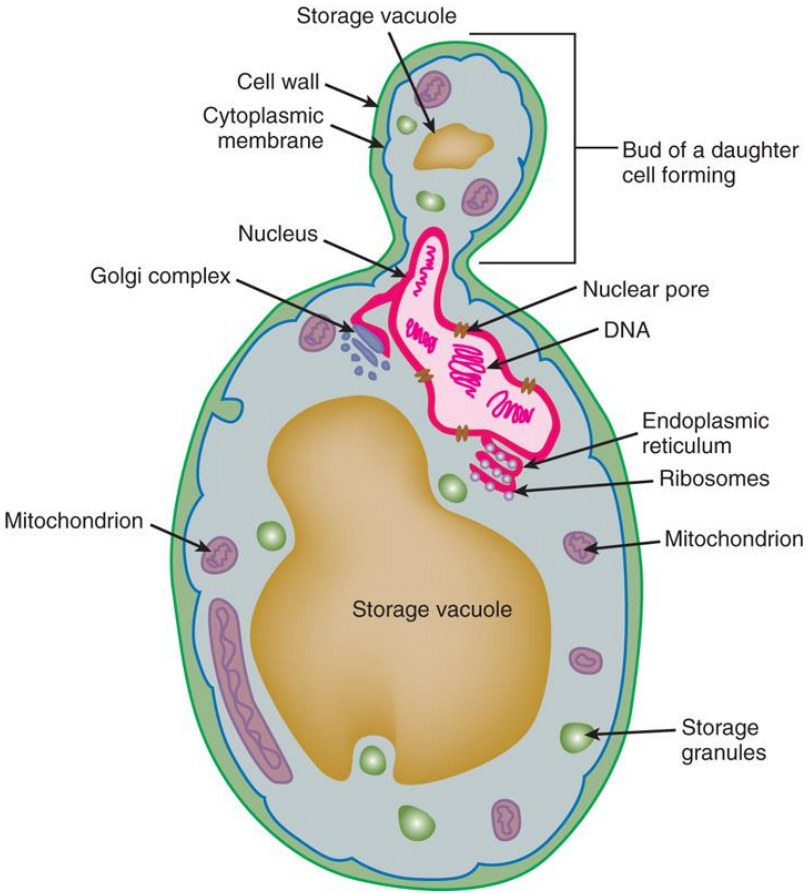
Yeasts and Fungi

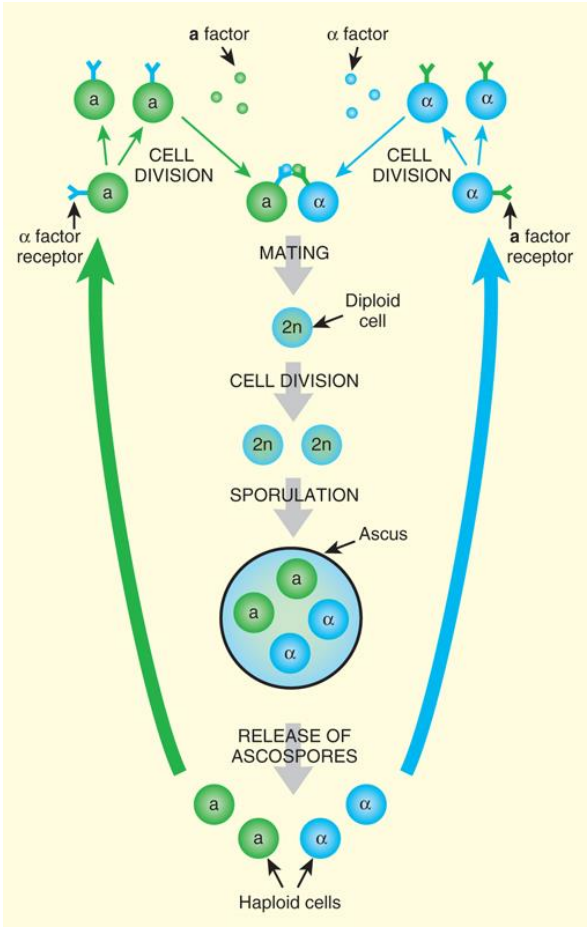
- Fungi are traditionally used in biotechnology - *Penicillium roqueforti* (Roquefort), *P. candidum*, *caseicolum* and *camembertri* (Camembert), *Aspergillus oryzae* (soy sauce), *Penicillium notatum* (Penicillin), *Aspergillus niger* (citric acid)
- Usually cultivated in bioreactors
- Yeasts have the advantages of both bacteria and eukaryotes
- The most commonly used yeast is *Saccharomyces cerevisiae*
- The yeast genome is separated by a nuclear membrane
- *S. cerevisiae* has 16 chromosomes containing telomeres and centromeres
- Some yeasts have extrachromosomal elements, the so-called 2.micron circle.



Yeasts

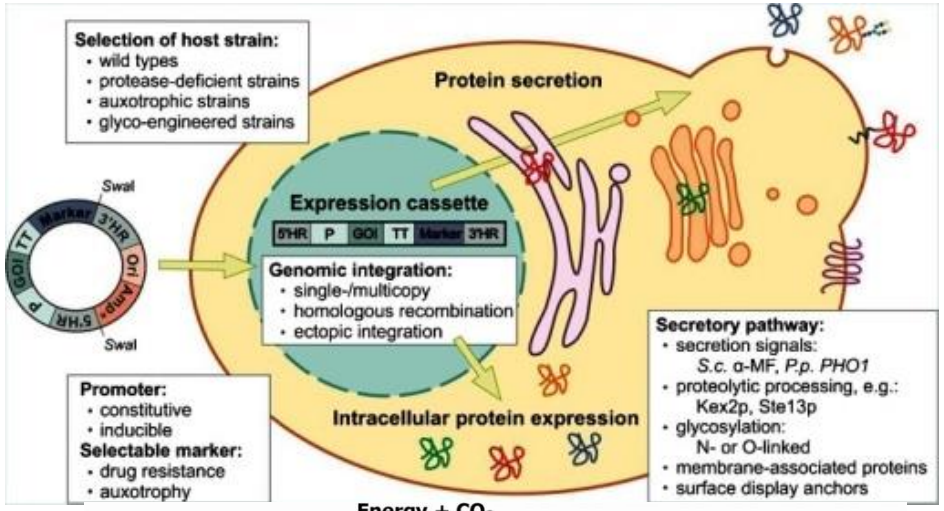
- Yeasts multiply by budding
- Budding produces identical cells - division by mitosis
- Yeasts have diploid and haploid phases within the life cycle
- Under critical conditions, yeast undergo meiosis - formation of haploid spores, called ascospores in the ascus)
- Under favorable conditions, spores germinate and conjugate to form diploid cells
- In yeast, conjugation can only occur between two different mating types (a, α)

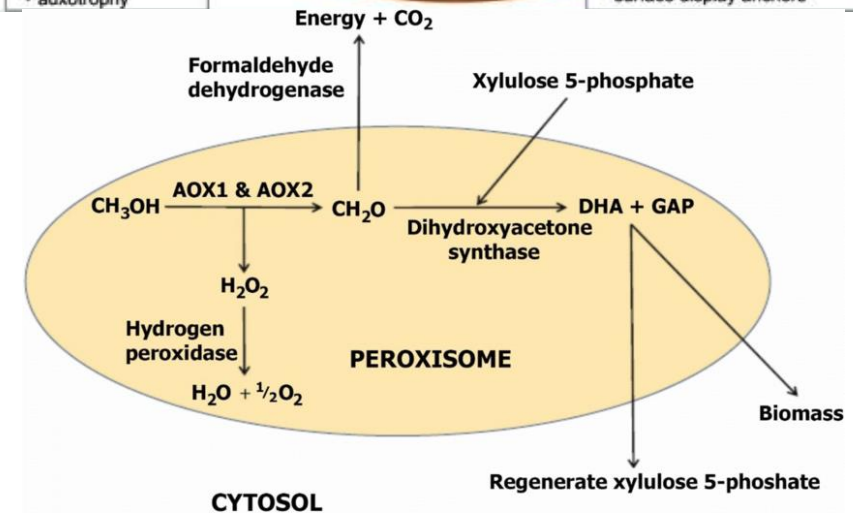


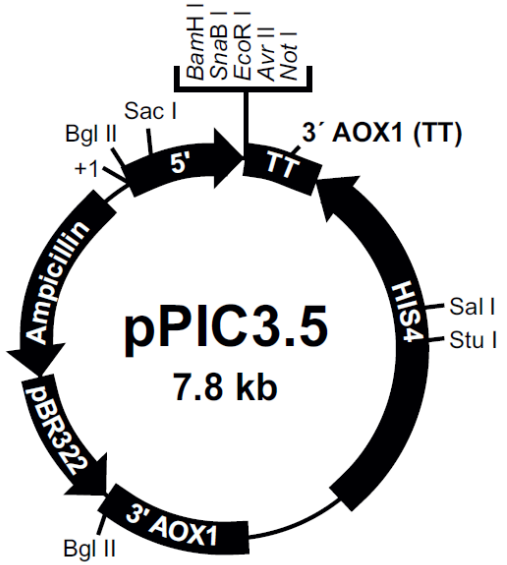


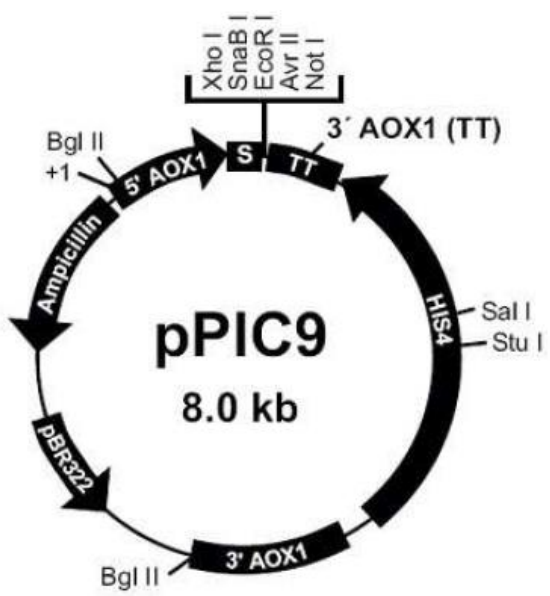
Clark and Pazdernik, 2016

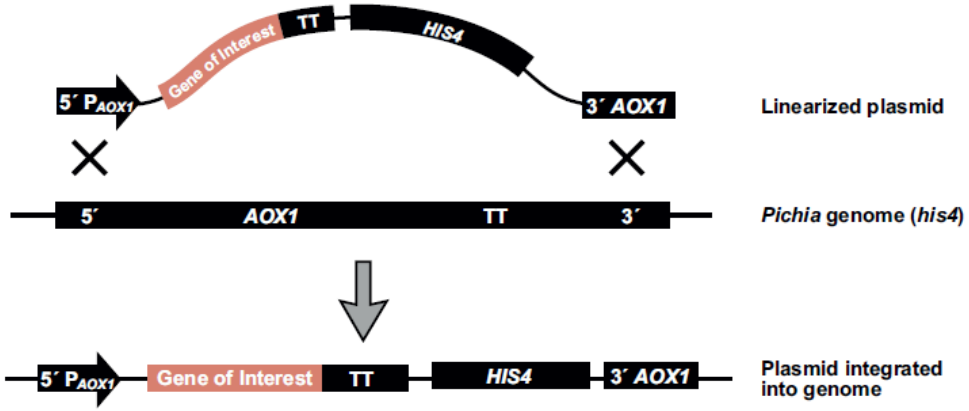
Pichia pastoris





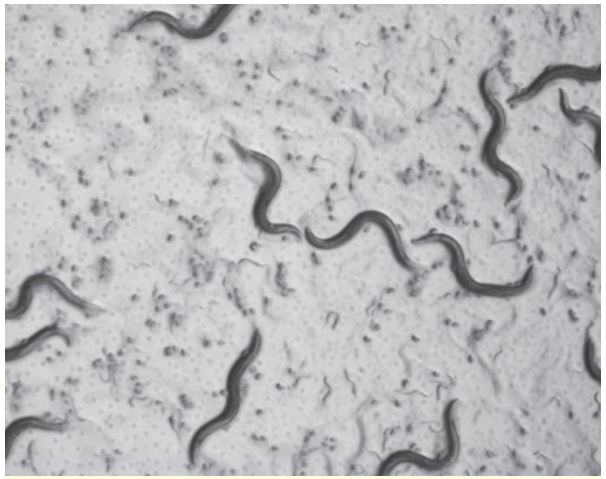


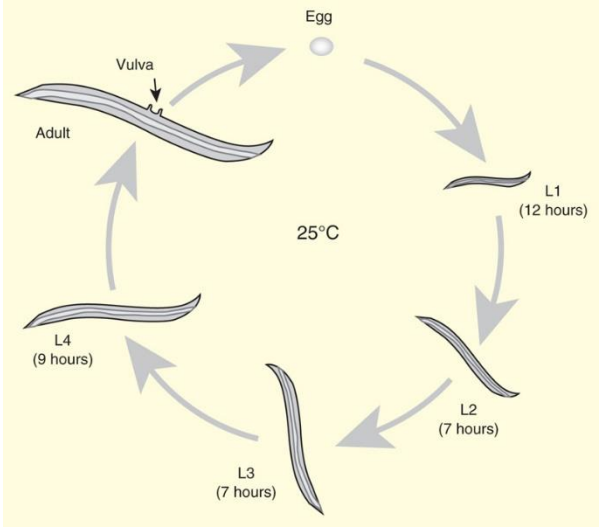




Caenorhabditis elegans

- Small nematode (nematodes) living in soil with mainly root vegetables
- It has two sexes - 99.9% hermaphrodite (self-fertilizing) and 0.1% male
- Body consists of a simple tube covered with a cuticle
- Inside the body - 959 somatic cells including about 300 neurons
- The head has a variety of sensory organs (taste, smell, temperature, touch)
- Body is translucent = easy to use fluorescence techniques, generation cycle 3 days
- RNA interference performed for the first time - ideal tool for reverse genetics
- First known complete genome of a multicellular organism (100 Mbp)

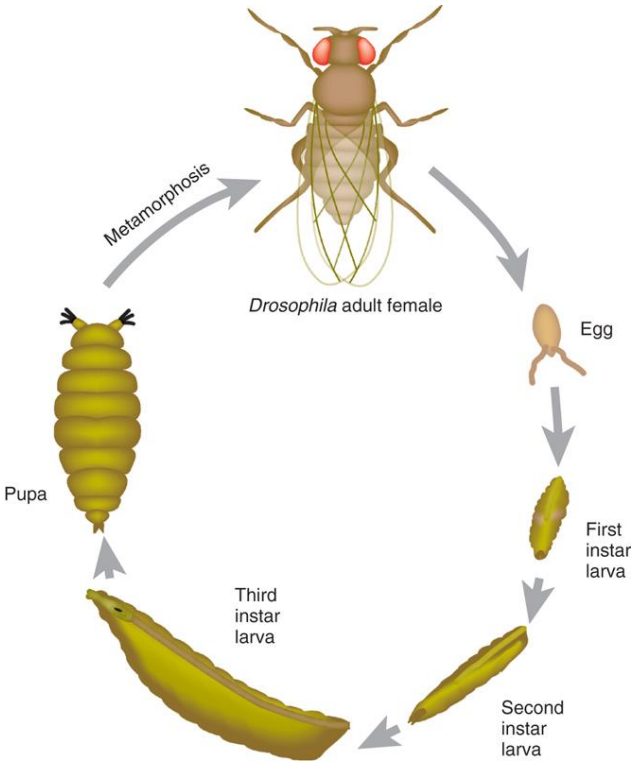


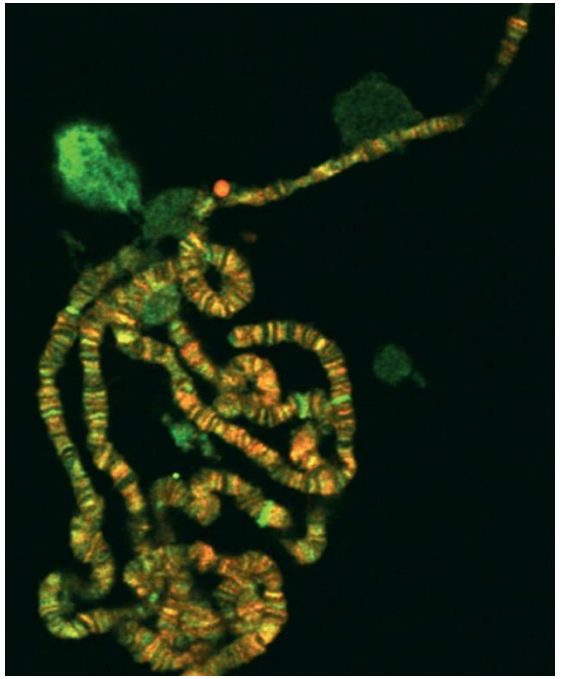


https://www.hsph.harvard.edu/mair-lab/c-elegans/

Drosophila melanogaster (fruit fly)

- A widely consumed organism in genetic studies
- Easy to grow, 2-week life cycle
- Egg hatches into a larva (24h), several larval stages after adult
- Many mutants available - identification of genes involved in development (homology with humans)
- Genome is 165 Mb - 3 pairs of autosomal and X/Y chromosomes
- Polytene chromosomes during rapid larval development





Clark and Pazdernik, 2016

Danio rerio (Zebra fish)

- A simple model vertebrate used in molecular biology
- Easy to grow and propagate in aquaria, availability of a wide range of mutants
- Embryonic development outside the mother's body, development from a single cell to an individual takes 24 hours
- Embryo is translucent - easy to monitor the effect of mutations on development
- Genome contains 25 pairs of chromosomes (1700 Mb), 70% of protein-coding genes in humans have orthologs in *Danio*
- Model for studying a range of human diseases
- Embryos are often used for screening new drugs

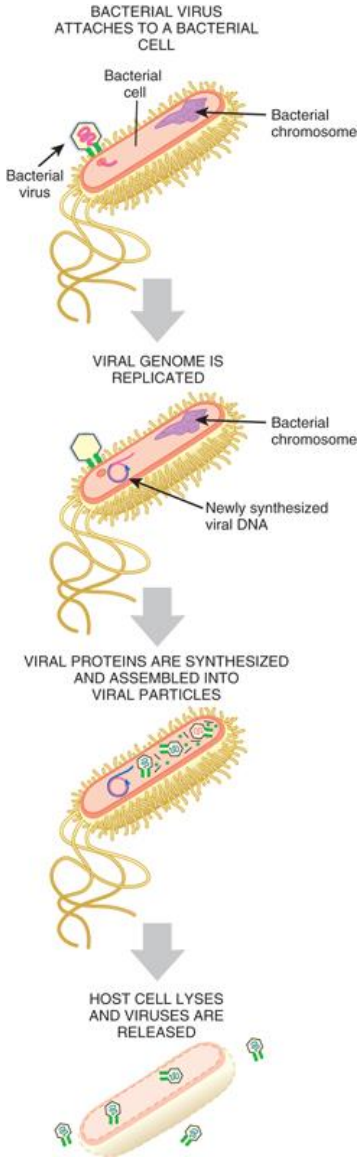
https://theconversation.com/animals-in-research-zebrafish-13804

Arabidopsis thaliana

- The most widely used model organism in plant genetics and molecular biology
- Similar response to stress factors and diseases as economic crops
- Many of the genes responsible for development and reproduction are identical to those of economic crops
- Easy to grow, space-saving, generation time 6-10 weeks, many seeds
- Can be maintained in a haploid state
- Small genome - five chromosomes (125 Mb), 25 000 genes
  - Rice (430 Mb), 40-50 thousand genes
  - wheat (17 Gb), tomato (950 Mb), tobacco (4.5 Gb)

Viruses

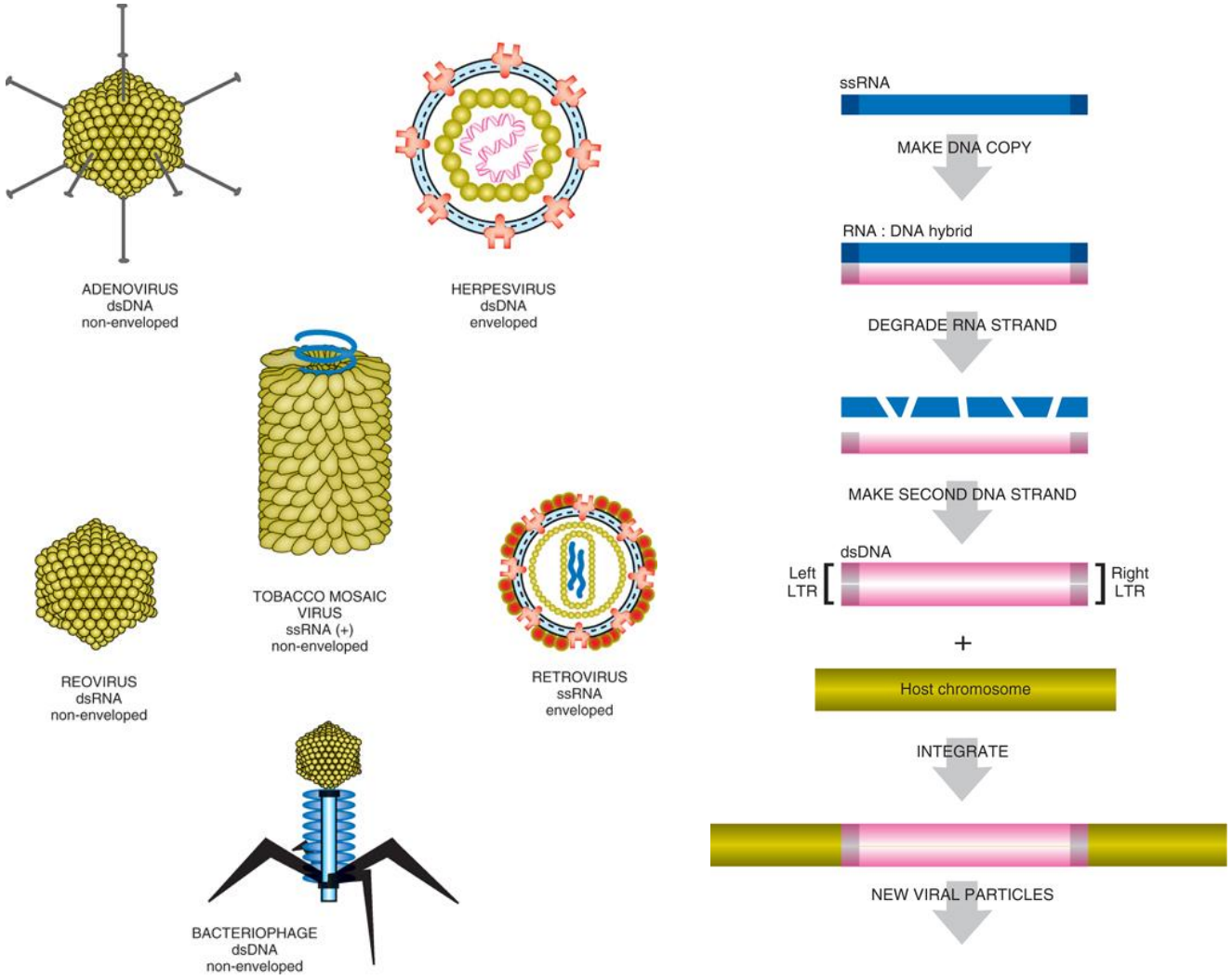
- Entities at the edge of the definition of life, pathogens attacking host cells
- Consists of a protein envelope called a capsid that encases the DNA/RNA genome
- Found in all living organisms (bacteria, plants, animals)
- Bacterial viruses = bacteriophages (phages)
  - attach to the host
  - entry of the viral genome
  - replication of the viral genome
  - production of new viral proteins
  - assembly of a new viral particle
  - release of virions from the host
- Many viruses go through a latent phase - lysogeny in bacteria
- Integration of the virion into the host genome often occurs - provirus (prophage) formation



Clark and Pazdernik, 2016

Viruses

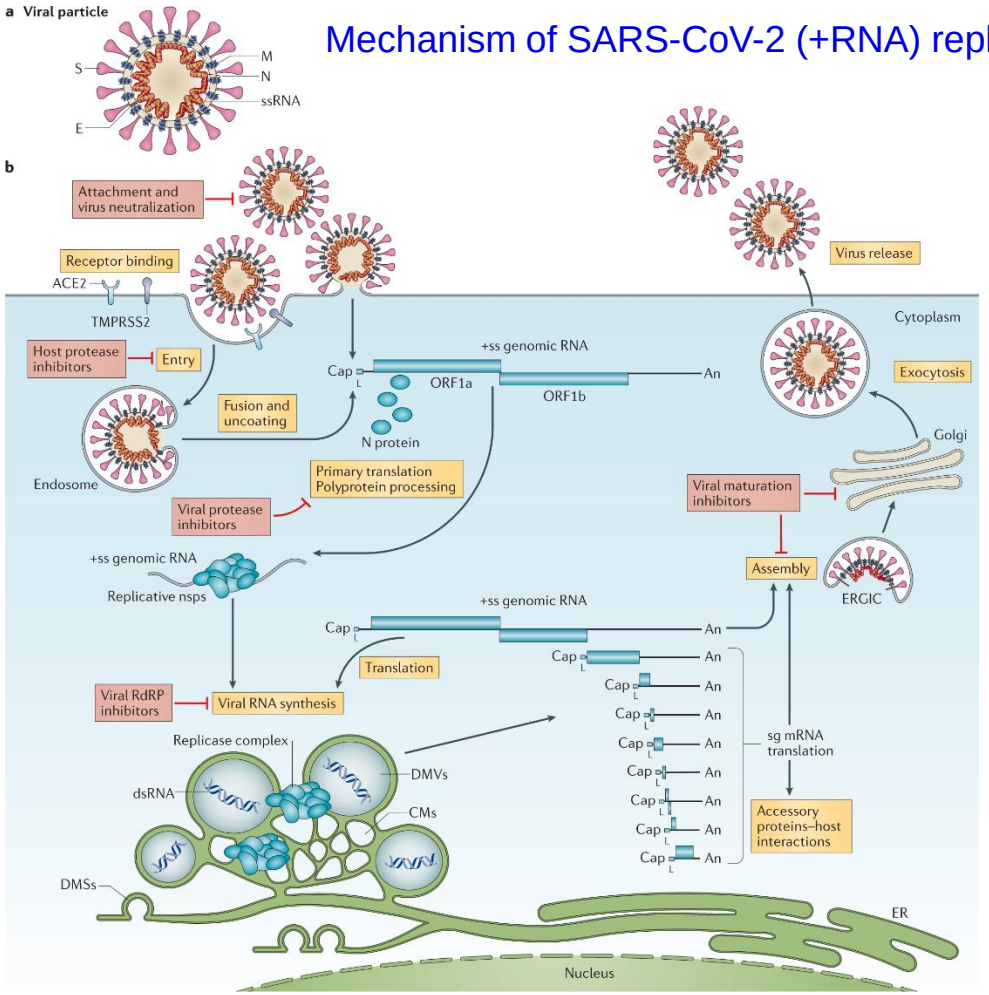
- We can divide based on the shape of the capsid (spherical, complex, fibrous)
- Complex = bacteriophages (T4, P1, Mu)
- ssRNA viruses have a positive (+) or negative (-) genome
- Retroviruses contain reverse transcriptase (transcription of RNA to DNA), integrate into the genome using long terminal repeats (LTRs)



Clark and Pazdernik, 2016

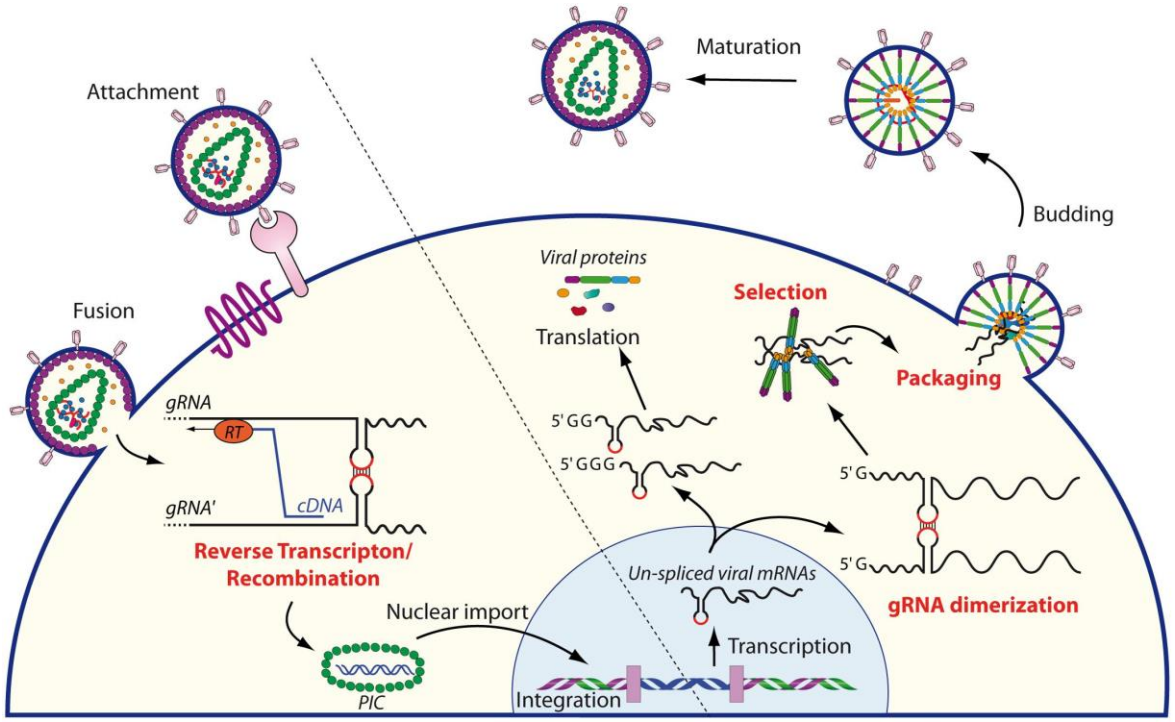
The Life Cycle of RNA Viruses

Mechanism of SARS-CoV-2 (+RNA) replication



V'kovski et al. 2021

Mechanisms of retroviruses replication



Dubois et al. 2018

MUNI
SCI

GENE TECHNOLOGIES

Manipulation of DNA, RNA, and proteins

Cell fractionation, isolation of proteins and nucleic acids.

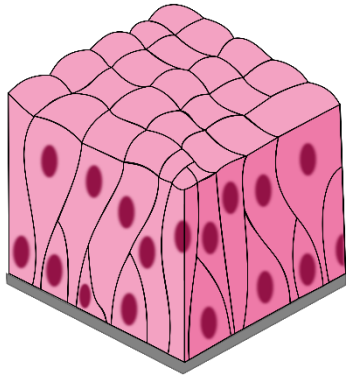
Isolation of DNA and RNA

- **Different types of samples = different strategies**

Plants tissues

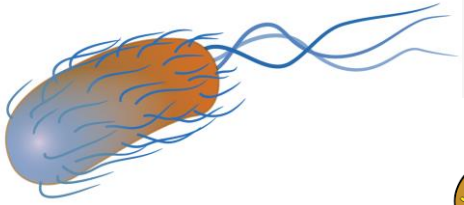
- seeds, leaves, roots, wood

Animal and human tissues



- blood, brain tissue, heart tissue, liver tissue
- stool, urine, swabs from the urethra, throat, vagina, rectum, conjunctiva, cerebrospinal fluid

Bacteria and environmental samples



- gram positive and negative bacteria, yeasts, fungi
- food (cheese, meat, egg, milk)
- soil, water, manure

Desintegration of sample

- Soft animal tissues - lysis at 50-60°C by Proteinase K
- Proteinase K
  - digests preferentially after hydrophobic amino acids
  - active in a wide range of temperatures (20 and 60°C), pH and buffers
  - activity is stimulated when up to 2% SDS or up to 4 M urea are included in the reaction
- Solid animal tissues and plant tissues - must be crushed mechanically
- Microorganisms - grinding with sea sand or garnet beads, lysozyme (G^+)
- Mechanical grinding

https://www.youtube.com/watch?v=Z8UvIQXRJFY
https://www.youtube.com/watch?v=k6mPWPuR8PY
https://youtu.be/OwoUAO7vaJA?list=TLGGIXBeSy4AvBcyODA5MjAyMg

Mechanical grinding

- Liquid nitrogen and mortar and pestle
- Retsch mill, Precellys, Cryomill
- Garnet beads

Lysis buffer

- The goal of lysis buffer is to suppress the activity of nucleic acid-degrading enzymes and to separate proteins from nucleic acids
- EDTA – chelating of Mg^{2+} ions = inhibition of nucleases
- RNAsin – inhibitor of RNAses
- Detergents - sodium and lithium salts of lauryl sulfate or Triton X-100 and Tween20 - nuclease inhibitors and at the same time release the nucleic acid from its binding to the proteins/histones

Deproteinization

- **Phenol** - one of the most effective denaturing agents, but phenol can degrade nucleic acids with repeated use.
- **Chloroform** mixed with isoamyl alcohol – effectively denatures proteins (chloroform denatures proteins and isoamyl alcohol reduces foaming)
- **Guanidine hydrochloride** – breaks the structure of proteins and biologically inactivates them. It can be used to isolate both DNA and RNA.
- **Sodium perchlorate** – removes detergents from extraction solutions by forming their complexes with proteins

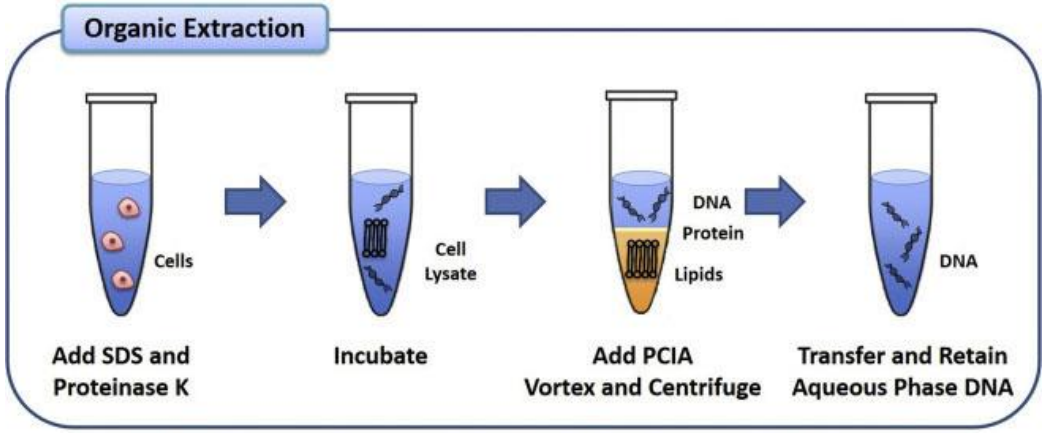
Removing of saccharides

- **Cetrimonium bromide (CTAB)** - can be used to precipitate DNA and RNA, while the saccharide remains in the liquid.
- **Tetraethylammonium bromide (TEAB)** –isolation of RNA from the saccharide of a 50% ethanol solution of TEAB. The saccharides precipitate and the RNA remains in the liquid. The saccharides are removed by centrifugation.
- **2.5 M LiCl** - LiCl precipitation is useful following RNA isolation or in vitro transcription, because RNA is efficiently precipitated, while protein, carbohydrates, and DNA are very inefficiently precipitated or are not precipitated at all

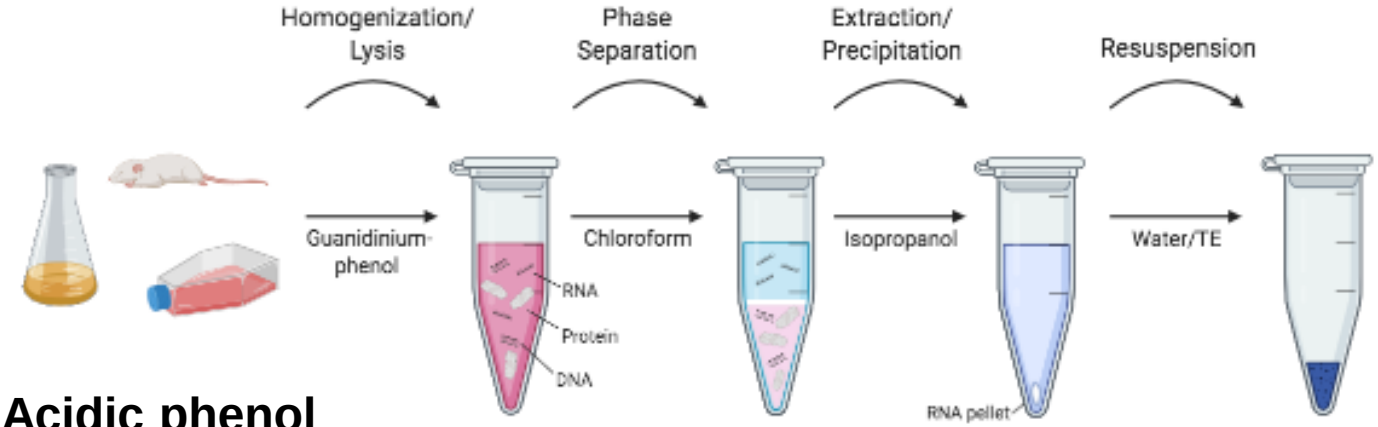
Phenol-chloroform isolation of NA

- The phenol-chloroform extraction method is most often used to isolate NA from plant tissues or enviromental samples and large amounts of DNA from blood.
- A mixture of phenol, chloroform and isoamyl alcohol is added to the sample.
- TriReagent, TRIZOL - A mixture of phenol, chloroform and GuHCl
- Chloroform does not mix with the aqueous solution of the cell lysate, so the mixture is divided into two phases - upper aqueous and lower chloroform. By shaking, the phases are mixed, during which the phenol precipitates the proteins present in the aqueous lysate.
- Using of acidic phenol (pH≈4) – isolation of RNA to upper aqueous phase/DNA in interphase
- Using of basic phenol (pH≈8) – isolation of DNA to upper aqueous phase
- DNA/RNA is precipitated from aqueous phase by isopropanol

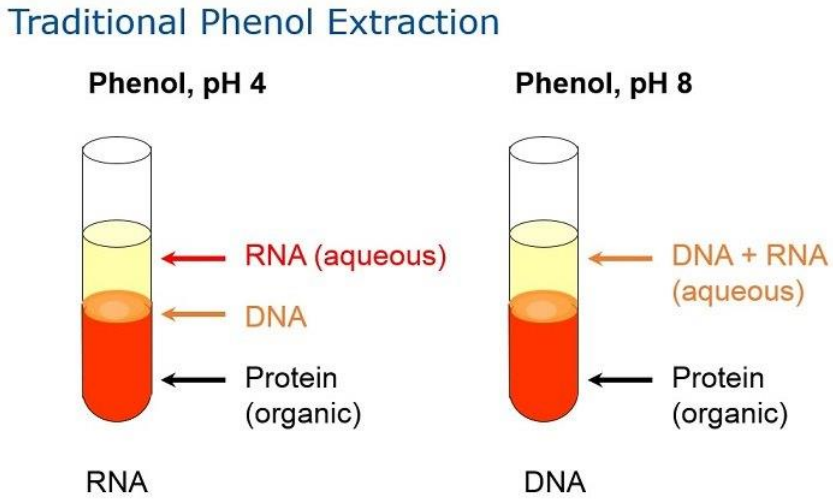
Phenol-chloroform isolation of NA

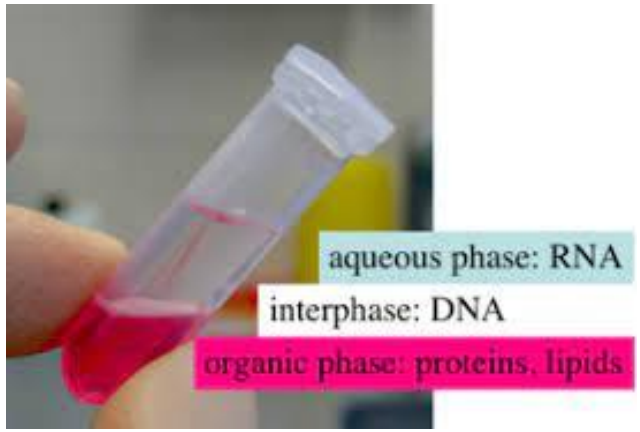


Basic phenol



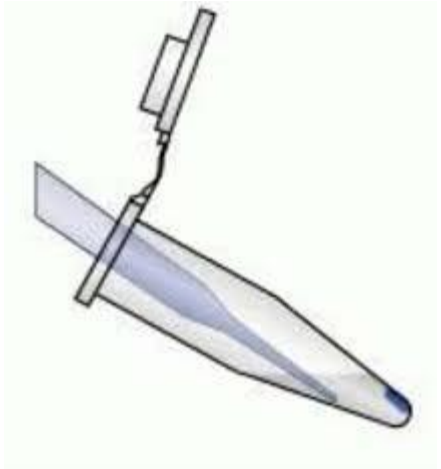
Acidic phenol





NA precipitation

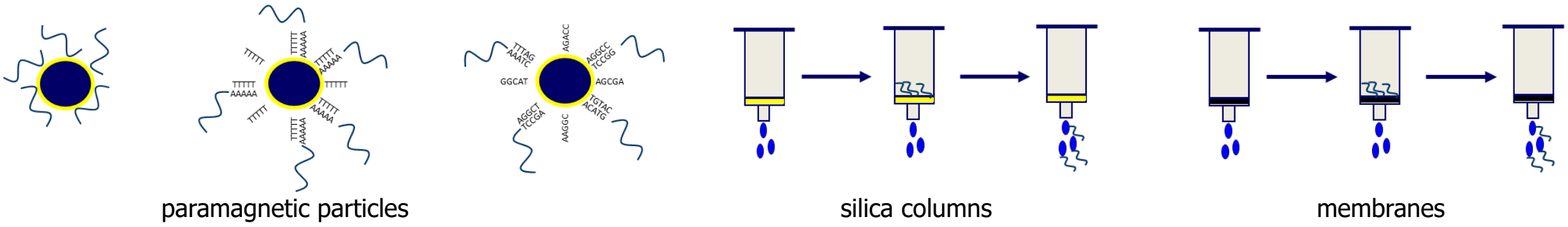
- Precipitation of RNA and DNA can be facilitated by addition of co-precipitant
  - Glycogen, GlycoBlue



GlycoBlue - dye covalently linked to glycogen, a branched chain carbohydrate, which is useful as a nucleic acid coprecipitant.

Isolation of NA using commercial kits

– Types of isolation techniques used by commercial kits:
  - **resins** - bind DNA specifically
  - **membranes** (filters)
  - **silica columns** – specific binding of nucleic acids
  - **paramagnetic particles** with a differently modified surface

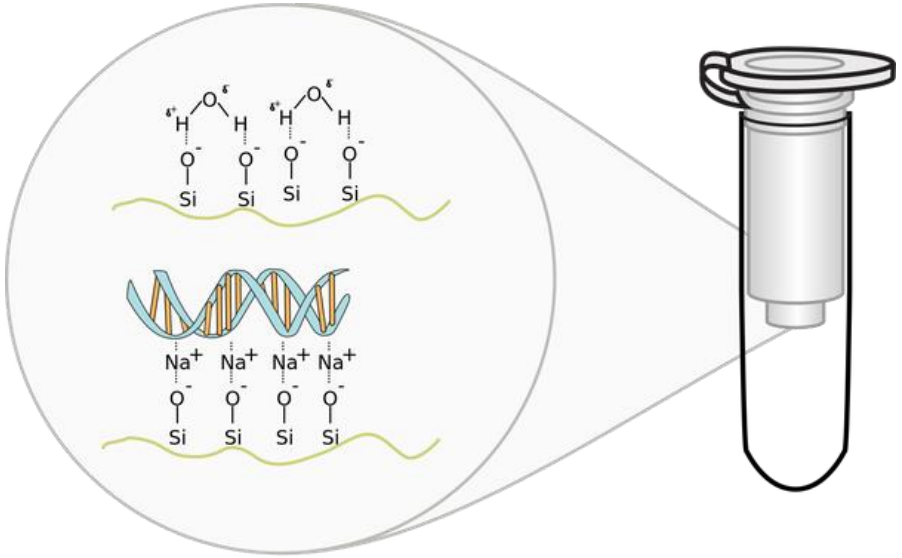


paramagnetic particles

silica columns

membranes

Silica columns



DNA/RNA

contaminants

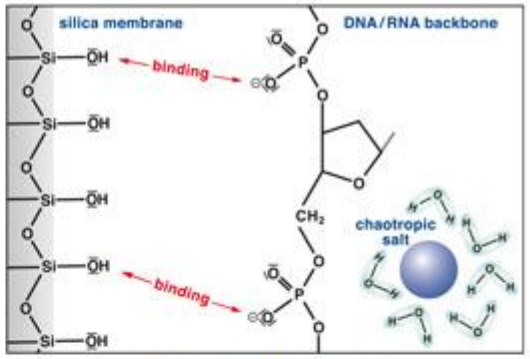
Sample lysis, release of DNA/RNA from cells, tissue, etc.

DNA/RNA is bound to the silica membrane under high-salt conditions

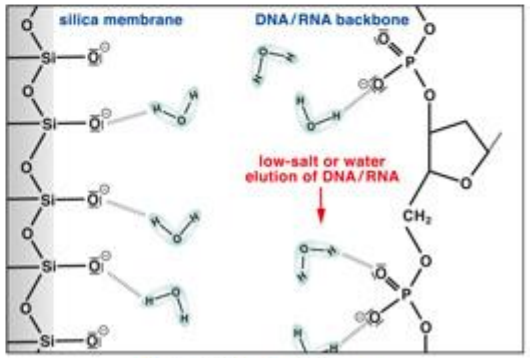
Interaction between DNA/RNA (hydrate shell is reversibly removed by chaotropic salt) and silica membrane

Contaminants are washed away under high-salt and/or ethanolic conditions to keep the DNA/RNA bound to the membrane

DNA/RNA is eluted in low-salt buffer or water, DNA/RNA is ready to use for downstream applications



Principle of binding

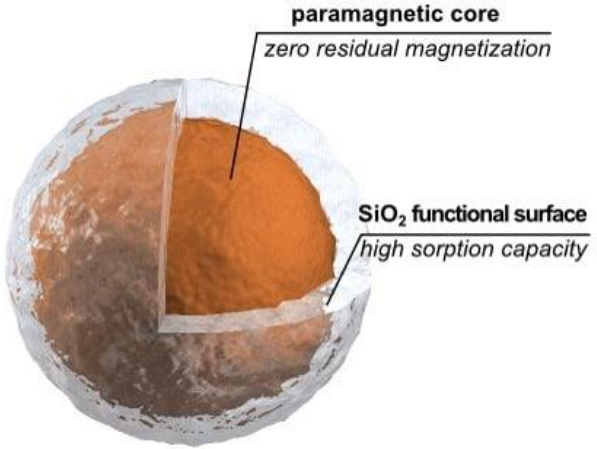


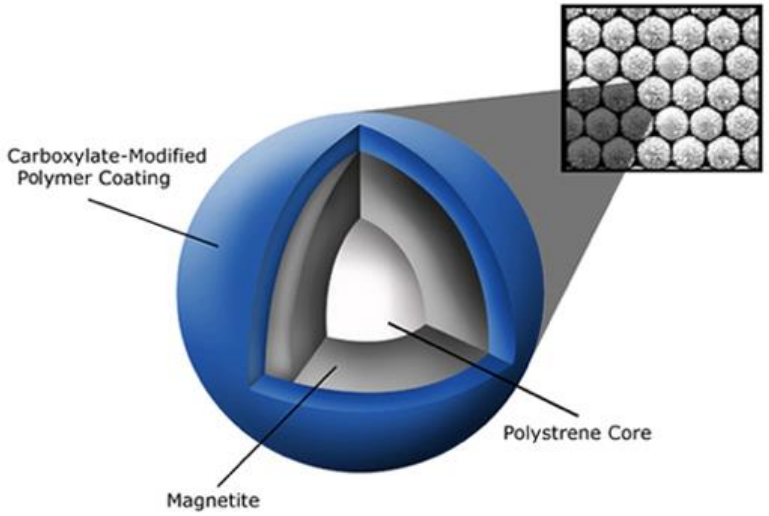
Principle of elution

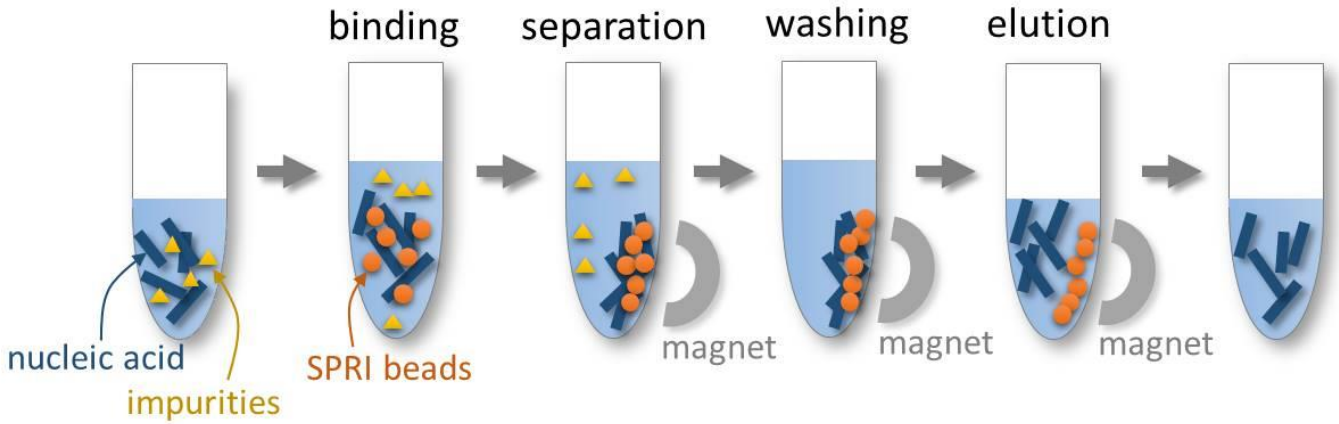
Paramagnetic particles (MPs)

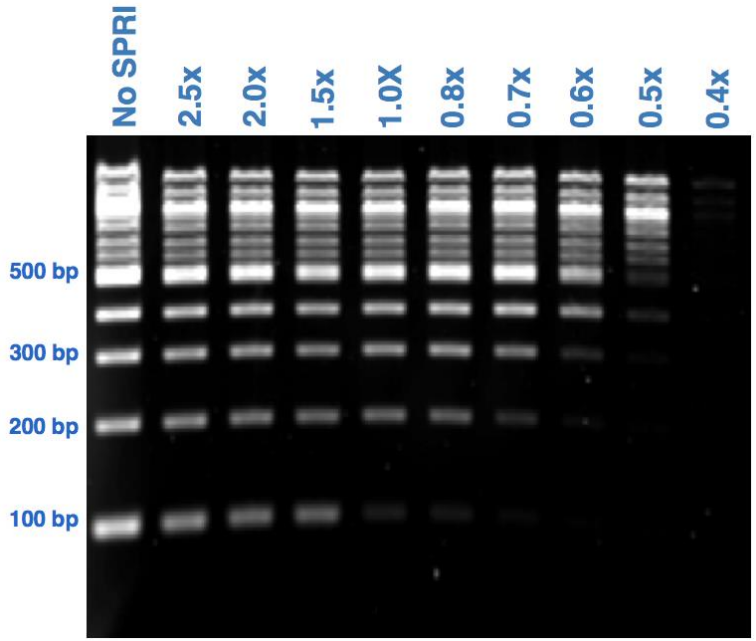
- One of the methods of isolation of nucleic acids, which has become more widespread
- MPs are particles with a size of 5 nm–100 μm formed from a metal core, which is most often gamma-Fe_2O_3 (maghemite) or Fe_3O_4 (magnetite).
- The core is covered by a layer that has a prepared specific surface. This can be adjusted according to which molecules we want to isolate from the given material.
- The size of MPs itself can be adjusted according to what we are isolating: 5-50 nm proteins; 20 –450 nm nucleic acids, viruses; 10–100 μm cells.
- The principle of isolation is based on the physico-chemical properties of MPs.

Paramagnetic particles









Binding of DNA fragments depends on the concentration of ethanol

https://www.beckman.com/resources/technologies/spri-beads?wvideo=kh244puadj

Purification of DNA from RNA

- For some applications it is necessary to have RNA without DNA contamination
- Precipitation of DNA with 1/10 volume of isopropyl alcohol - DNA precipitates and RNA remains in solution; however, the method is not 100%
- Treatment of sample with DNAse I (RapidOut DNA removal kit)
  - DNase I binds to Inhibition reagent (beads)
  - special DNase I with lower Km

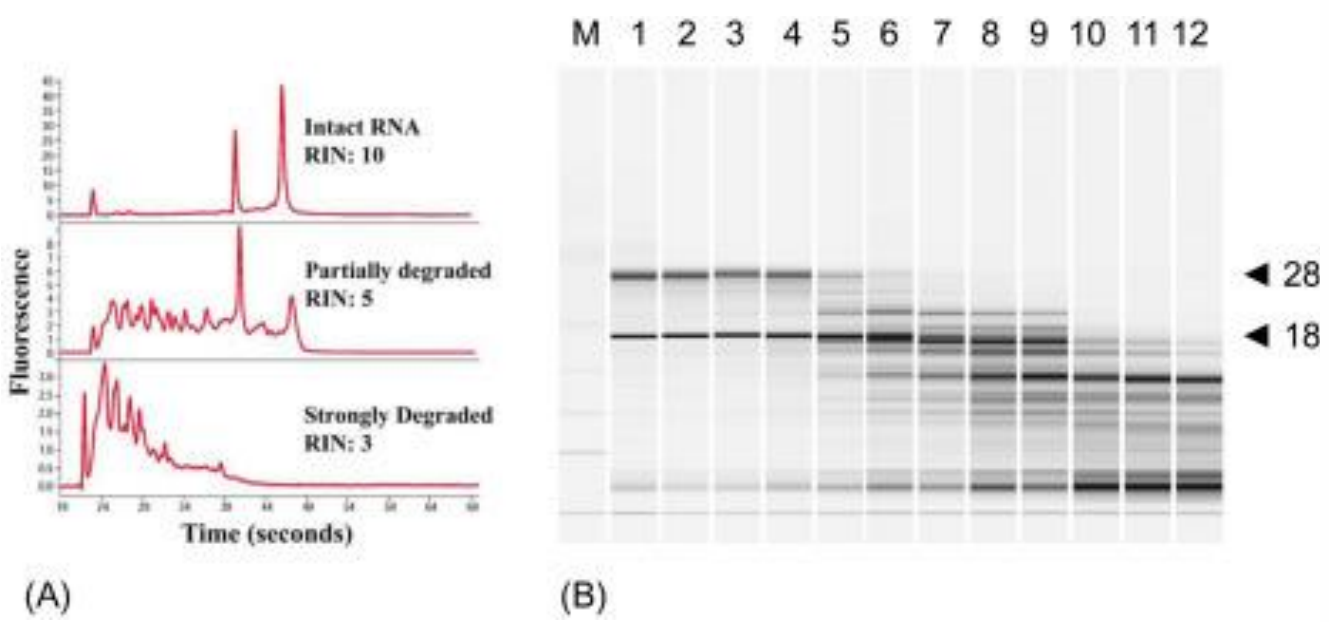
Quantification and Purity

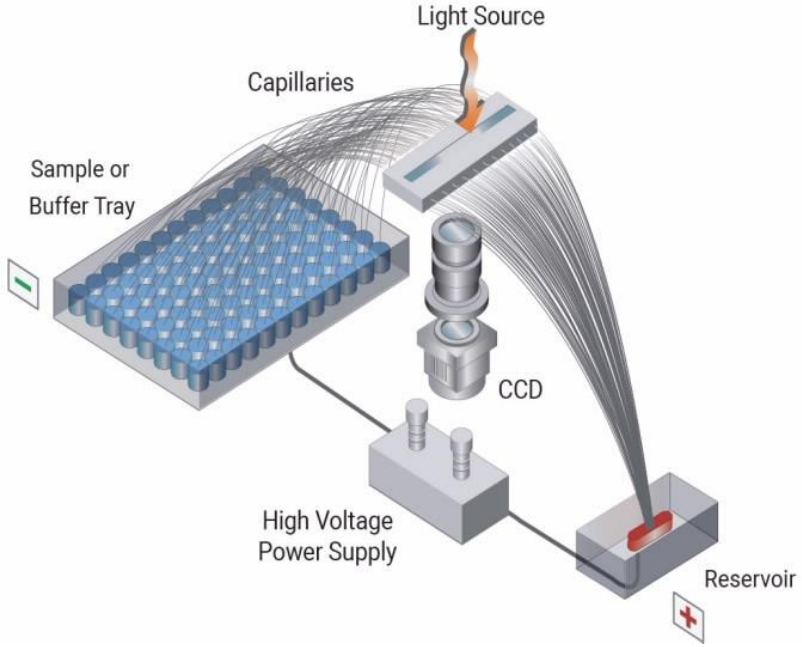
- Measure of concentration and purity by spectrophotometer (NanoDrop)
  - RNA: $A_{260/280}$ = 2.0, $A_{260/230}$ >1.5, ε = 40 $(\mu g/mL)^{-1}cm^{-1}$
  - DNA: $A_{260/280}$ = 1.8, $A_{260/230}$ >1.5, ε = 50 $(\mu g/mL)^{-1}cm^{-1}$

- Measure of concentration by Qubit (fluorometry)

- Measure of RNA integrity by Fragment Analyzer or TapeStation (Electrophoresis)
  - RIN (RNA integrity number) > 7

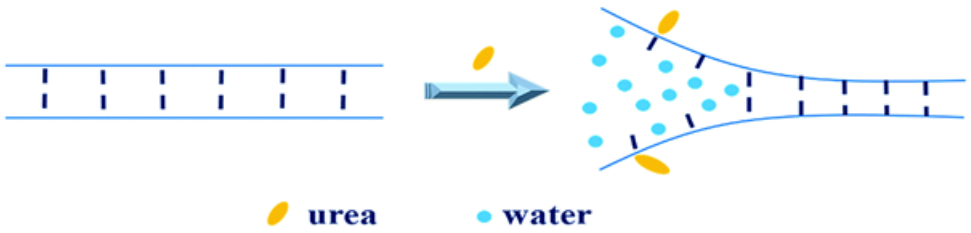
Fragment Analyzer (RIN)





Protein isolation

- **RIPA buffer (from tissue cultures)** - 30mM HEPES, pH 7.4,150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 5mM EDTA, 1mM $NaV0_4$, 50mM NaF, 1mM PMSF, 10% pepstatin A, 10 µg/ml leupeptin, and 10 µg/ml aprotinin
- **Homogenization in SDT buffer -** 4% SDS, 0.1M DTT, 0.1M Tris-HCl pH=7.6
- **Homogenization in Urea buffer** - 9M Urea, 20mM HEPES pH 8.0



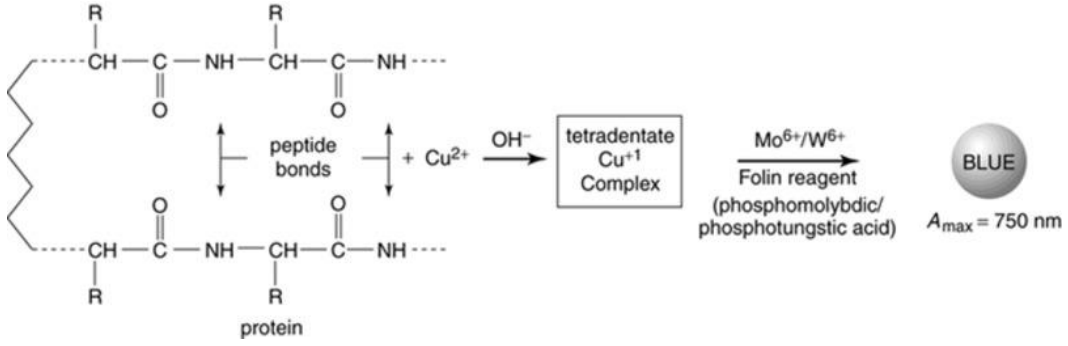
The hydrogen bond interaction between urea and the peptide groups opens the entrance for water and contributes to the unfolding denaturation of protein.

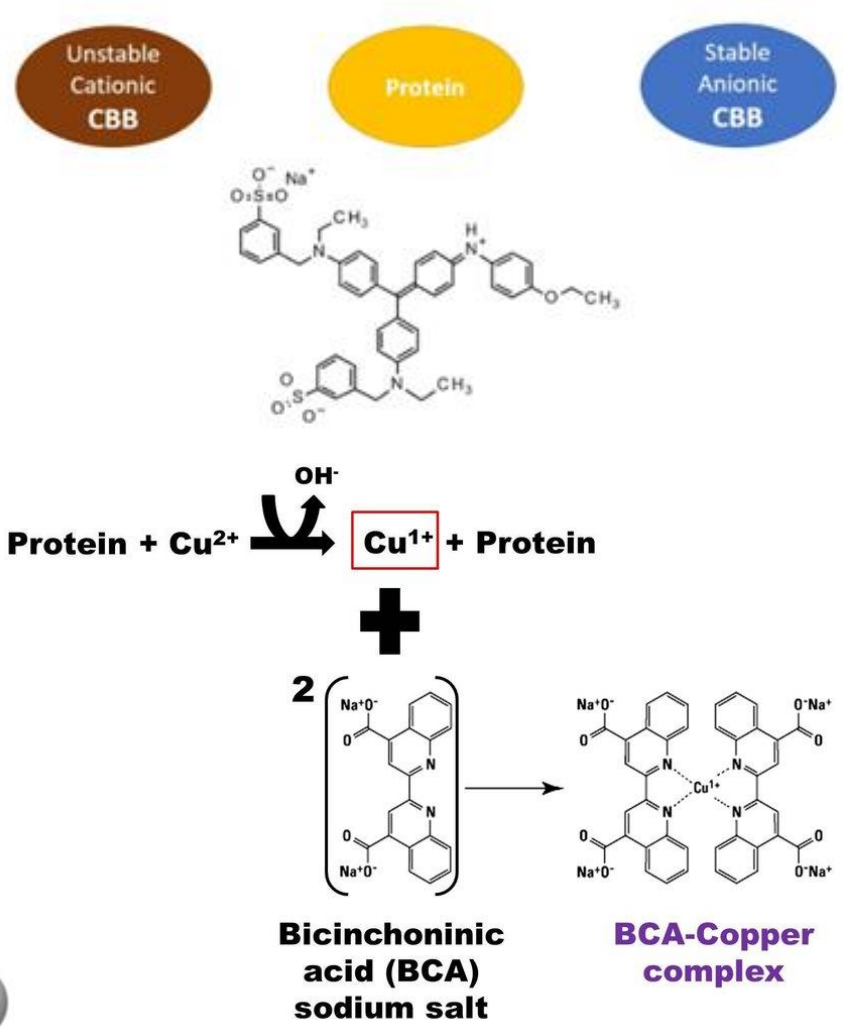
Proteins quantification

- **Bradford assay** (A_{595}) - interferuje SDS
- **Bicinchoninic assay (BCA)** (A_{562})

  - strong interference –SH group and EDTA

  - no interference with SDS (up to 5%)
- **Folin assay** (A_{750})
- **Measurement of Trp fluorescence** (280/350 nm)





MUNI
SCI

Manipulation of DNA, RNA and proteins

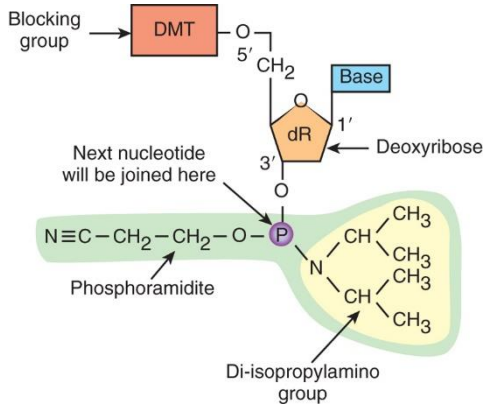
PCR techniques. DNA sequencing, high-throughput sequencing methods

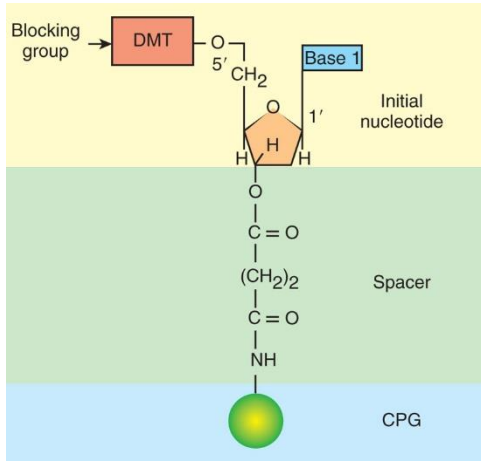
Chemical synthesis of DNA

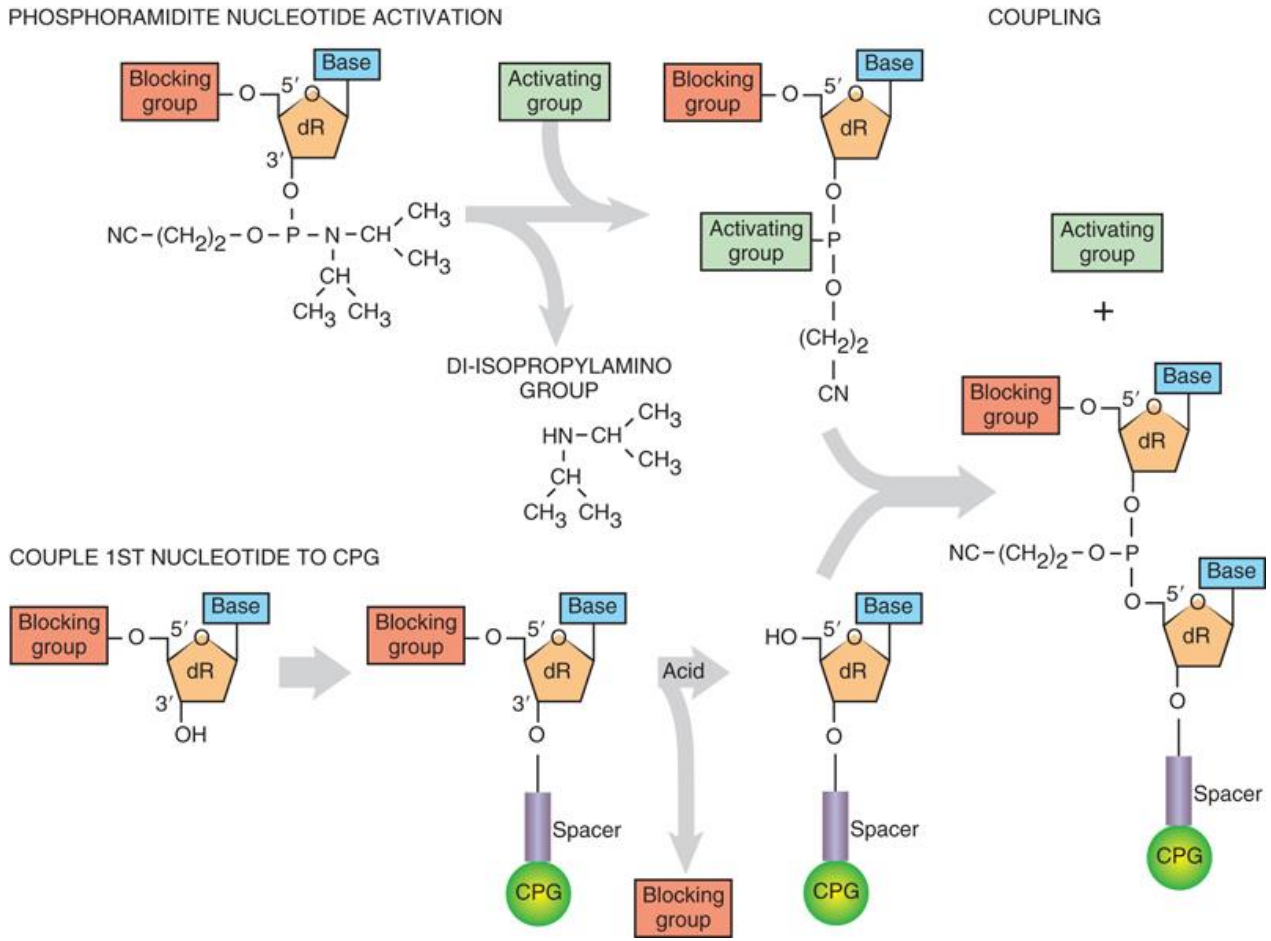
- H. Gobind Korana synthesized the first active tRNA molecule of 72 nucleotides (1970)
- Artificial DNA synthesis is in the 3' → 5' direction
  - attaching the first base to CPG (controlled pore glass)
  - the 5' end is blocked with DMT (dimethyloxytrityl)
  - the DMT group is removed using a weak acid (TCA)
  - another nucleotide is added in the form of so-called phosphoramidite activated by tetrazole
  - 5'- OH ends of unreacted nucleotides are acetylated using acetic anhydride
  - repeating the process

(*from the left*) **Har Gobind Khorana**, **Robert W Holley**, Luis W Alvarez, **Marshall W Nirenberg**, Lars Onsager and Yasunari Kawabata at the awarding of the Nobel Prize in 1968.

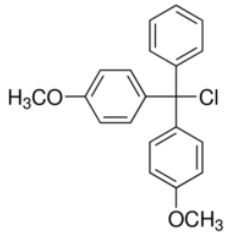
Chemical synthesis of DNA



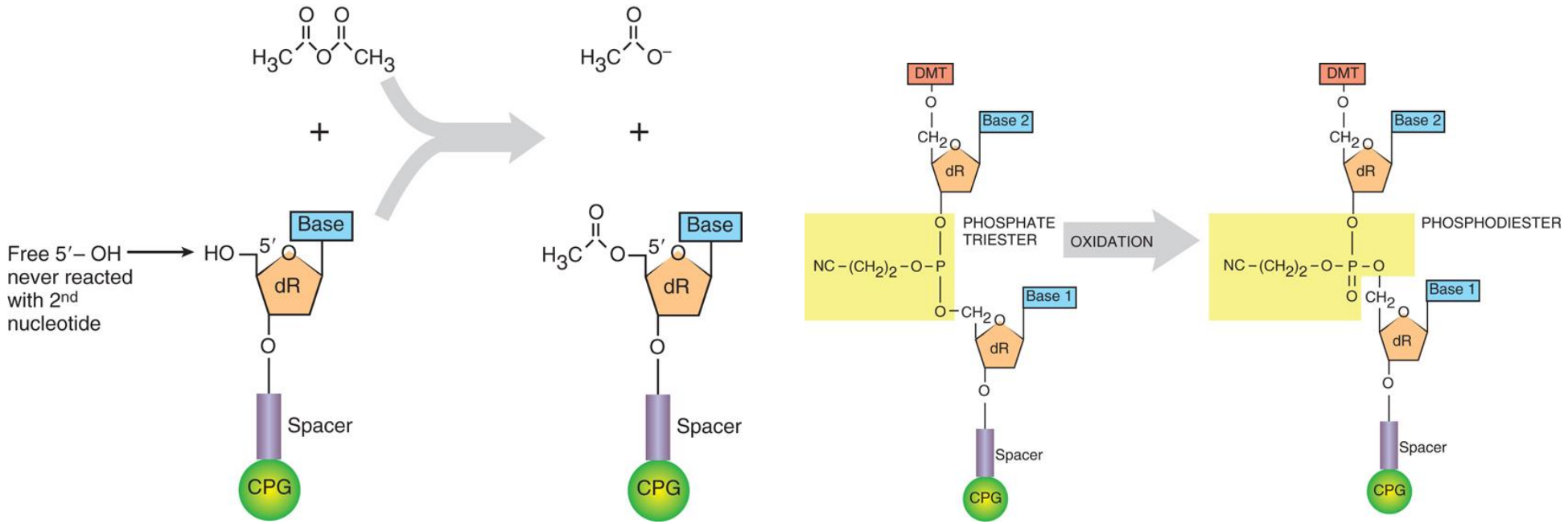


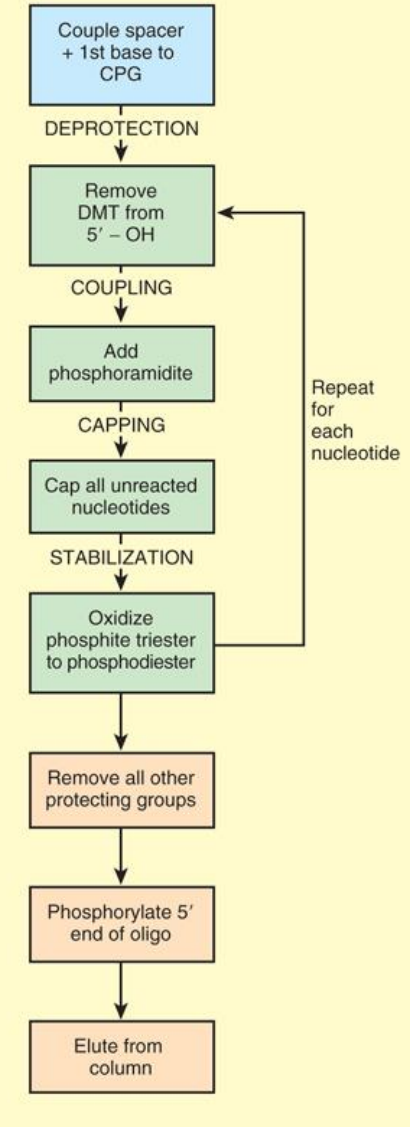


DMT-Dimethoxytrityl



Chemical synthesis of DNA





https://www.youtube.com/watch?v=1S0x3aRCviM

Polymerase Chain Reaction

K Mullis, F Faloona, S Scharf, R Saiki, G Horn, H Erlich.
Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.
Cold Spring Harb Symp Quant Biol;**1986**;51

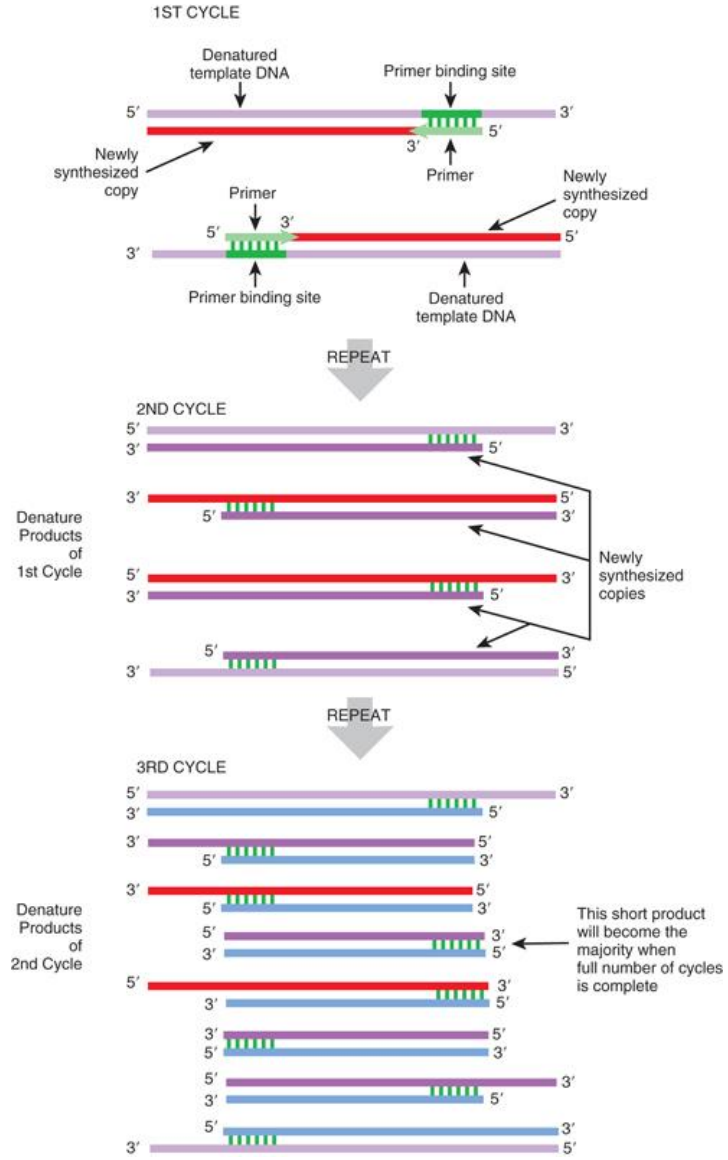
..the idea of PCR came to him while driving with his girlfriend on a highway..

"It was quiet and something just went, Click!"

KARY B MULLIS

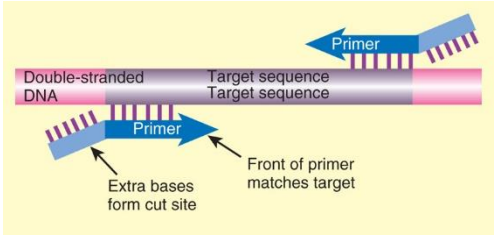
1944-2019

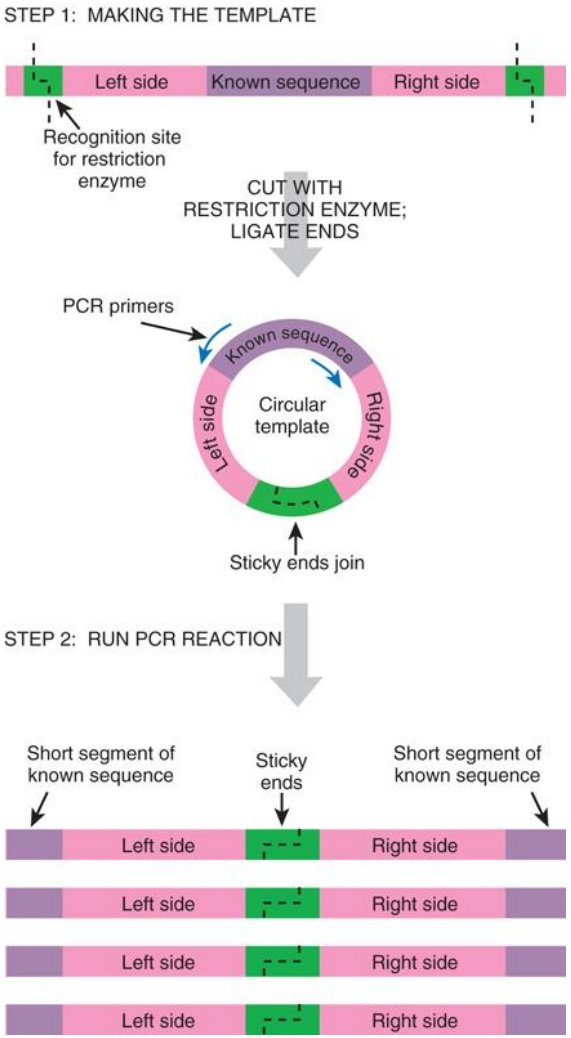
Inventor of PCR Technique

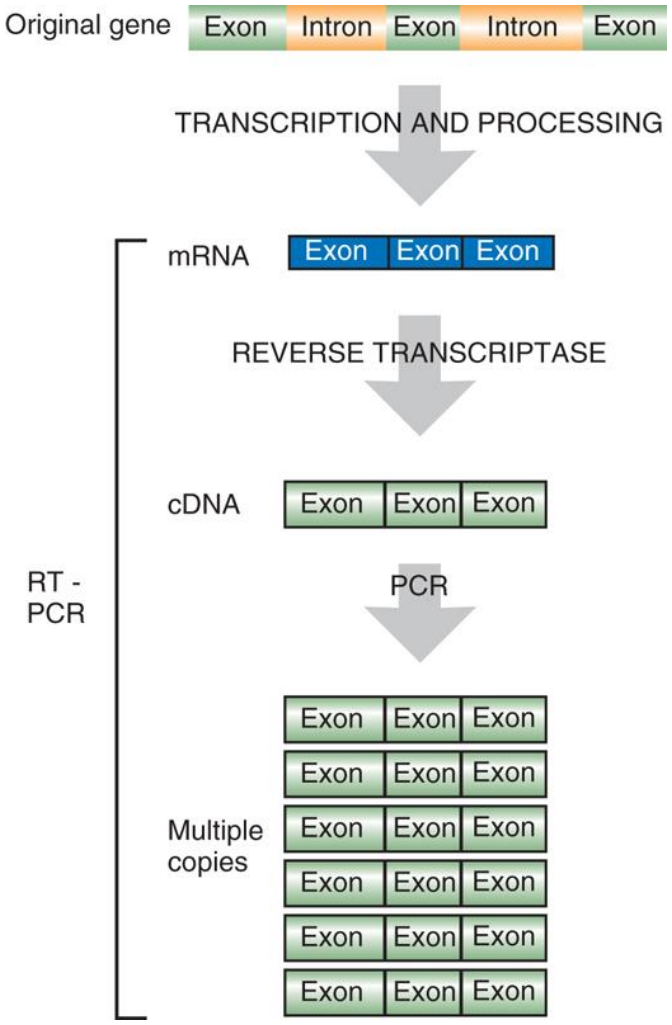


Modifications of PCR

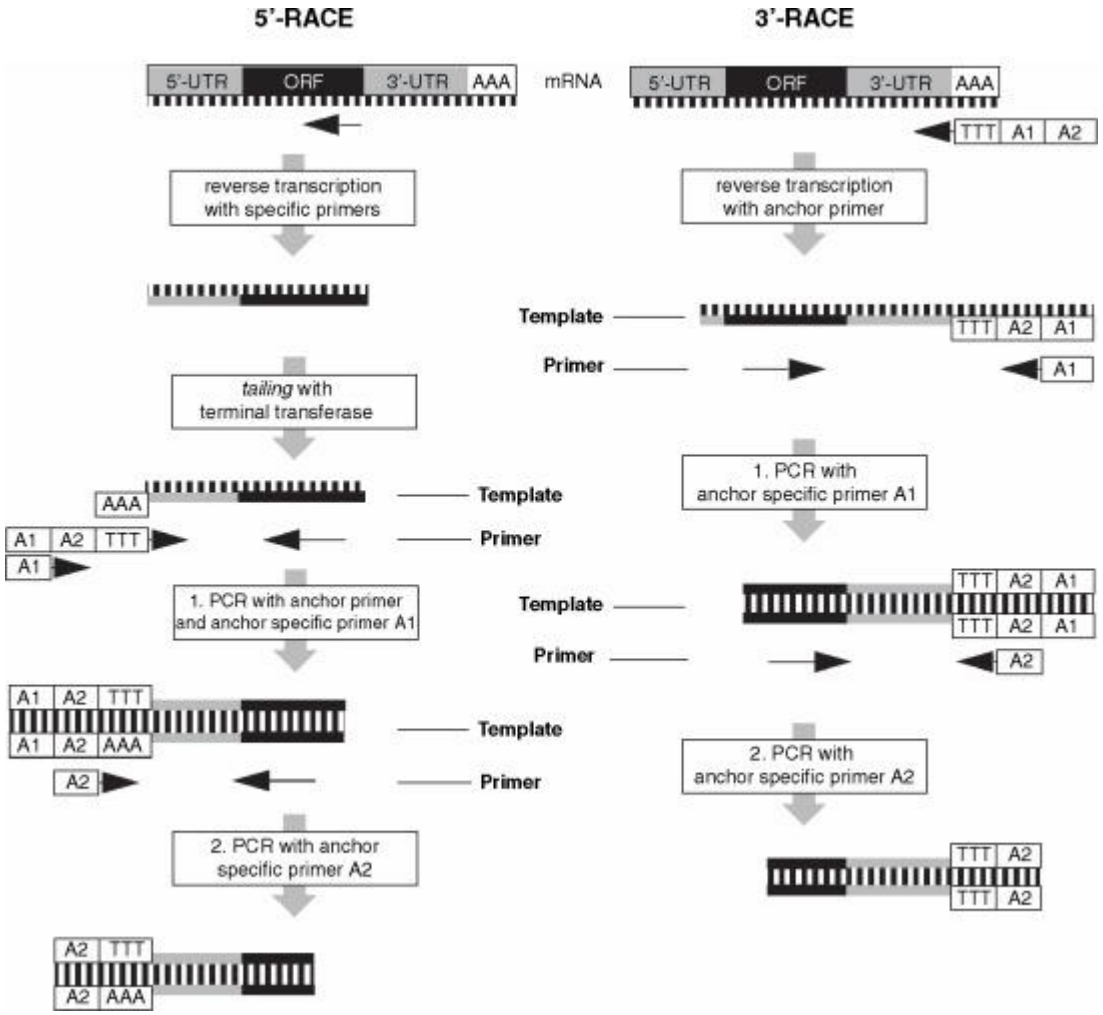
- Inversion PCR
- Reverse Transcription PCR (RT-PCR)
  - 5´RACE, 3´RACE
- PCR mutagenesis
- Emulsion PCR
- Droplet Digital PCR







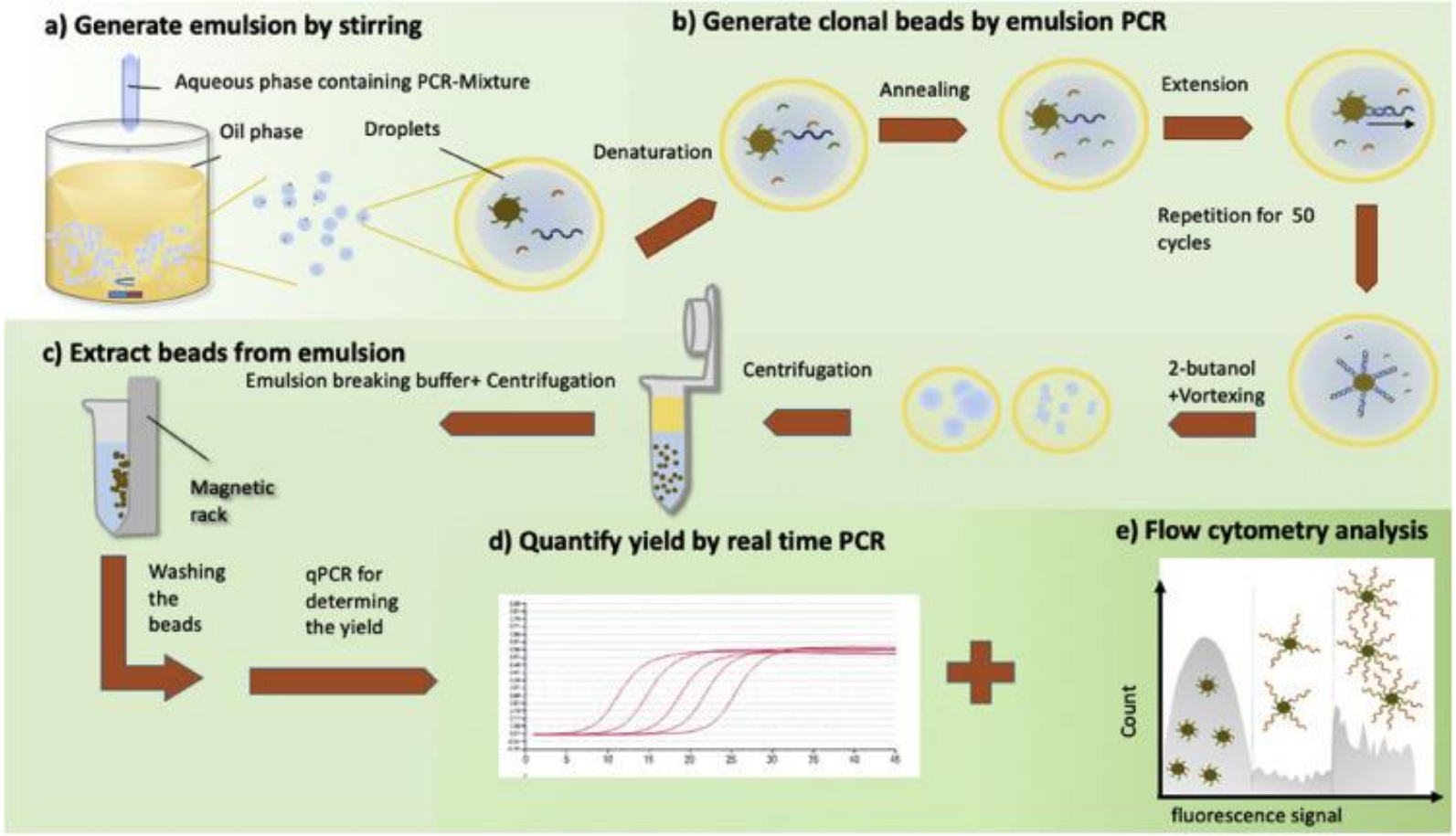
RACE PCR



Rapid Amplification of cDNA Ends

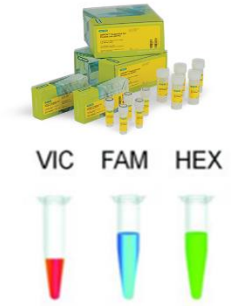
Emulsion PCR

Used in NGS technology (454, ion torrent)



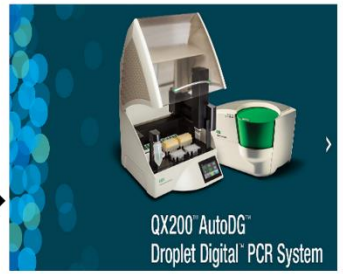
Droplet digital PCR

DNA isolation

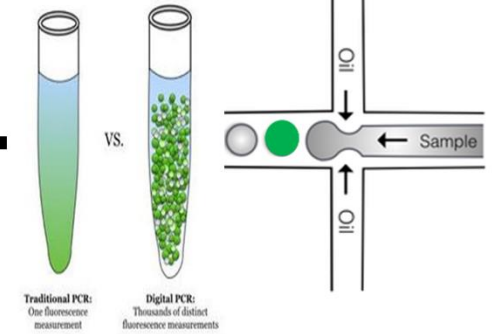


Prepare ddPCR ready mix

Load mix and sample into a plate

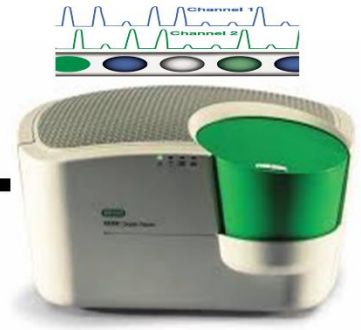


Place plate in a droplet generator

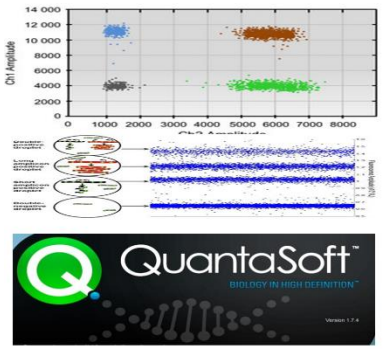


Generate droplets

Amplify targets in droplets by PCR



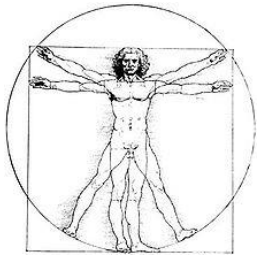
Place amplified droplets in a droplet reader



Read and analyze results

https://www.youtube.com/watch?v=lAVVoyZxlTU

Sequencing of human genome

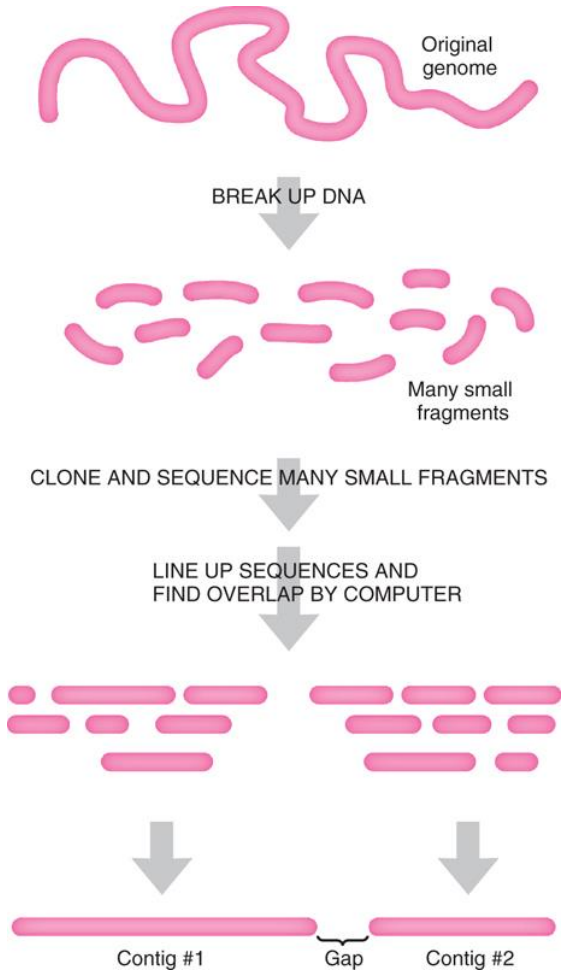


- at the time of the beginning (the year 1990) a monumental task
- started in 1990 with the participation of the DOE and NIH
- sequencing done using contig maps and BACs
- the initial plan envisaged duration of 15 years
- finally, sequencing using the Sanger method was almost completed already in 2000
- resulting sequence map published on April 14, 2003, with 99.99% accuracy (National Human Genome Research Institute)
- total cost of the project 3 billion dollars
- in 2000, President Bill Clinton asserted the unpatentability of DNA

https://https://www.genome.gov/25019885/online-education-kit-how-to-sequence-a-human-genome//

Celera Genomics Project

- founded by scientist Craig Venter and started a sequencing project in 1998
- the total cost of 300 million dollars was fully covered by private sources
- the "whole genome shotgun sequencing" method was used for the first time
- used an approach developed by Gene Myers to analyze the sequencing data
- this approach required extreme computational demands
- final calculation performed on 7000 processors to obtain 1000 times the speed of Pentium computers
- this innovative approach allowed sequencing to be completed in just 9 months



The strong role of diplomacy

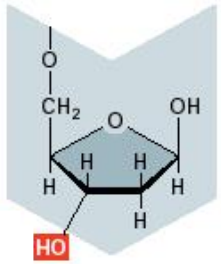
It is hard to imagine today's politicians reminding scientists that cooperation has as much value as competition.

In 26 June 2000, US President Bill Clinton and UK Prime Minister Tony Blair presided over a carefully choreographed piece of scientific theatre. Through a video link connecting Washington DC and London, they announced to the world that scientists had completed a rough first draft of the human genome sequence.

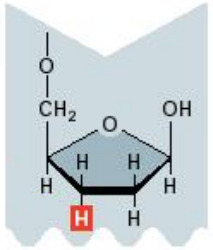
Craig Venter (left), Francis Collins, Bill Clinton (right)

Sanger sequencing

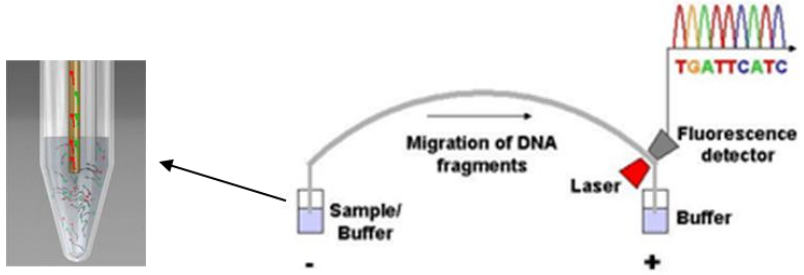
- Synthesis of DNA in-vitro using "terminators" - dideoxynucleotides that prevent further elongation after being incorporated into DNA.



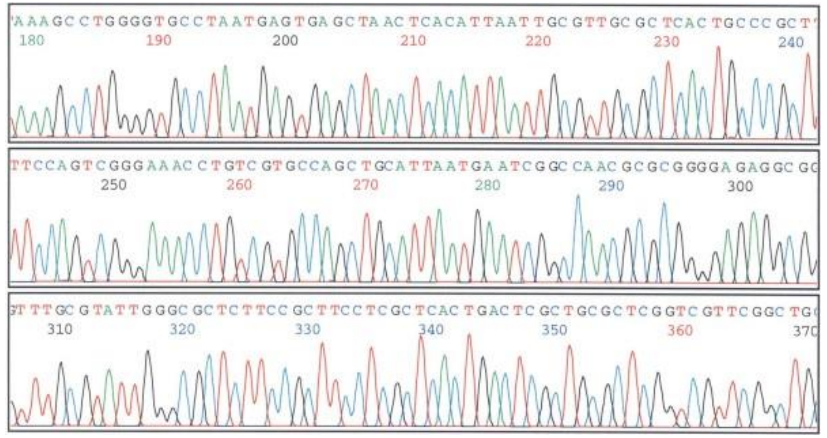
Deoxyribosa



Dideoxyribosa



- It requires the use of an initial primer, DNA polymerase and a mixture of dNTPs with labeled ddNTPs
- The synthesized strands are separated using polyacrylamide gel electrophoresis or capillary electrophoresis
- Possibility of fully automated separation using fluorescently labeled ddNTPs



Sanger sequencing

Throughput/Performance by Run Module

XLRseq: 768 samples per day (690 Kbases)

LongSeq: 1152 samples/day (980 Kbases)

StdSeq: 2304 samples/day (1550 Kbases)

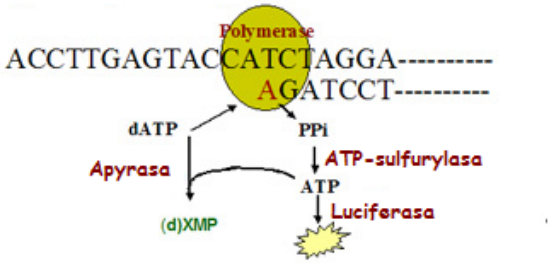
FastSeq: 2304 samples/day (1600 Kbases)

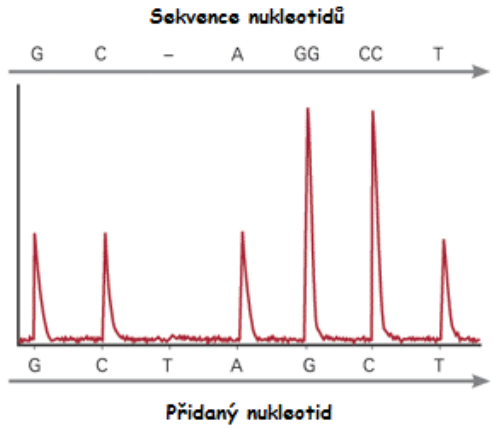
RapidSeq: 3840 samples per day (2100 Kbases)

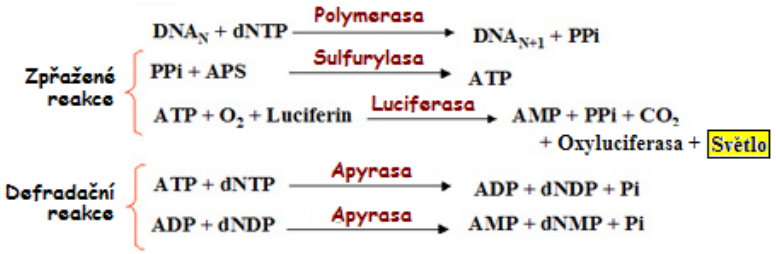
Pyrosequencing (1990)

- it enables rapid sequencing of short stretches of DNA - sequencing of 30 to 50 bases takes approximately 30 to 45 minutes.

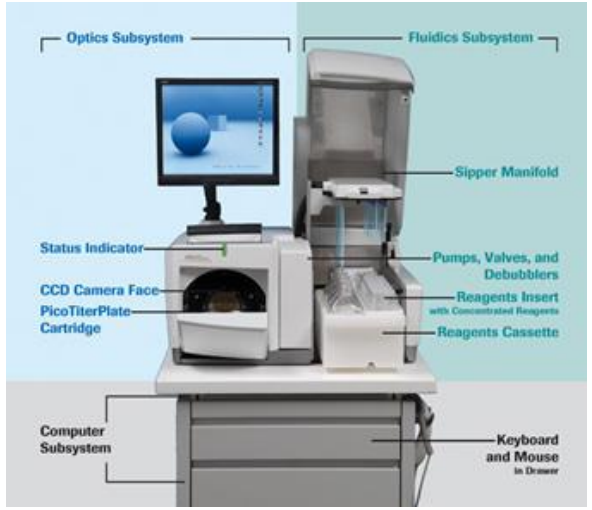
- it is bio-luminometric DNA sequencing based on the detection of inorganic pyrophosphate (PPi) released during nucleotide incorporation.

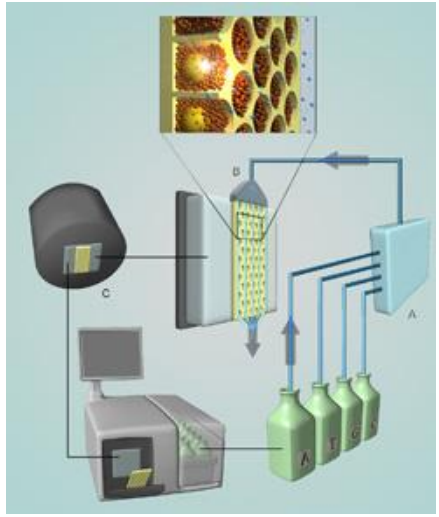




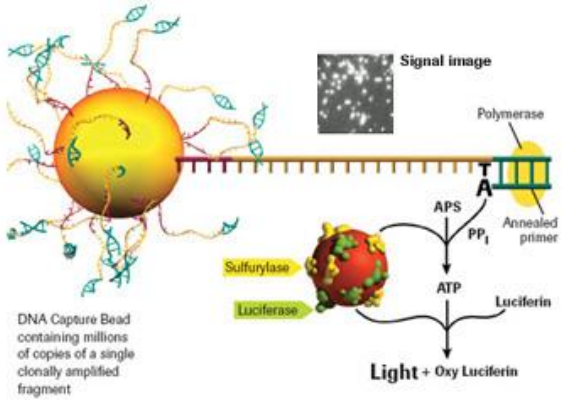


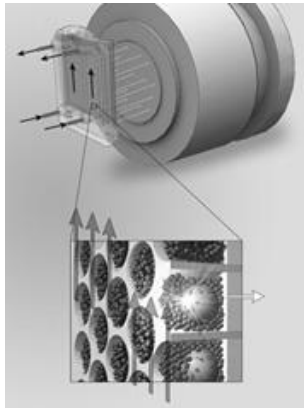
454 a GS Junior system





Průchodnost	1 miliarda bazí za den
Doba analýza	10.0 hodin
Délka čtení	400
Počet čtení/analýzu	1 000.000
Správnost	>99.0% správnost jednoho čtení na 400 bazích
Potřebné množství DNA	Méně než 100 ng DNA
Multiplexování	Až 192 vzorků/běh

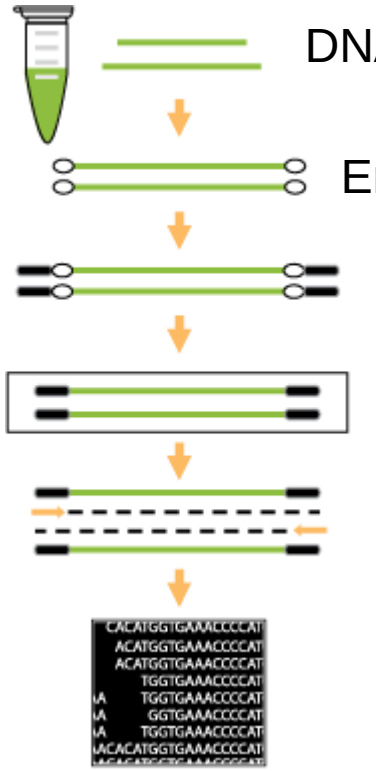




Qiagen – PyroMark instruments

- https://www.labtube.tv/video/MTAxNzE1
- https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/explainer-videos-and-demos/pyrosequencing-cascade-reaction

Preparation of Sequencing Library



DNA sample – fragmentation (Covaris, fragmentase)

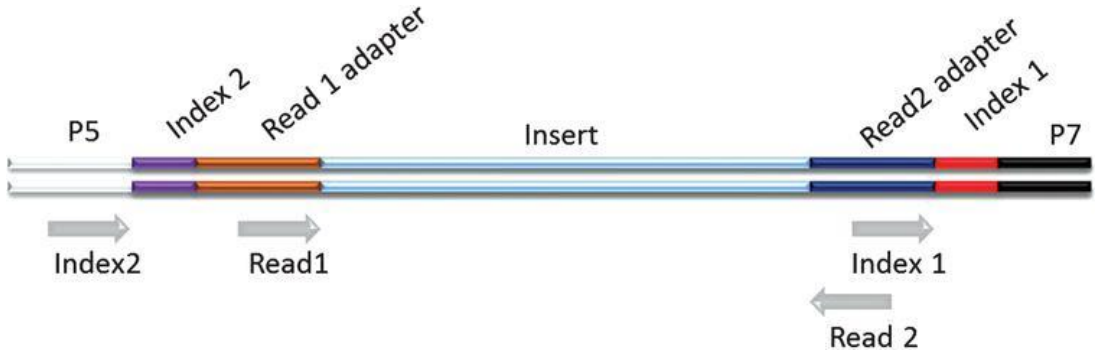
End-repair (DNA polymerase)

Adaptor ligation (ligase)

Selection of fragments (SPRI beads)

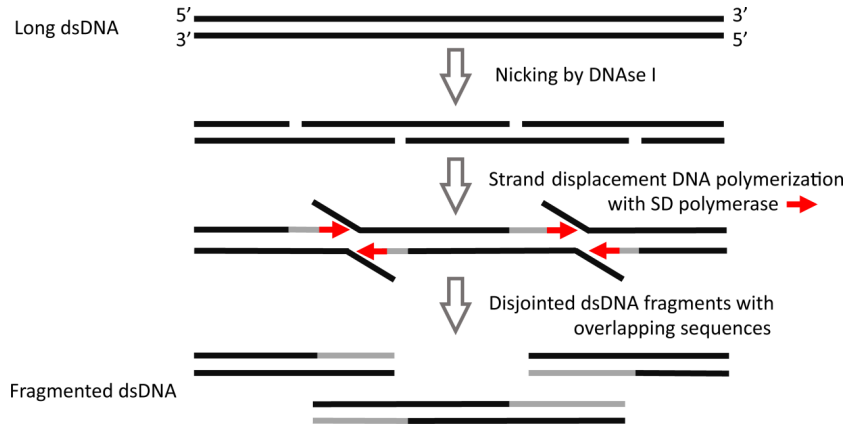
Amplification of fragments

Sequencing (Illumina, IonTorrent, Nanoballs)



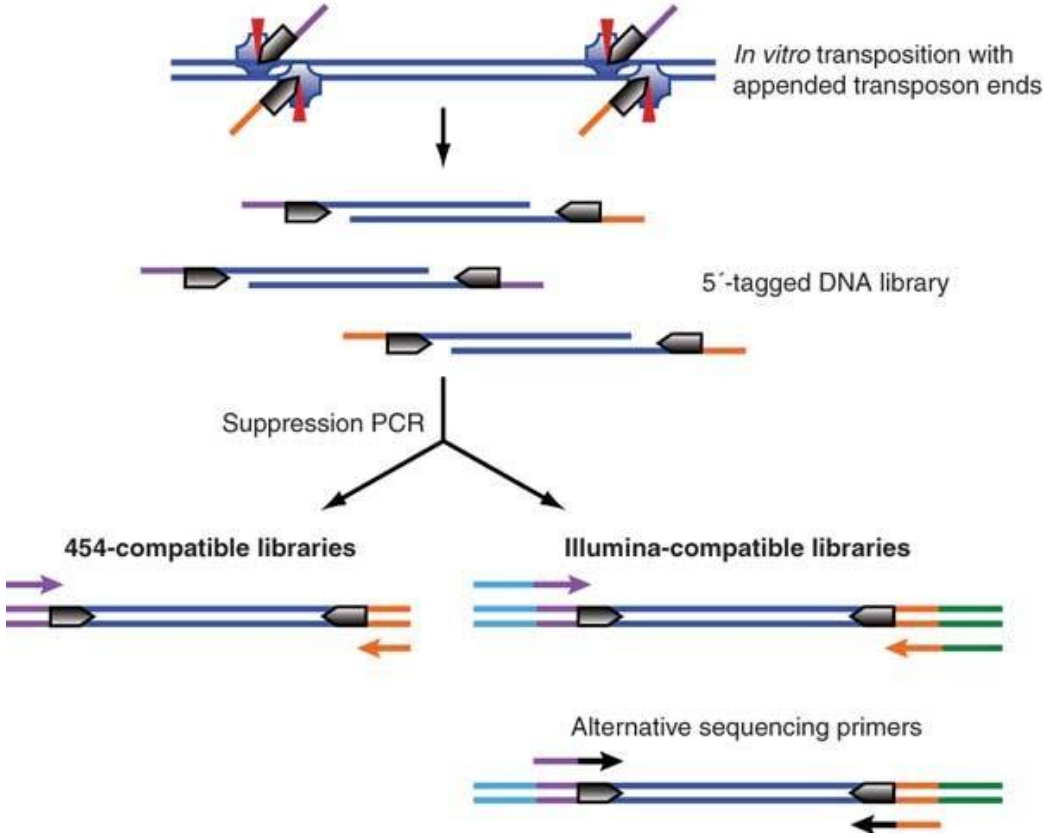
Fragmentase

- A mixture of endonucleases (NEases) cleaving one strand and then the opposite one
- A mixture of two enzymes (DNase I and SD (strand-displacement) polymerase)



Nextera technology

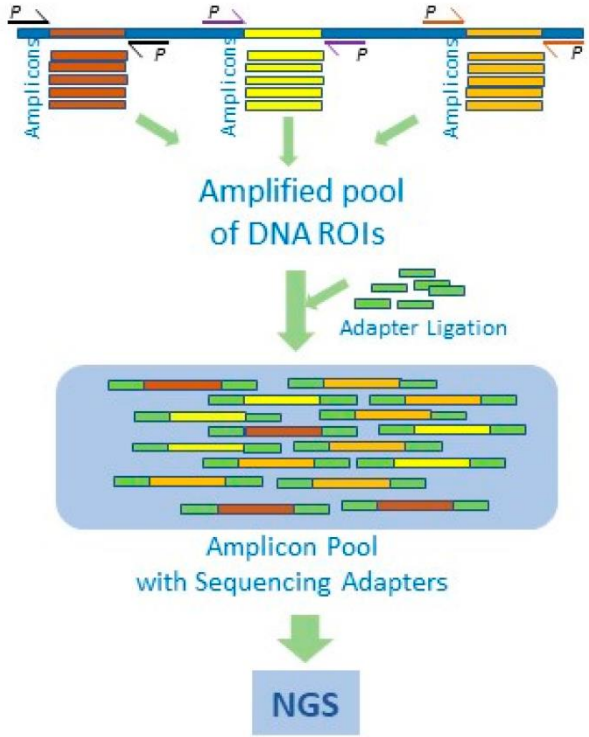
- uses *in vitro* transposition
- transposases catalyze the random insertion of excised transposons
- transposase makes random, staggered double-stranded breaks in the target DNA and covalently attaches the 3′ end of the transferred transposon strand to the 5′ end of the target DNA.
- for integration only free transposon ends are sufficient



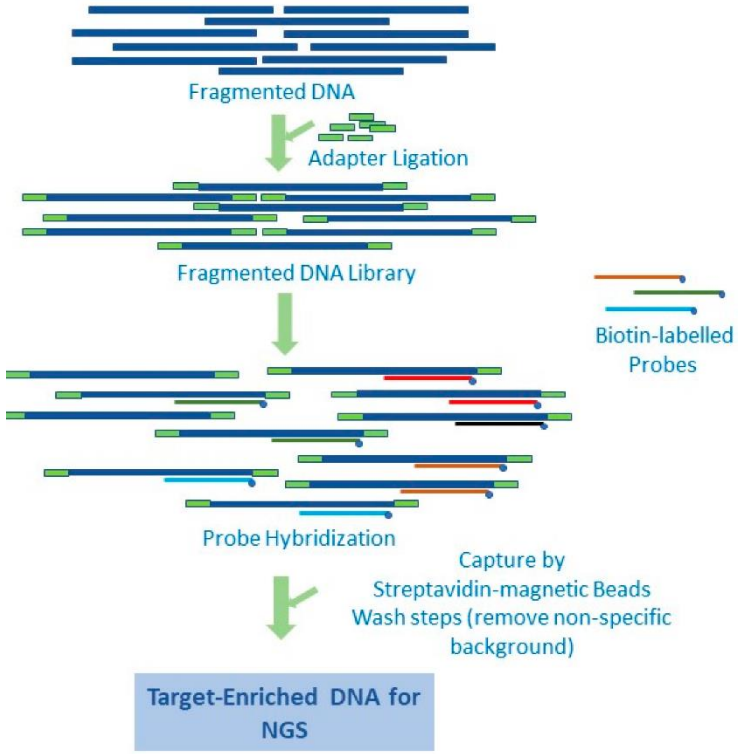
https://doi.org/10.1038/nmeth.f.272

Targeted Enrichment

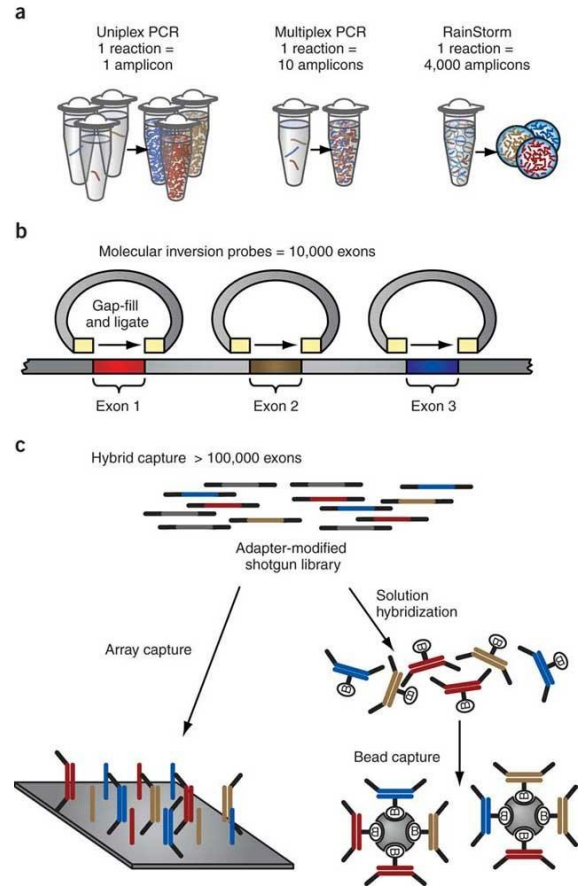
PCR enrichment



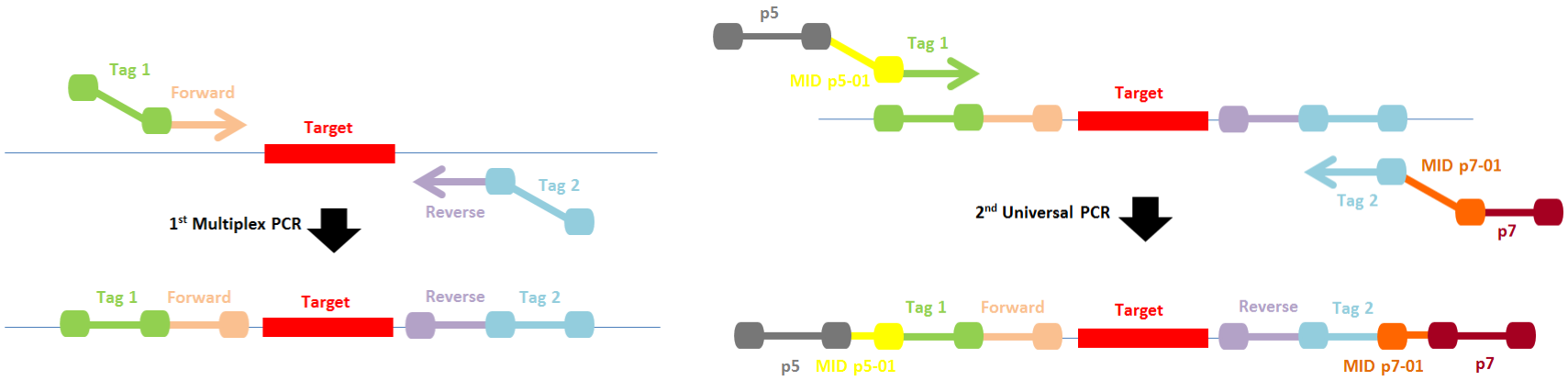
DNA capture

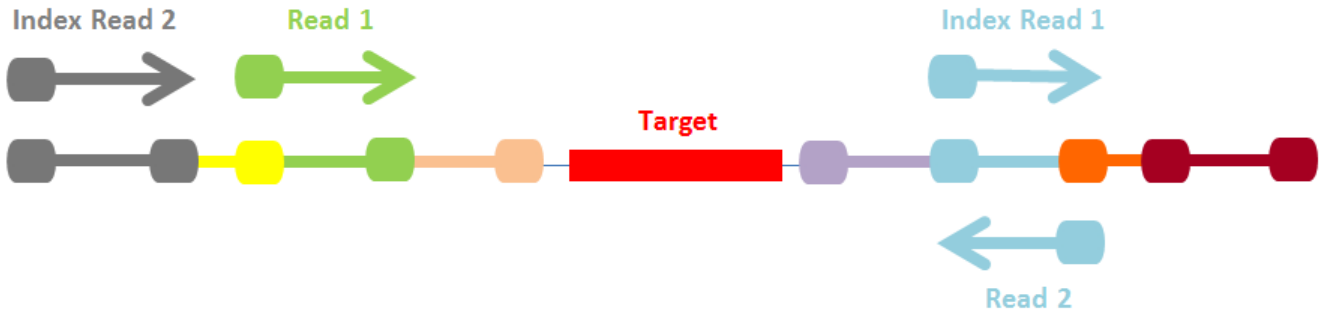


Inversion probes



Amplicon sequencing





Quantification of NGS library

Electrophoretic methods

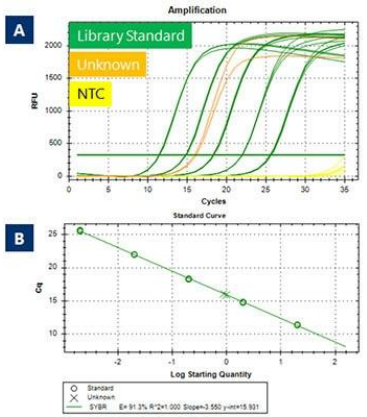
- Fragment Analyzer (Adv. Anal.)
- TapeStation (Agilent)
- BioAnalyzer (Agilent)

Fluorometric methods

- Qubit (Thermo Scientific)
- Quantus (Promega)

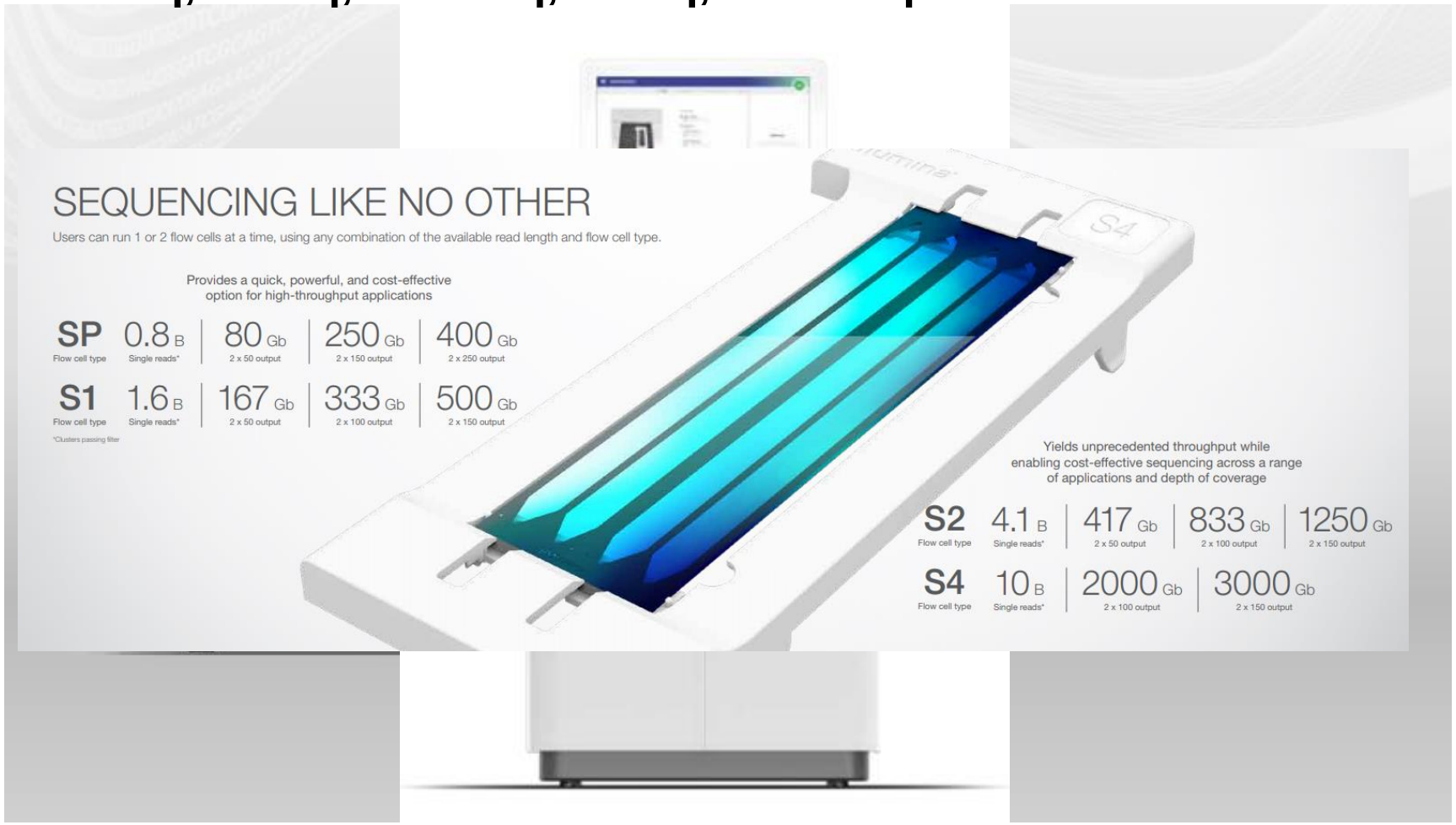
Real-Time PCR

- KapaBiosystem
- NEB



Illumina sequencing system

MiniSeq, MiSeq, NextSeq, HiSeq, NovaSeq

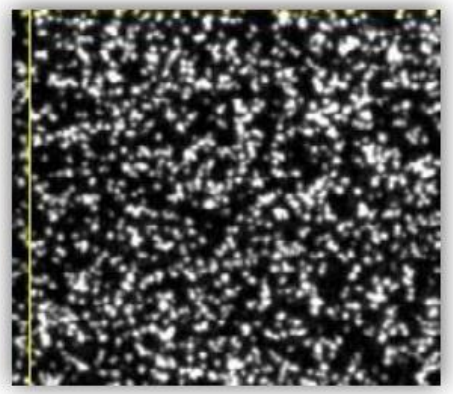


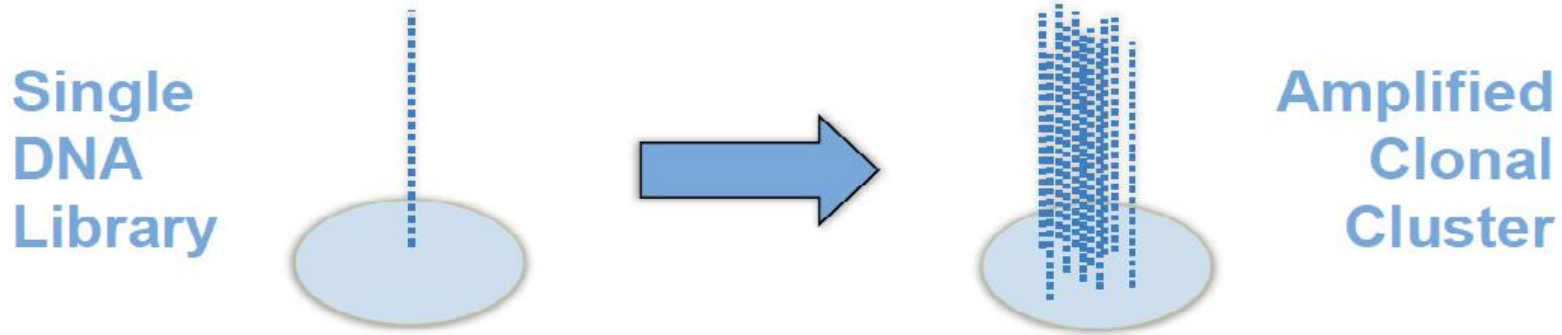
Sequencing in clusters

What is a cluster?

Clusters are bright spots on an image

Each cluster represents thousands of copies of the same DNA strand in a 1–2 micron spot



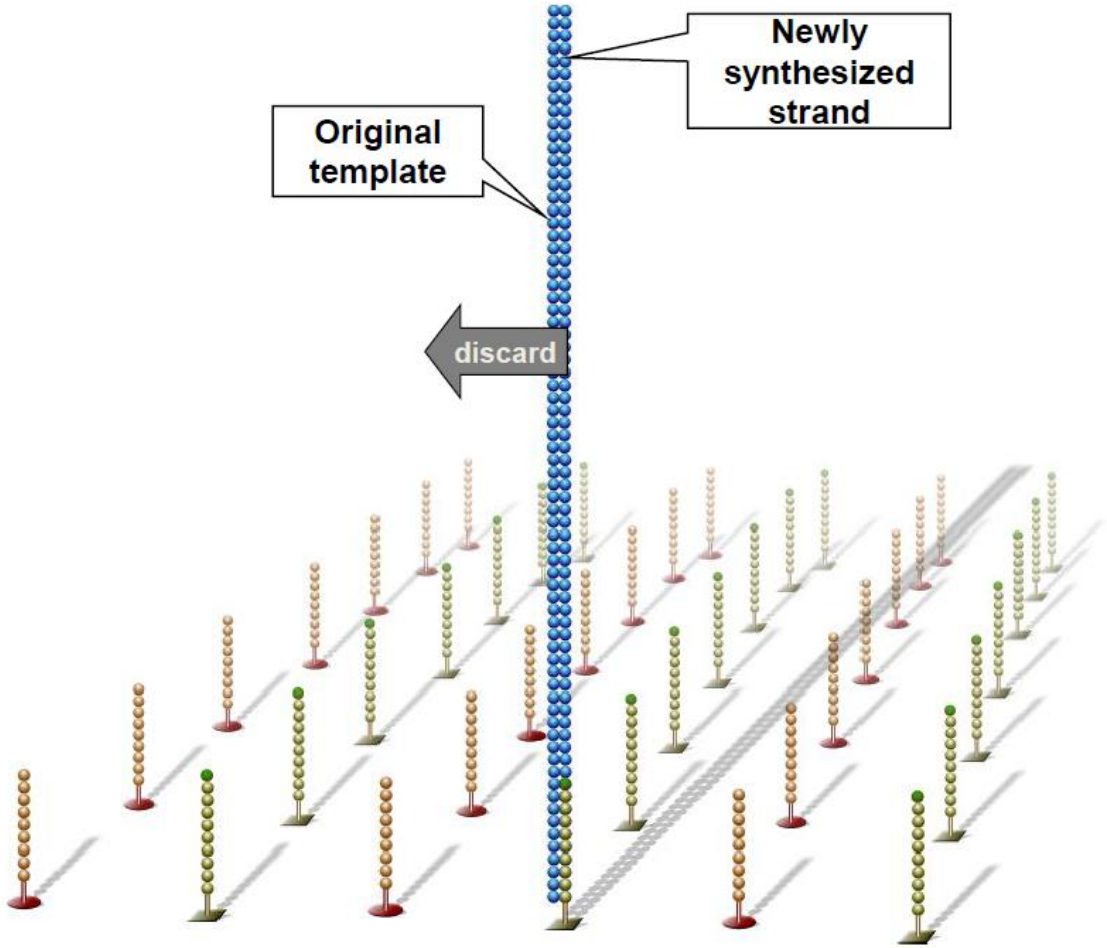


1st step – hybridization on flow-cell

Double-stranded molecule is denatured

Original template washed away

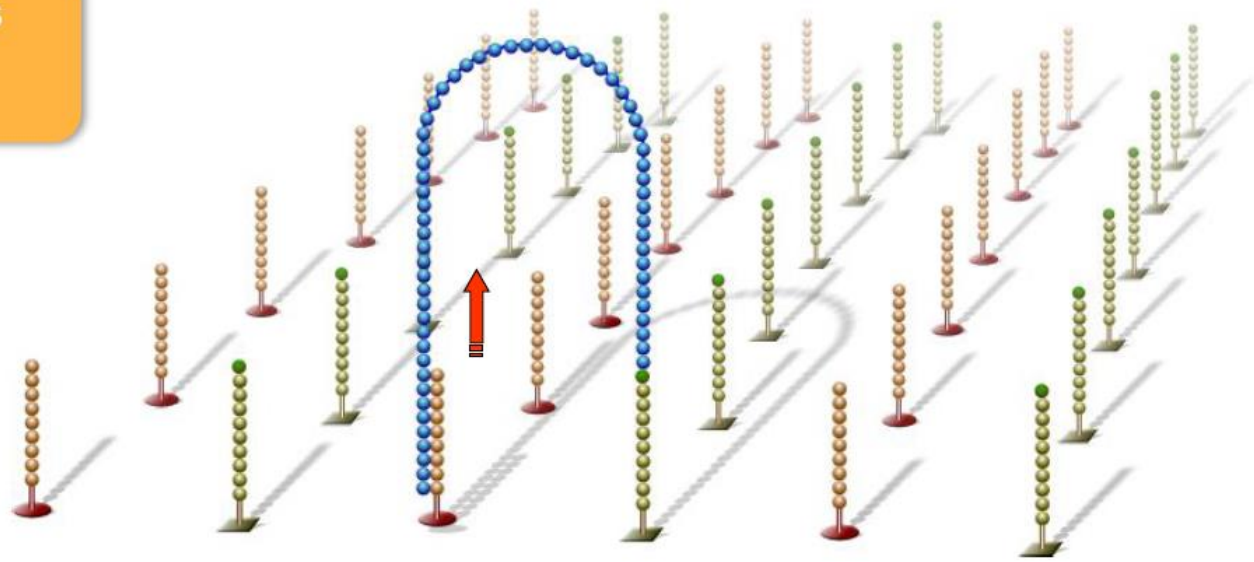
Newly synthesized strand is covalently attached to flow cell surface



2nd step – bridge PCR

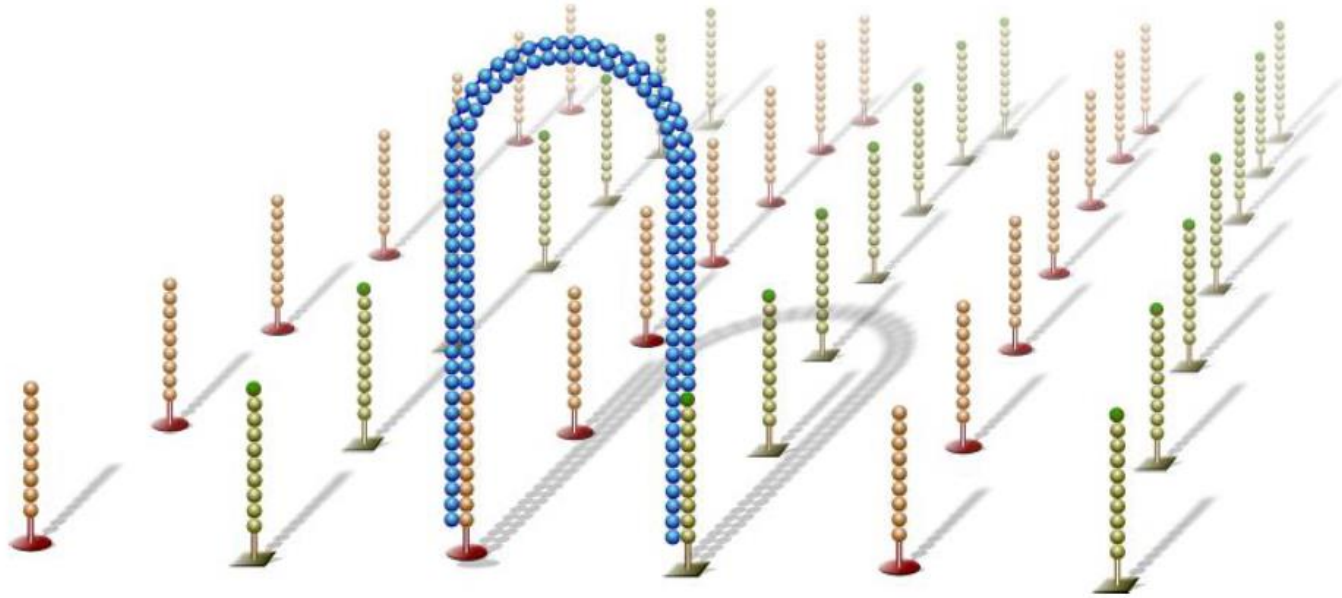
Single-stranded molecule flips over and forms a bridge by hybridizing to adjacent, complementary primer

Hybridized primer is extended by polymerases



2nd step – bridge PCR

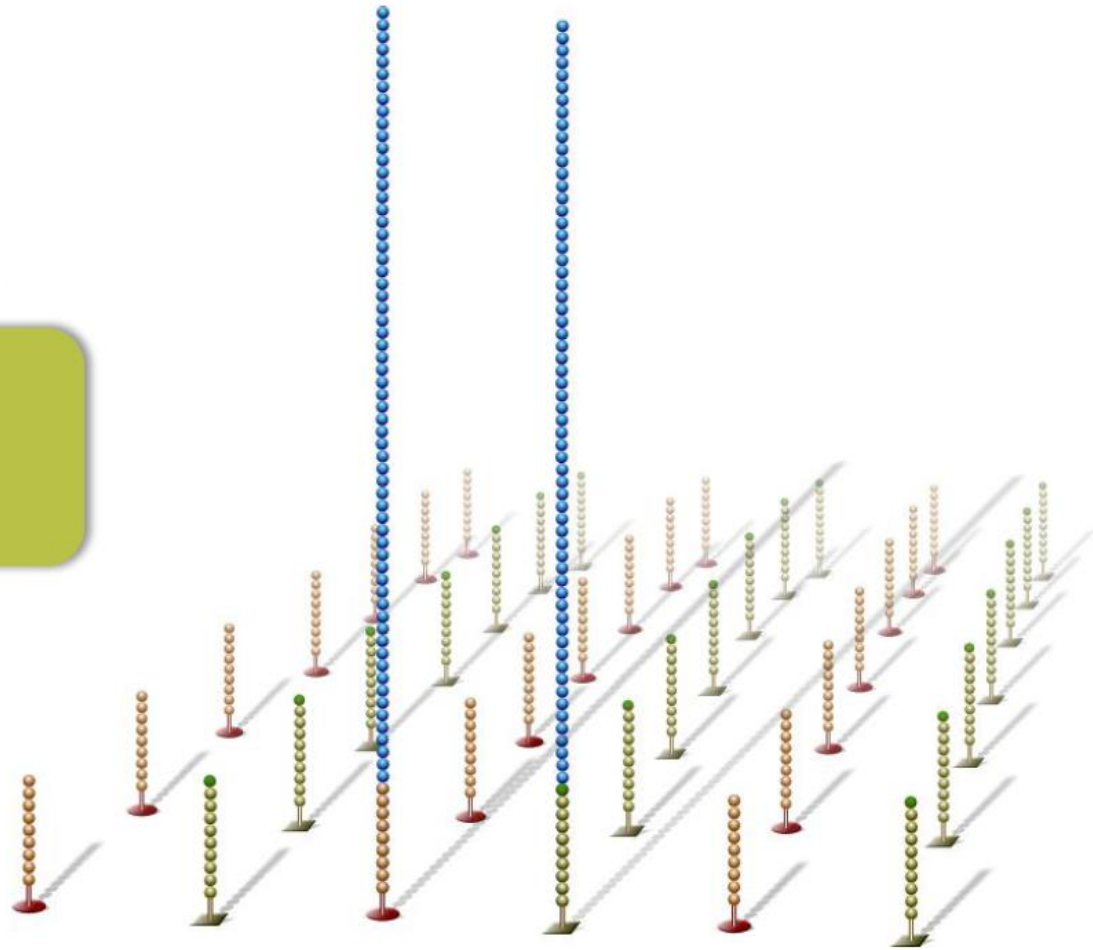
Double-stranded bridge is formed



2nd step – bridge PCR

Double-stranded bridge is denatured - 1st cycle denaturation

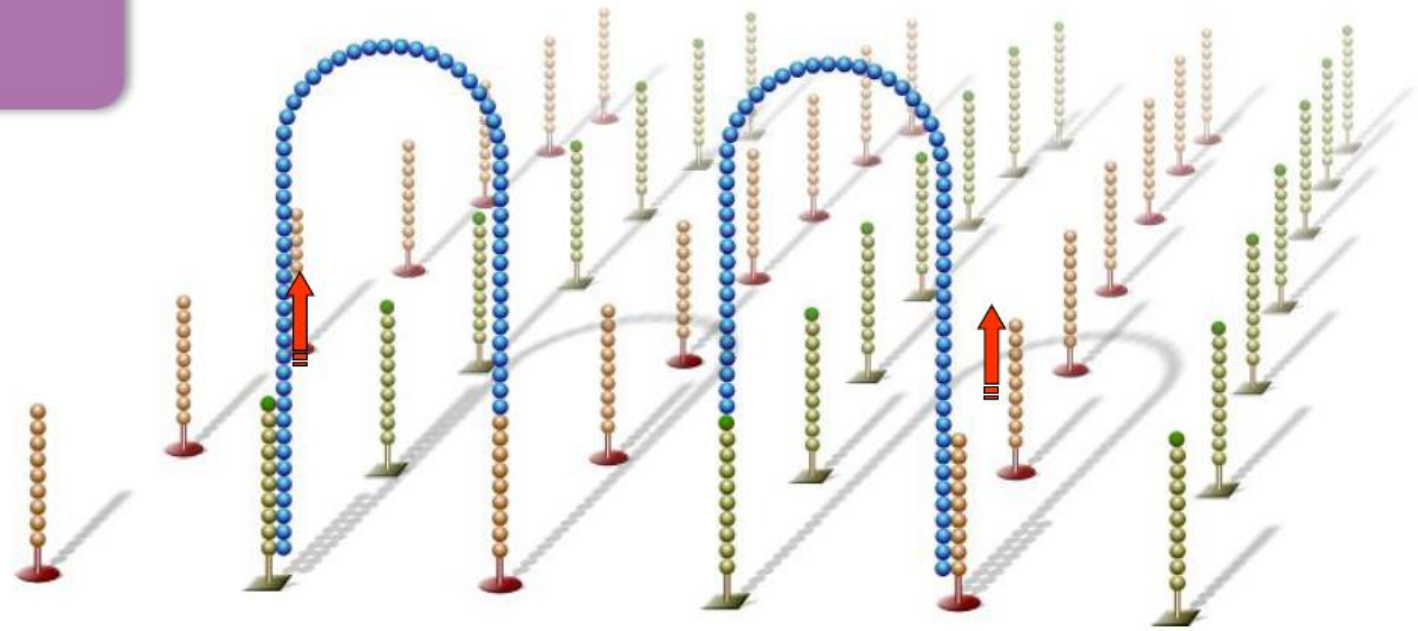
Result:
Two copies of covalently bound single-stranded templates



2nd step – bridge PCR

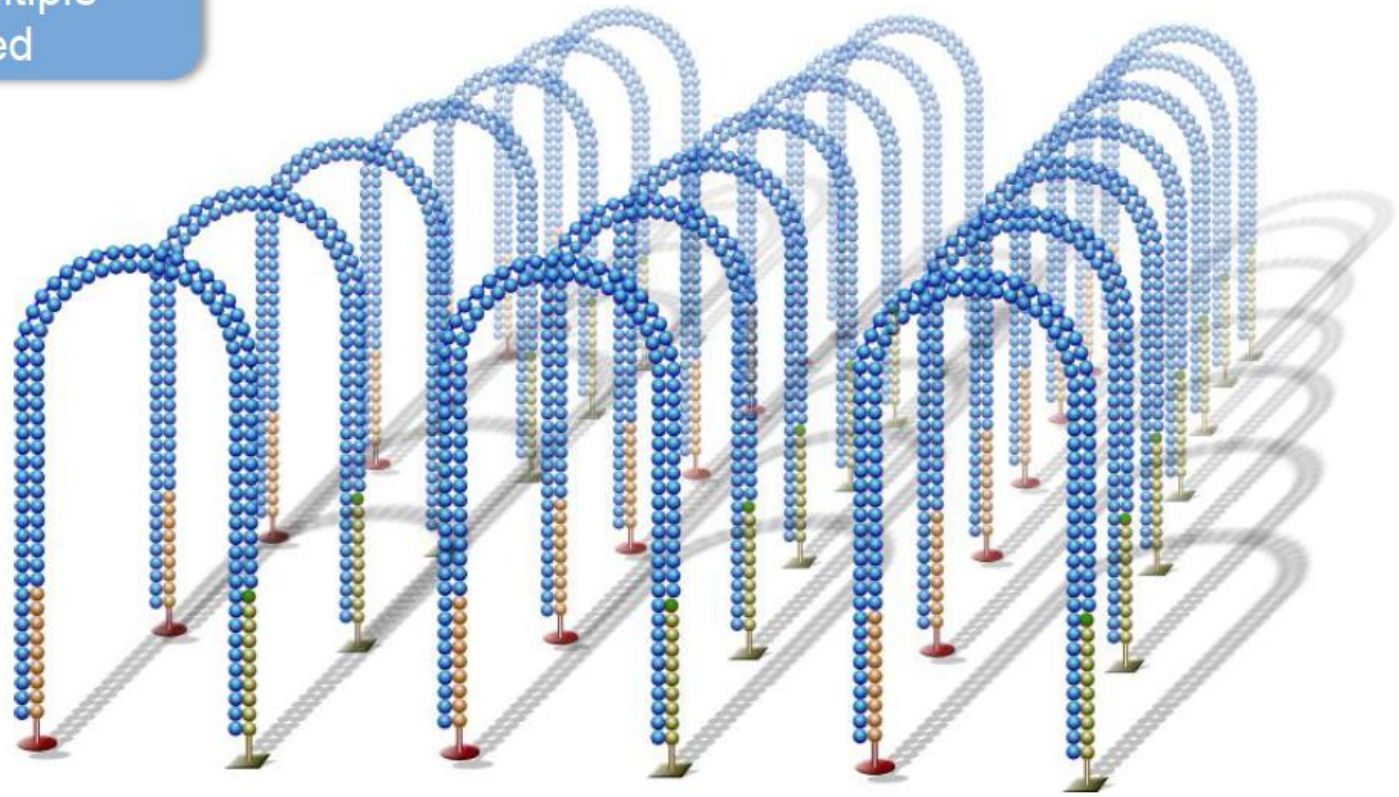
Single-stranded molecules flip over to hybridize to adjacent primers

Hybridized primer is extended by polymerase



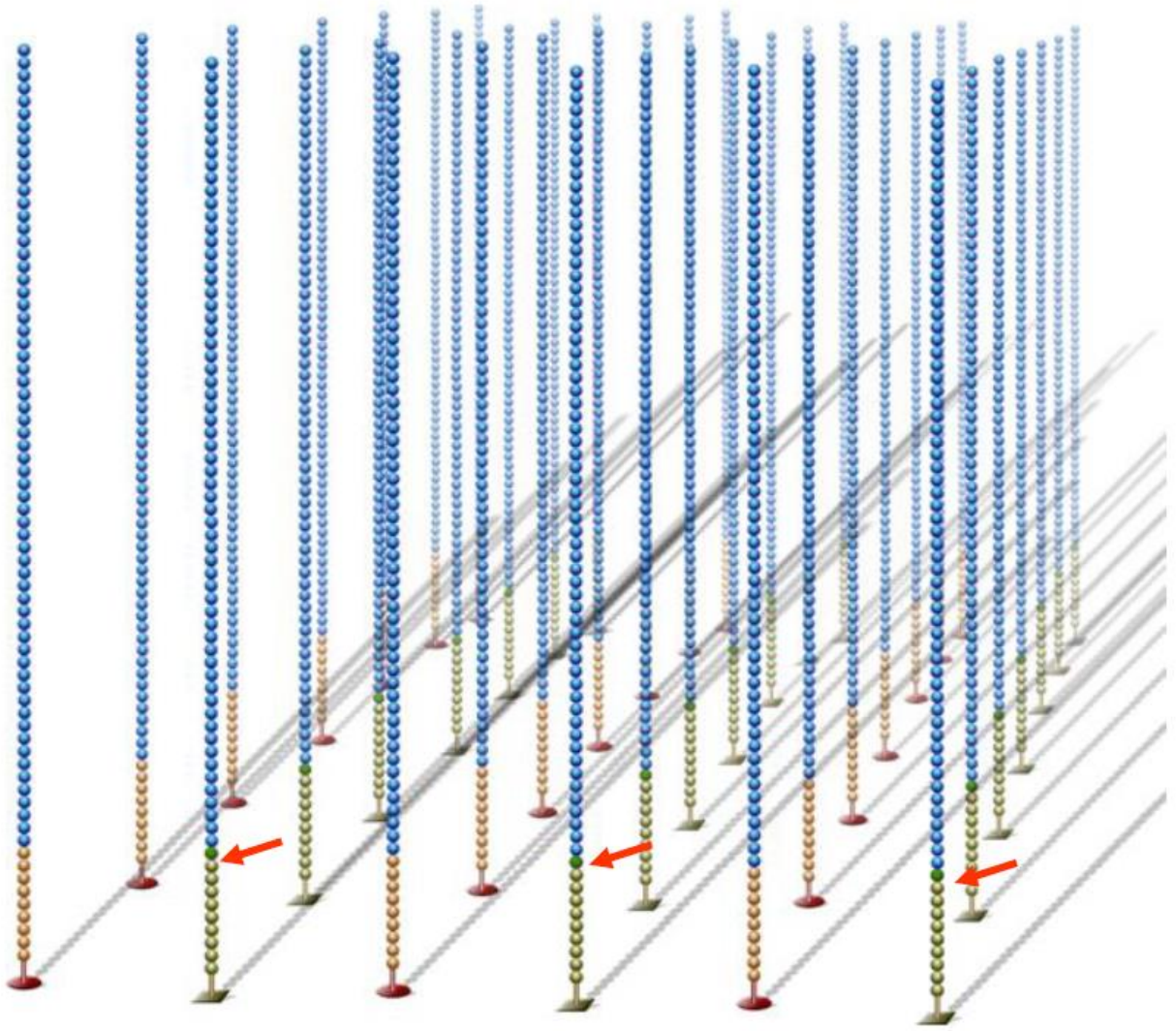
2nd step – bridge PCR

Bridge amplification cycle repeated until multiple bridges are formed



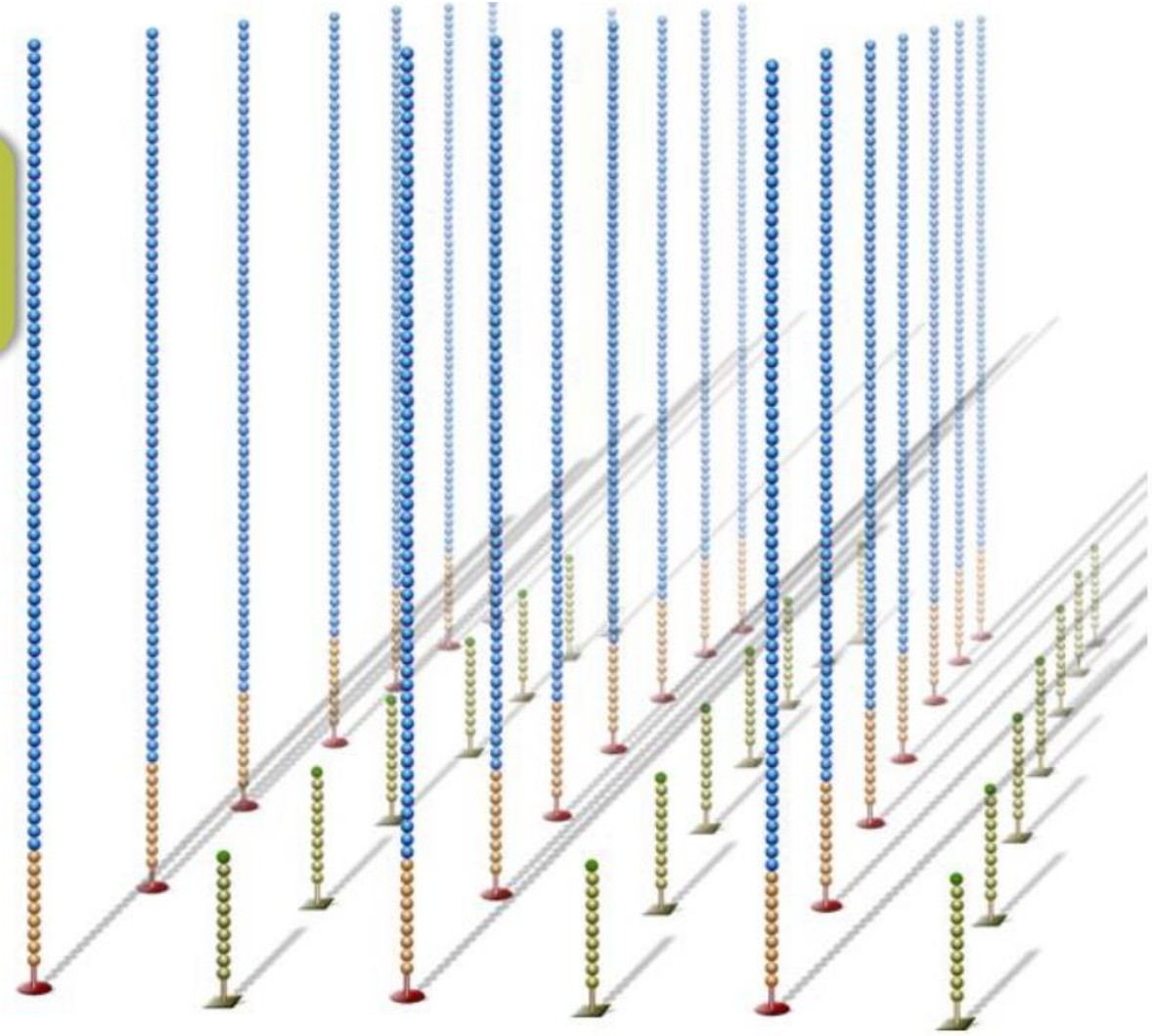
3rd step - linearisation

dsDNA bridges are denatured



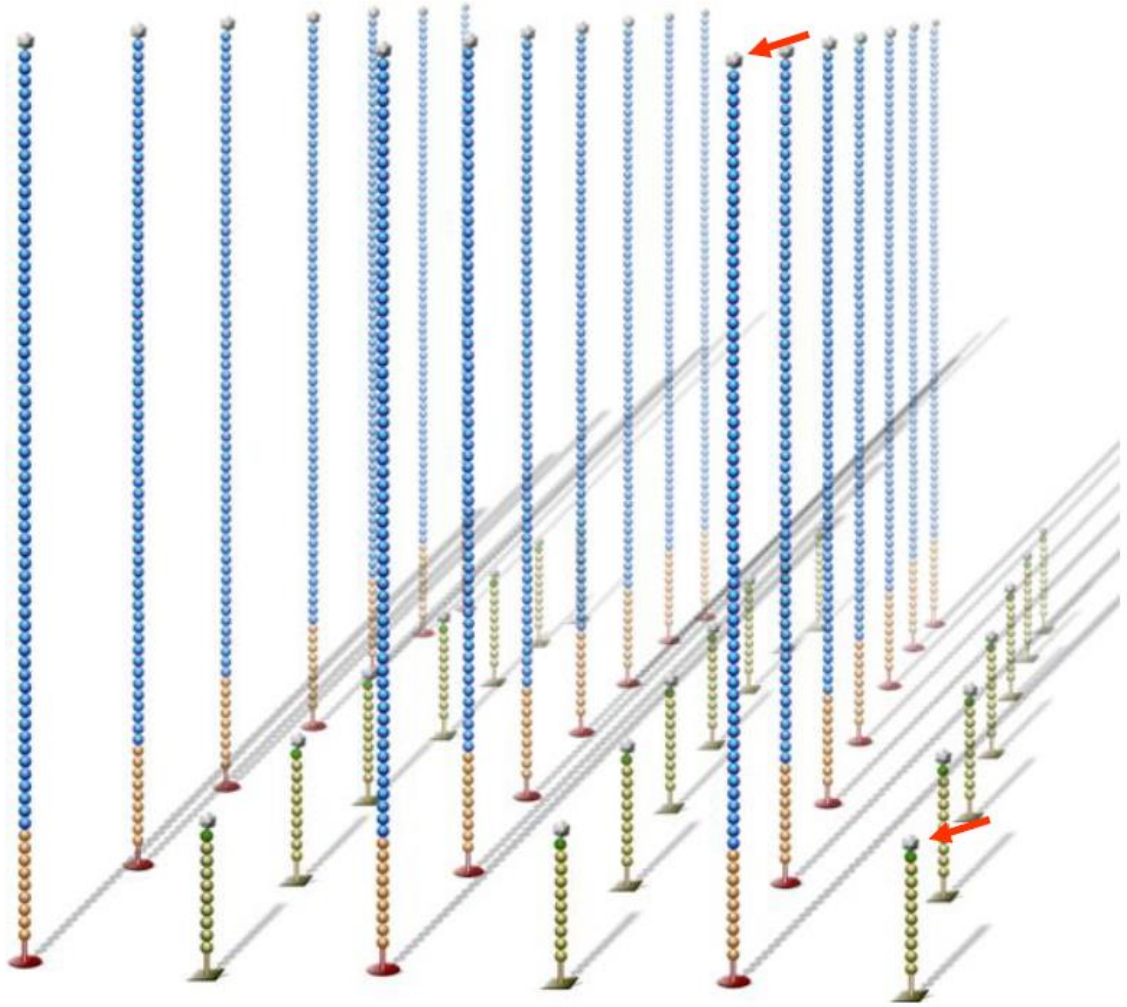
4th step – separation of reverse strand

Reverse strands cleaved and washed away, leaving a cluster with forward strands only



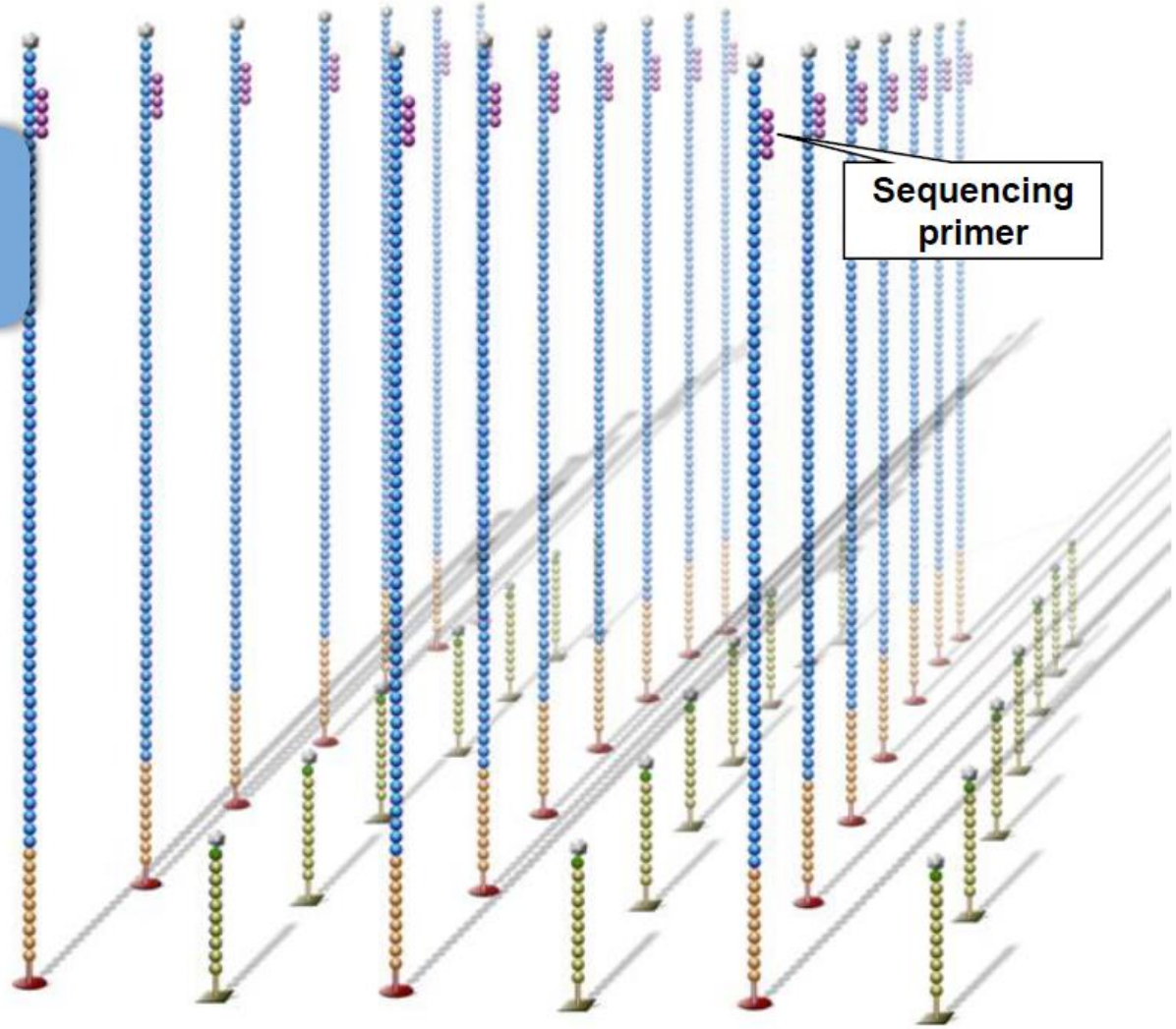
5th step – blocking of 5´end

Free 3’ ends are blocked to prevent unwanted DNA priming



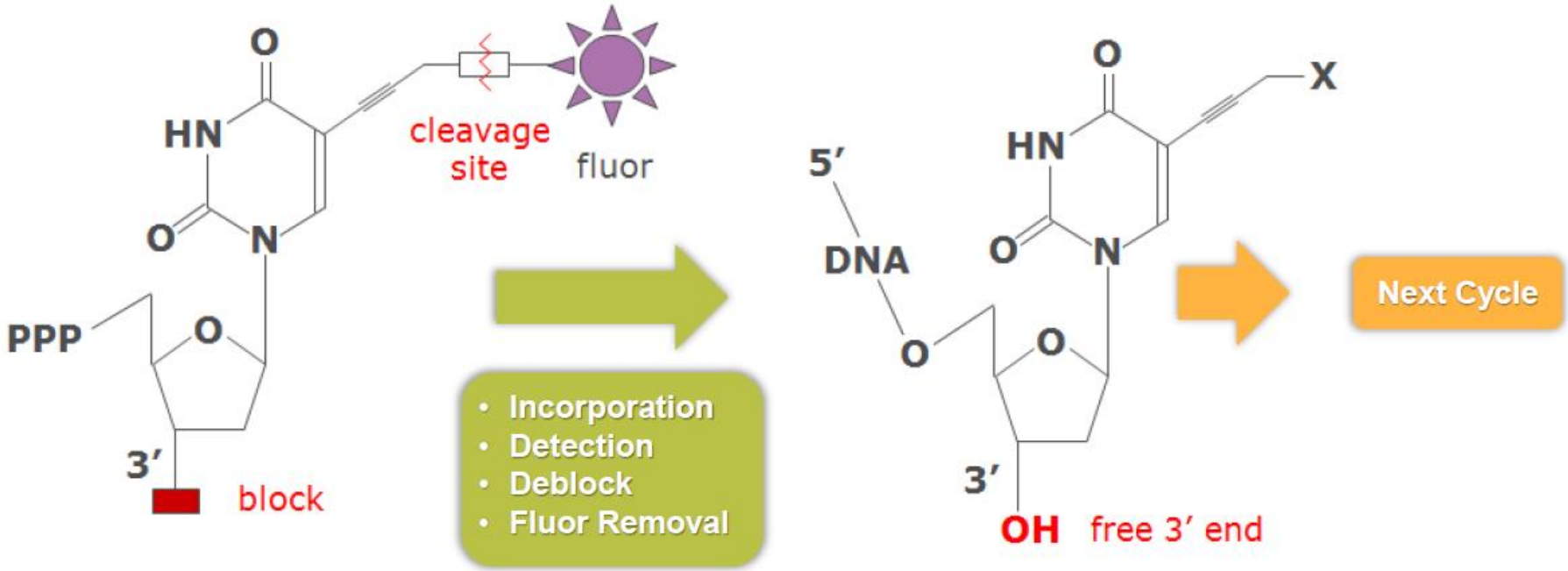
6th step – hybridization of seq. primer

Sequencing primer is hybridized to adapter sequence

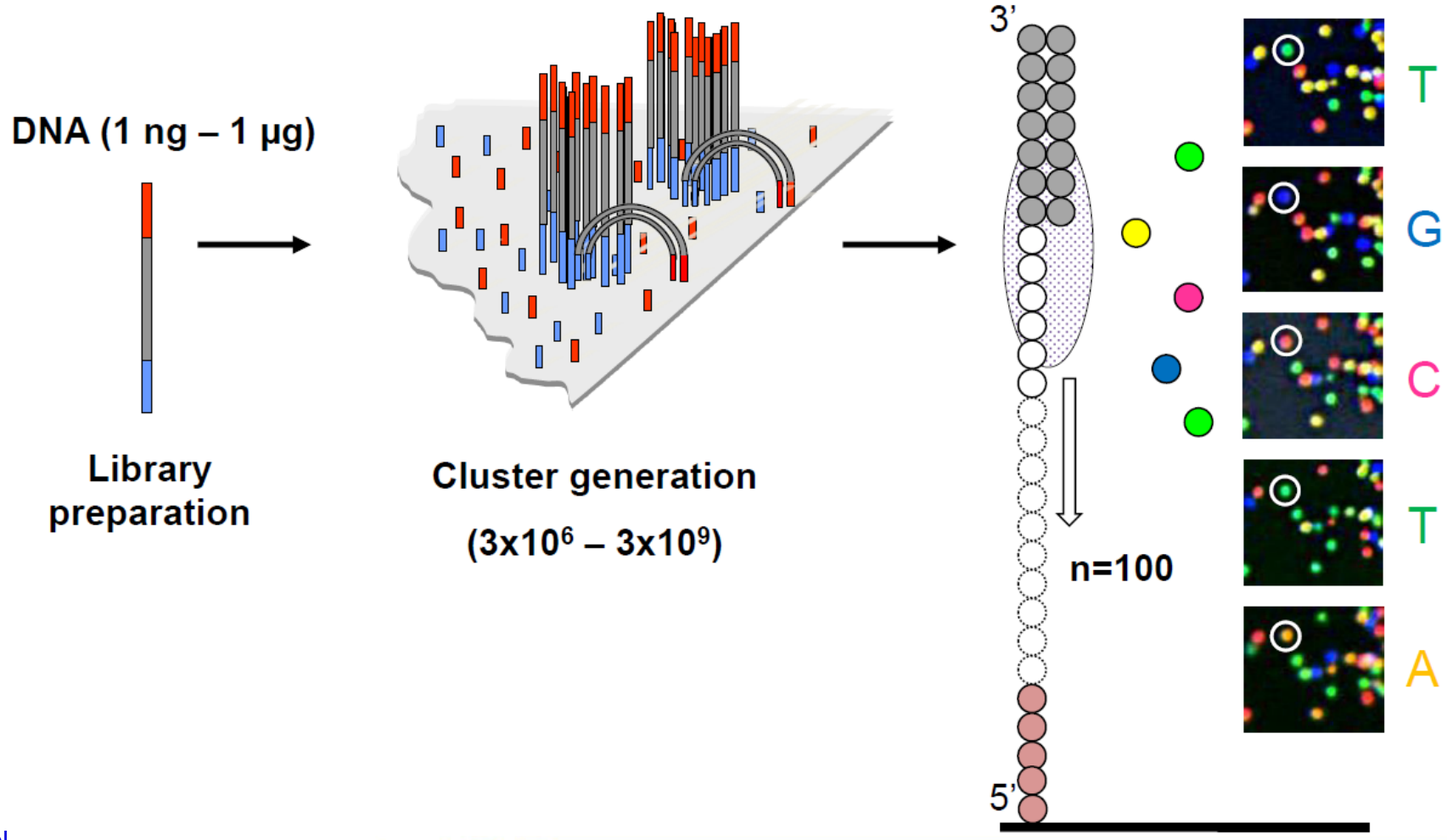


Reverse terminators

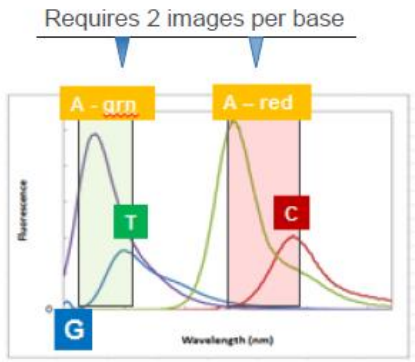
- All 4 nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats

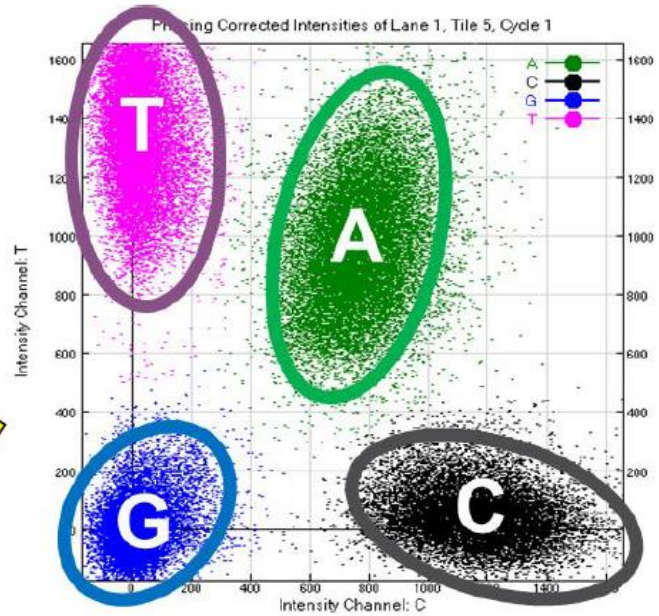


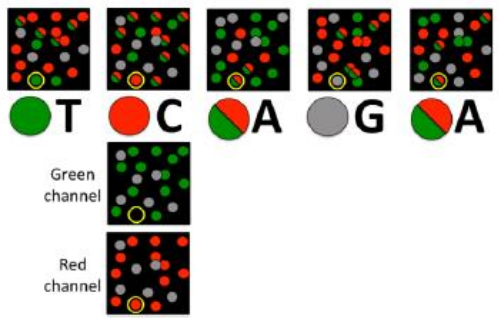
Sequencing by synthesis (SBS)



2-Channel SBS Chemistry: MiniSeq, NextSeq







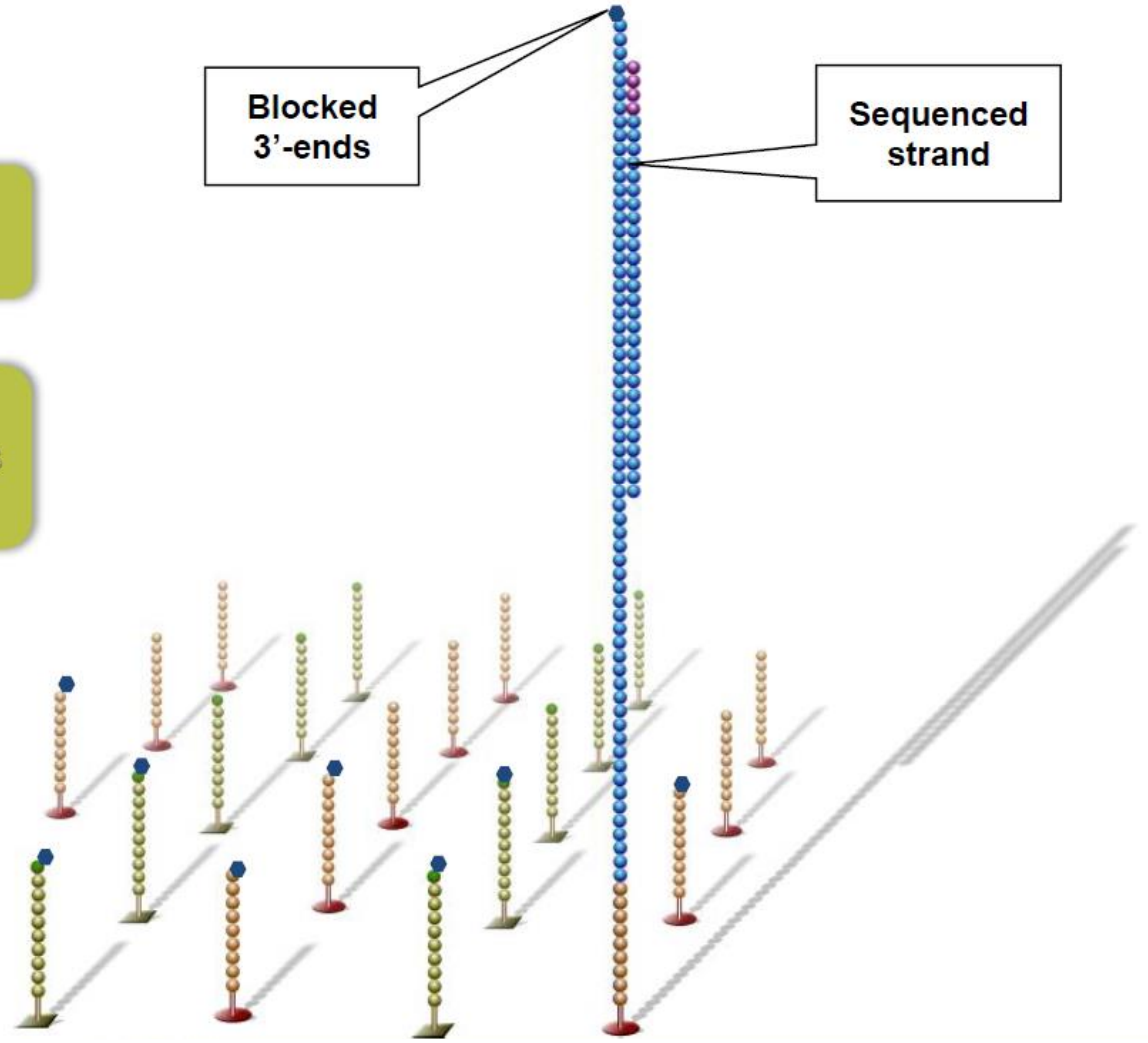
High performance with half the pictures

50% fewer images

7th step – pair-end sequencing

Sequenced strand is stripped off

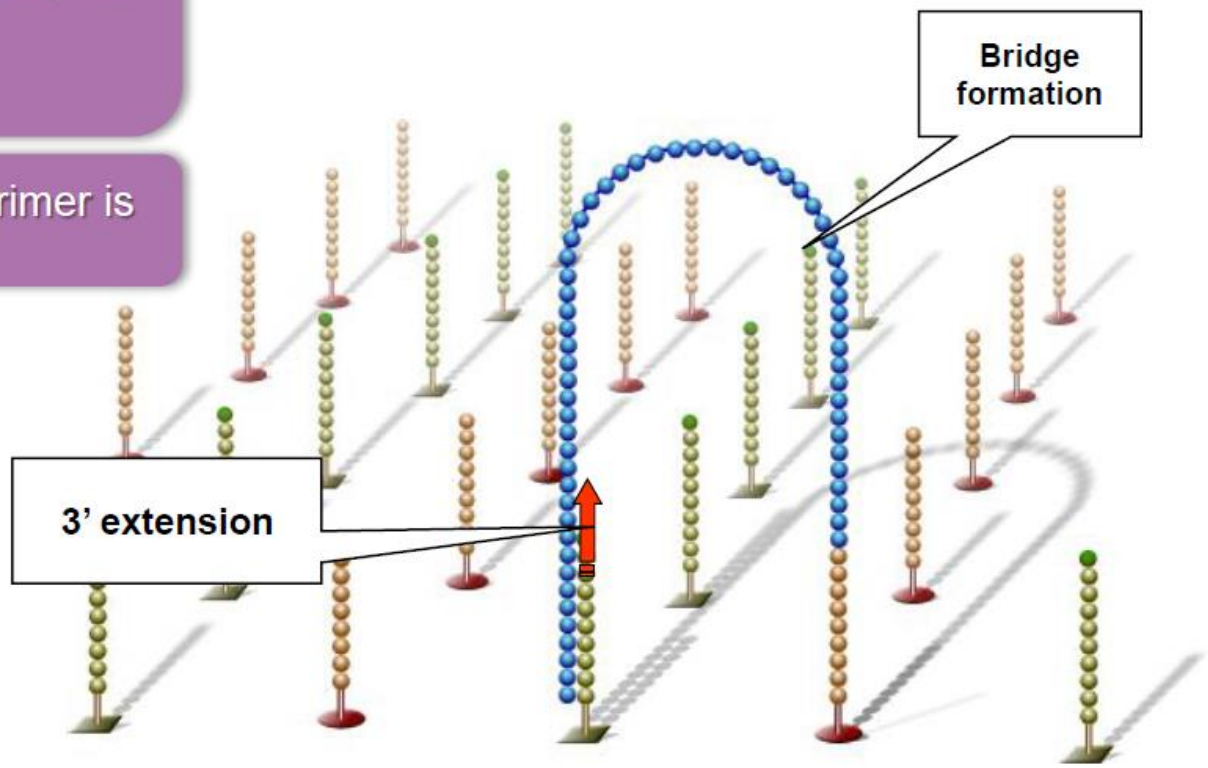
3’-ends of template strands and lawn primers are unblocked



7th step – pair-end sequencing

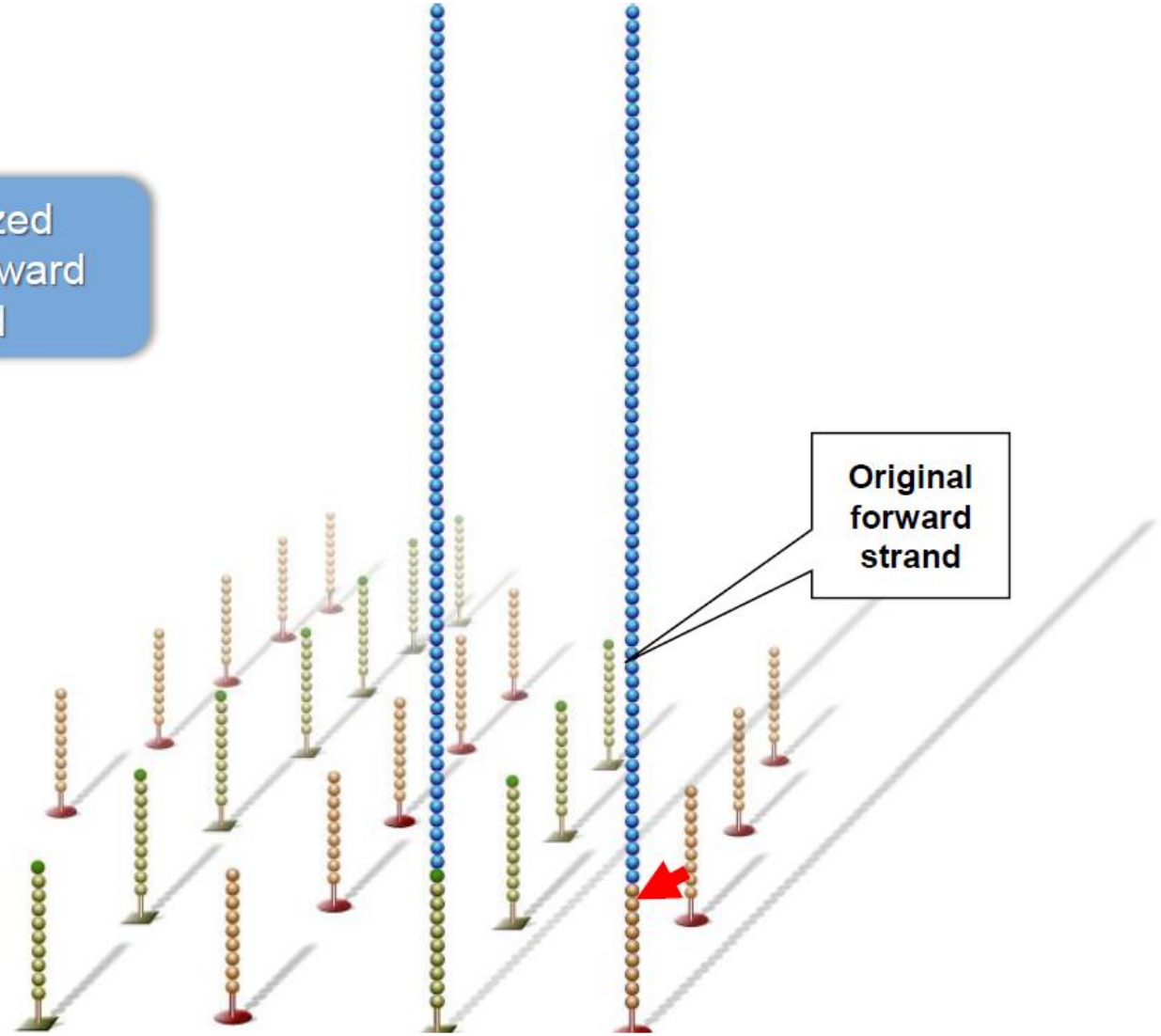
Single-stranded template loops over to form a bridge by hybridizing with a lawn primer

3'-ends of lawn primer is extended



7th step – pair-end sequencing

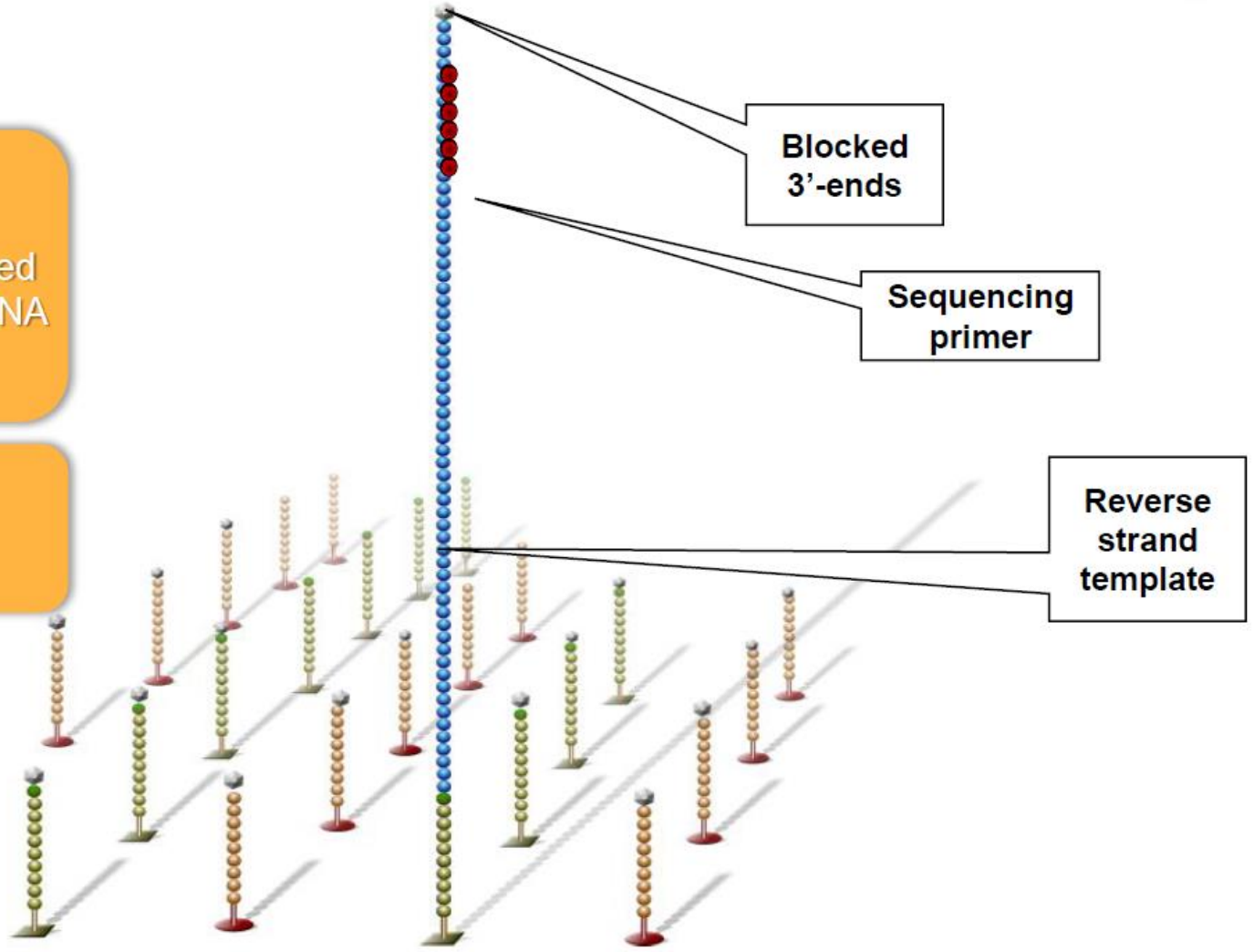
Bridges are linearized and the original forward template is cleaved



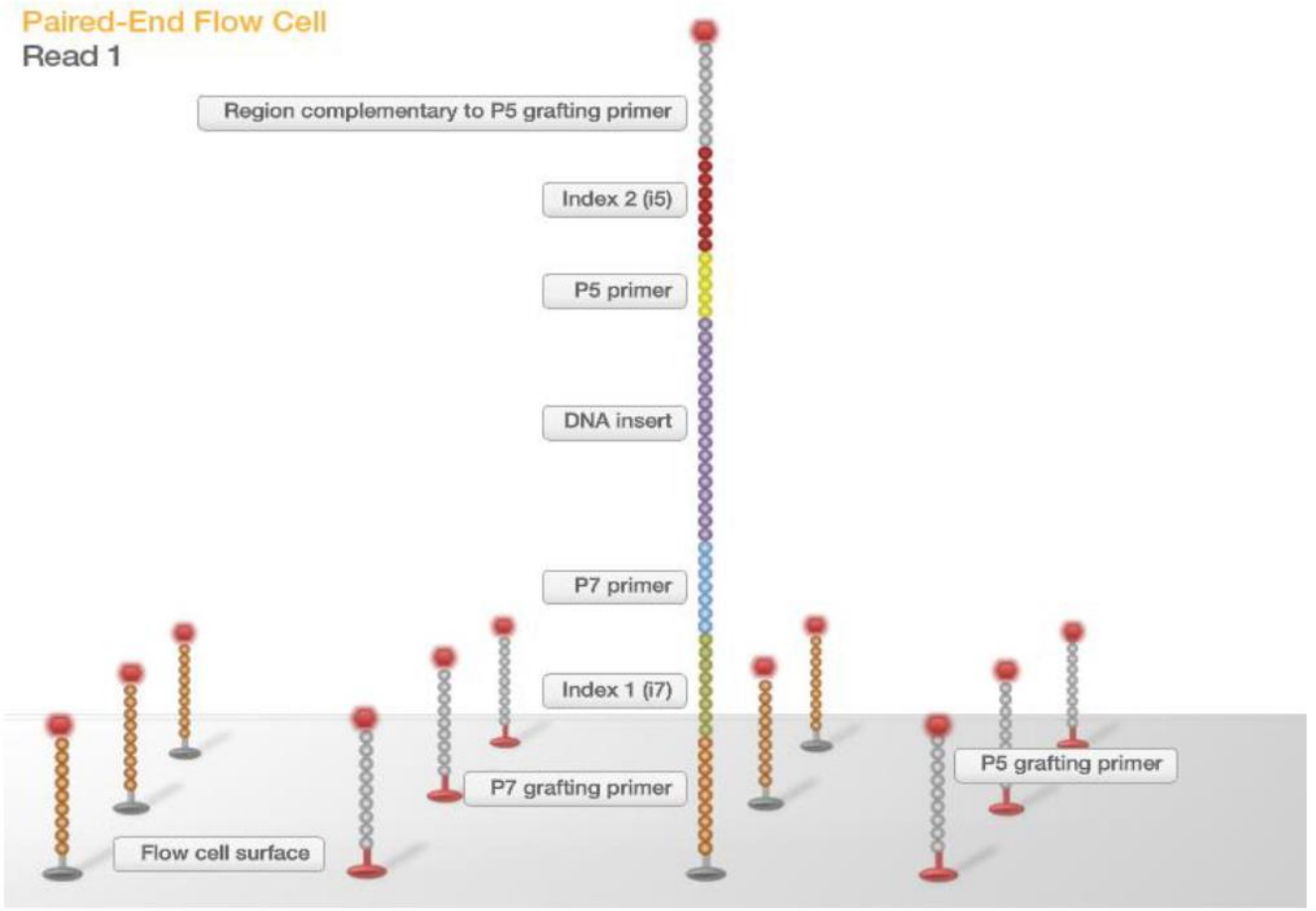
7th step – pair-end sequencing

Free 3’ ends of the reverse template and lawn primers are blocked to prevent unwanted DNA priming

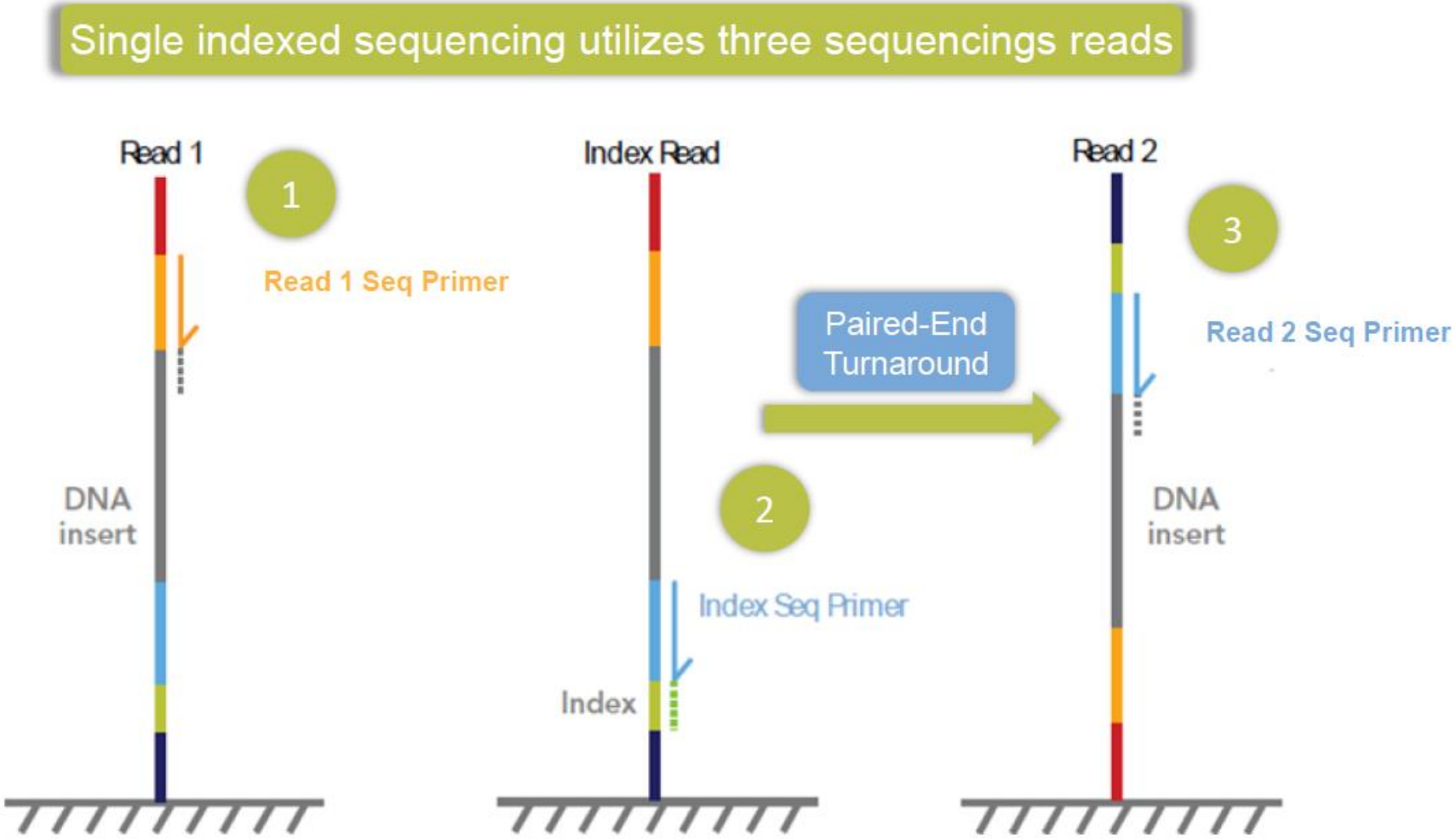
Sequencing primer is hybridized to adapter sequence



Index read

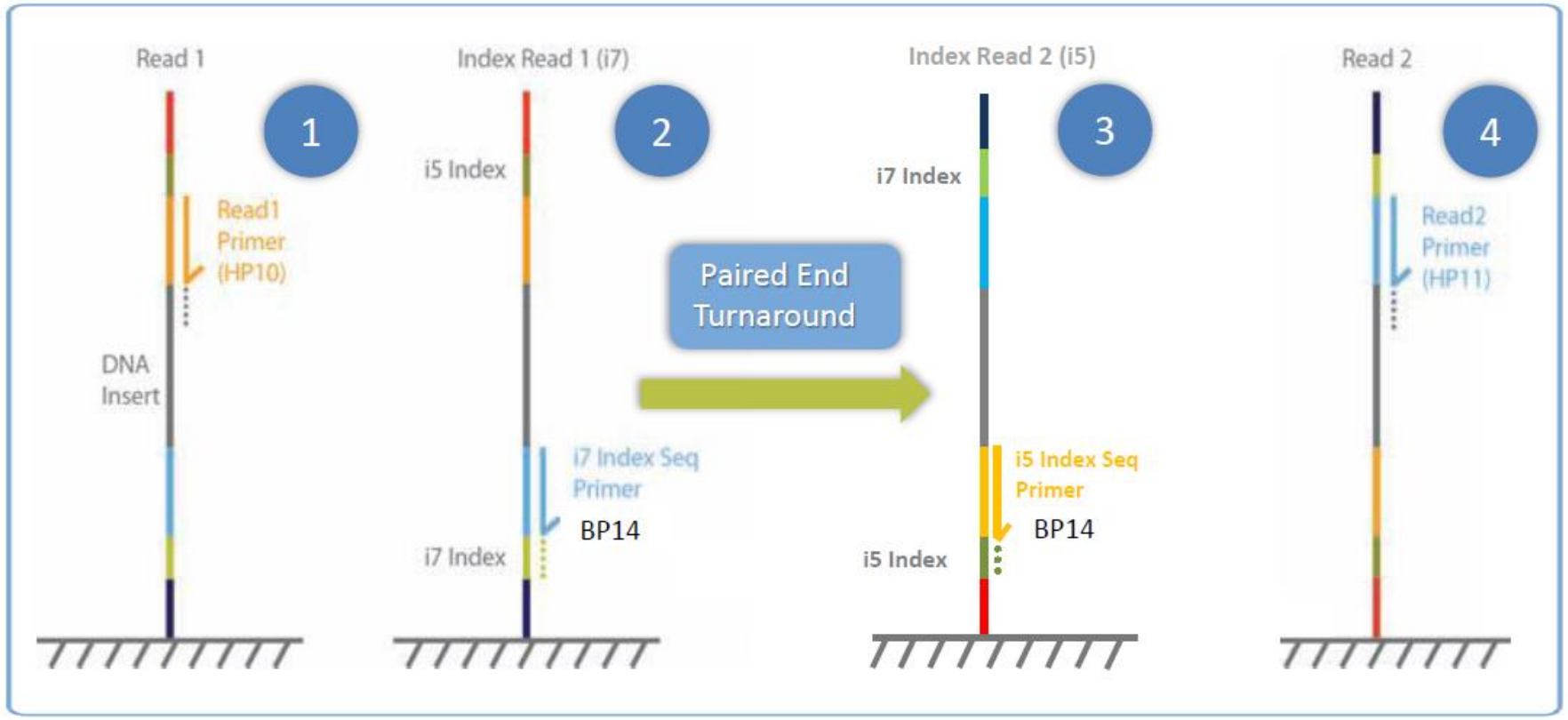

Paired-End Flow Cell
Read 1
Region complementary to P5 grafting primer
Index 2 (i5)
P5 primer
DNA insert
P7 primer
Index 1 (i7)
P7 grafting primer
P5 grafting primer
Flow cell surface

Single index read



Dual index read

iSeq, MiniSeq, NextSeq,HiSeq



PGM analyser (ion torrent)

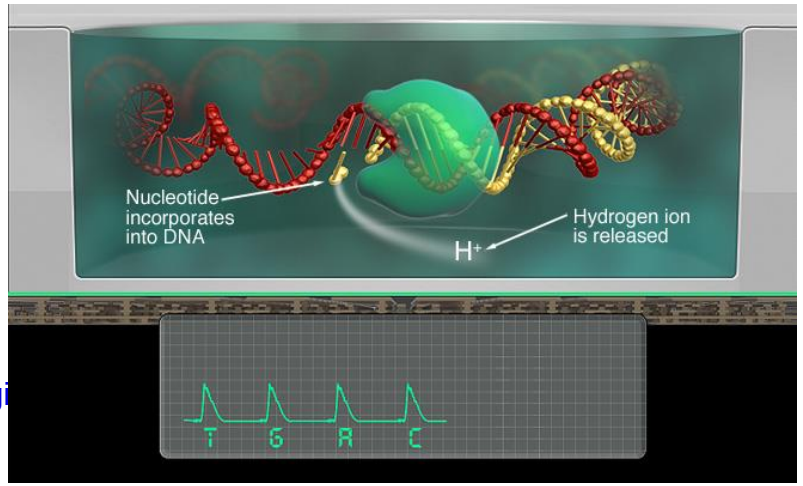
Application

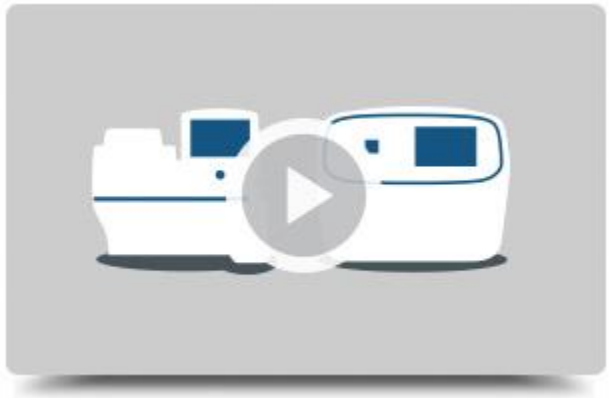
Targeted

Exome

Transcriptome

Genome

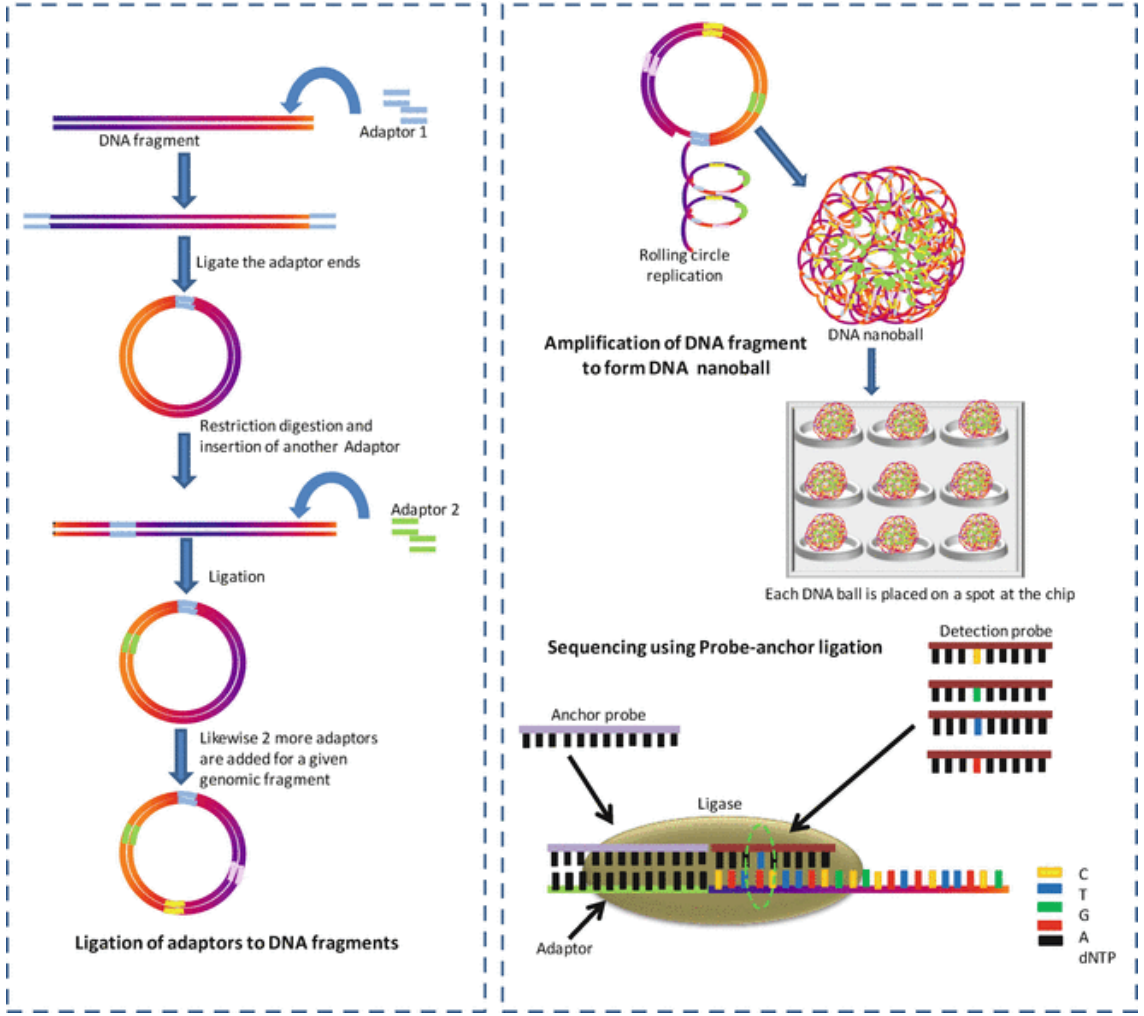




DNBSeq (MGI)

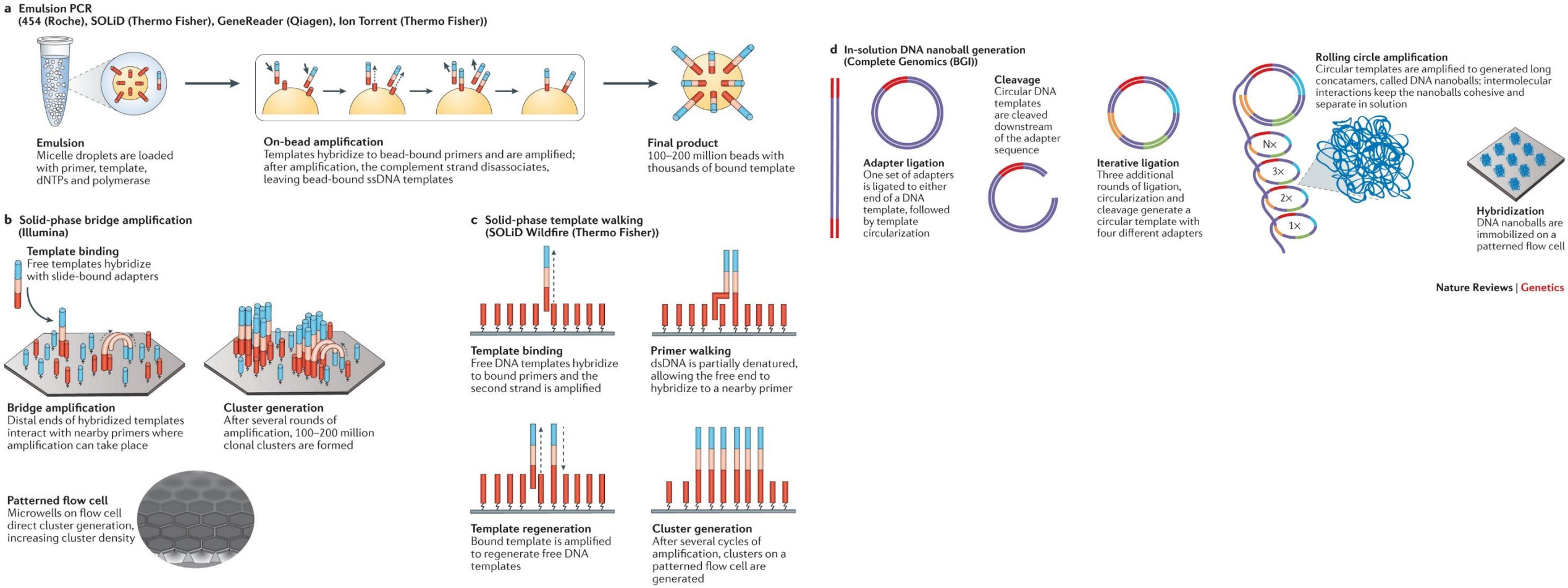
- it uses phi29 polymerase for amplification of one-strand template
- this process creates nanoballs
- sequencing cell contains regions with positive charge for binding of nannoballs
- different technology of sequencing

https://en.mgi-tech.com/products/



https://youtu.be/xUVdJN0m38c

Current techniques of 2nd generation



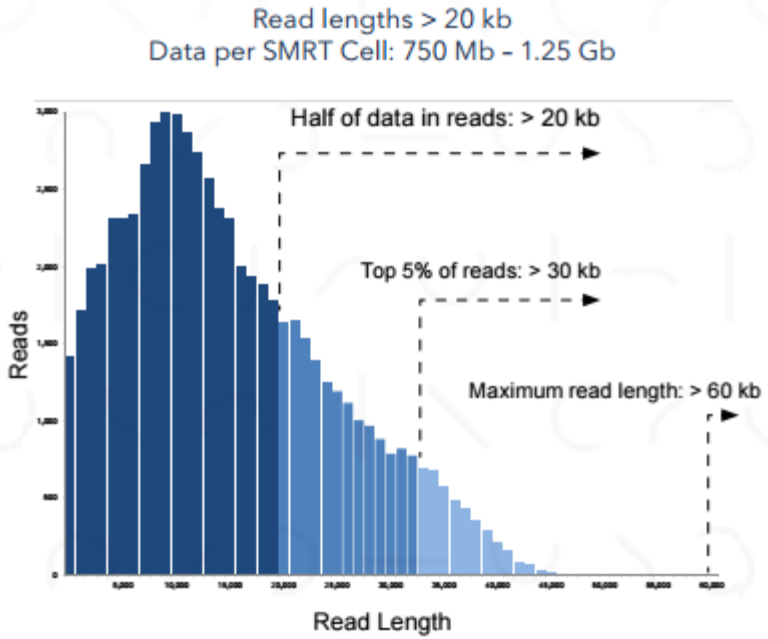
3rd generation of sequencers (PACBIO)

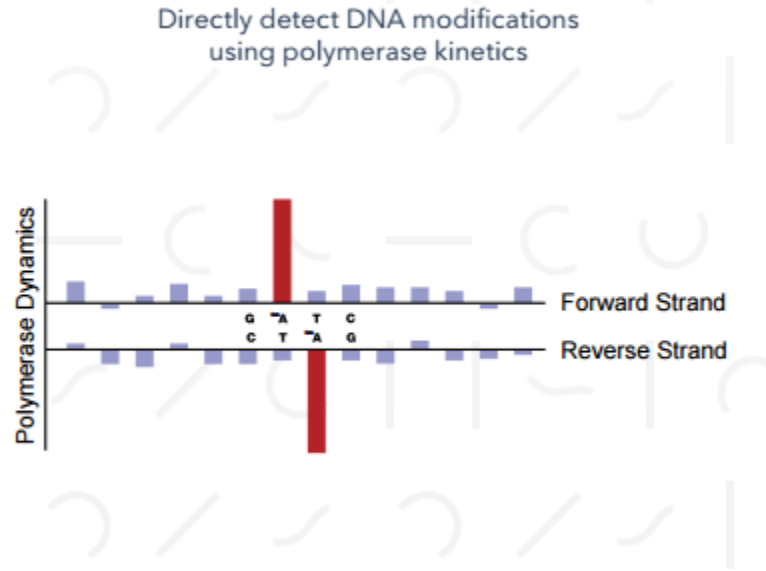
Sequel System

PacBio RS II

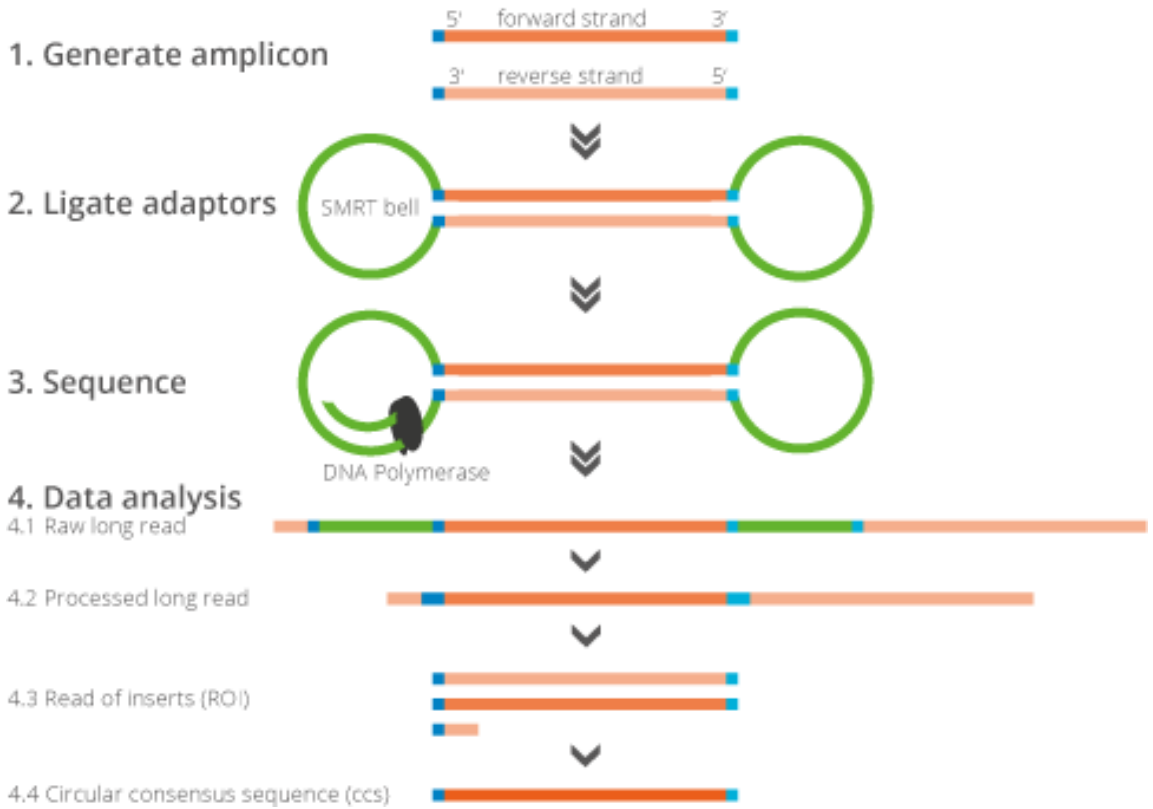
PACBIO

- Sequencing based on Single Molecule, Real-Time (SMRT®) technology
- It uses so-called Zero-Mode Waveguides (ZMWs) enabling the illumination of only the lower part of the well, in which the DNA polymerase is immobilized at the bottom
- The main advantage is the possibility of long reads (up to 20 kb)
- Another advantage is the possibility of direct detection of methylated bases (epigenome)





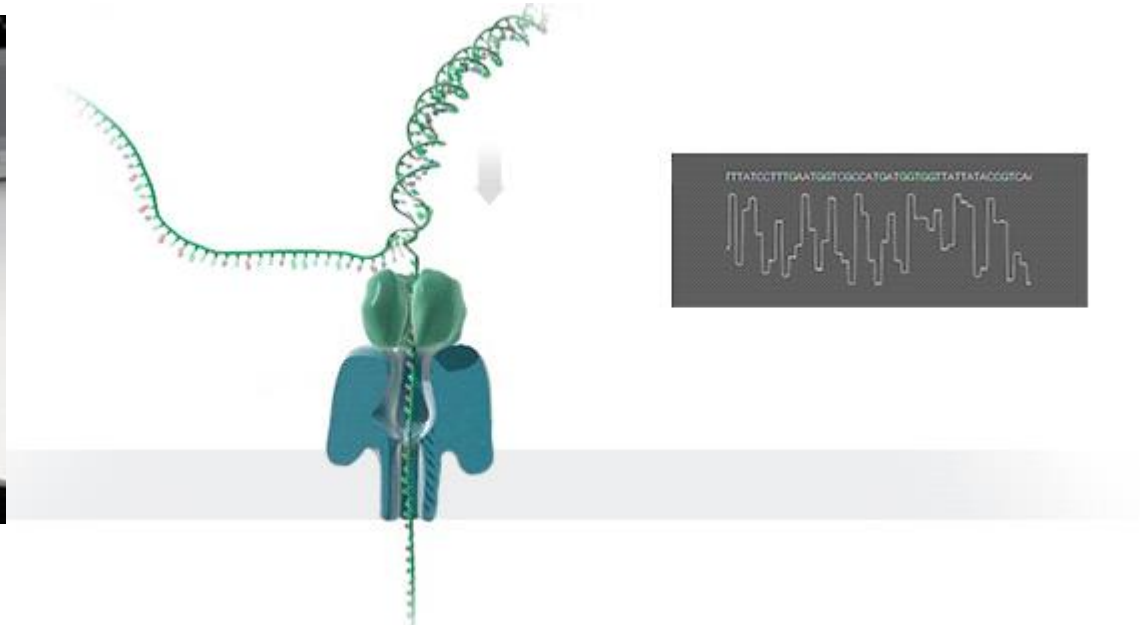
Library preparation



https://www.youtube.com/watch?v=v8p4ph2MAvI

3rd generation – Oxford Nanopores

- The technology is based on nanopores
- At the beginning of sequencing, NA is bound to a nanopore formed by a protein
- It is then denatured and passes through the nanopore, generating a change in current
- Based on the observed change, individual bases are read in real-time
- Enables sequencing of very long chains (tens to hundreds of kilobases)
- The disadvantage is a higher error rate, correctness >95%



MUNI
SCI

GENE TECHNOLOGIES

Methods of studying gene expression and function
Mapping techniques, DNA libraries, gene expression, metagenomics

Mapping techniques

- Genome maps provide a series of markers for assembling sequence data:
- Creation of a genome map:

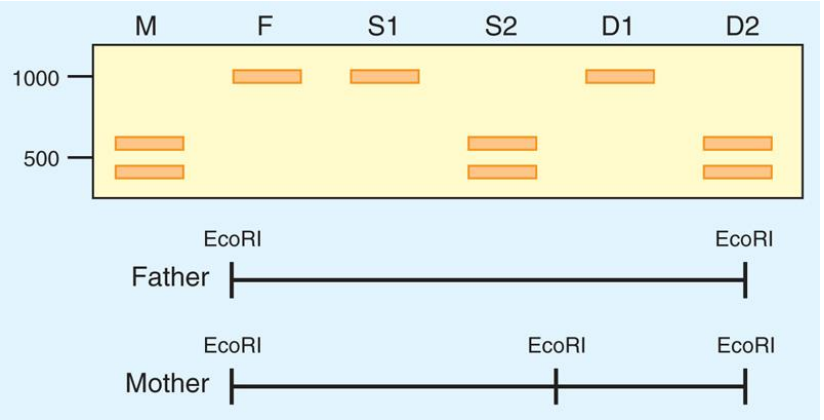
  - genetic maps (crossbreeding, pedigree analysis, gene transfer) - linkage maps
  - physical maps (radiation hybrid panel, FISH)

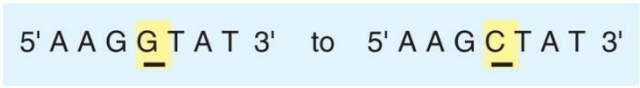
- Genetic maps based on linkage = the probability that two mapped markers will separate from each other in a cross
- To determine the relative distance of markers, the percentage of times they are found together is crucial
- A variety of markers are used today

Type of mapping	Markers	Methods of localization
Genetic	Gene, biochemical properties, DNA markers (RFLP, VNTRs, microsatellite, SNPs)	Linkage analysis using crossing or mating Kinship analysis
Physical	STSs, EST, VNTRs, microsatelites	Restriction analysis, Radiation hybrid panel, FISH, Cytogenetic mapping

Genetic markers

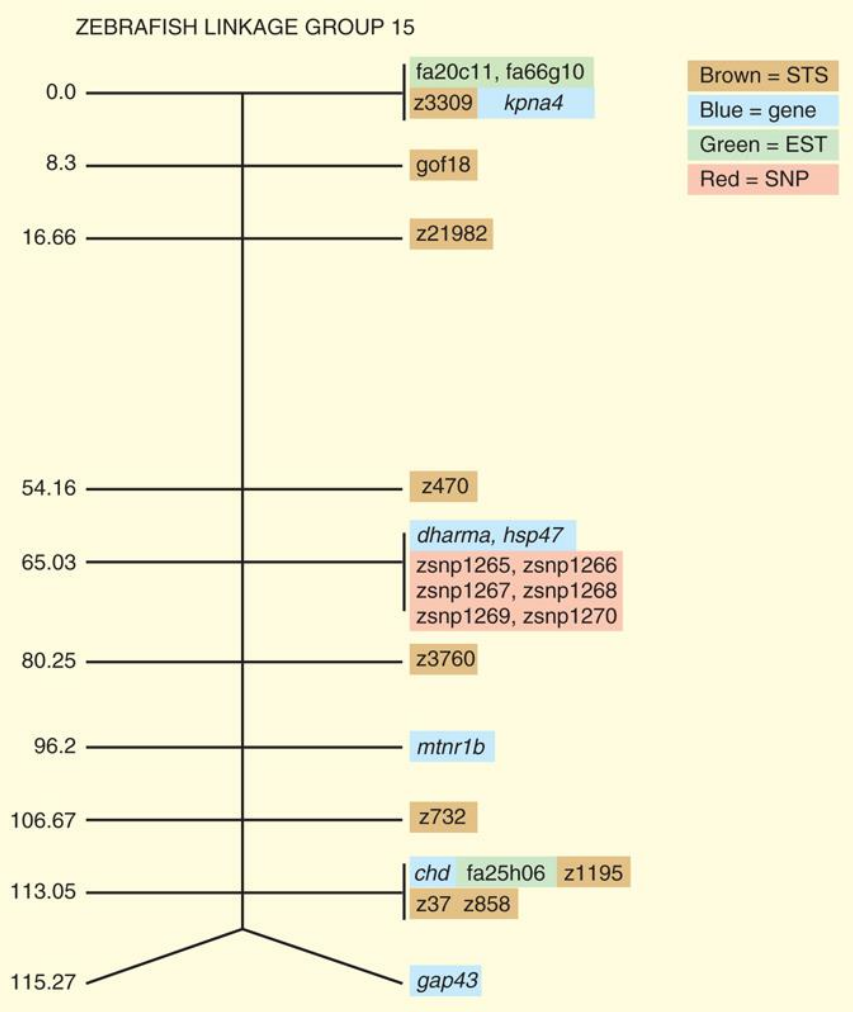
- RFLP analysis of related individuals, easy identification
- Variable Number Tandem Repeat (VNTR, minisatellites) – tandem repeats with a length of 9-80bp (forensic testing, paternity tests)
- Microsatellite polymorphism – tandem repeat of 2-5bp length
- Single Nucleotide Polymorphism (SNP)
- SNPs, VNTRs RFLPs are also used in physical mapping
- For large genomes we need additional markers
  - STSs (Sequence Tagged Sites) – the unique sequence of 100-500 bp
  - ESTs (Expressed Sequence tags) – identification in cDNA libraries
- Digestion of gDNA using restriction enzymes - physical mapping method

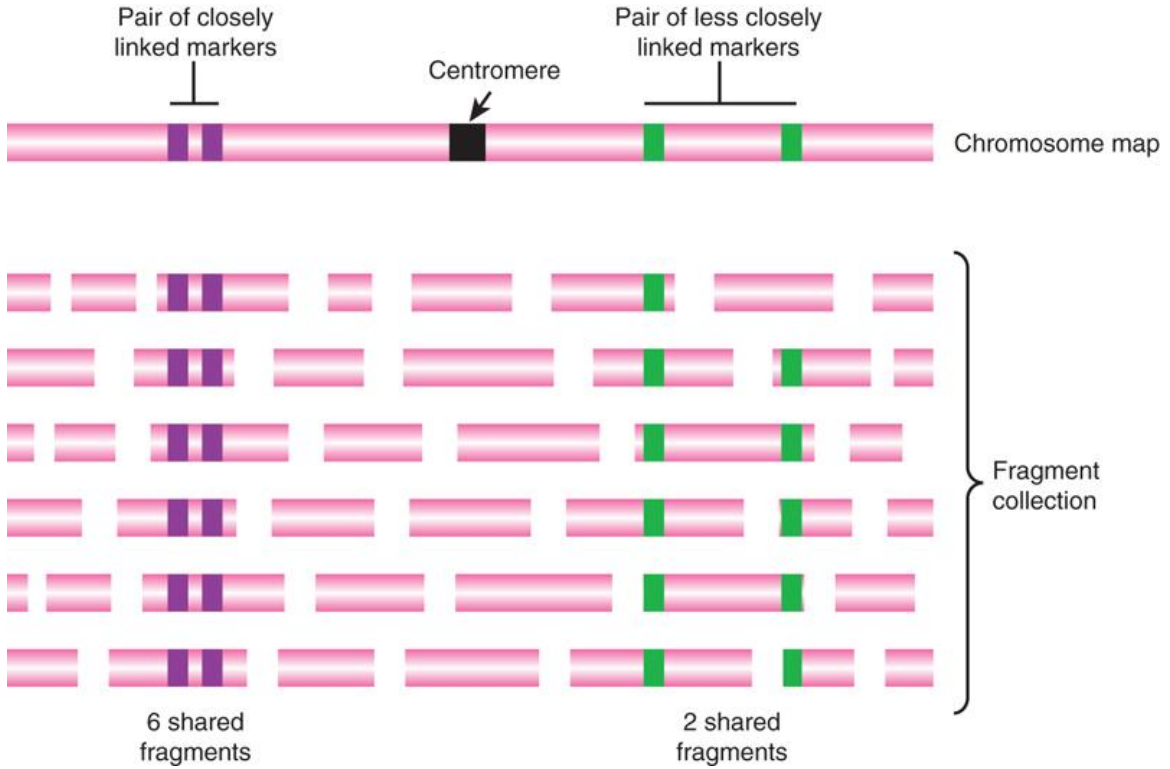




Clark and Pazdernik, 2016

Genetic markers

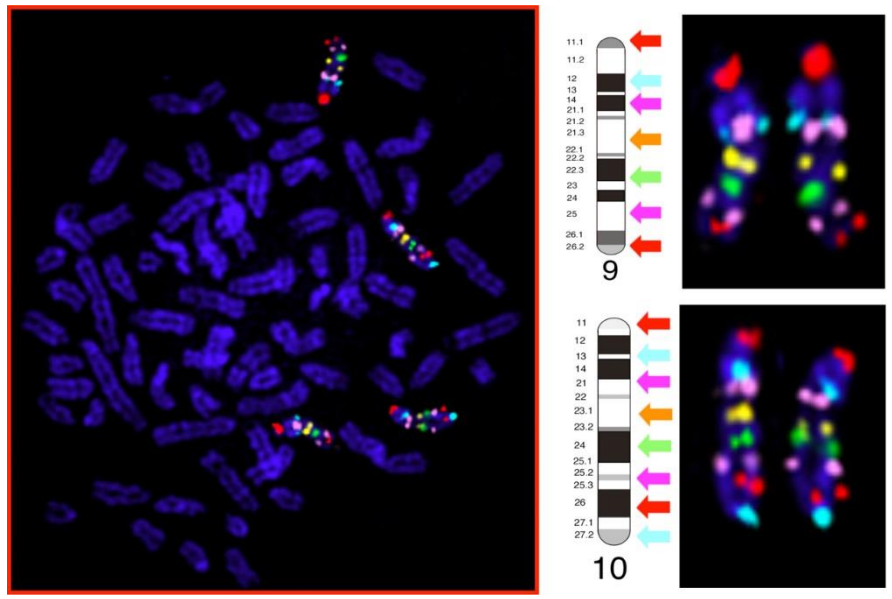




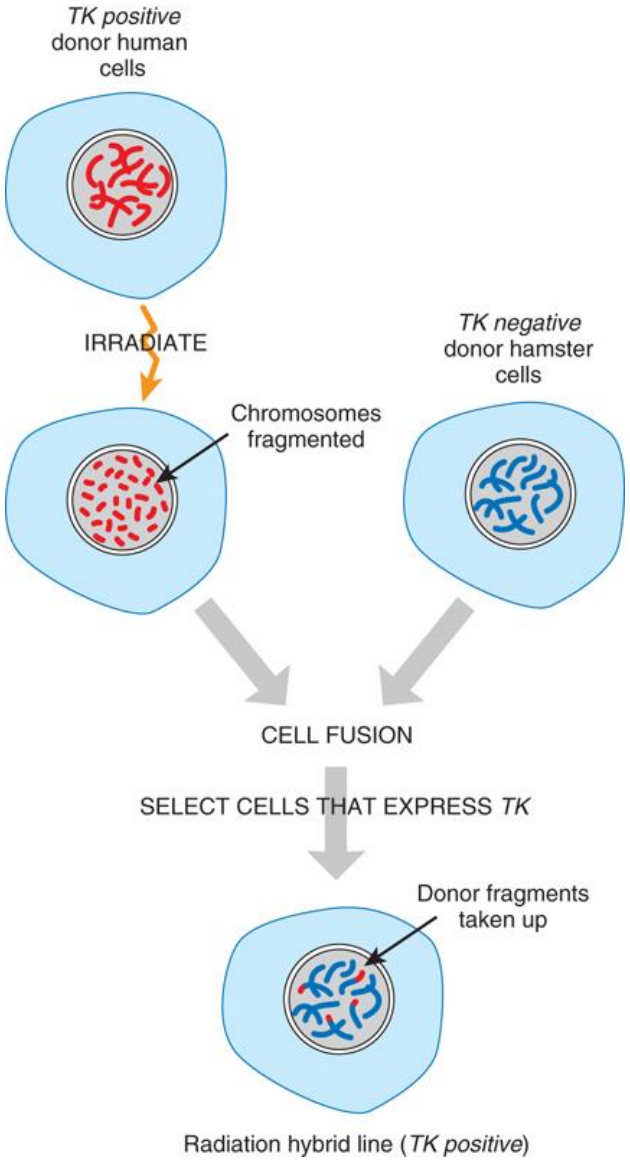
Clark and Pazdernik, 2016

Physical mapping techniques

– FISH (Fluorescence in-situ hybridization) – the location of a specific DNA sample on chromosomes in metaphase relative to banding (chromosome painting)



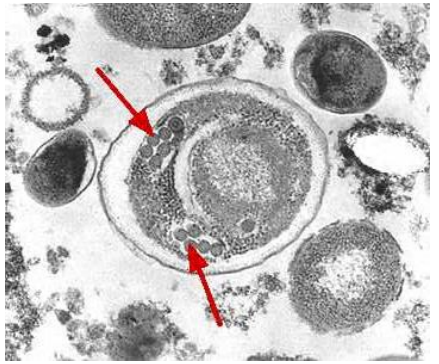
– radiation hybrid mapping – large segments of the cloned genome may contain two fragments from different parts of the genome



Clark and Pazdernik, 2016

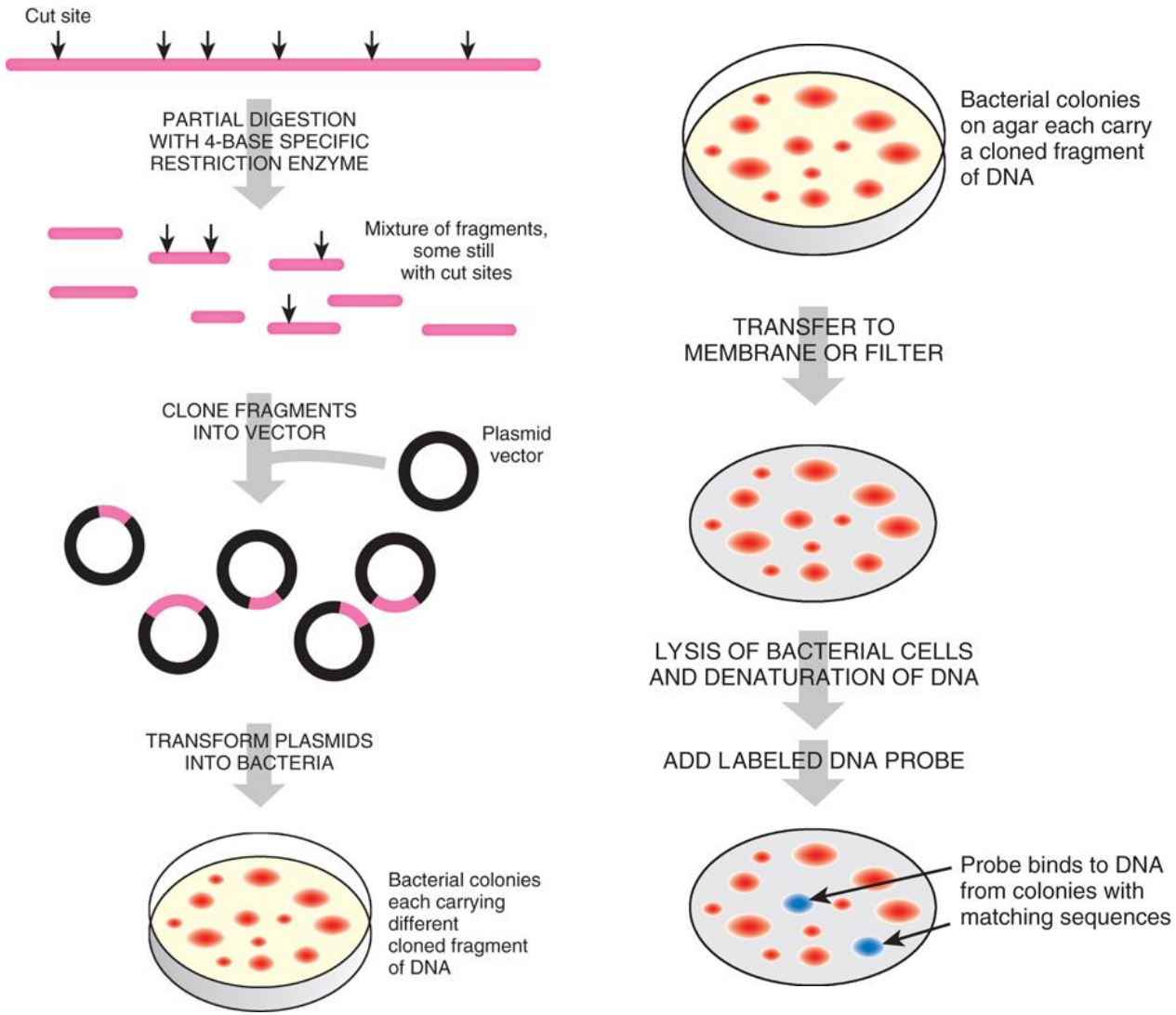
Number of genes x Genome

Organismus	Velikost genomu (Mbp)	Počet protein-kódujcích genů
Wheat	17 000	95 000
Rice	520	45 000
Paris Japonica (Pieris japonský)	149 000	26 000
Trichomonas vaginalis	160	46 000
Encephalozoon intestinalis	2.25	1833
Marbled lungfish	130 000	?
Human	3200	21 850
Nematode	97	20 493
Fruit fly	180	13 600
Streptomyces coelicolor	8.7	7800
E. coli	4.6	4300
Mycoplasma genitalium	0.58	470



DNA libraries

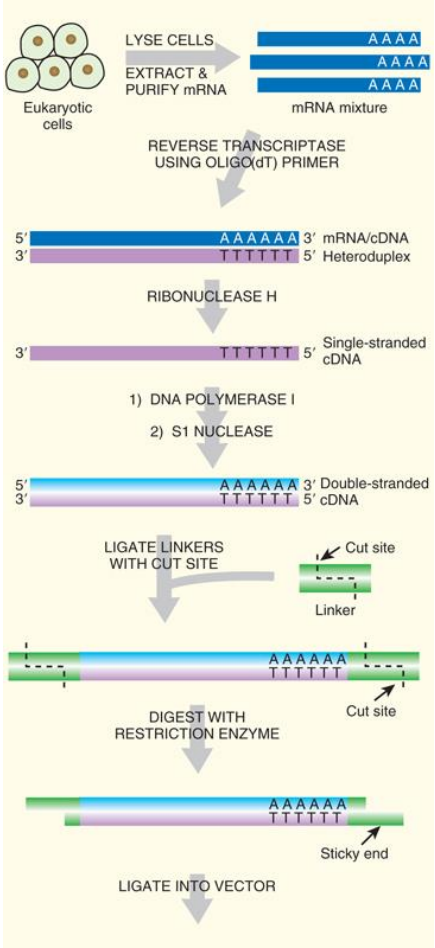
- Used for:
  - finding new genes
  - genome sequencing
  - comparison of genes from different organisms
- Basic steps in creating a library:
  - isolation of chromosomal DNA
  - cleavage of DNA with a restriction enzyme
  - linearization of the vector
  - insertion of fragments into the vector
  - transformation into *E. coli*

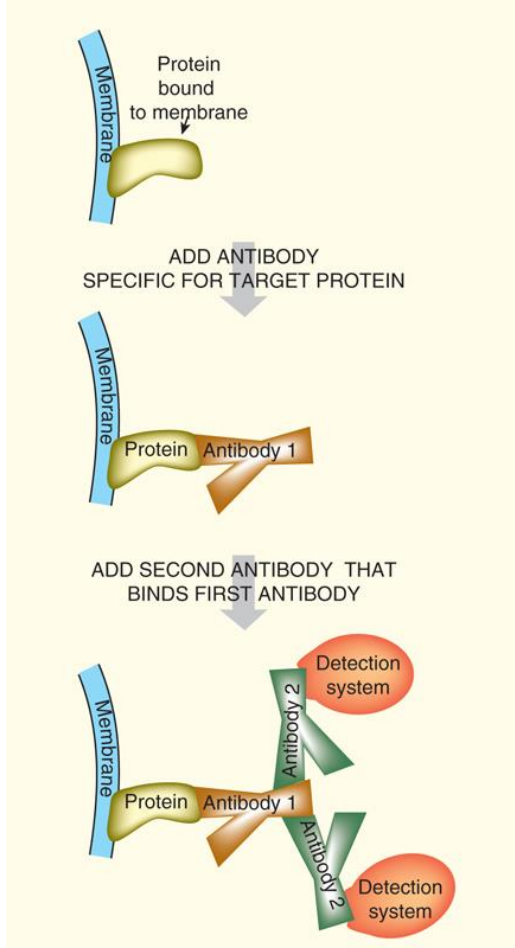


Clark and Pazdernik, 2016

Eukaryotic expression libraries

- The vector contains the sequence necessary for transcription and translation
- Constructed from complementary DNA (cDNA)
- Identification of new genes, splicing variants

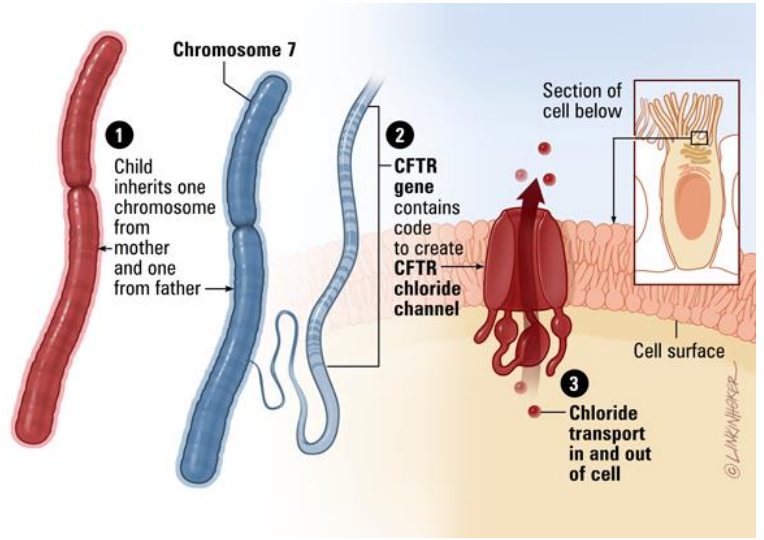


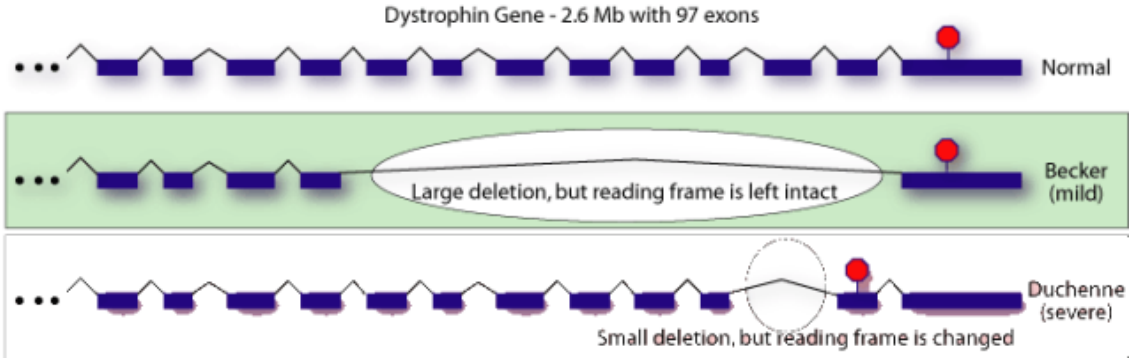


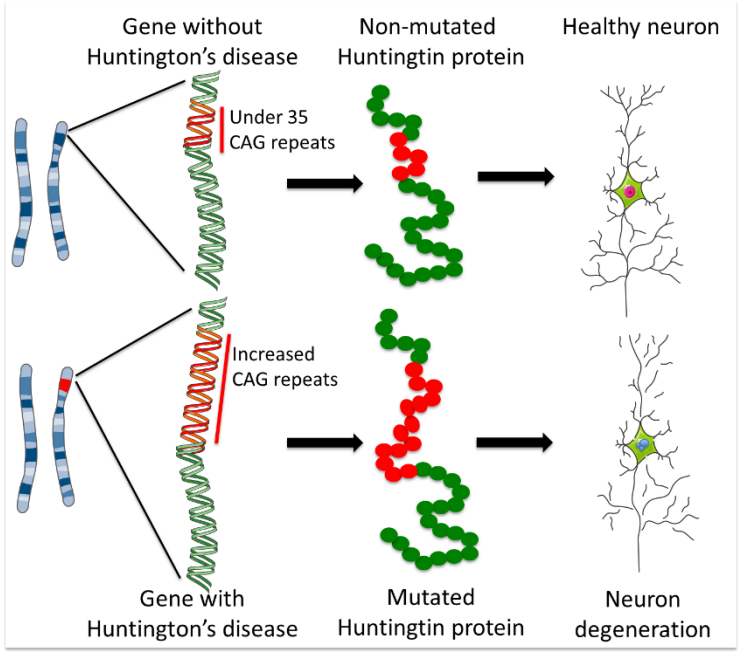
Clark and Pazdernik, 2016

Medical genomics

- The largest application of genomic data in disease diagnosis
- Genetic testing – determination of the presence of a gene associated with the disease:
  - muscular dystrophy (dystrophin gene)
  - cystic fibrosis (CFTR gene)
  - Huntington's disease (HTT gene)

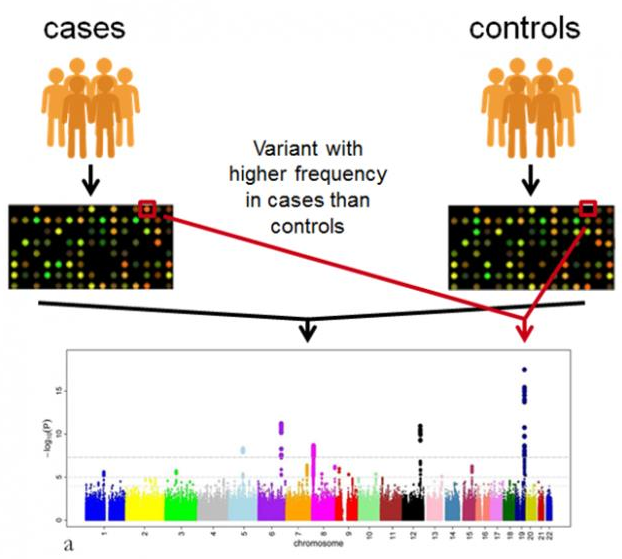






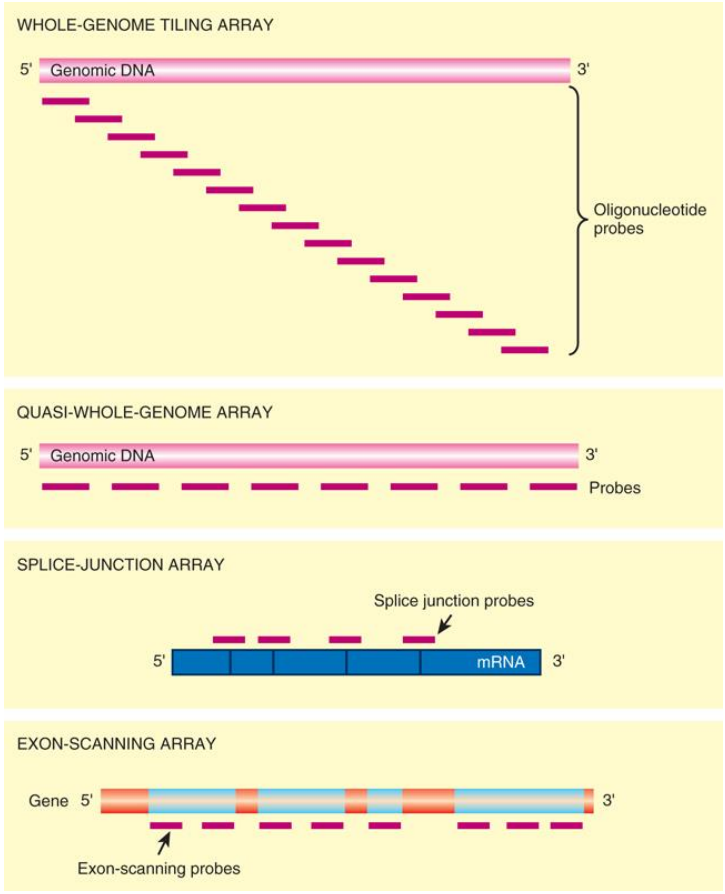
Medical genomics

- To identify causal mutations, it is more advantageous to sequence the exome (2%) than the genome
- Currently, more than 3,000 diseases have been identified using genomics and pedigree analysis
  - the so-called Mendelian disease (a mutation in one gene leads to the disease)
- Many diseases are polygenic (contribution of multiple genes to the development of the disease)
  - Crohn's disease
  - autoimmune disease
  - psychiatric disorders (schizophrenia, AD, mild cognitive impairment)
- Within these diseases, the use of GWAS (genome-wide association study)
  - analysis of single point polymorphisms (SNPs)
  - frequency lower than 1%
  - influence of genotype and environment on disease development

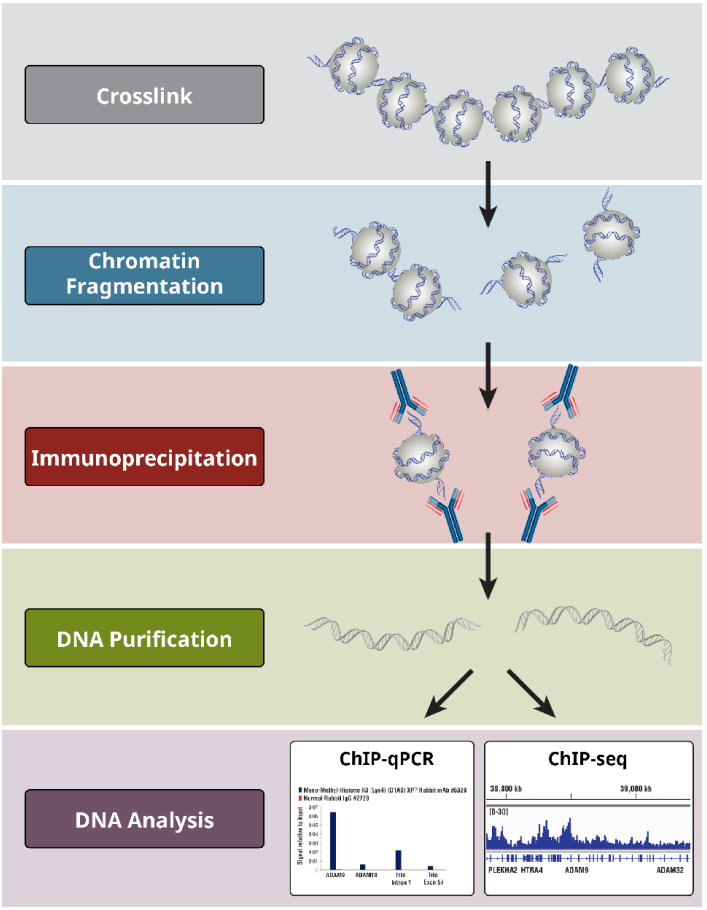


Gene expression – WGAs, ChIP

- WGAs (whole-genome tiling arrays cover all genome
- Firstly, in Arabidopsis (25-mer oligonucleotides)
- Discovery of new genes, splicing variants
- ChIP (chromatin immunoprecipitation):
  - analysis of DNA regions of individual transcription factors
  - DNA analysis of regions associated with histone PTMs



Clark and Pazdernik, 2016



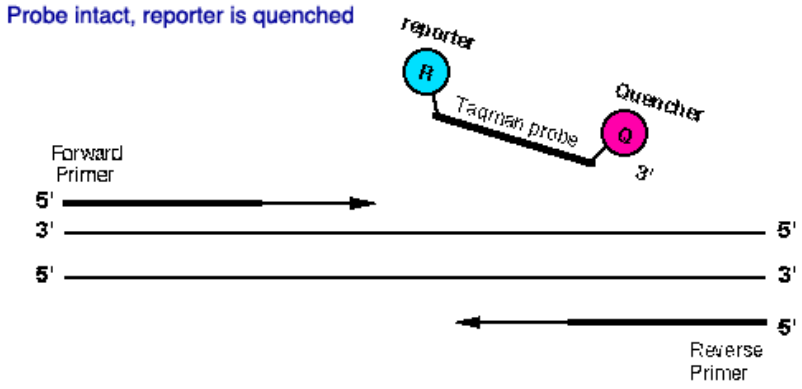
Gene expression – RT-qPCR

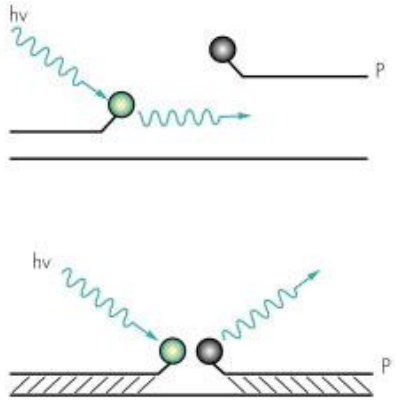
Using of specific fluorescence probes

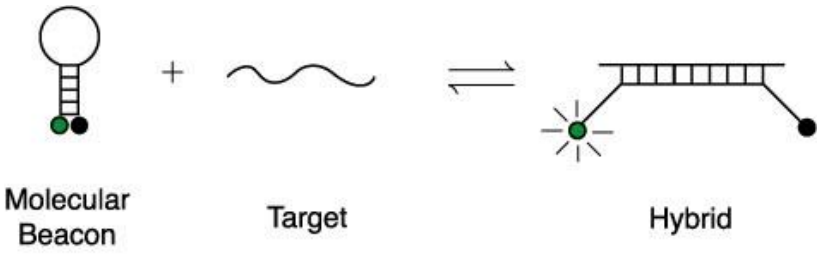
Using intercalating dye

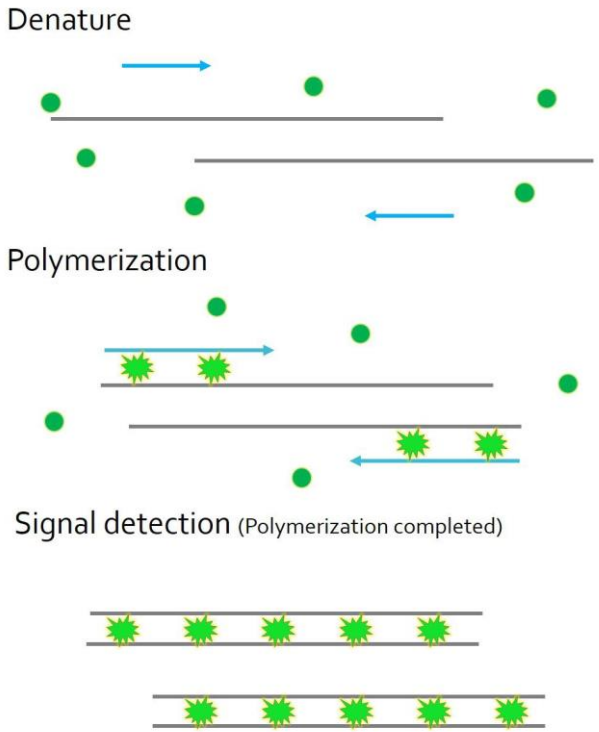
Hydrolysis probe
TaqMan

Hybridization probes
Molecular beacons, FRET



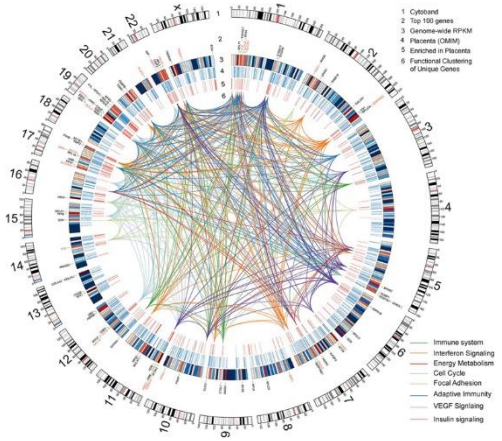


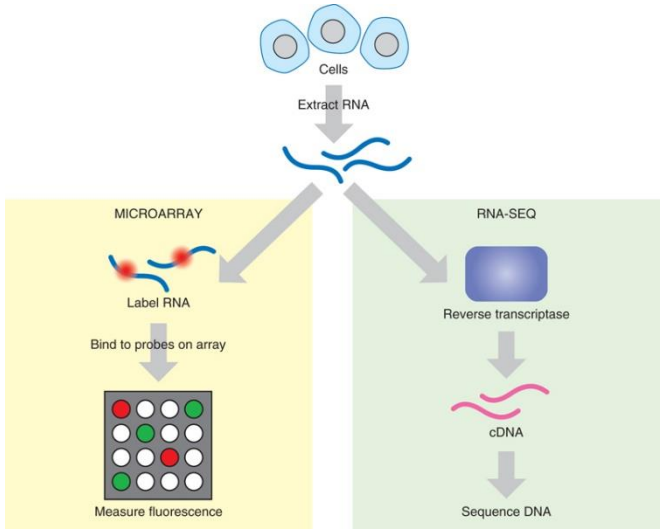


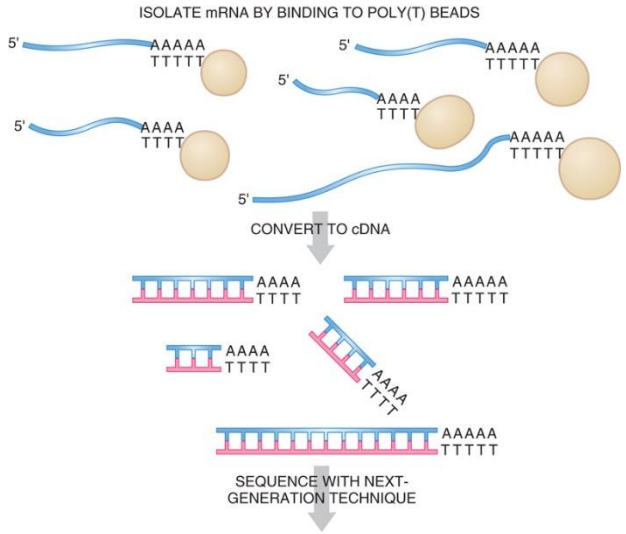


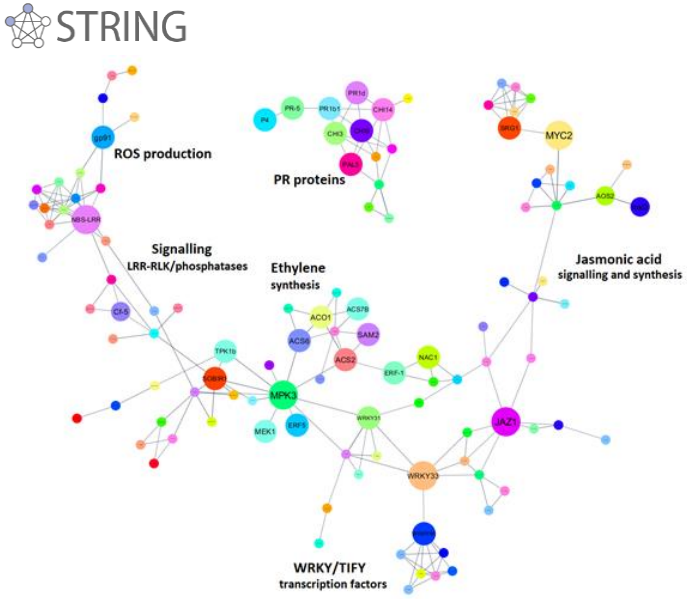
Gene expression - RNAseq

- Advantages of the RNAseq method:
  - does not depend on probes (more correct quantification of given RNA molecules)
  - large dynamic range
  - detection of alternative splicing and the possibility of their quantification
  - the possibility of analysis without knowledge of the genome sequence
  - the possibility of analysis from one cell





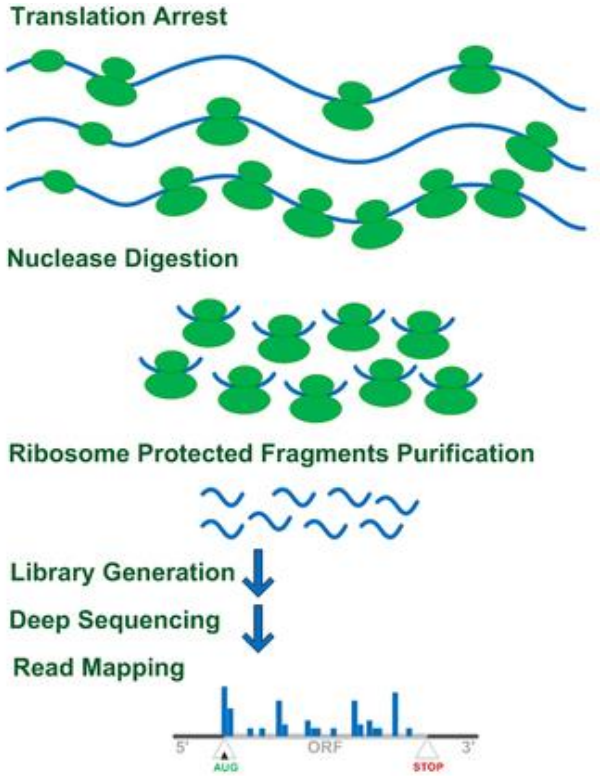


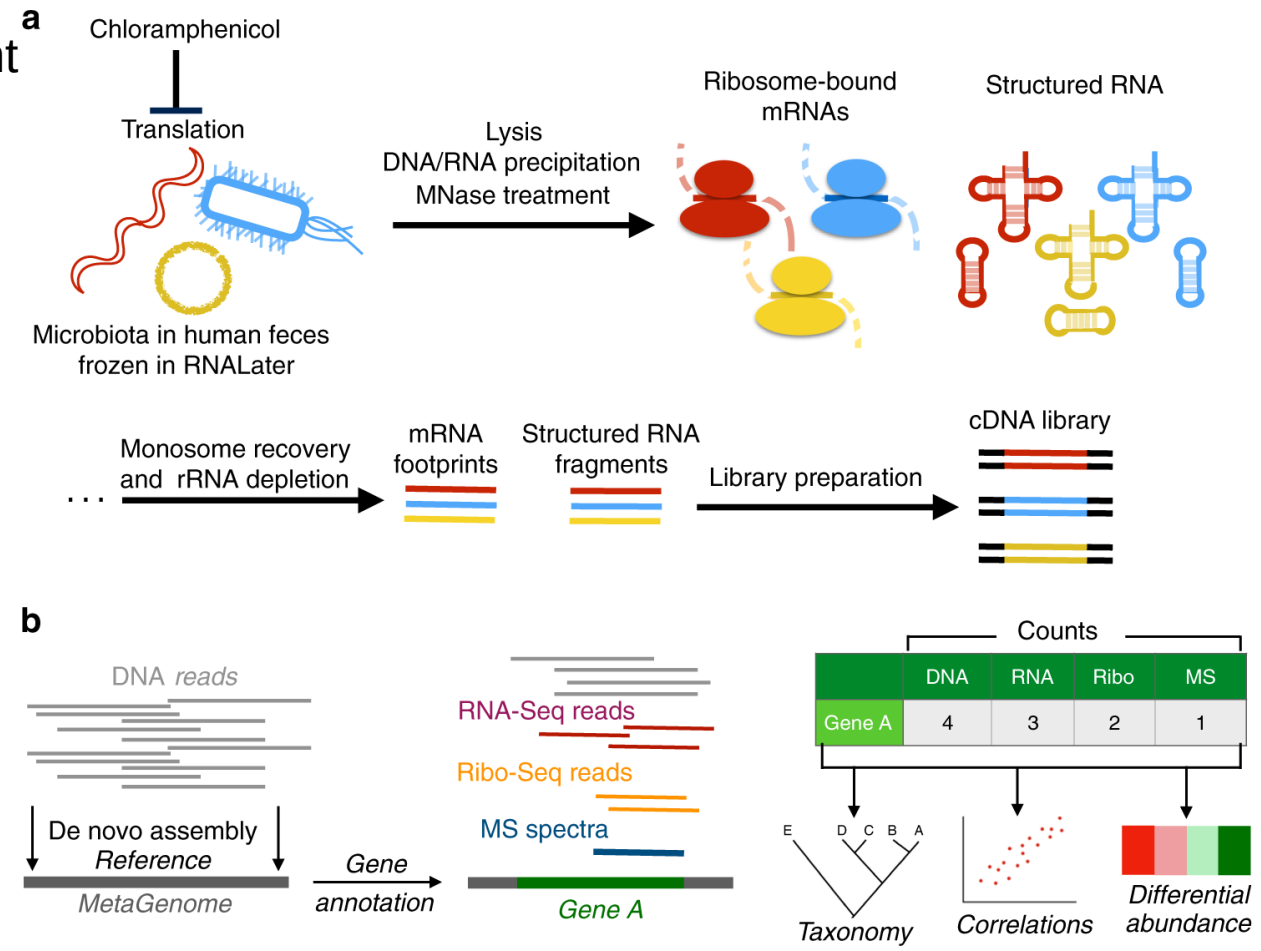


Clark and Pazdernik, 2016

MetaRibo-Seq

– Riboseq – translation arrest and subsequent sequencing of the translatome

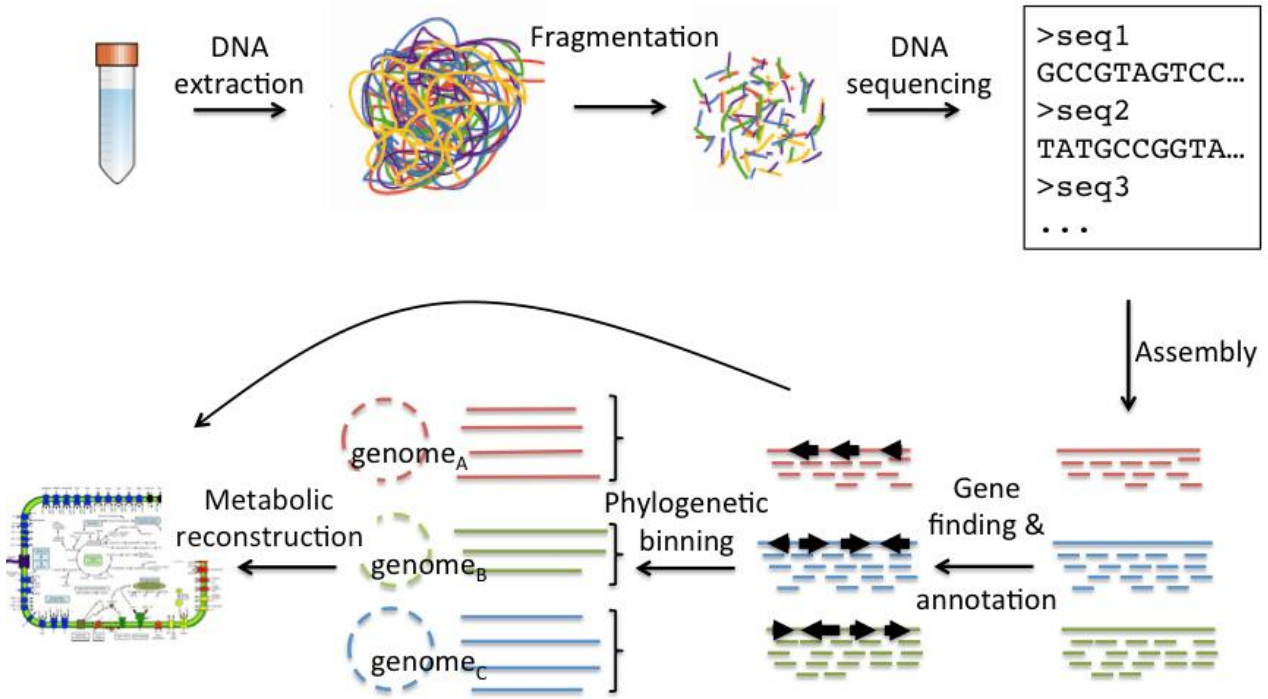




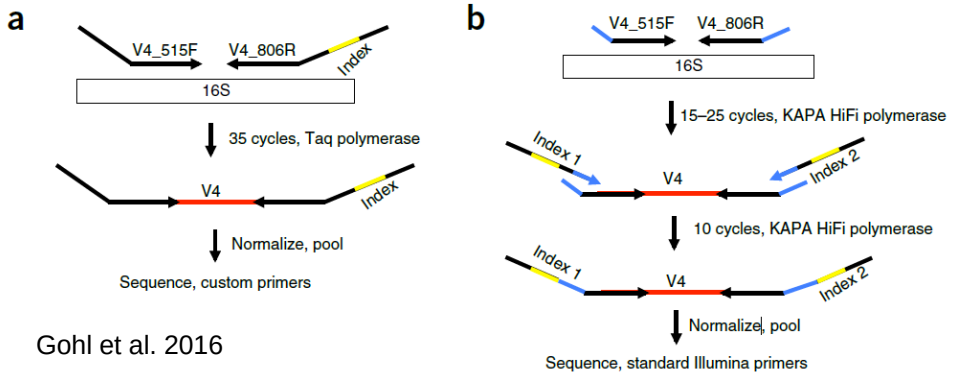
Fremin et al. 2020

Metagenomics

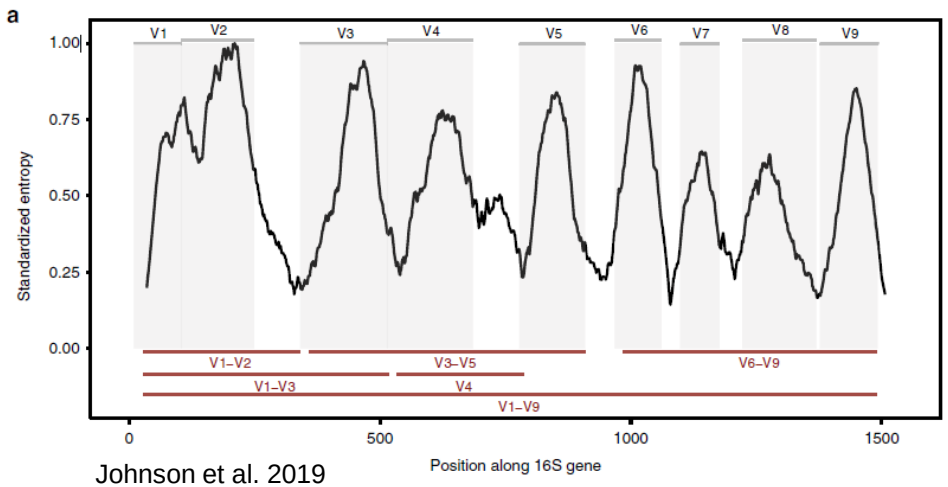
- A study of the genetic material contained in a sample
- ShotGun approach X sequencing of specific phylogenetic regions (16S, 18S, ITS, mcrA)



Microbiome	Microbiota	Metagenome
Microorganisms (and their genes) living in a specific environment	Microorganisms (by type) living in a specific environment	The genes of microorganisms in a specific environment



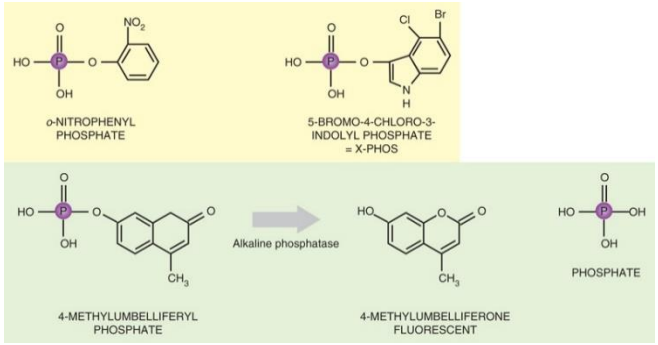
Gohl et al. 2016



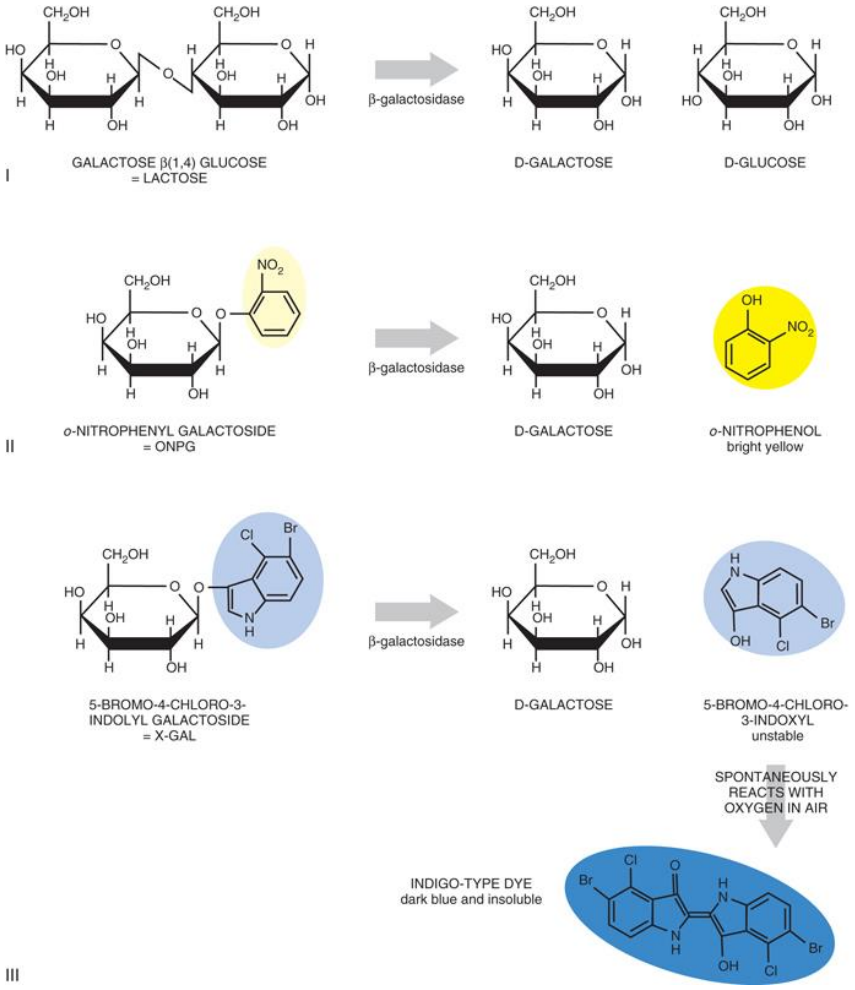
Johnson et al. 2019

Monitoring of gene expression

- A whole range of details about a gene obtained using reporter genes
  - adding a reporter gene behind the promoter
  - adding a reporter gene behind the CDS
- Using the following genes:
  - lacZ gene (β-galactosidase)
  - phoA gene (alkaline phosphatase)
  - lux/luc gene (luciferase)
  - gfp gene (Green Fluorescent Protein)

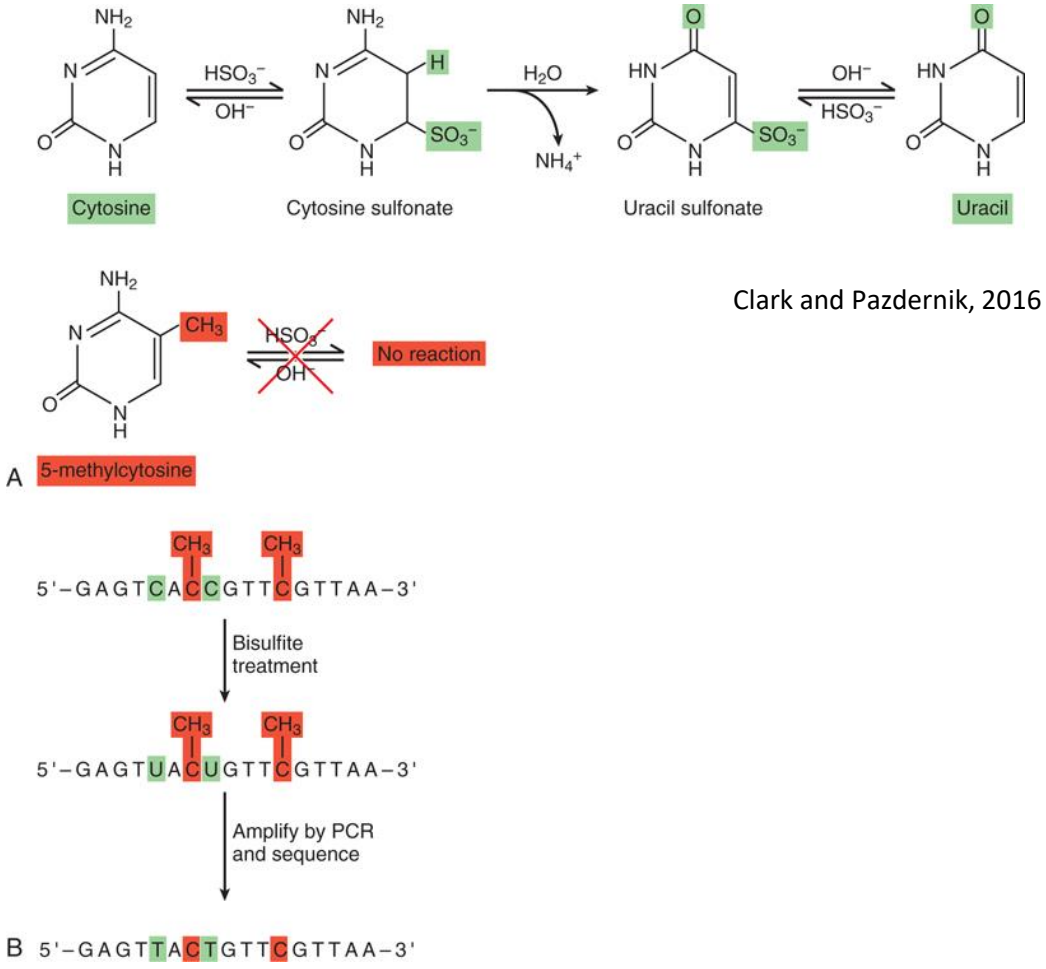


Clark and Pazdernik, 2016



Analysis of methylome

- Analysis of gDNA methylation sites
- Methylation usually silences transposon elements
- Silencing of one copy of the X chromosome in females
- Analysis using the bisulfite method
  - the addition of sodium sulfite leads to the conversion of non-methylated cytosines to uracil
  - subsequent sequencing without and with the addition of sulfite leads to the detection of methylation sites
- 3rd generation sequencers (Nanopores, PacBIO) are able to directly read cytosine methylation



Clark and Pazdernik, 2016

MUNI
SCI

GENE TECHNOLOGIES

Gene Cloning Strategies

Restriction endonucleases, plasmids, and cloning vectors, optimization of gene expression, expression in foreign hosts

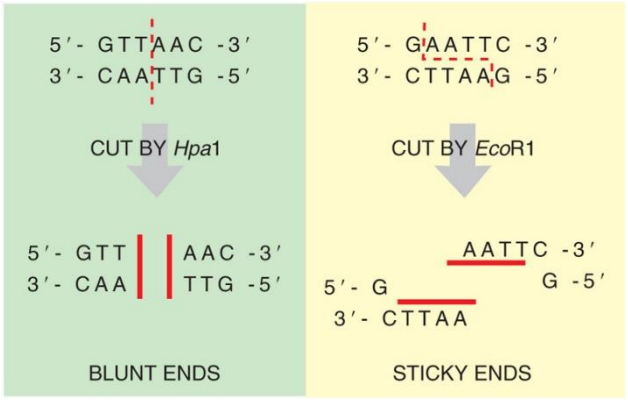
Restriction enzymes

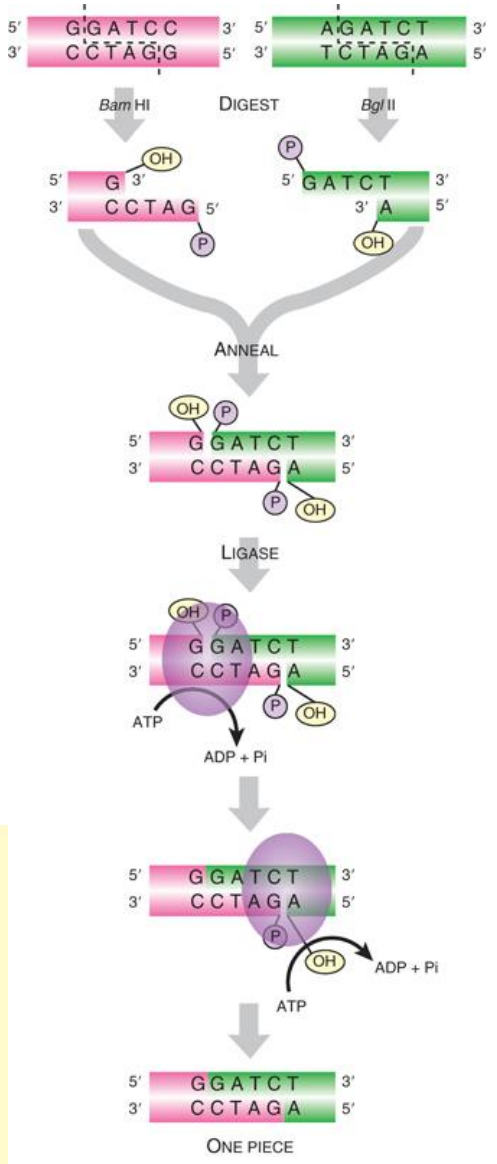
- Bacterial enzymes binding to a specific sequence and cleaving both strands
- Protection of bacteria from foreign DNA (viruses)
- Sensitive to DNA methylation
- Two basic types:

  Type I - cleaves the DNA strand 1000 or more bases from the recognized sequence

  Type II - cleaves the DNA strand at the location of the recognized sequence (blunt, sticky ends)

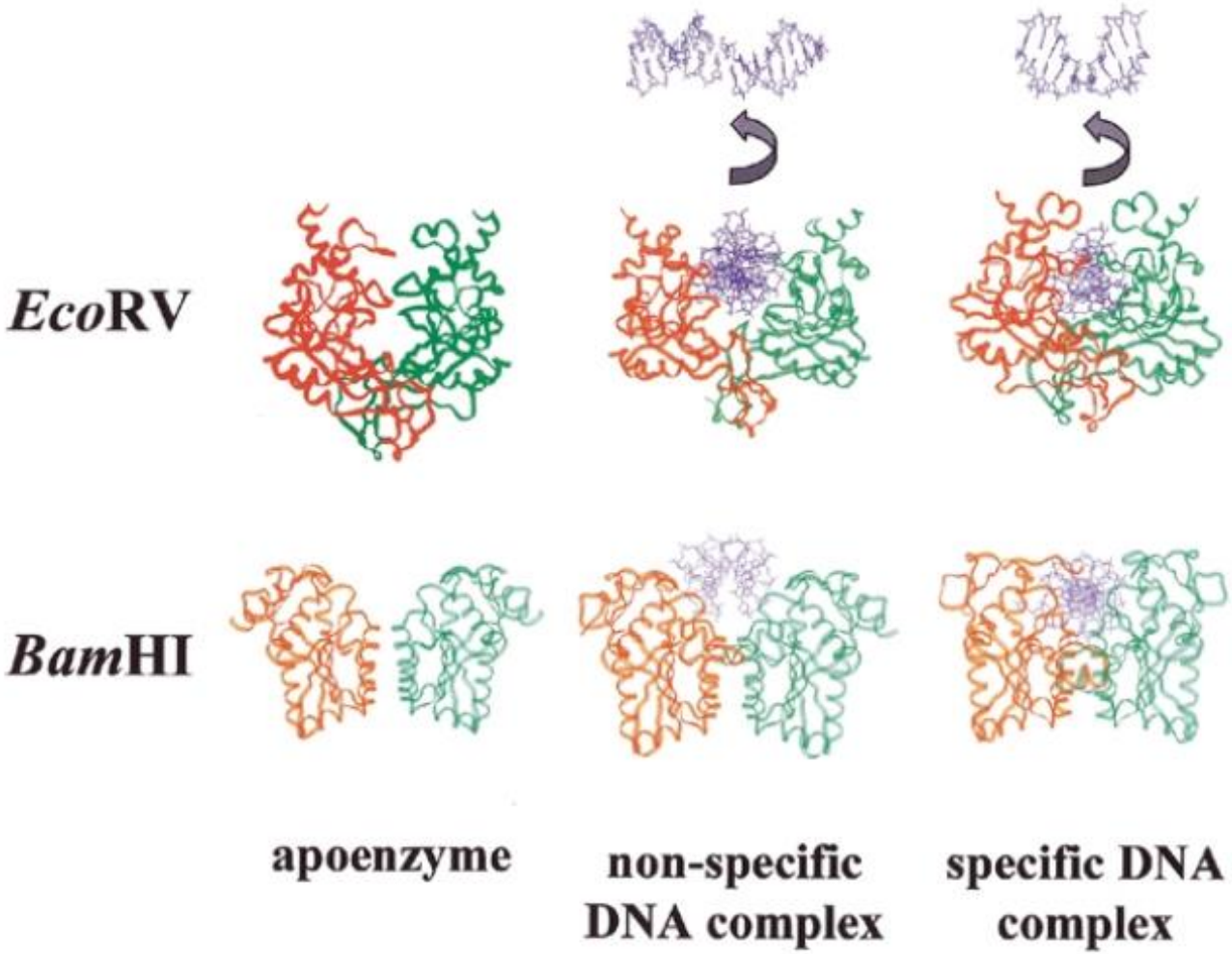
- The number of bases recognized = the degree of DNA fragmentation
- Joining fragments - ligase (T4 ligase)

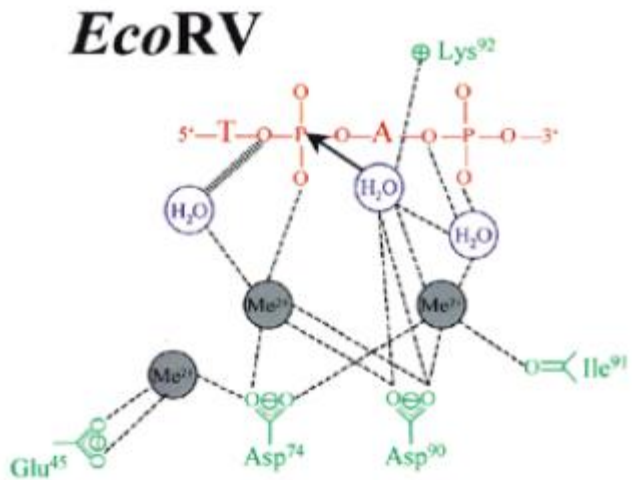


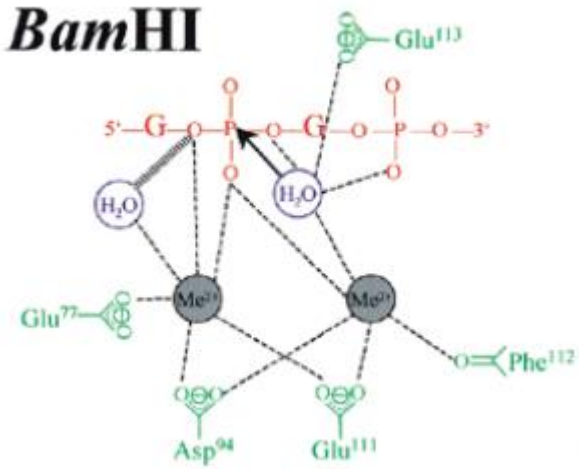


Clark and Pazdernik, 2016

Restriction enzymes (structure)



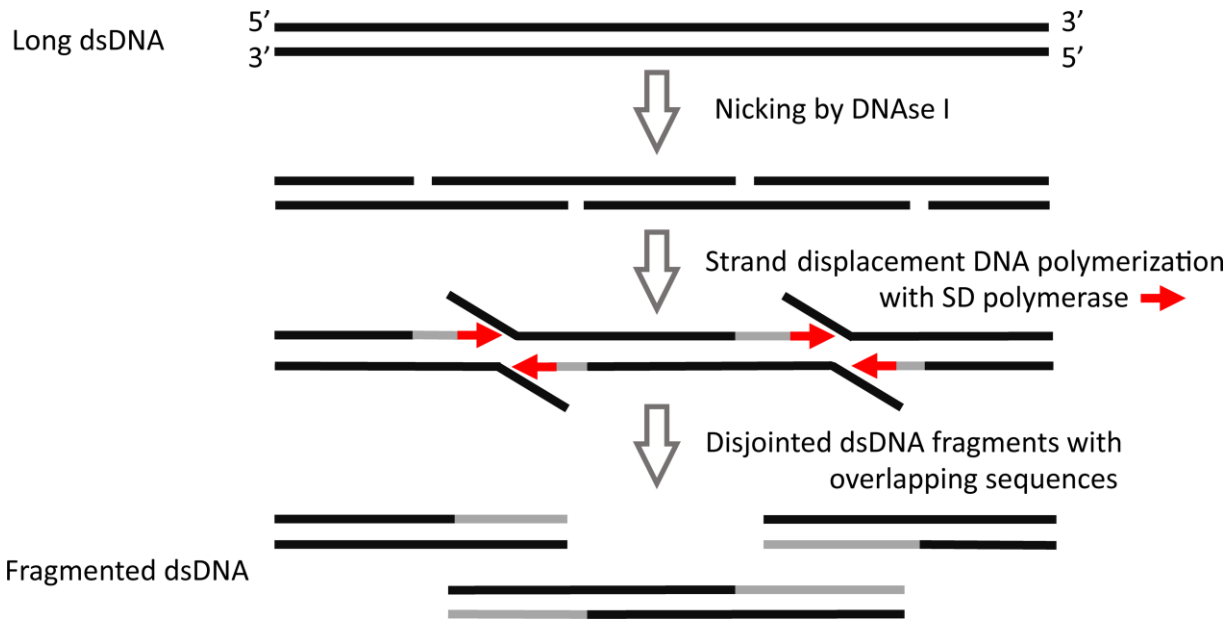


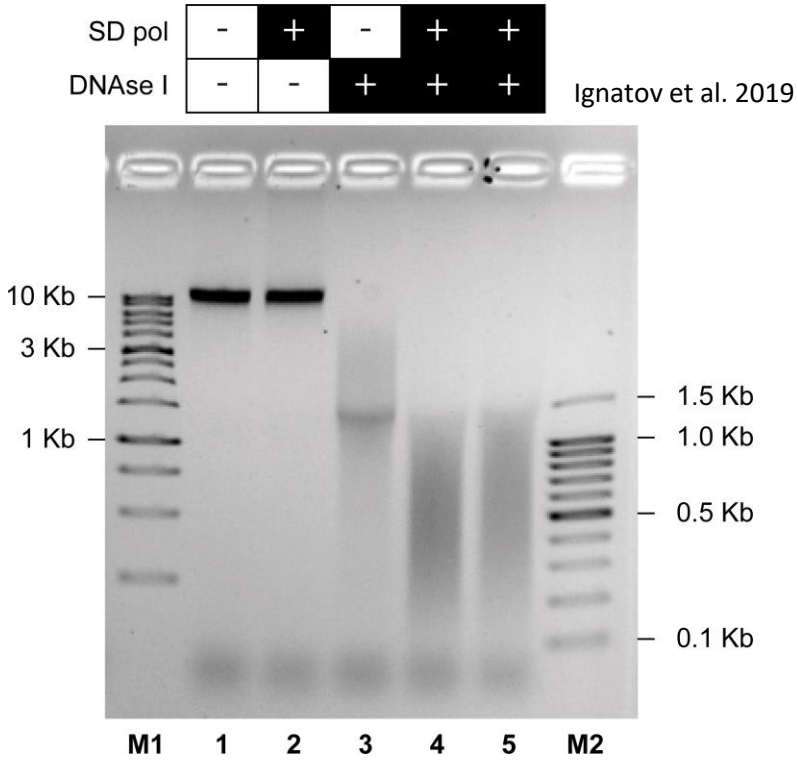


Pingoud and Jeltsch, 2001

Fragmentase

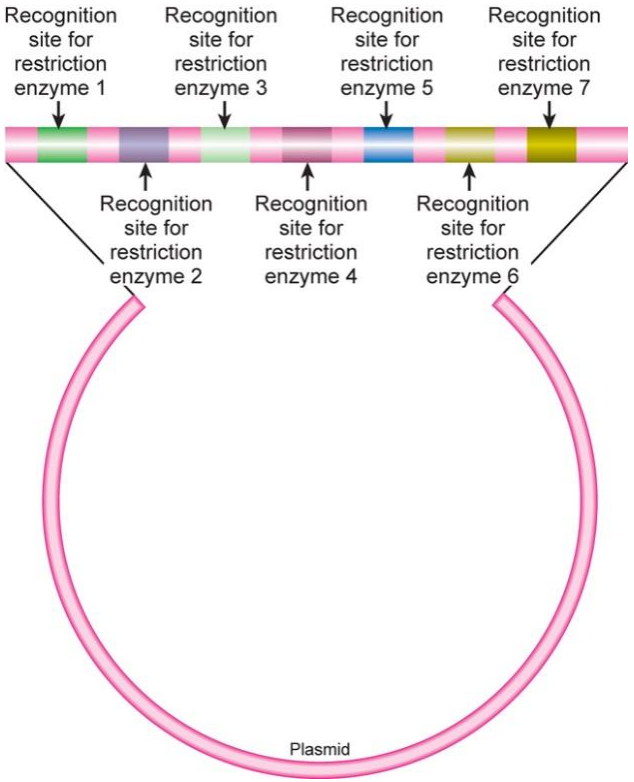
- Used for DNA fragmentation in NGS
- A mixture of endonucleases (NEas) cleaving one strand and then the opposite one
- A mixture of two enzymes (DNase I and SD (strand-displacement) polymerase)





Cloning vectors

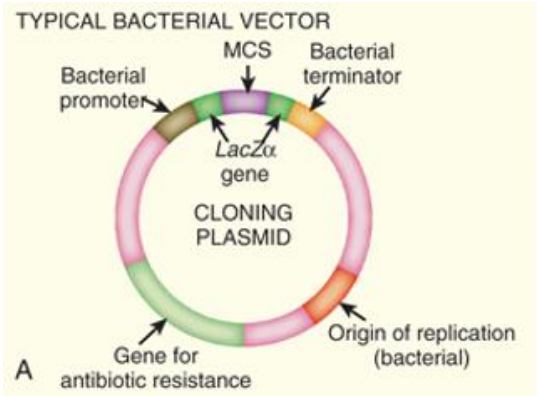
- Specialized plasmids (other elements) carrying foreign DNA for study/manipulation
- Currently, we also use artificial chromosomes and viruses
- Basic properties of cloning vectors:
  - small size (easy handling and isolation)
  - easy transfer between cells by transformation
  - easy isolation from the host organism
  - easy detection and selection
  - occurrence in a larger number of copies (ori site)
  - multiple cloning sites for insertion of cloned DNA
  - method confirming the presence of inserted DNA in the vector

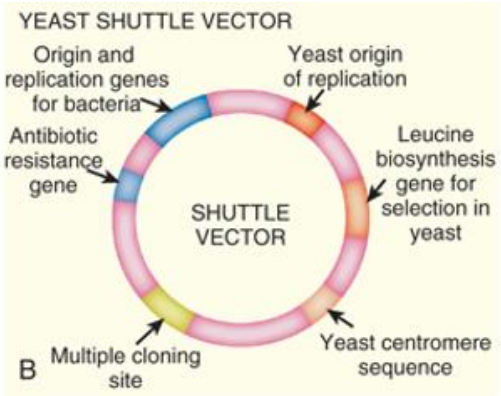


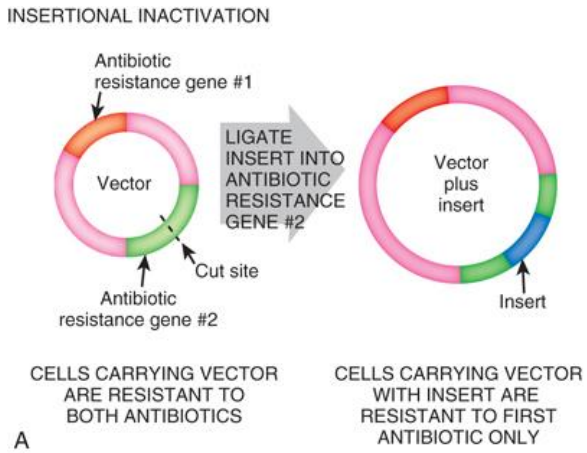
Clark and Pazdernik, 2016

Cloning vectors

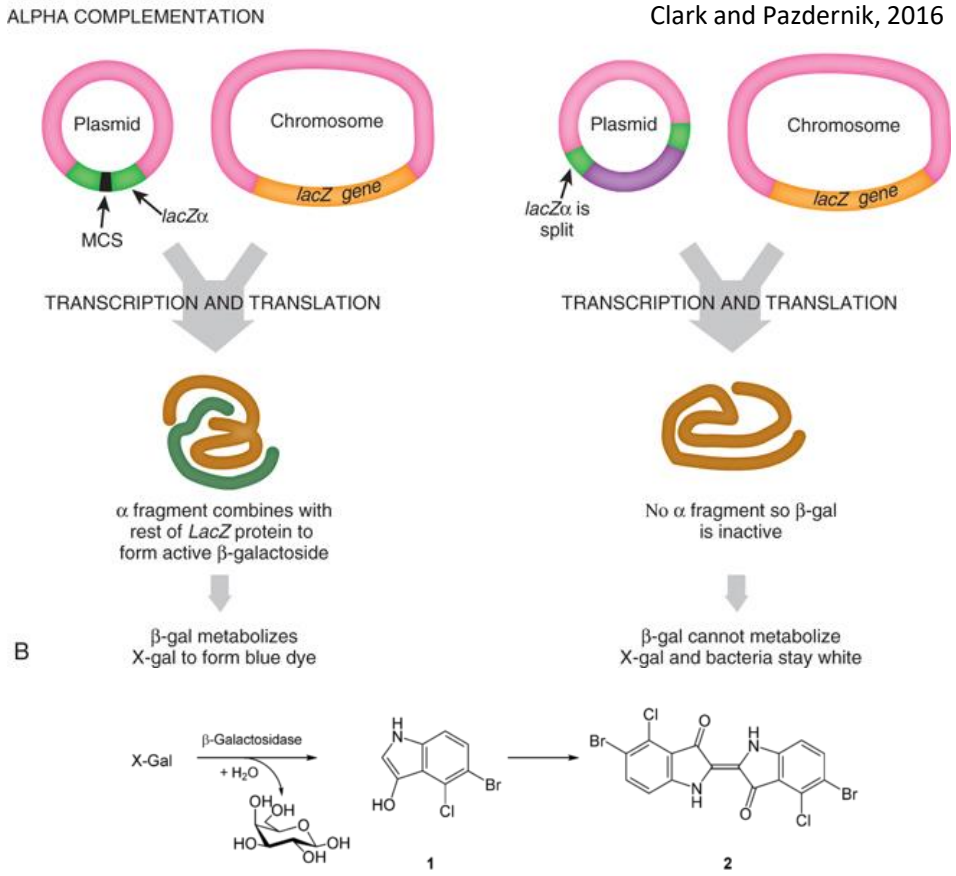
- **DNA insertion control options**
  - insertional inactivation (ATB resistance gene)
  - *ccdB* gene (death gene interfering with DNA gyrase activity) https://link.springer.com/article/10.1007/BF00280310)
  - alpha complementation (β-galactosidase)
- **Yeast vectors based on a 2μ circle**
  - *ori* site from two organisms, the *Cen* sequence
  - selection based on AA synthesis







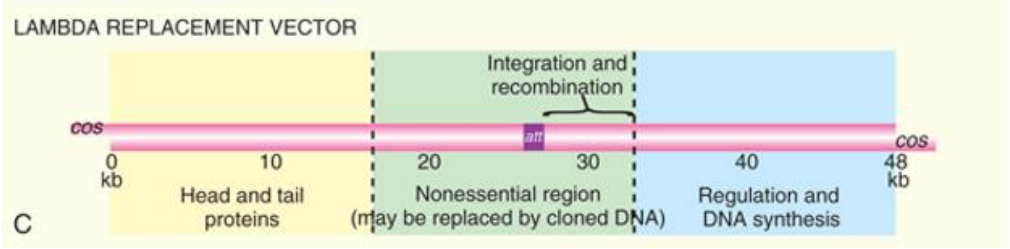
Clark and Pazdernik, 2016

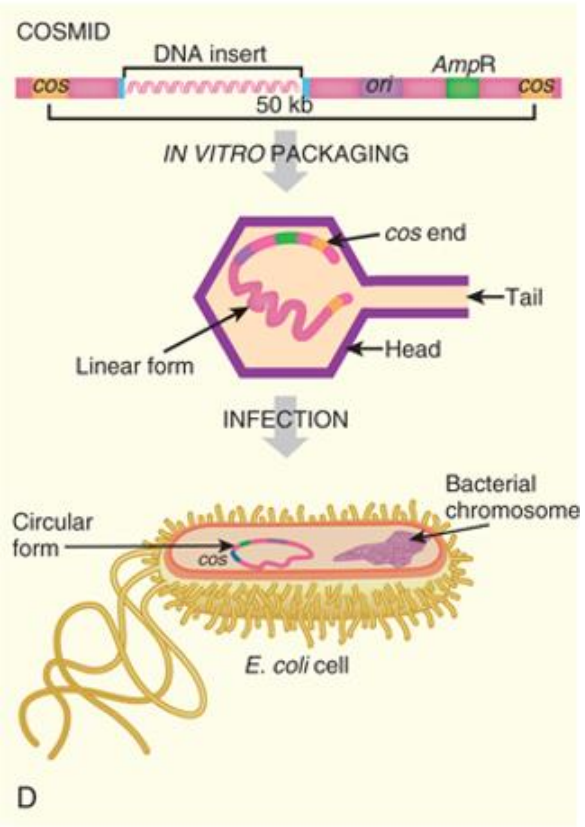


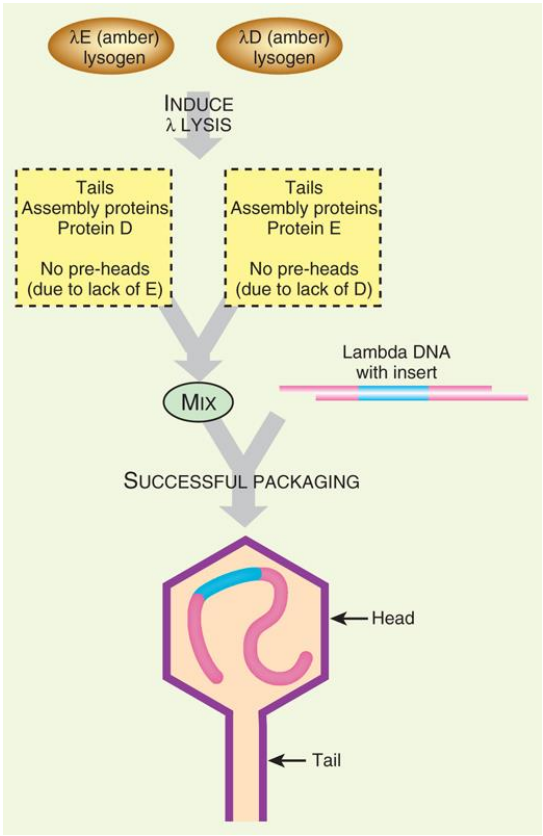
Virus vectors

- **Bacteriophage vectors**
  - modified to carry non-viral DNA in the capsid
  - connection of *cos* sequences = formation of a replication form (RF) replicated by a rolling circle
  - an insert with a size of 37 to 52 kb can be used
  - use of helper viruses to package DNA into virus capsid

- **Cosmids**
  - a highly modified lambda vector having only *cos* sites
  - the necessity of packaging by helper phage







Clark and Pazdernik, 2016

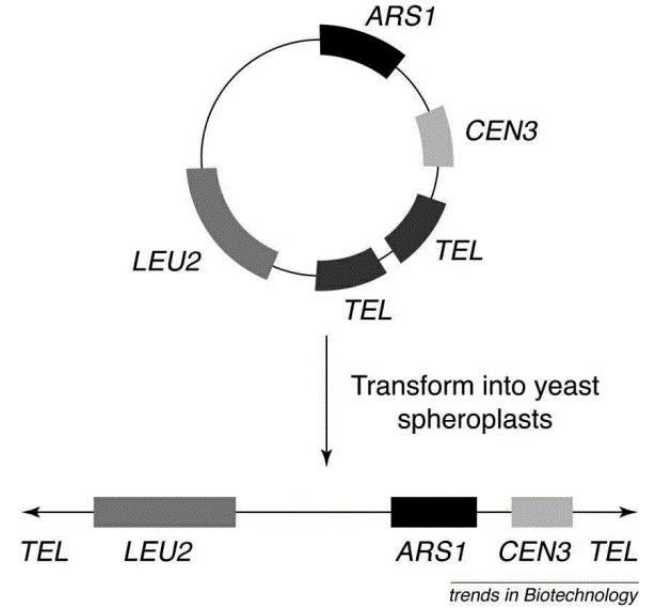
Artificial chromosomes

- Used for handling large pieces of DNA (150 – 2000 kb)
- **Include**

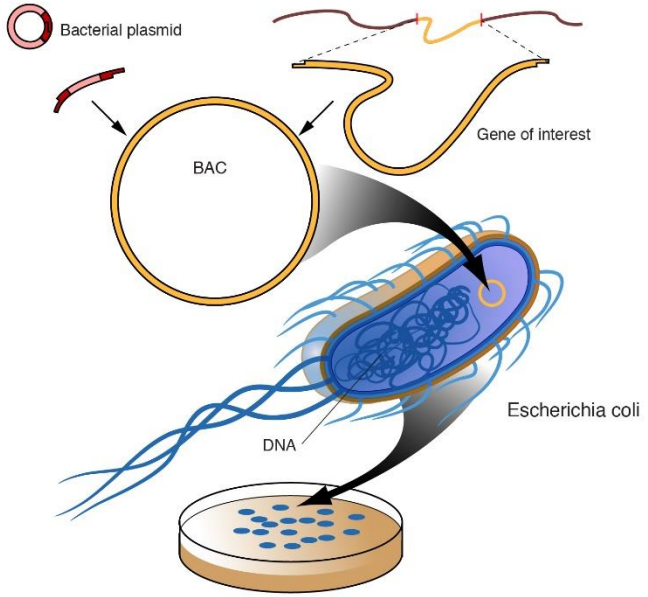
  - yeast artificial chromosomes (YACs)

  - bacterial artificial chromosomes (BACs)

  - P1 bacteriophage artificial chromosomes (PACs)
- **YACs** contain a centromere and telomeres for permanent maintenance in yeast
- **BACs** are circularized and propagated in bacteria (*ori* site and resistance gene)

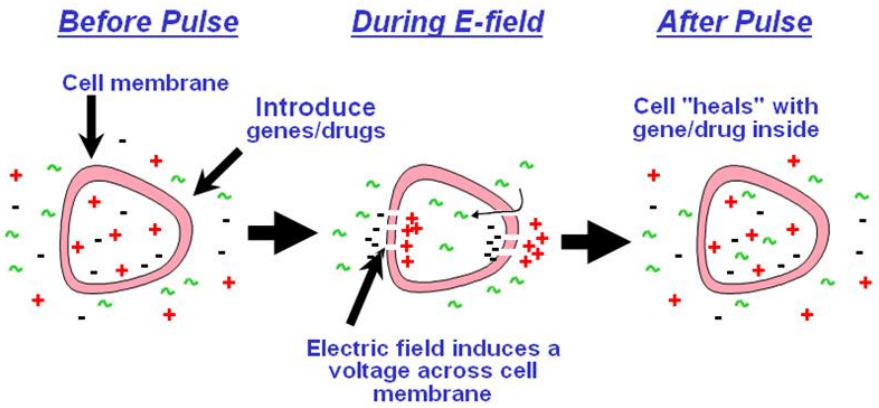


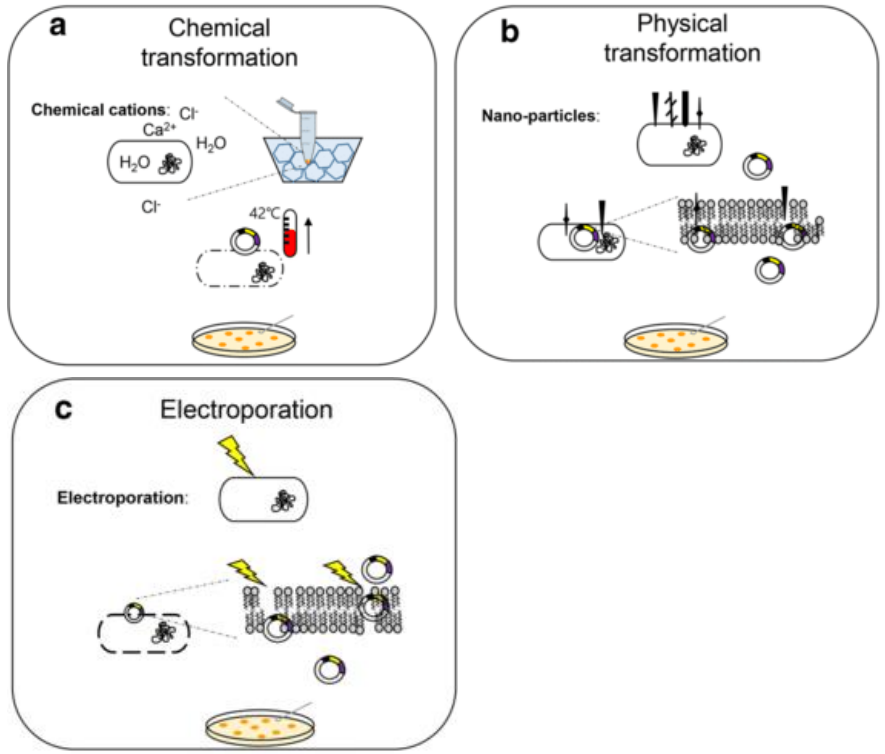
Trends in Biotechnology 2000, DOI: (10.1016/S0167-7799(00)01438-4)



DNA transformation

- Transformation is the process by which foreign DNA is introduced into a cell.
- **Competent *E. coli* cells**:
  - the use of calcium ions and thermal shock to increase the permeability of the cell wall and membrane
  - use of electroporation to open the cell wall and membrane
- **Competent yeast**:
  - a combination of lithium acetate, single-stranded carrier DNA and polyethylene glycol (PEG)

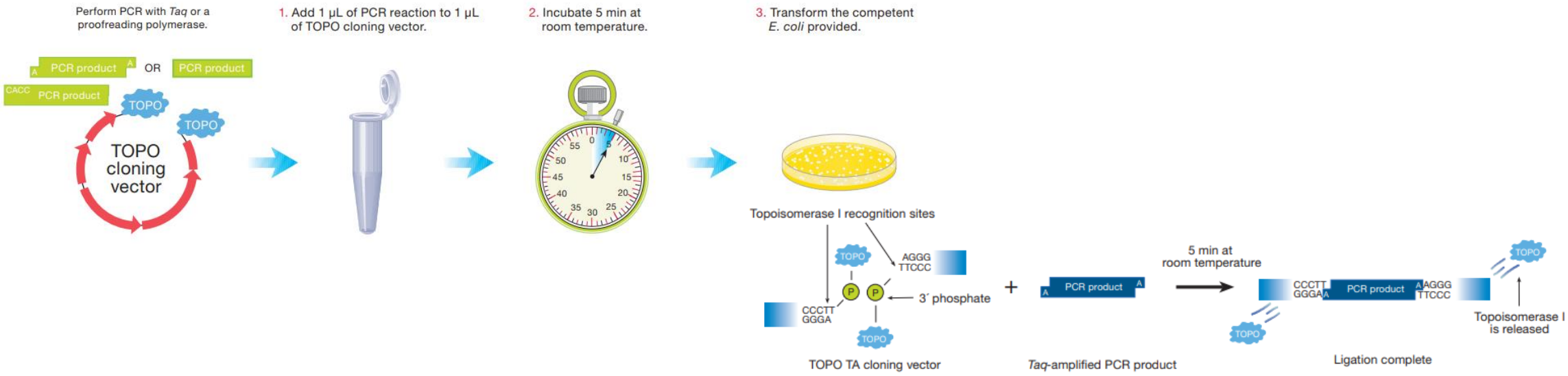




Cloning strategies

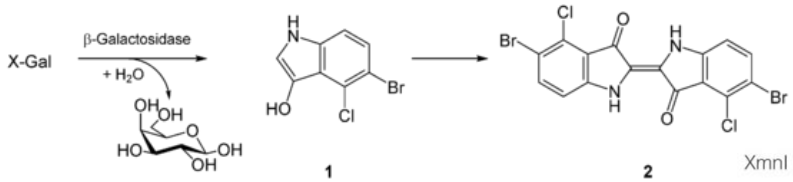
– TOPO Cloning (Thermo)

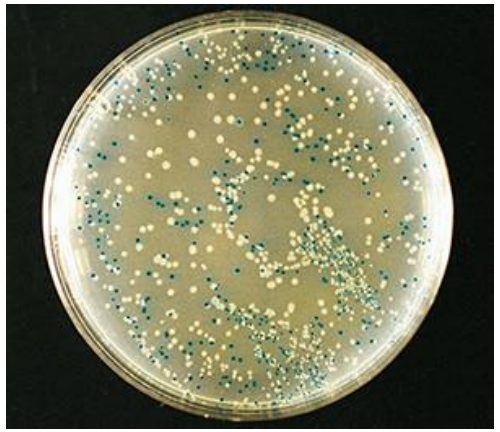
- use of topoisomerase I
- *Vaccinia* virus topoisomerase I specifically recognizes the sequence 5'-(C/T)CCTT-3’
- topoisomerase is covalently attached to the 3' end of the vector

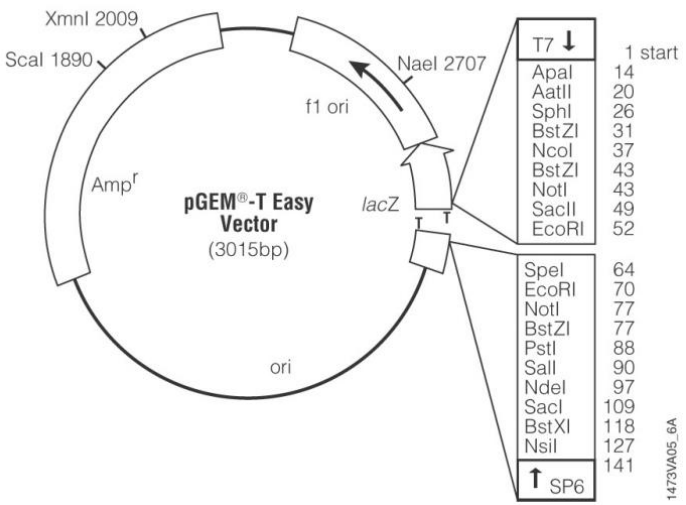


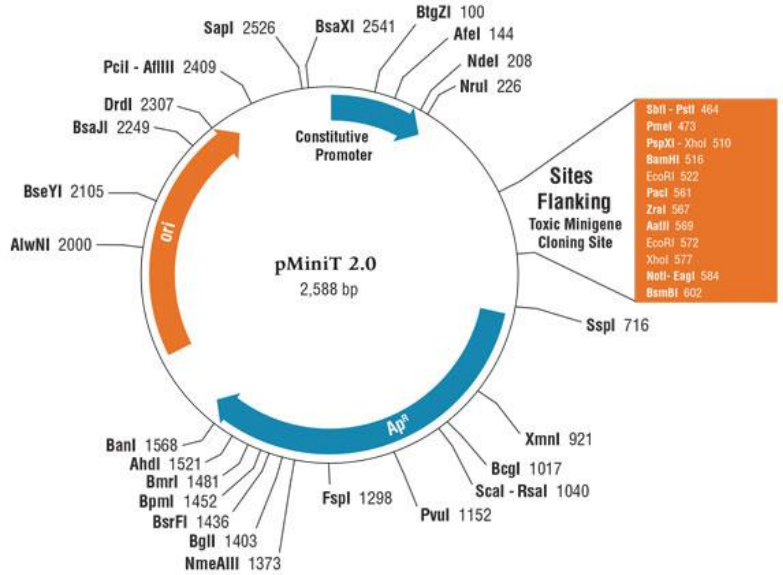
Cloning strategies

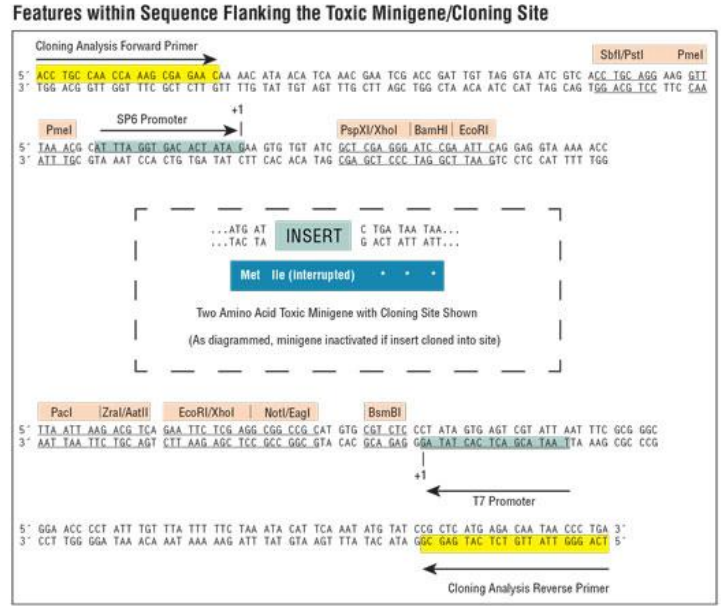
- TA cloning
  - using the property of Taq DNA polymerase to add A to the 3' end
  - **pMiniT 2.0** (toxic mini-genes) (NEB)
  - **pGEM-Teasy** (blue-white selection) (Promega)





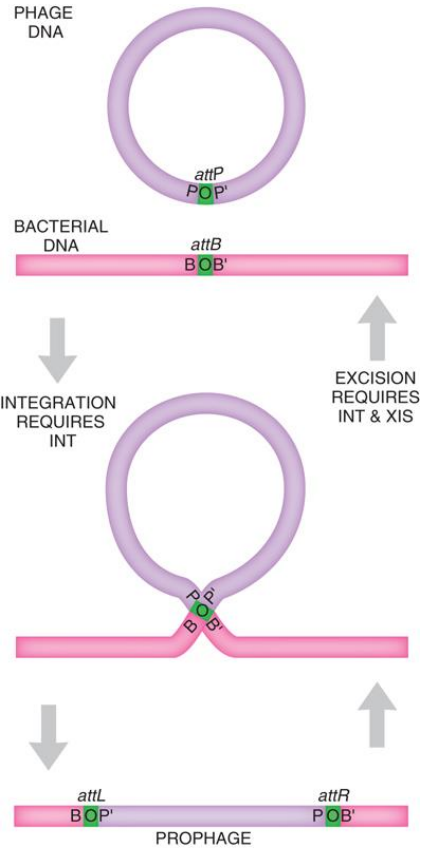


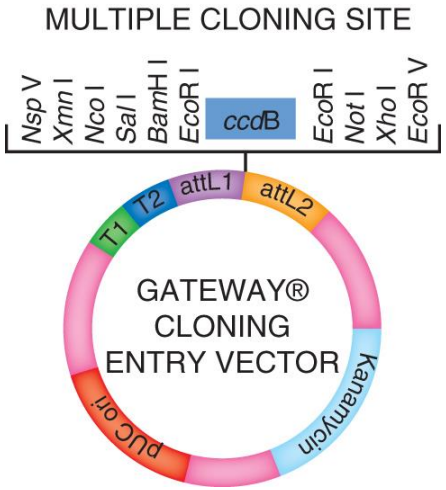


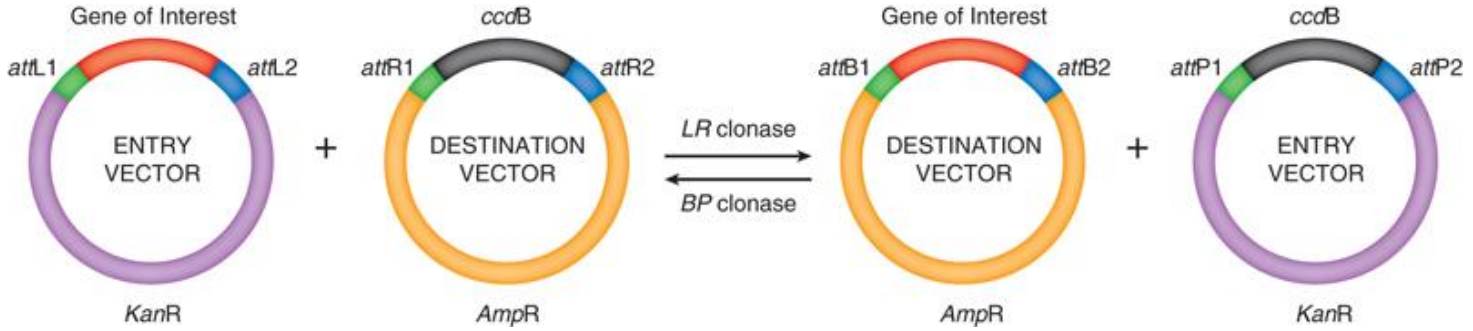


Cloning strategies

- **GATEWAY** cloning vectors (Invitrogen-Thermo)
  - use of phage lambda integrase and excisionase enzymes
  - use of **ENTRY** and **DESTINATION** vectors
  - the BP reaction removes the gene of interest from attR sites and inserts it into attL sites.
  - the LR reaction removes the gene of interest from attL sites and inserts it into attR sites



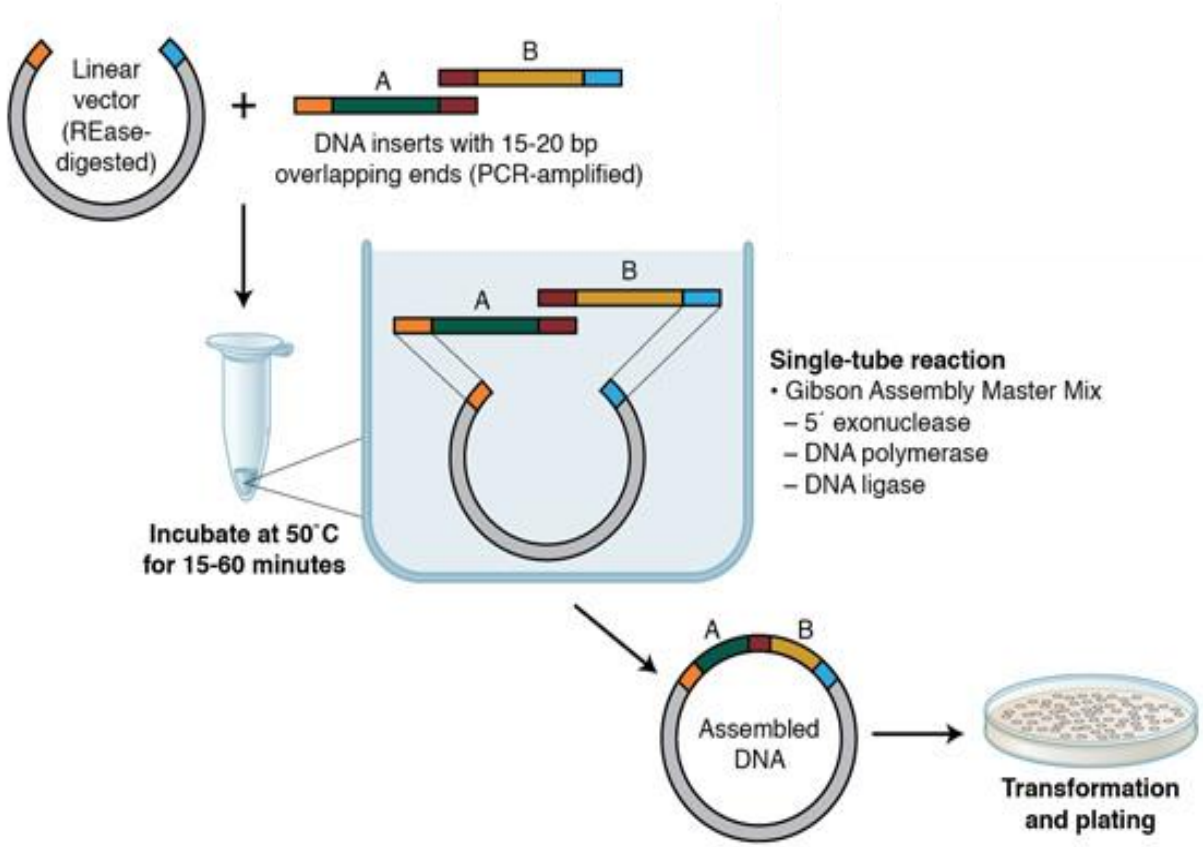




Clark and Pazdernik, 2016

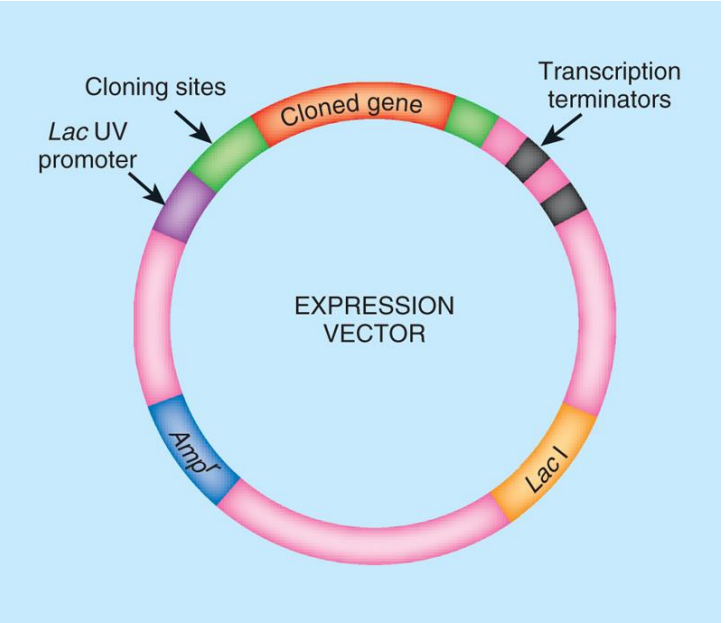
Cloning strategies

- In 2009 Dr. Daniel Gibson and colleagues at the J. Craig Venter Institute developed a new method to easily assemble multiple linear DNA fragments
- **Advantages**
  I. There is no need for specific restriction sites.
  II. Join any fragments regardless of order.
  III. The reaction takes place in one tube.
- **Gibson's Mix consists of three different enzymes**
  I. T5 Exonuclease
  II. Phusion DNA Polymerase
  III. Taq DNA ligase



https://www.youtube.com/watch?v=tlVbf5fXhp4

Expression vectors

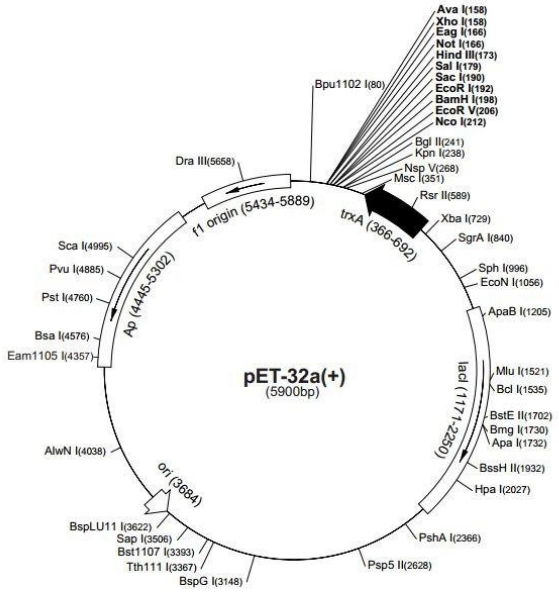


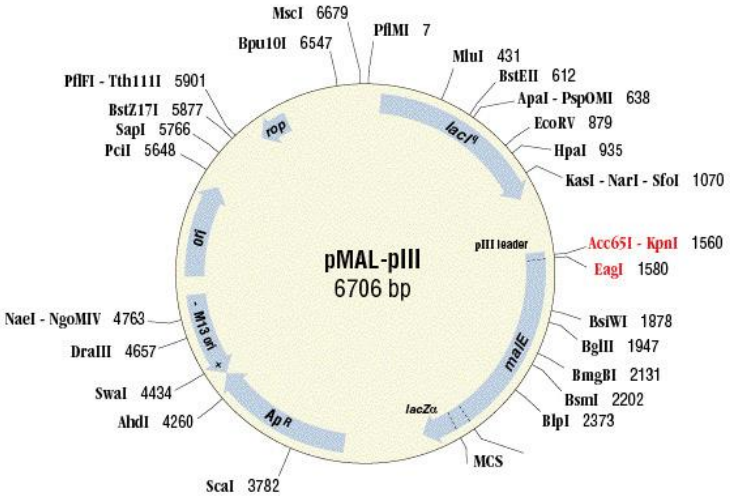
Clark and Pazdernik, 2016

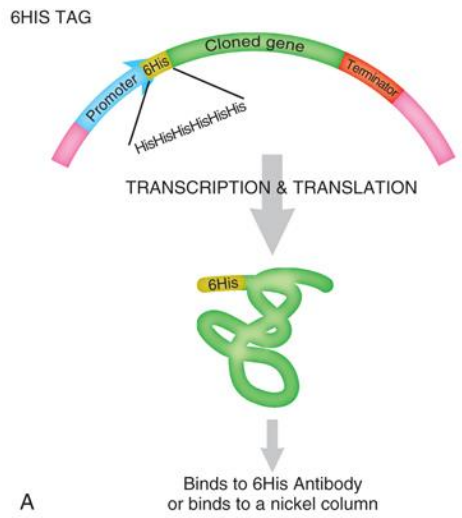
- The most commonly used *lacUV* promoter (modified *lac* promoter)
  - RNA polymerase binding site
  - *lacI* repressor site
  - transcription start site
  - transcription termination site
- Another frequently used promoter is the lambda left promoter (P_L)
  - lambda repressor binding site
  - most frequent activation by increased temperature (42°C)
- Expression systems also use a promoter binding only bacteriophage T7 RNA polymerase
  - *E. coli* strains carrying T7 RNA polymerase after inducer control
- Expression vectors often contain sequences for various tags (6xHis, Myc, FLAG, S-tag, MBP)

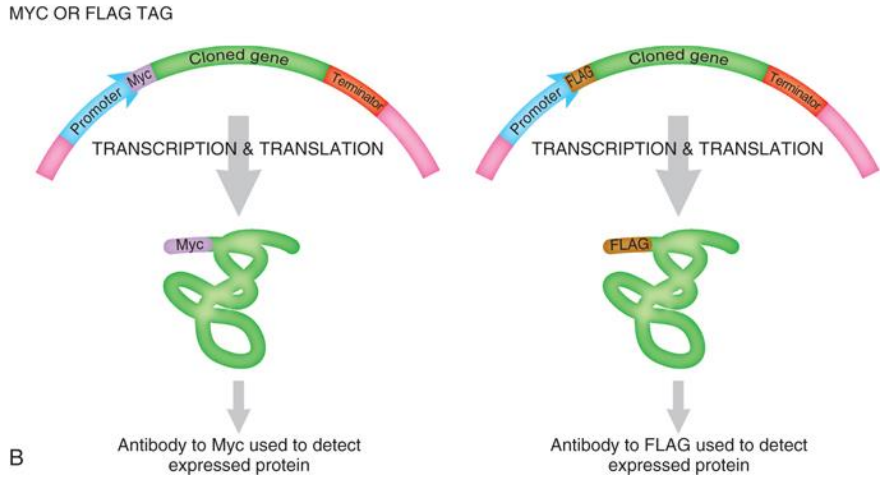
Bacterial Expression Vectors

- pET, prSET *E. coli* T7 expression vectors
  - expression in BL21(DE3)pLysS cells
- pMAL expression vectors
  - carry maltose-binding protein (MBP)



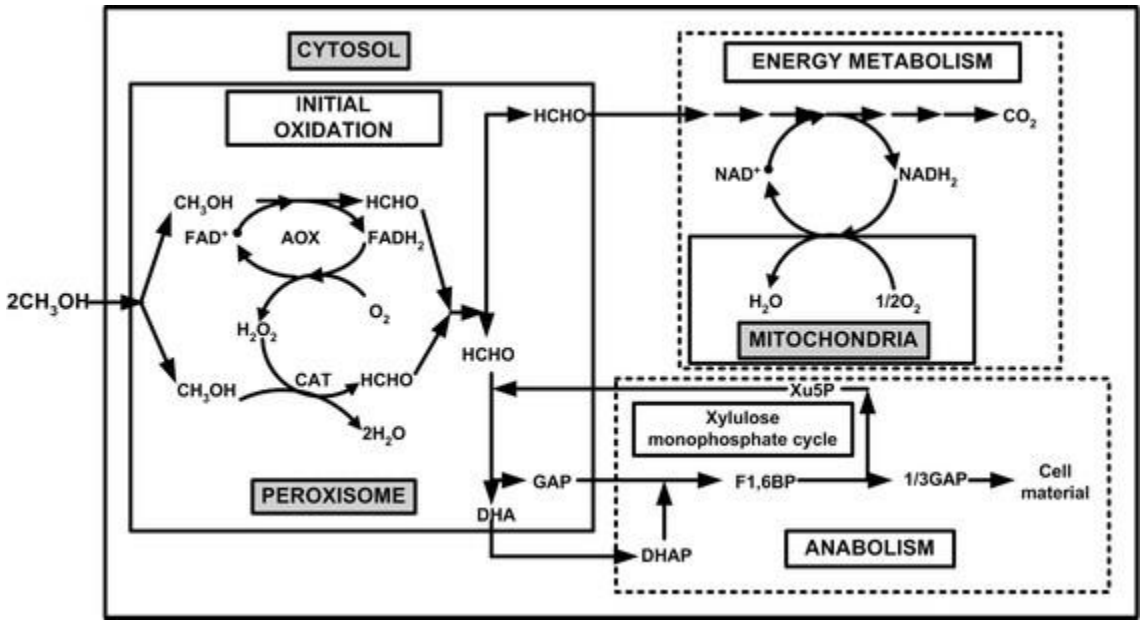


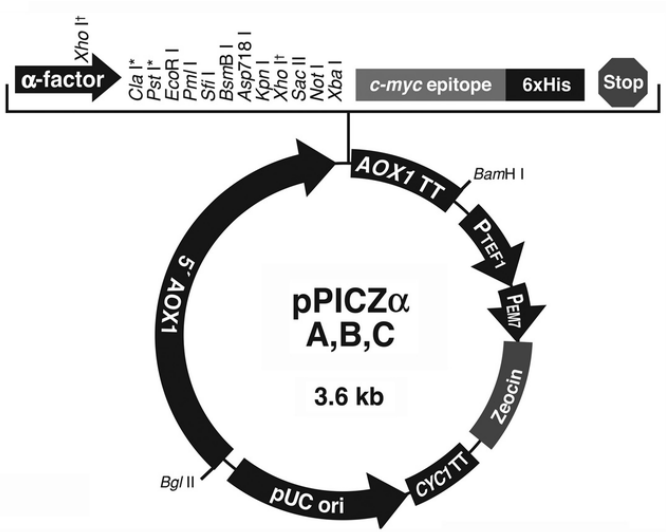


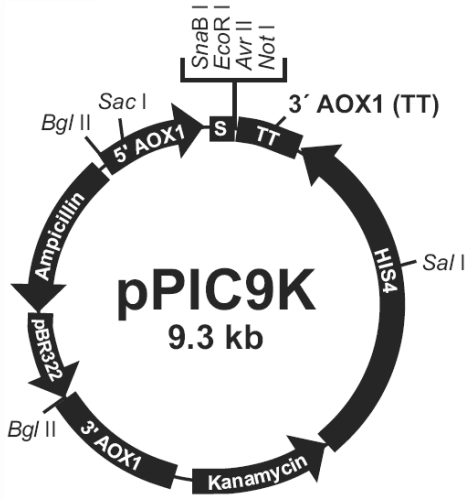


Yeast Expression Vectors

- Inducible AOX promotor (methanol)
- Possibility of intra- and extracellular expression
- Expression in yeasts *P. pastoris* and *S. cerevisiae*







Comments for pPICZα A
3593 nucleotides

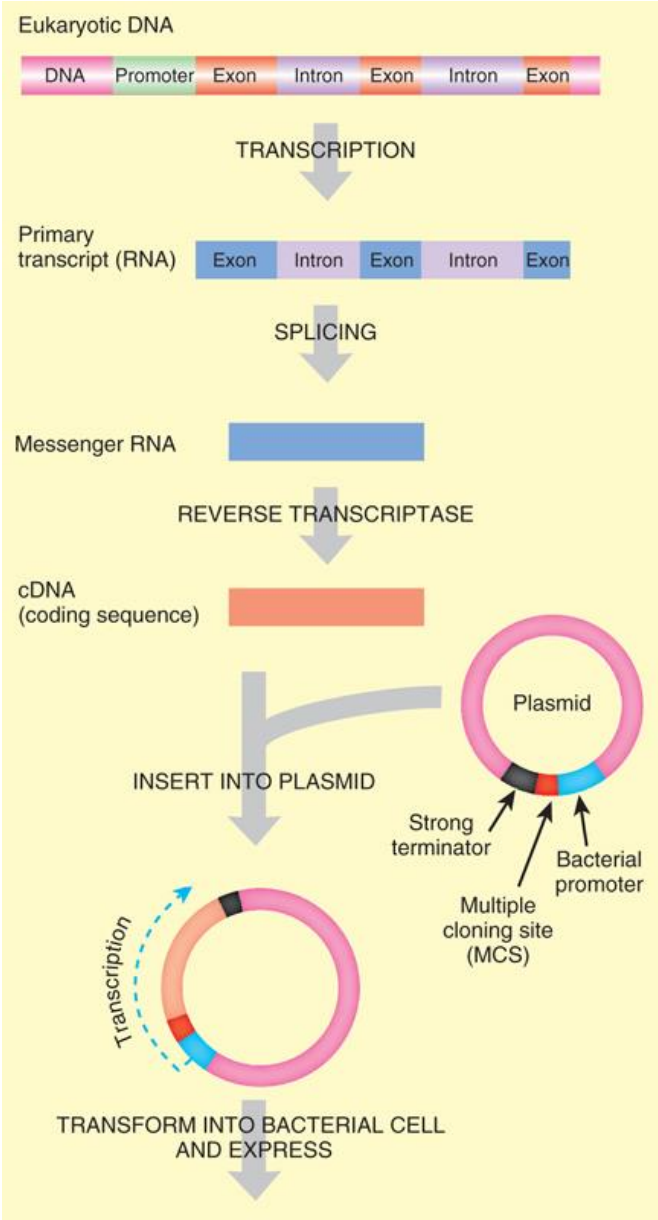
5´ *AOX1* promoter region: bases 1-941
5´ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3´ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

* *Pst* I is in Version B only
Cla I is in Version C only

†The two *Xho* I sites in the vector allow the user to clone their gene in frame with the Kex2 cleavage site, resulting in expression of their native gene without additional amino acids at the N-terminus.

Expression in Bacteria

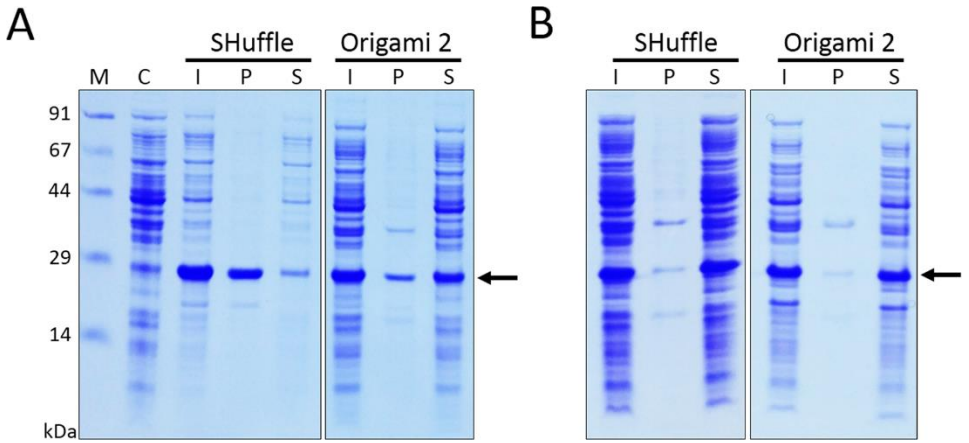
- Special plasmids (expression vectors) are used to increase proteins expression
  - strong promoter, adequate *ori* site, selection marker for antibiotic
- Expression of eukaryotic proteins is more problematic
  - promoter modification, absence of splicing, low rate of translation
  - weak interaction of the ribosome with the RBS site, mRNA instability, limited amount of tRNA
- The necessity of using specially modified vectors



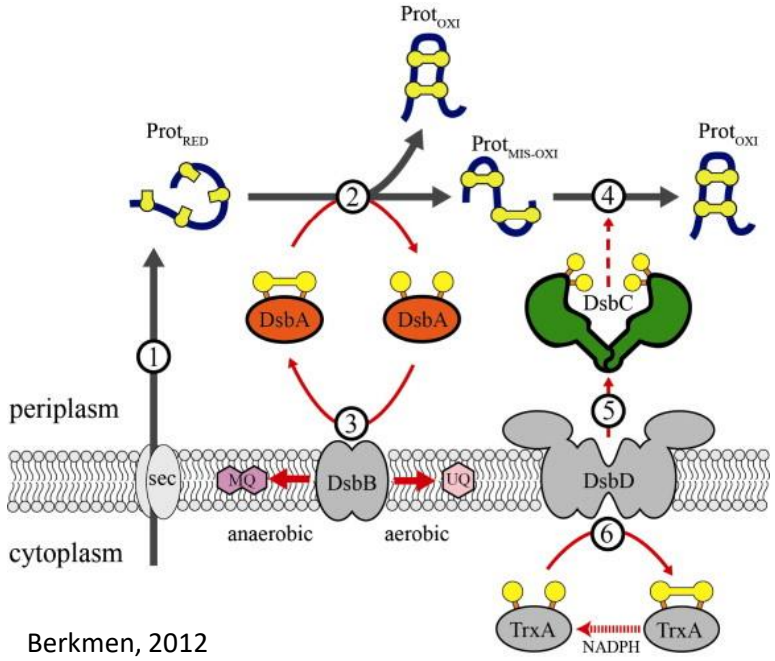
Clark and Pazdernik, 2016

E. Coli Origami™ 2

- They carry a mutation in the gene thioredoxin reductase (*trxB*) and glutathione reductase (*gor*)
- Increase in the formation of disulfide bonds in the cytoplasm of *E. coli*
- Suitable for proteins requiring the formation of S-S bridges for proper composition



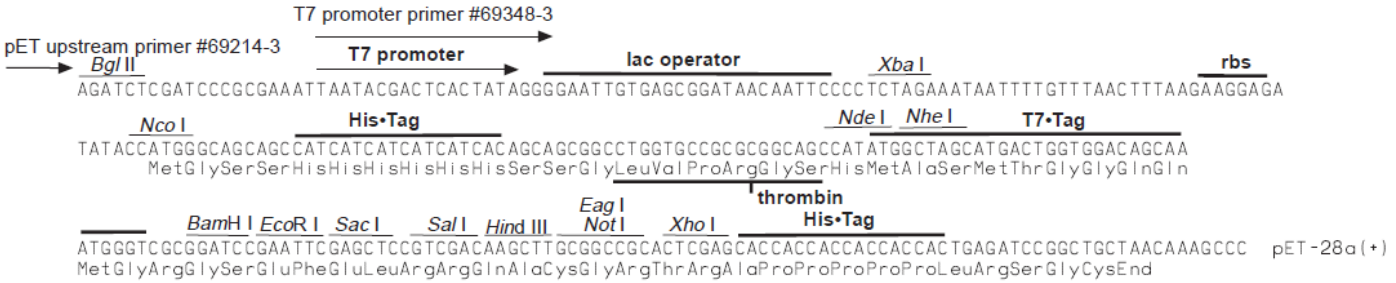
Exprese Oncostatinu M (OSM): A (37°C), B (18°C). C-kontrola bez IPTG, I-lyzát, P-pelet, S-solubilní frakce (Nguyen et al., 2019, SciRep)

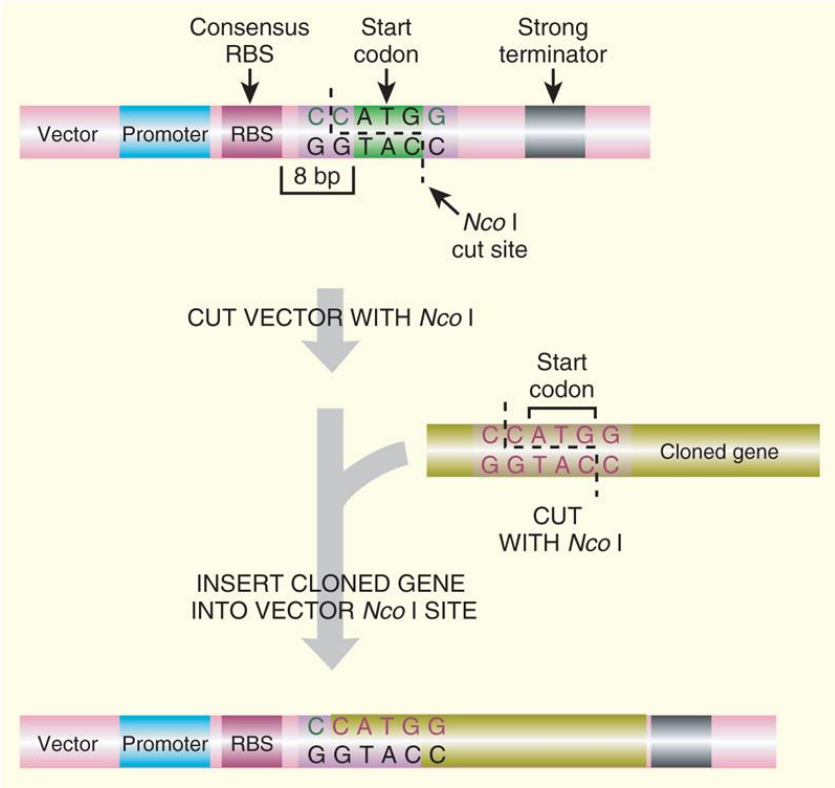


Berkmen, 2012

Translational Expression Vectors

- Designed for protein expression (pET, pRSET)
  - maximum translation initialization
  - consensus RBS site
  - ATG codon at an optimal distance of 8 bases from the RBS
  - cloning site directly in the ATG codon (*Nco* I)
- The possibility of further complications in protein folding

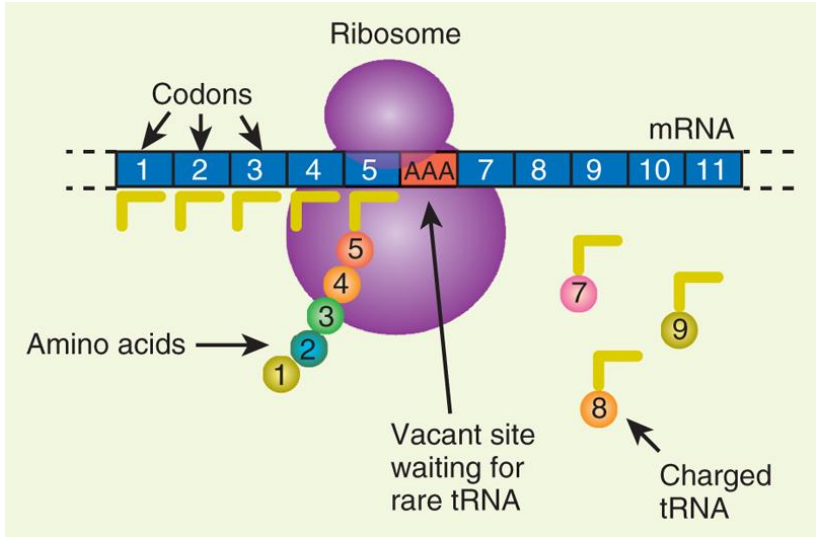




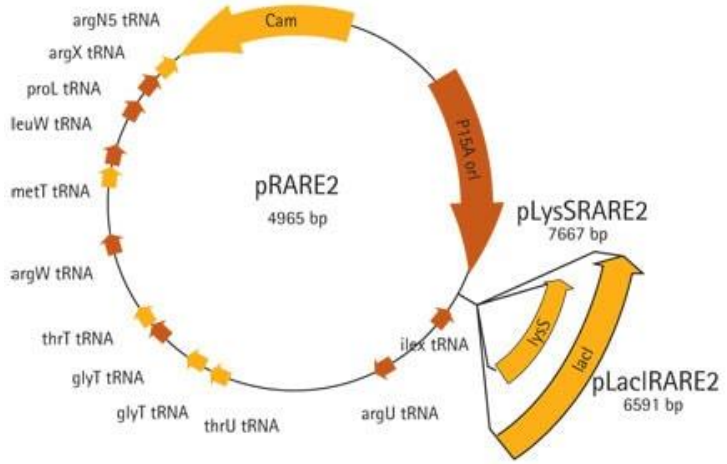
Clark and Pazdernik, 2016

Codons Effect

- Protein expression in other organisms (eukaryotic in bacteria)
- Different organisms prefer different codons for a given AA
  - optimization of the codons used in gene synthesis
  - up to a 10-fold increase in production
  - delivery of tRNA carrying rare codons to the organism
  - ***E. coli* ROSETTA** – seven tRNAs for rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG)

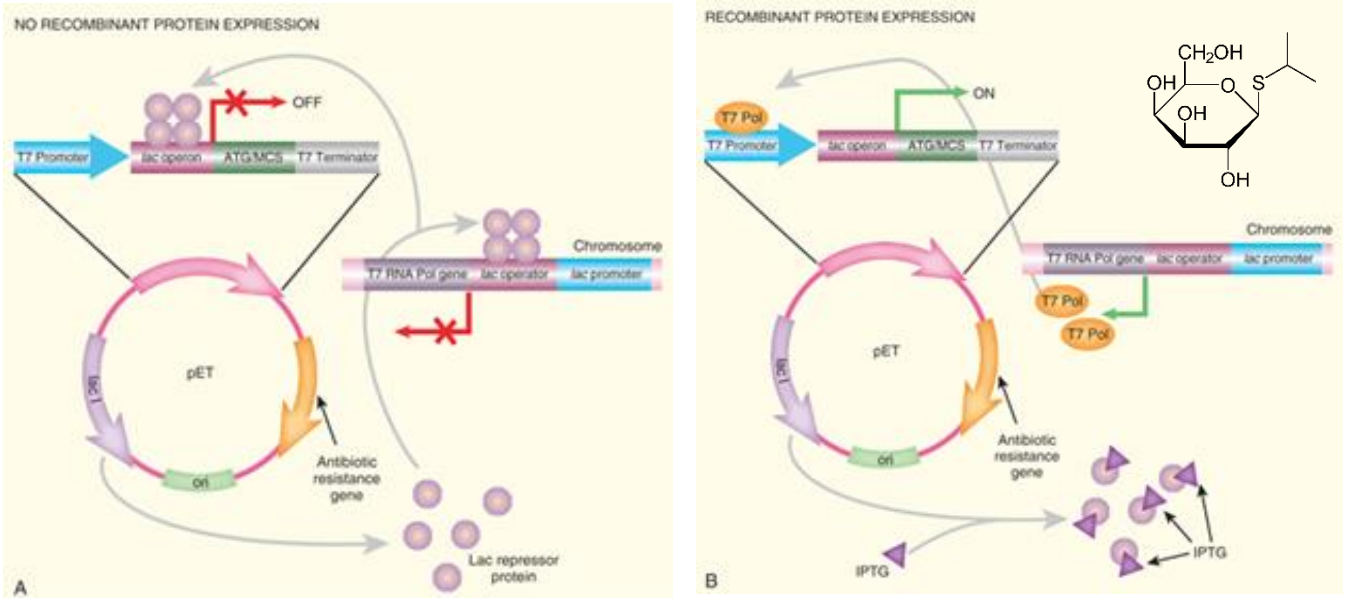


Clark and Pazdernik, 2016

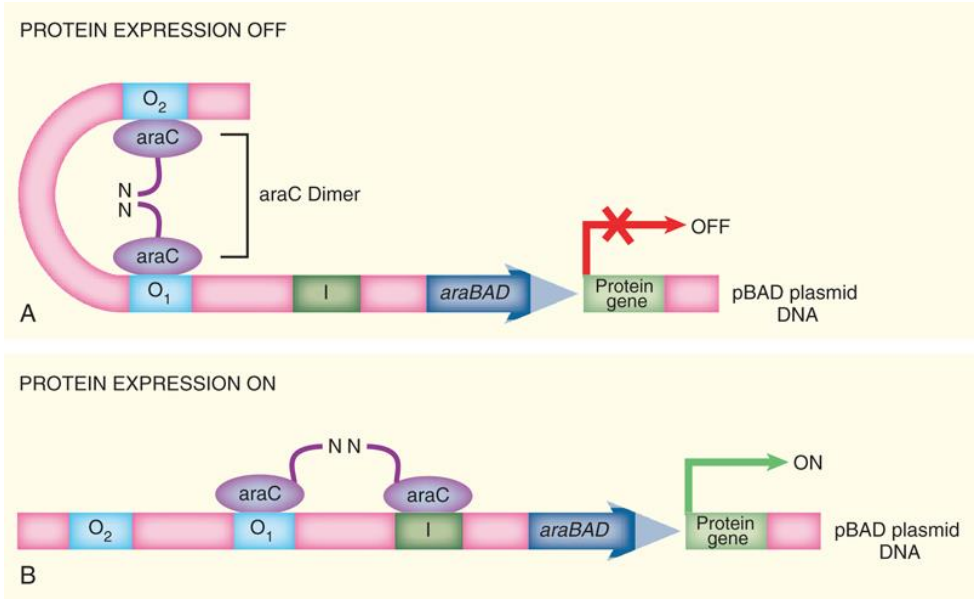


Toxic effect of overexpression

Lactose operon

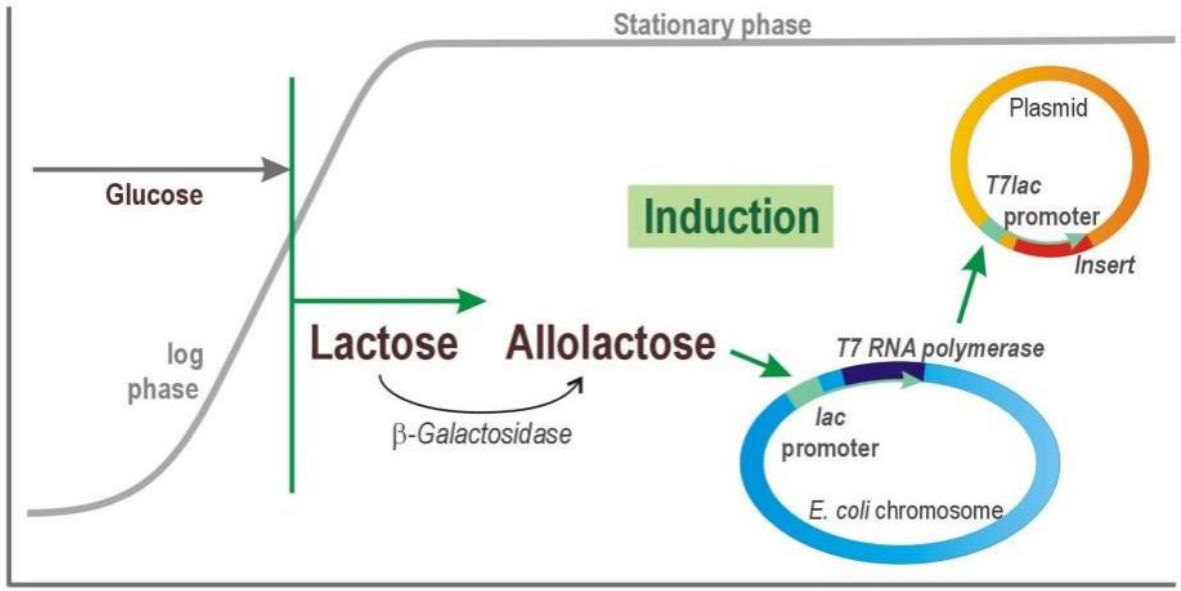


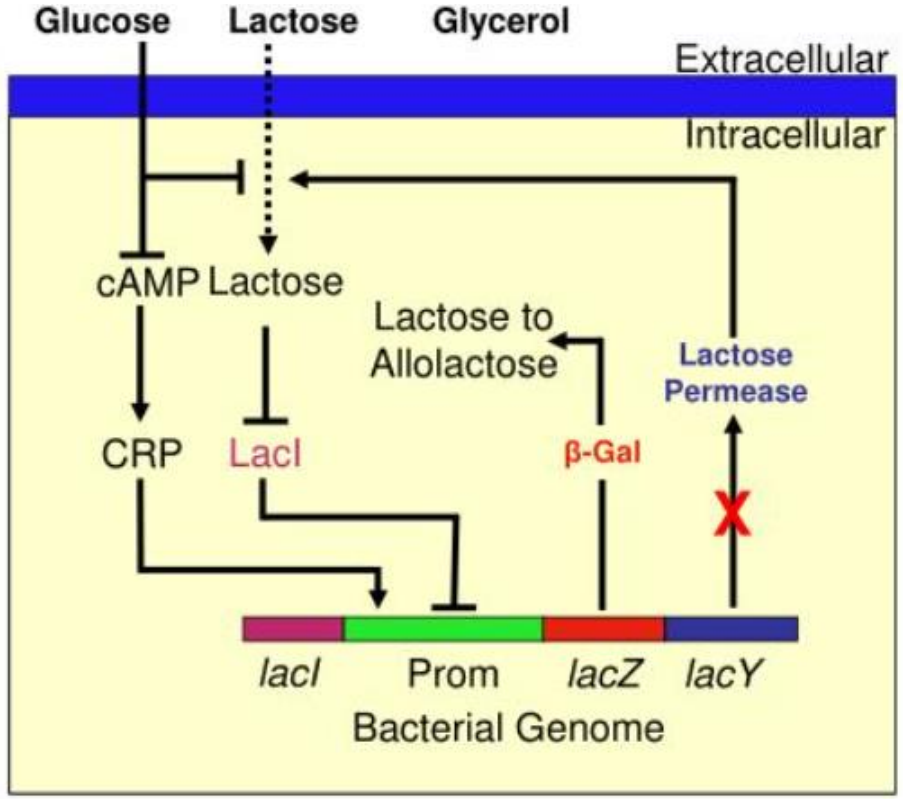
Arabinose operon



Clark and Pazdernik, 2016

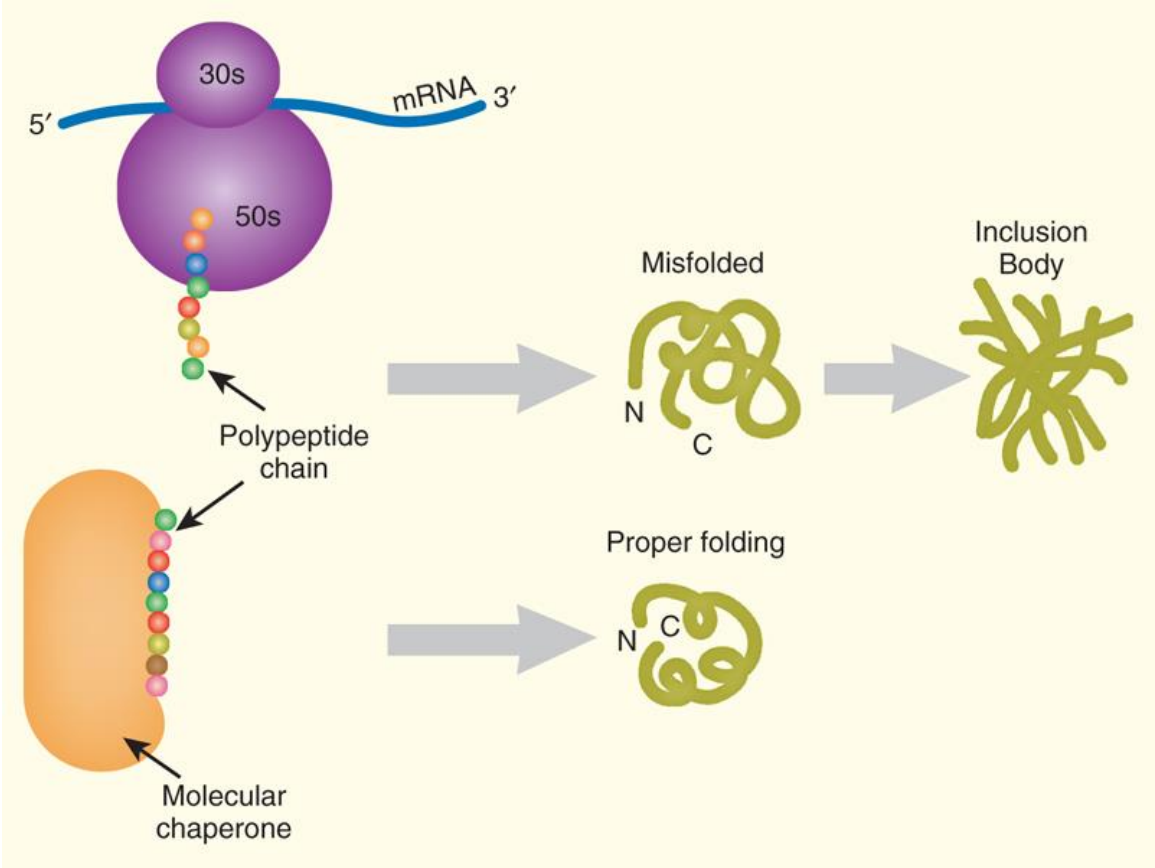
Autoinduction Medium





Inclussion Bodies

- Misfolded proteins accumulate in inclusion bodies
- Molecular chaperones – they help with proper packing
- Possible secretion of proteins into the periplasm or medium
- Proteins can be solubilized from inclusion bodies with a chaotropic agent and renaturation



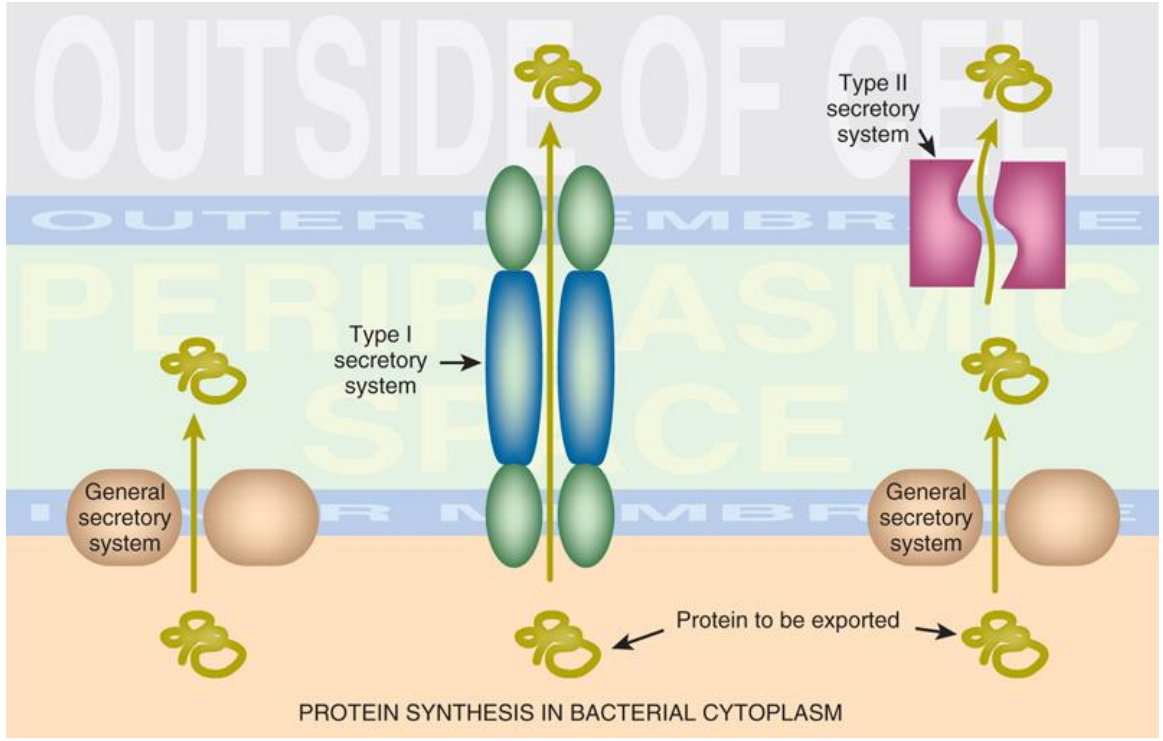
Clark and Pazdernik, 2016

Secretion of Proteins

- Possible expression into the **periplasm** or **medium**
- Secretion controlled by a hydrophobic sequence at the N-terminus cleaved by signal peptidase
  - possible addition of a signal sequence to the protein (risk of inclusion bodies)
  - possible fusion with a naturally secreted protein (maltose-binding protein in *E. coli*)
  - possible secretion in gram-positive bacteria (*Bacillus*)
  - use of a special **Type I** secretion system (hemolysin secretion, *E. coli*) or **Type II** (Endotoxin A, *Pseudomonas*)
  - use of autotransport proteins

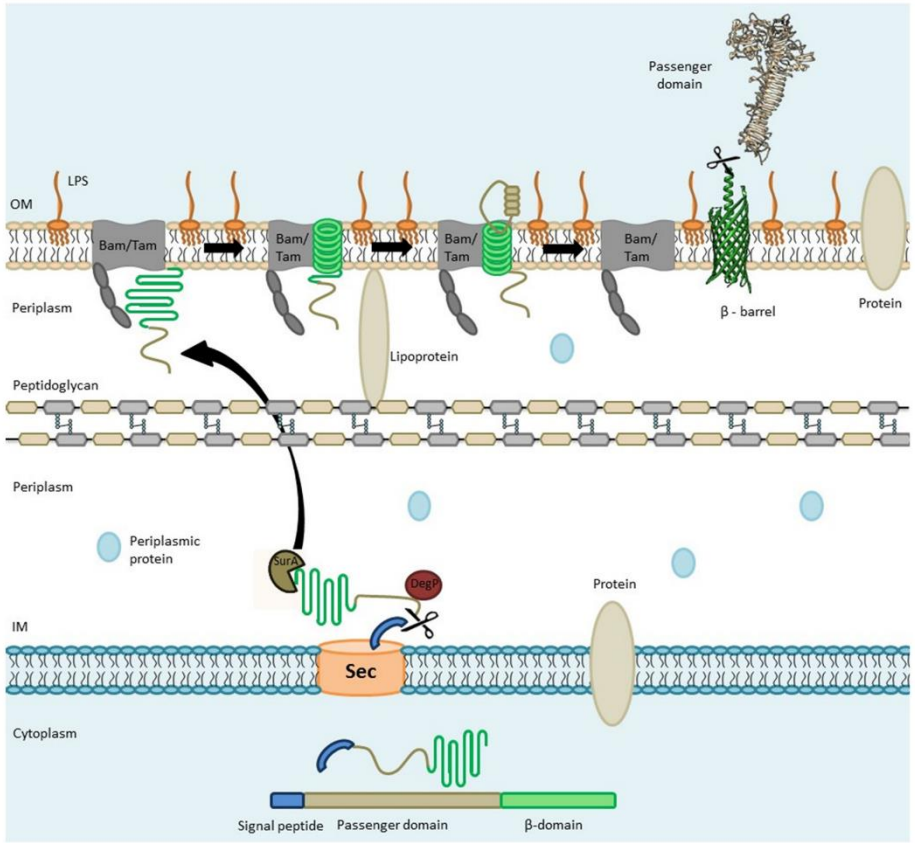
Secretion of Proteins

Secretion system of type I and II



Clark and Pazdernik, 2016

Autotransporter proteins



Protein glycosylation

- A whole range of proteins in higher organisms is **glycosylated**
- Glycosylation is necessary for proper function - e.g. membrane proteins
- The bacterium carries out **O-glycosylation** (N-glycosylation was also discovered in the genus *Campylobacter*)
- Eukaryotic organisms mostly have **N-glycosylation**
- Insect cells are the solution for the expression of glycosylated proteins

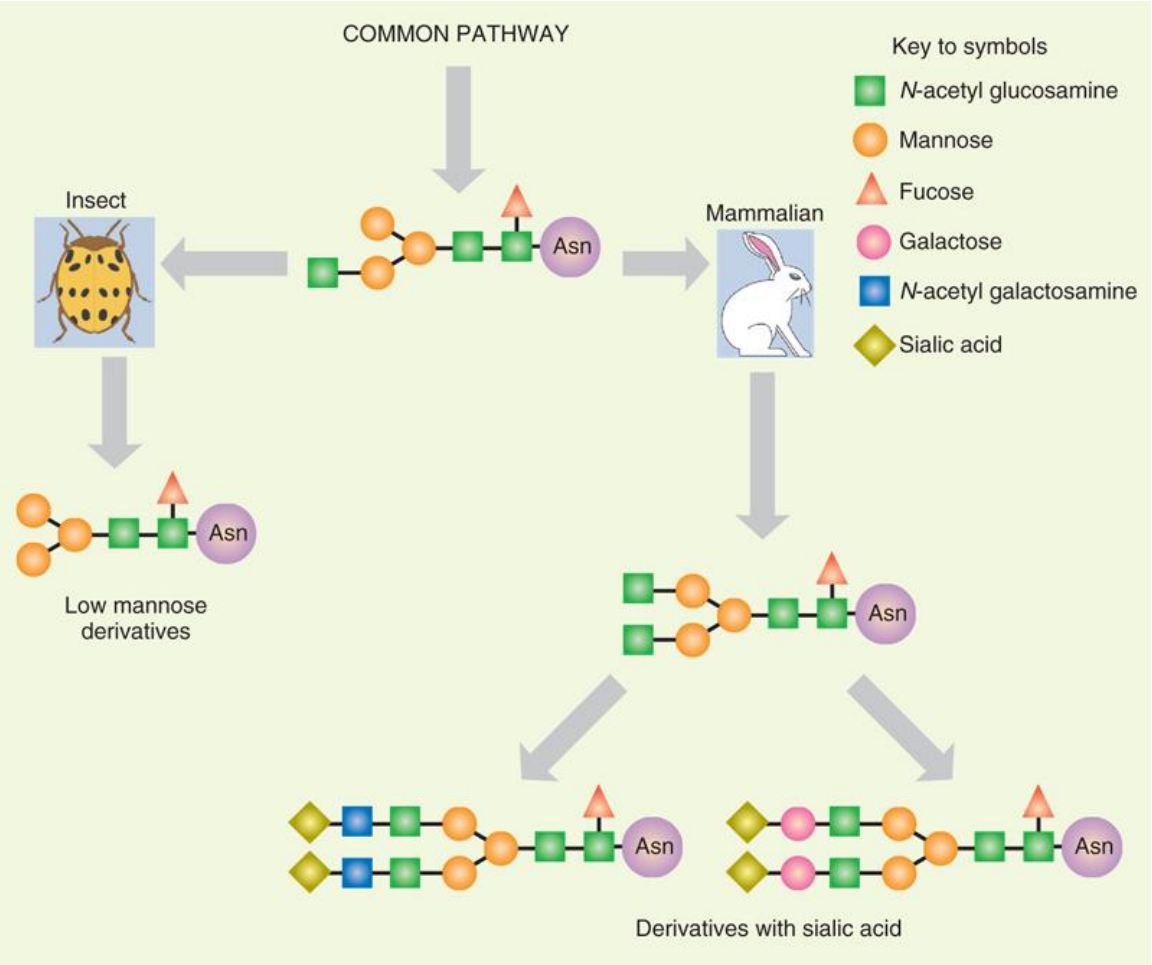
  - a different pattern of glycosylation compared to mammals

  - the solution is modified insect cells with a mammalian glycosylation pathway
- A change in the glycosylation pattern can affect the properties of the protein

  - recombinant human erythropoietin contains an extra N-glycosylation site (Asn-Xxx-Ser/Thr)

  - lower affinity to the receptor, but a longer half-life prolongs the overall clinical activity

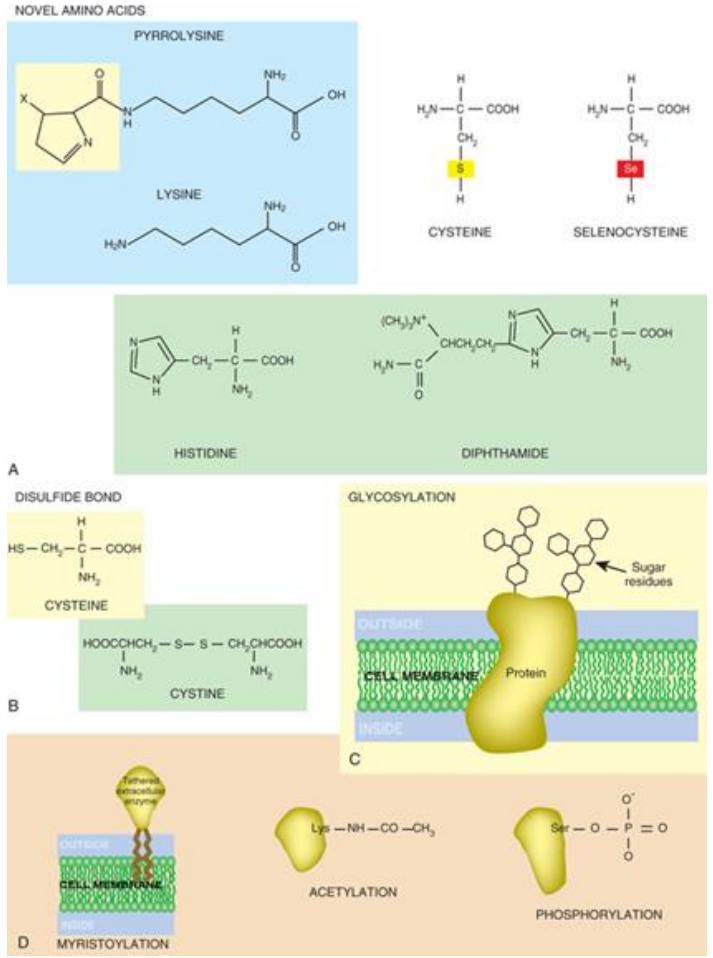
Protein glycosylation



Clark and Pazdernik, 2016

Protein Expression in Eukaryotic Cells

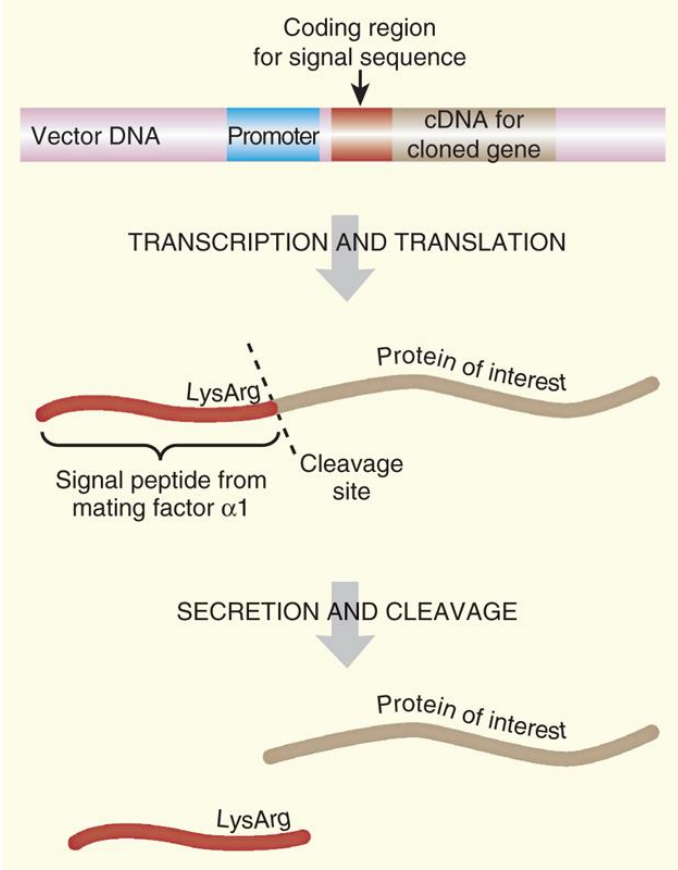
- A number of eukaryotic proteins are more efficiently expressed in eukaryotic cells
- Possibility of post-translational modifications
  - chemical modifications forming new amino acids
  - formation of disulfide bridges
  - glycosylation
  - addition of functional groups (fatty acids, acetylation, phosphorylation, methylation, sulfurization)
  - cleavage of pre-cursor proteins required for secretion, assembly, and/or activation



Clark and Pazdernik, 2016

Yeasts

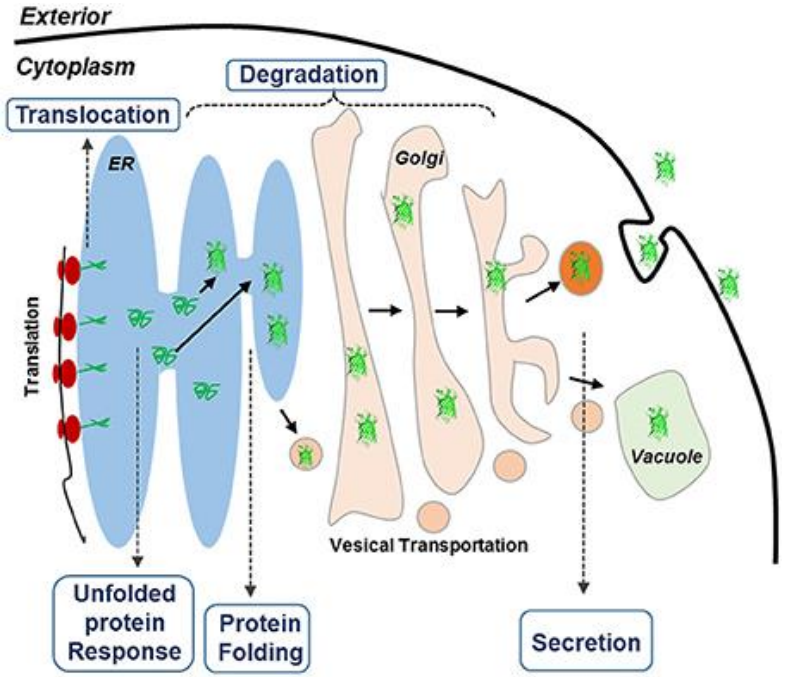
- **A whole range of advantages**
  - easy cultivation on a small and large scale
  - the yeast *S. cerevisiae* is considered a safe organism
  - yeasts secrete very few of their own proteins – an advantage in secreting the expressed protein
  - DNA can be easily transformed (chemically, enzymatically, electroporation)
  - characterization of a whole series of promoters for targeted expression
  - capable of a whole range of post-translational modifications characteristic of eukaryotic organisms
  - glycosylation takes place only in secreted proteins
- **Frequent secretion** of recombinant proteins by the signal sequence of the mating factor α gene
- The signal peptidase recognizes the Lys-Arg sequence



Clark and Pazdernik, 2016

Yeasts

- Currently expressed in the yeast *S. cerevisiae* and *P. pastoris*
  - insulin
  - clotting factor VIIIa
  - various growth factors
  - viral proteins for the production of vaccines or diagnostics (HIV, HBV, HCV)
- The most common **expression problems** in yeast
  - loss of expression plasmids in large-scale cultivations
  - secreted proteins remain between the PM and the cell wall
  - **hyper-glycosylation** of secreted proteins occurs (solution by strain modification)



Sheng et al. 2017

MUNI
SCI

GENE TECHNOLOGIES

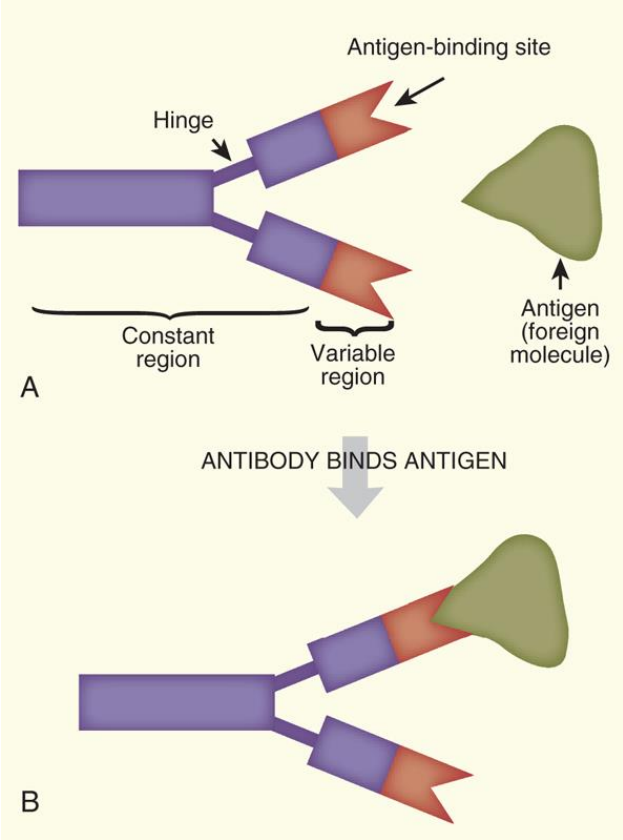
Technologies in Immunology

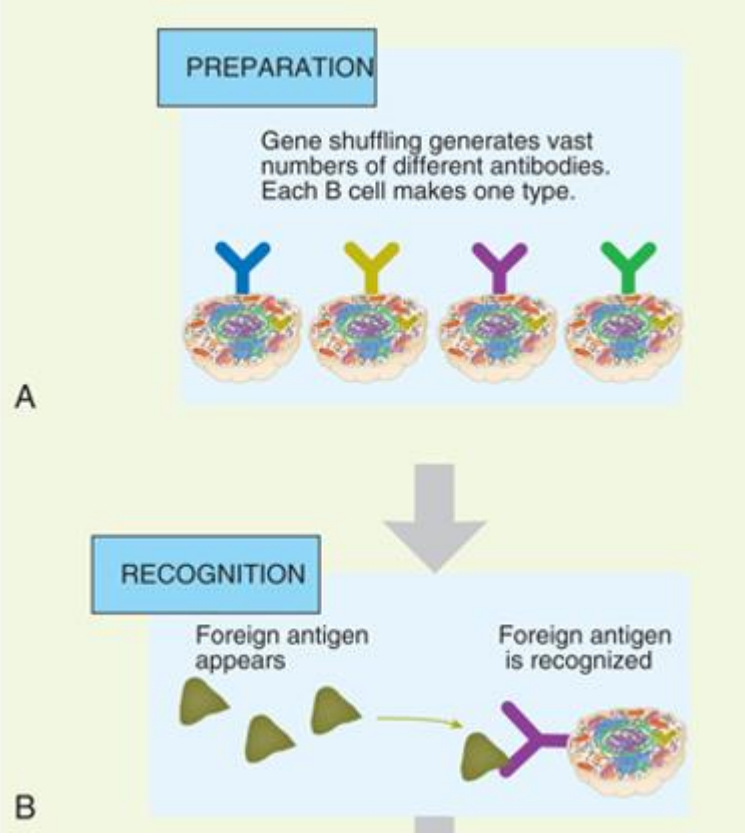
Antibodies (structure, function), targeted antibody design, monoclonal antibodies, ELISA, vaccines (design and production, identification of potential new antigens, DNA vaccines)

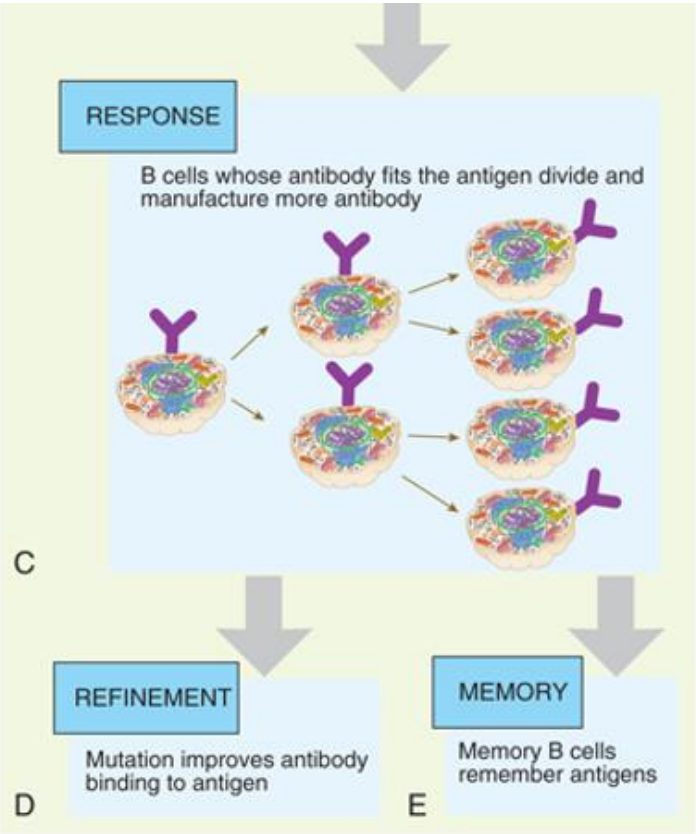
Introduction

- The surrounding environment is full of infectious microorganisms and virusesOchrana organismu pomocí buněk imunitního systému
- Protection of the body by the cells of the immune system
- Antigens - mostly proteins on the surface of microorganisms = activation of immune response
- Antibodies - recognize and bind to antigens = produced by B-cells of the adaptive immune system
- Antibodies mostly secreted into the lymph, some bind to surface = B-cell receptors
- Massive proliferation of B-cells producing antibodies recognizing a given antigen
- Immune system records all successfully used antibodies = faster and more massive response

Introduction

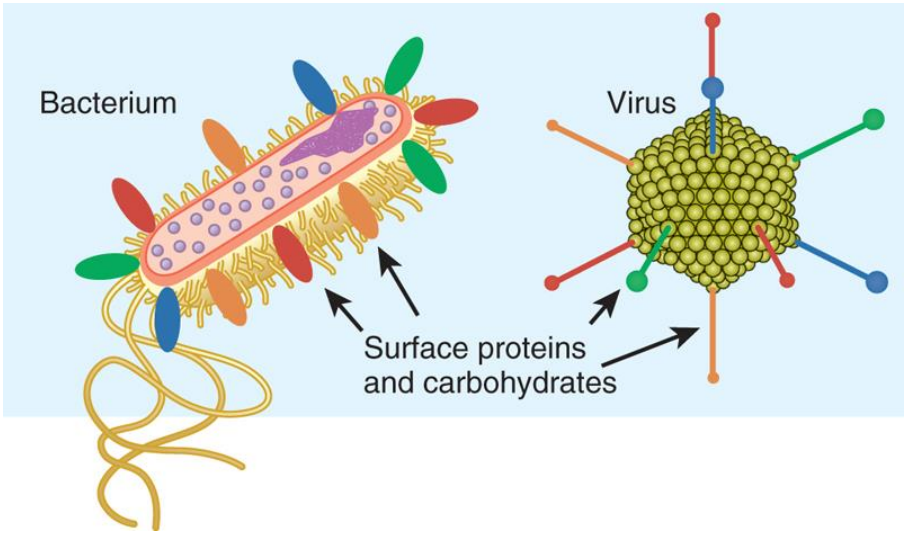


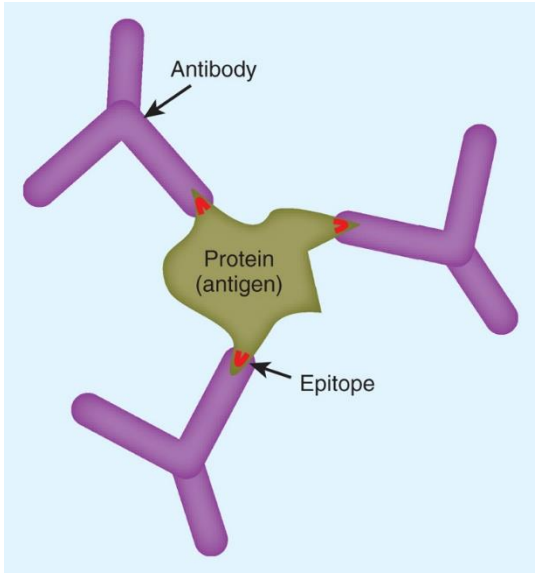




Antigen, antibody, epitope

- Antigen - a foreign molecule that activates the immune system
- Strongest immune responses = glycoproteins and lipoproteins
- Very often polysaccharides on the surface of microorganisms serve as antigens
- DNA can also serve as an antigen
- The animal immune system is based on specific (acquired) immunity divided into:
  - humoral immunity (mediated by immunoglobulins)
  - cell-mediated immunity (T-lymphocytes = T_H and T_C)
- Antibody = binding to whole proteins
- T-lymphocytes = binding to protein fragments
- Epitope - region of protein recognized by antibody

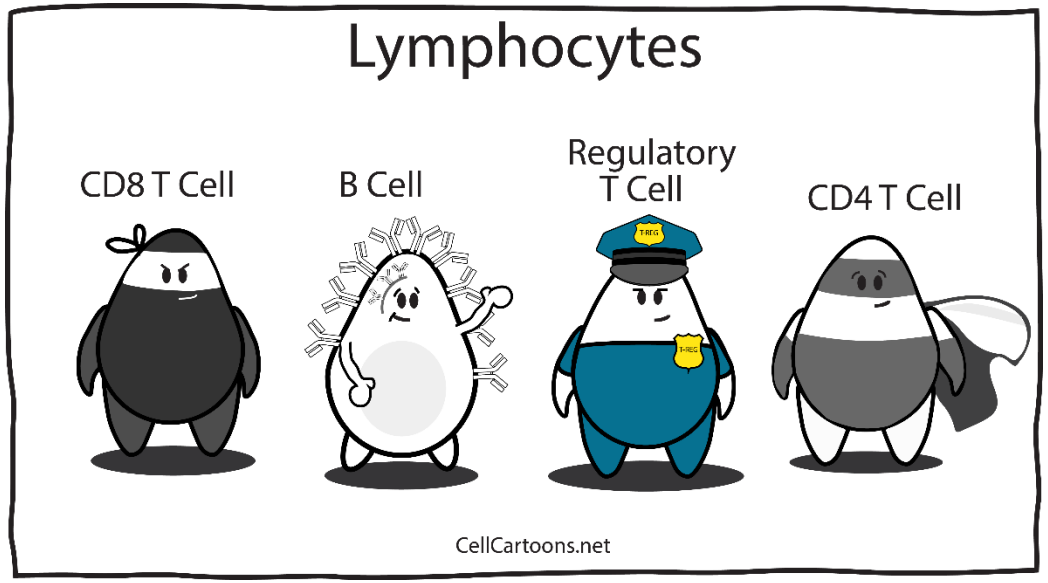


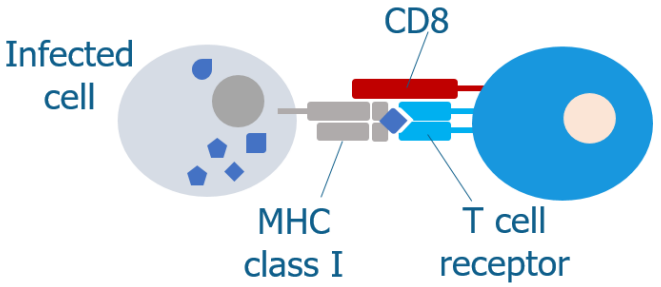


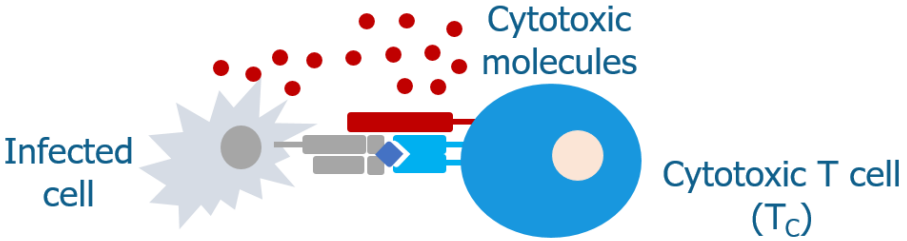
T-lymphocytes

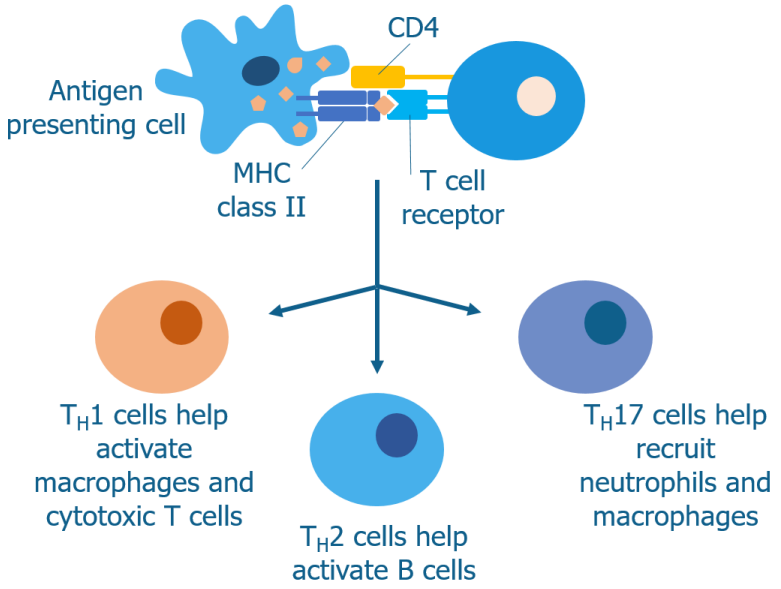
- recognize only antigens expressed on the surface of other cells, mainly macrophages, virus-infected cells or B-lymphocytes
- T-lymphocytes recognise these cells via class I and II major histocompatibility complex (MHCs) receptor proteins
- Class I activates T_H cells and class II activates T_C cells
- MHC receptors are encoded by a family of genes specific to each individual
- MHC receptors are also called major histocompatibility complexes HLA

T-lymphocytes



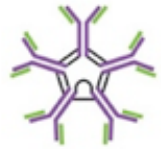






Structure and Function of Immunoglobulins

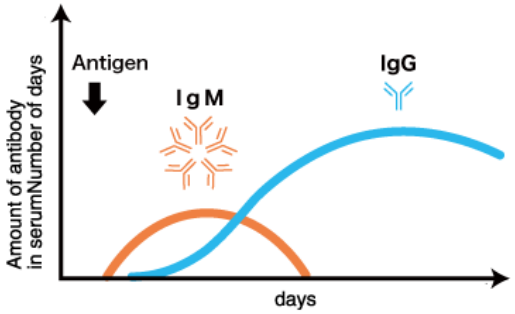
- Antibodies divided into 5 basic classes
- The most abundant are IgG in serum
- Only IgG antibodies cross the placenta
- IgA - secretory antibodies important in suppressing respiratory and gastrointestinal infections
- IgM - 10 binding sites = coating microorganisms and stimulating cells
- IgE - on the surface of mast cells, stimulation of allergic response by histamine release





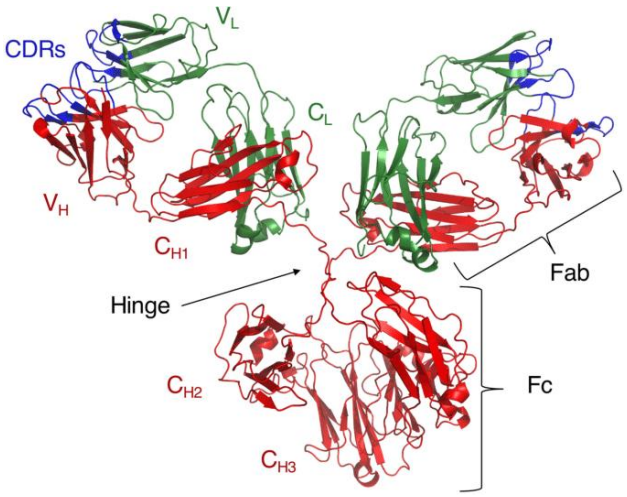
	IgM	IgG	IgA	IgE	IgD
Heavy Chain	μ (mu)	γ (gamma)	α (alpha)	ε (epsilon)	δ (delta)
MW (Da)	900k	150k	385k	200k	180k
% of total antibody in serum	6%	80%	13%	0.002%	1%
Fixes complement	Yes	Yes	No	No	No
Function	Primary response, fixes complement. Monomer serves as B-cell receptor	Main blood antibody, neutralizes toxins, opsonization	Secreted into mucus, tears, saliva	Antibody of allergy and anti-parasitic activity	B cell Receptor

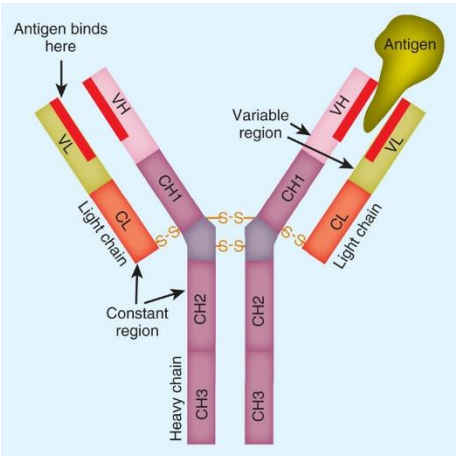
Levels of circulating antibodies to a specific antigen

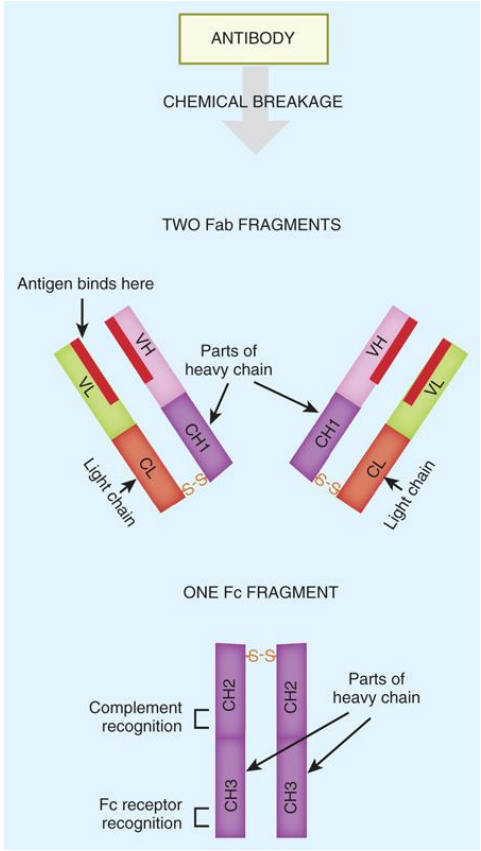


Structure and Function of Immunoglobulins

- IgG antibody consists of two light and two heavy chains
- Light chains encoded by one of two gene loci κ or λ
- Each of the light and heavy chains consists of one to four constant regions and one variable region
- The variable regions form the so-called paratope-antigen binding
- We have millions of different variable regions
- In the Pant region, antibodies can be divided chemically (by papain) into Fc and two Fab fragments

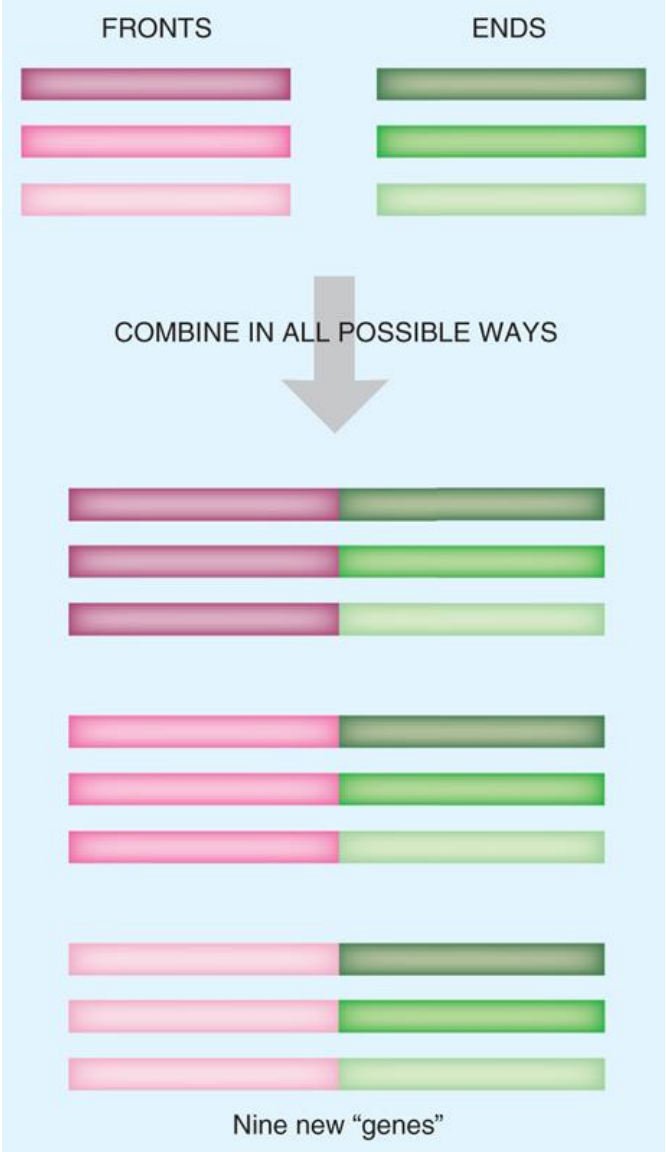






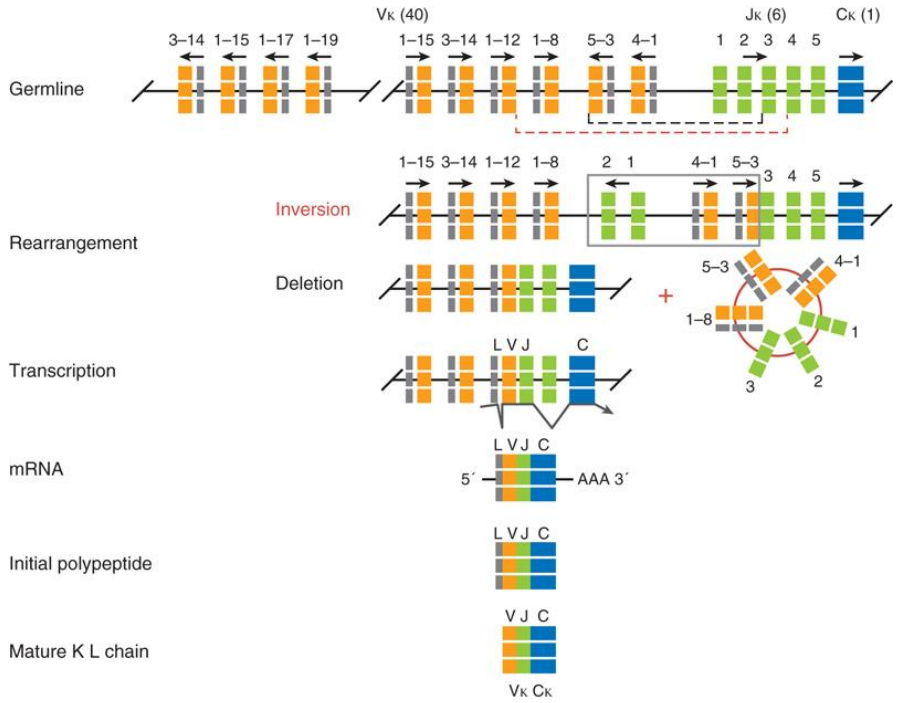
Diversity of Antibodies

- There are an almost infinite number of antigens = an almost infinite number of antibodies are needed
- Genetic problem concerning the number of genes encoding each antibody
- The entire human genome would encode only a few million antibodies
- The immune system generates a large number of sequences from a relatively small number of genes in the process of V(D)J recombination
- The immune system assembles genes for antibodies from collections of short DNA segments
- V(D)J recombination occurs in the bone marrow during B-cell development and is initiated by RAG1 and RAG2 proteins followed by NHEJ

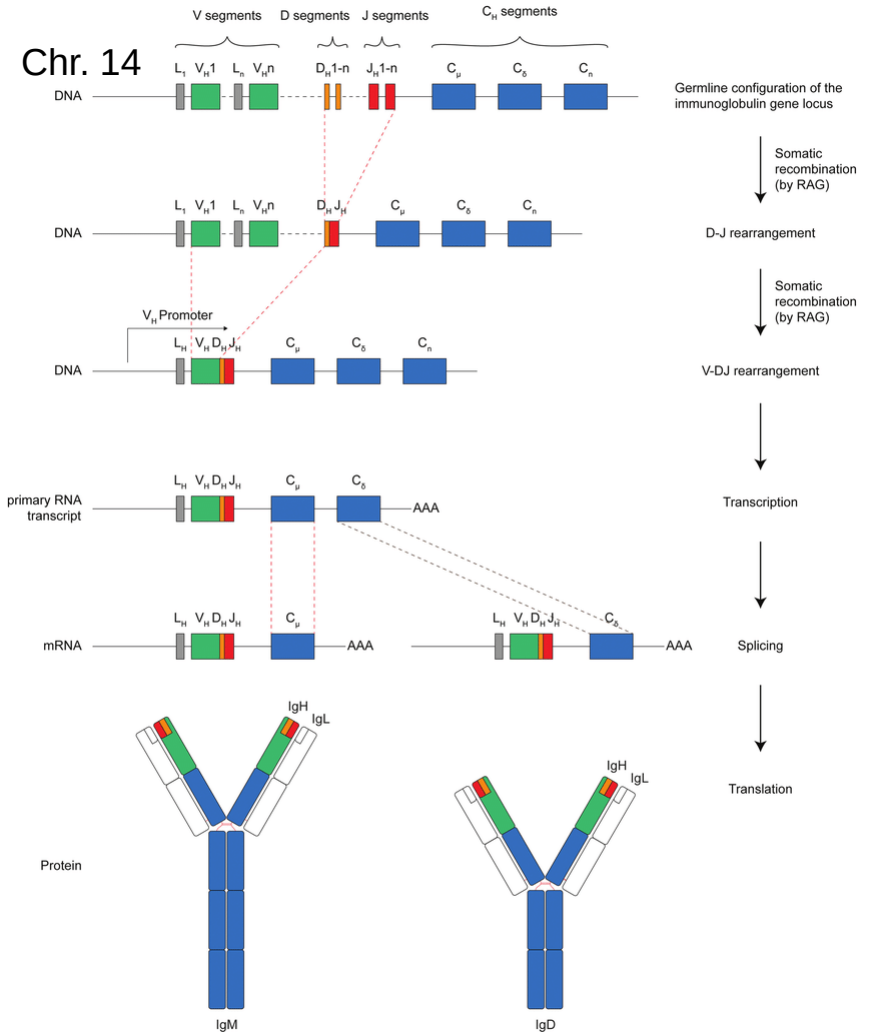


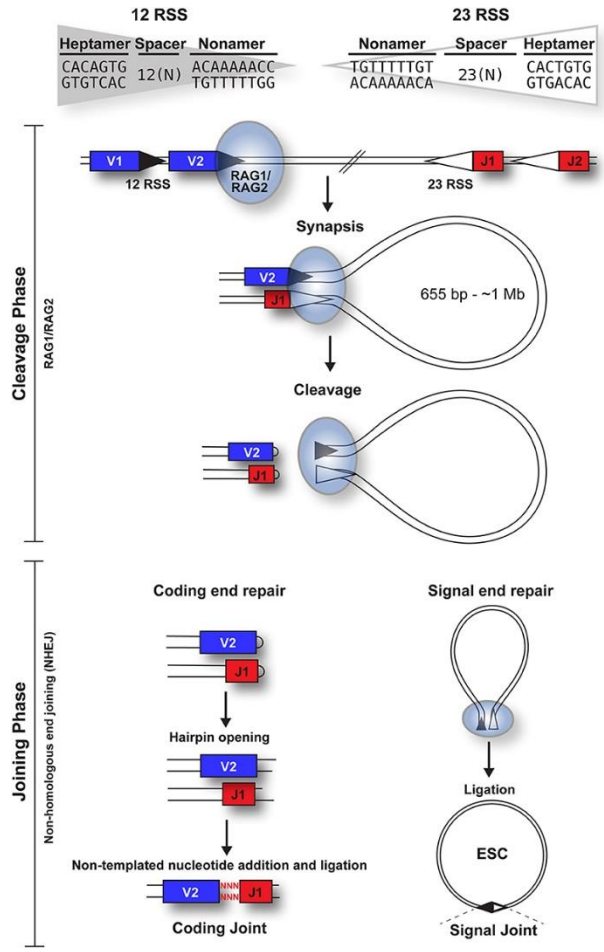
V(D)J Recombination

Chr. 2



Chr. 14





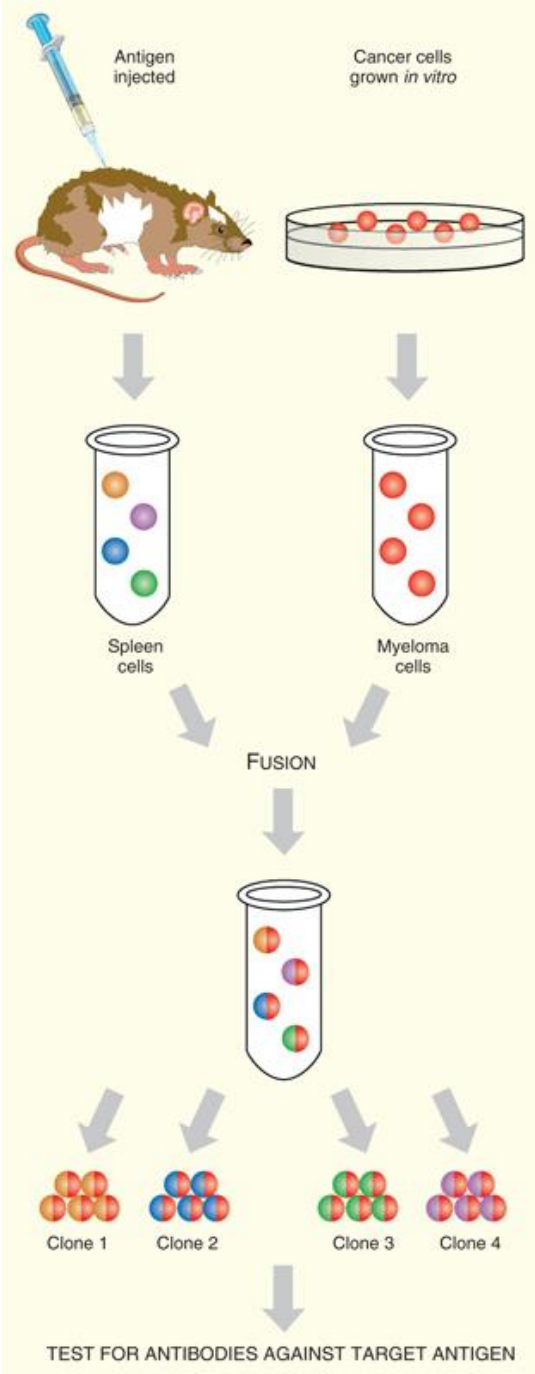
Smith et al. 2019

https://youtu.be/QTOBSFJWogE

Backhaus et al. 2018

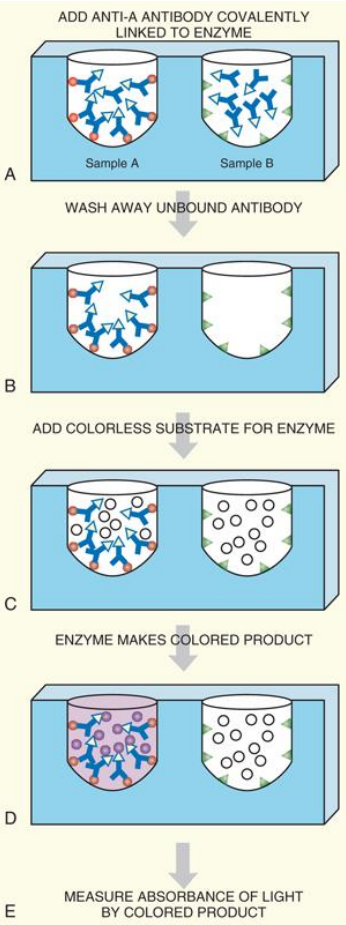
Monoclonal Antibodies

- Antibodies find wide clinical use
- Need for one specific antibody against an antigen
- One antigen has many epitopes = polyclonal antibodies
- Polyclonal antibodies = mixture of antibodies with different degrees of specificity and binding
- Monoclonal antibody = one specific antibody from one B-cell
- Viability of B-cells outside the body is very low = fusion with myeloma cells
- The resulting cell is called a hybridoma = a forever living cell producing the targeted antibody

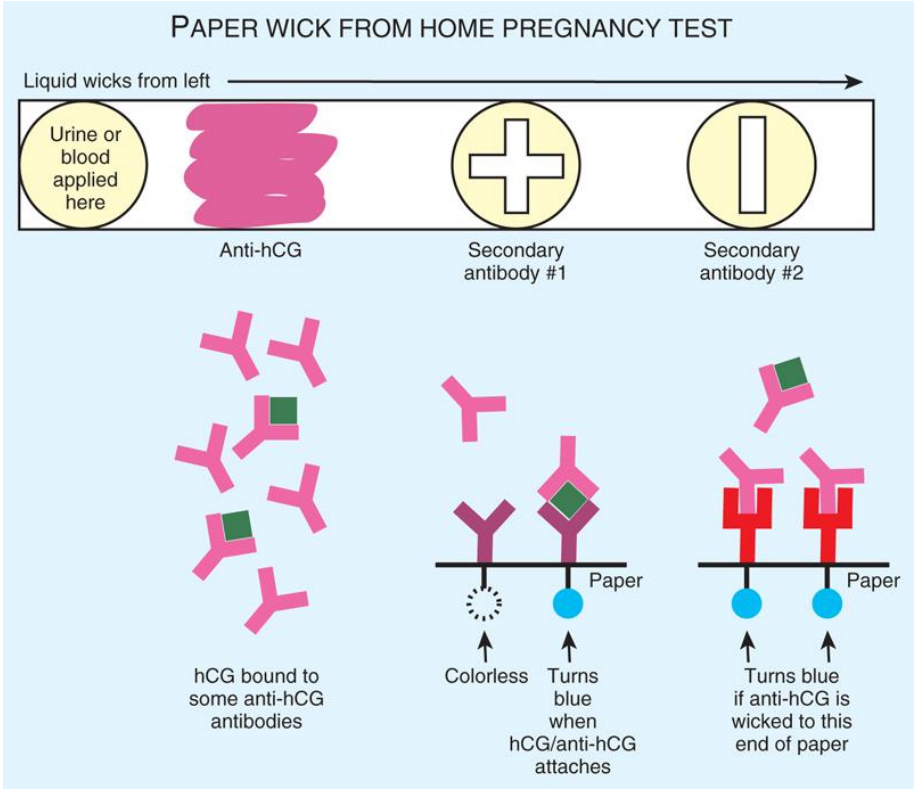


Use of Antibodies

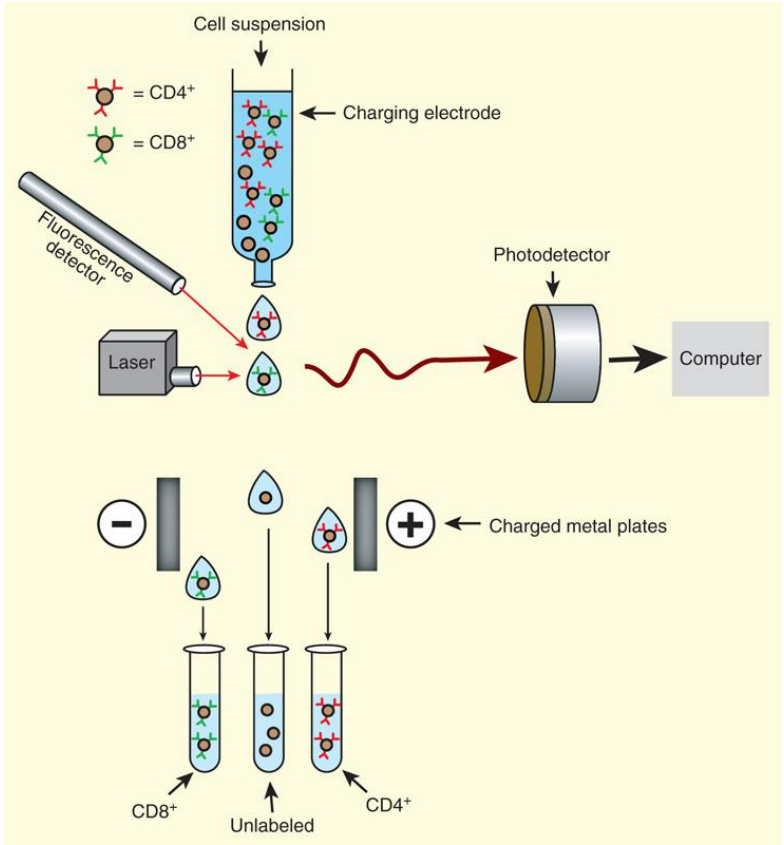
ELISA



Rapid tests



FACS (Fluorescence-activated cell sorting)

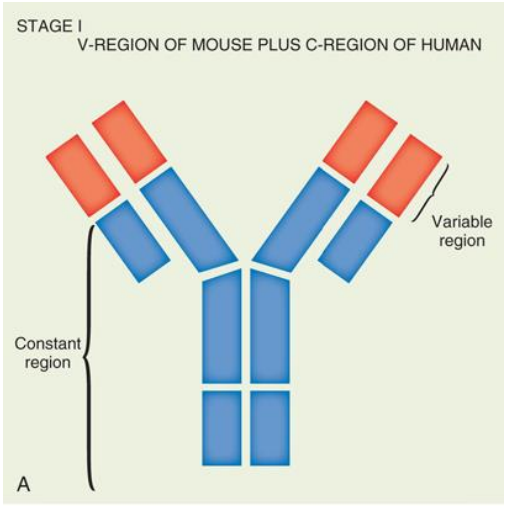


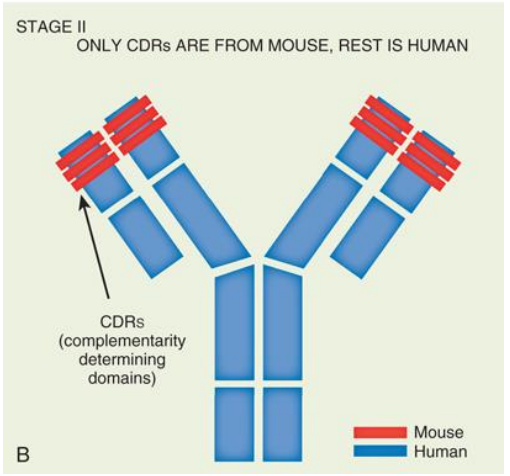
"Humanization" of Monoclonal Antibodies

- Human immune system recognises mouse antibodies
- Several solutions:
  - Replacing the C-region with a human variant of the antibody
  - Replacement of V-regions not involved in antigen recognition with a human variant
  - Complementarity Determining Region (CDR) - hypervariable region recognizing Ag

Names	Trade Names	Target	Antibody Format	Malignancy
Bevacizumab	Avastin	VEGF	Humanized IgG1	Glioblastoma, NSCLC, metastatic colon and kidney cancer
Cetuximab	Erbitux	EGFR	Chimeric IgG1	Head and neck squamous cell carcinoma, mCRC
Cixutumumab	IMC-A12	IGF1R	Fully human IgG1	Thymic carcinoma, soft tissue sarcomas, osteosarcoma, breast cancer, Ewing's sarcoma
Panitumumab	Vectibix	EGFR	Fully human IgG1	Metastatic colon cancer
Pertuzumab	Perjeta	HER2	Humanized IgG1	Metastatic breast cancer
Ramucirumab	Cyramza	VEGFR2	Human IgG1	Gastric cancer
Trastuzumab	Herceptin (Herclon)	HER2	Humanized IgG1	Breast cancer, gastric adenocarcinoma, gastroesophageal junction adenocarcinoma
Trastuzumab emtansine	Kadcyla	HER2	Humanized IgG1	Advanced breast cancer

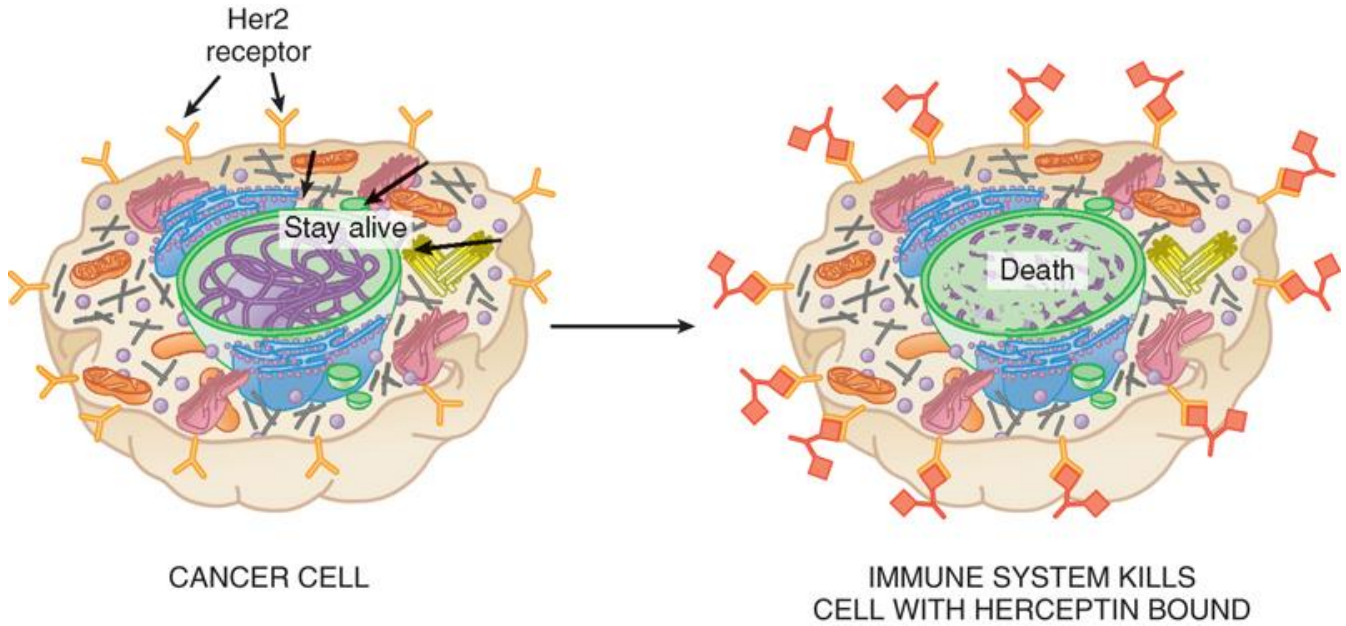
VEGF: vascular endothelial growth factor, NSCLC: non-small cells lung carcinoma, EGFR: epidermal growth factor receptor, mCRC: metastatic colorectal carcinoma, IGFR: insulin growth factor receptor, HER: human epidermal growth factor receptor, VEGFR: vascular endothelial growth factor receptor.





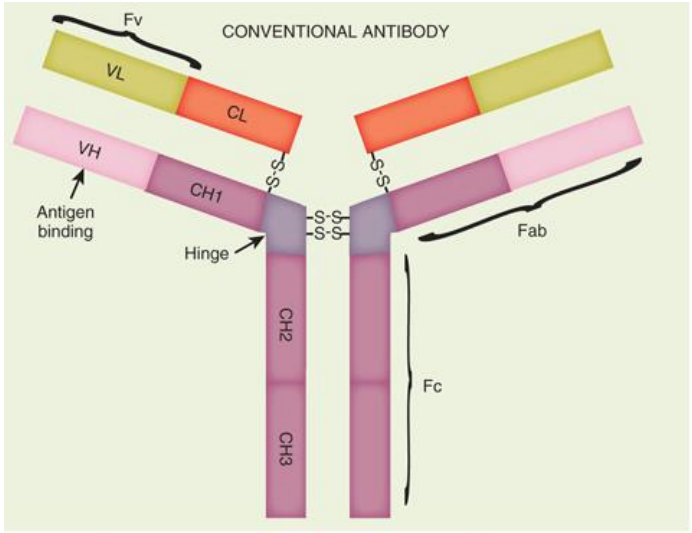
Herceptin and Casirivimab

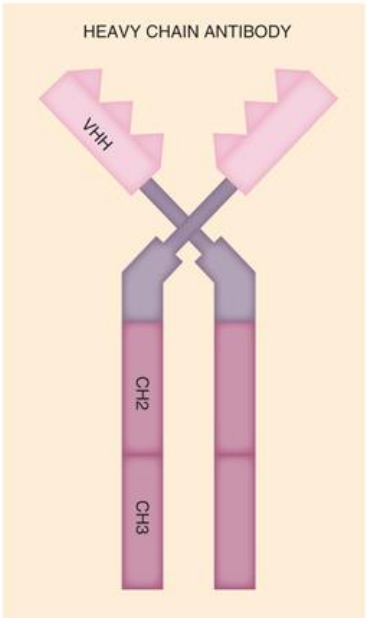
- Monoclonal antibody recognises the epidermal growth factor receptor type 2 (HER2)
- In breast cancer patients, HER2 overproduction is associated with resistance to chemotherapy
- Binding of antibodies to the receptor prevents its internalization = better efficacy of chemotherap
- Casirivimab - a monoclonal antibody that recognizes the SARS-CoV-2 coronavirus spike protein



Nanobodies

- Antibodies from camels, alpacas and llamas have only heavy-chain antibodies (hcAb)
- The antigen is bound by the terminal variable region of the heavy chain called the VHH (12-15 kDa)
- Recombinant antibodies containing only this part are called nanobodies (Nb)
- The VHH region has a very high affinity for the antigen
- Nanobodies can cross into the brain





chromotek
new tools for better research
part of Proteintech Group

Vaccines

- The immune system remembers foreign antigens - immune memory
- Special memory B-cells mediate immune memory
- Vaccines consist of derived infectious agents that can no longer cause disease but are still antigenic

  Vaccines:

  - Attenuated = pathogens still alive but no longer producing disease-causing toxins or proteins

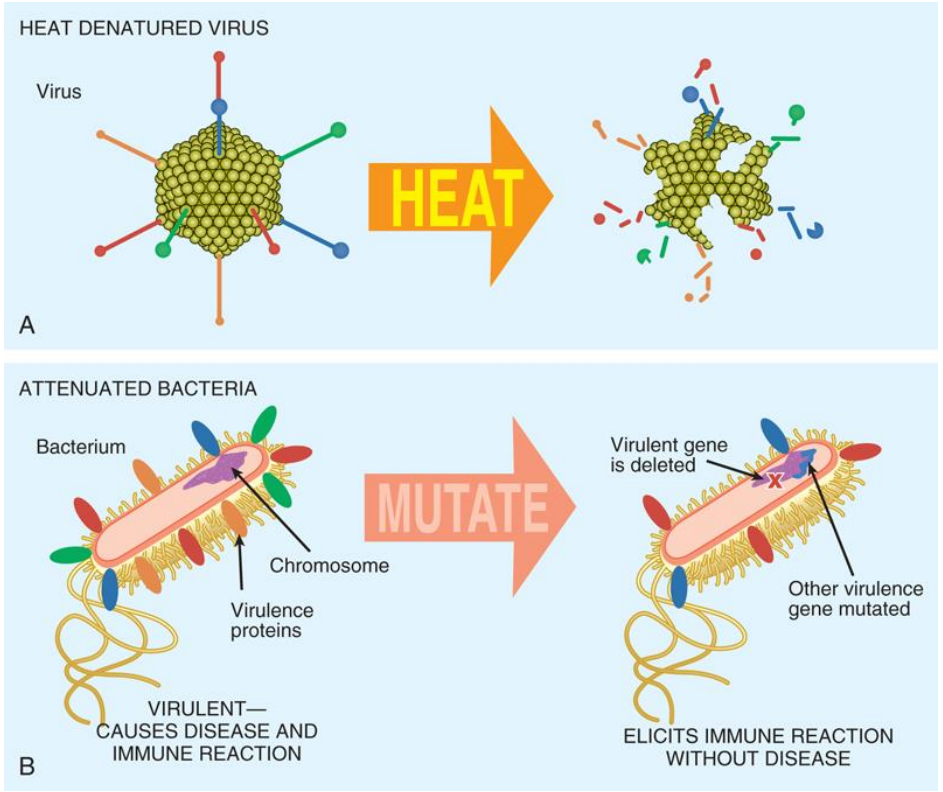
  - Subunits = effective against only one component of the pathogen, often requires the use of adjuvants

  - multivalent = targets several proteins from one or more viruses

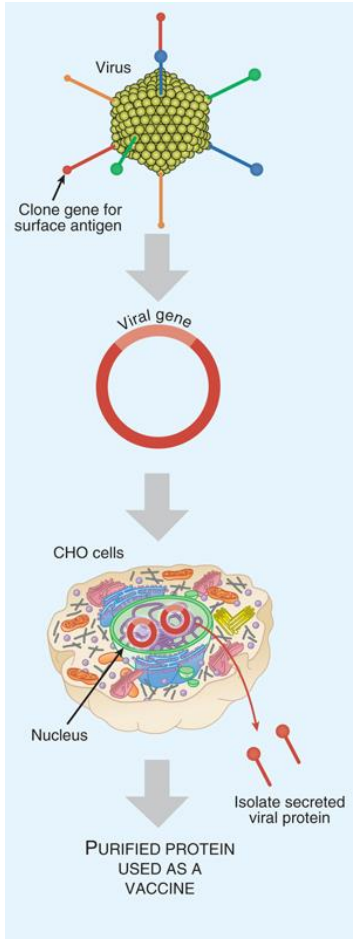
- Vaccines from attenuated microorganisms usually induce best immune response

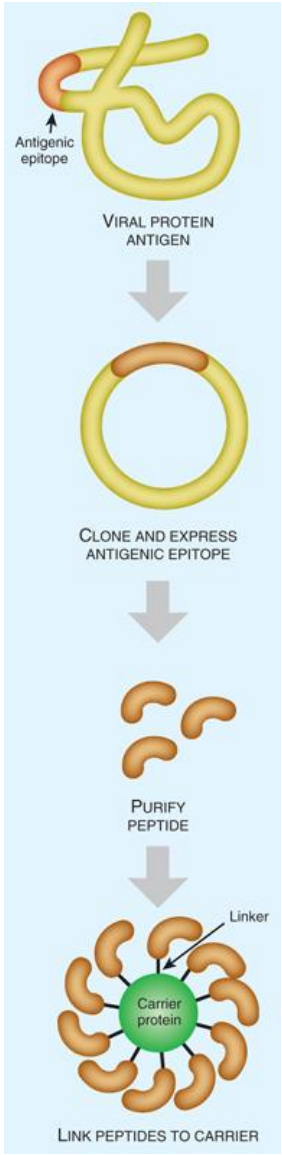
Vaccines

Attenuated vaccines



Subunit vaccines





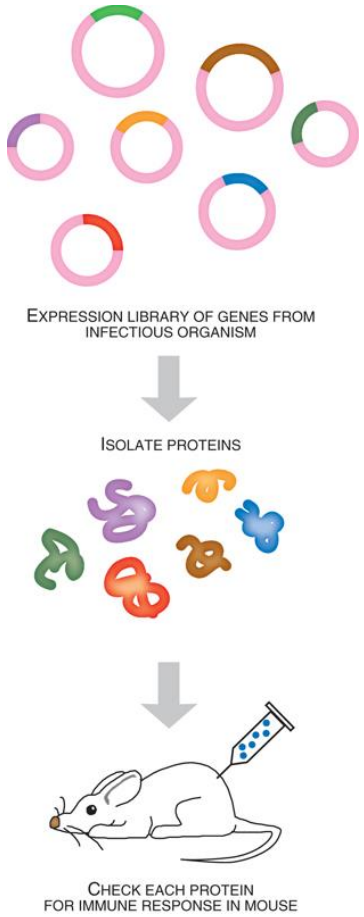
Search for Suitable Antigens and Adjuvants

– Reverse vaccinology = sequential cloning of pathogen genes and expression of proteins used for immunization (vaccine for Neisserie meningitidis serogroup B)

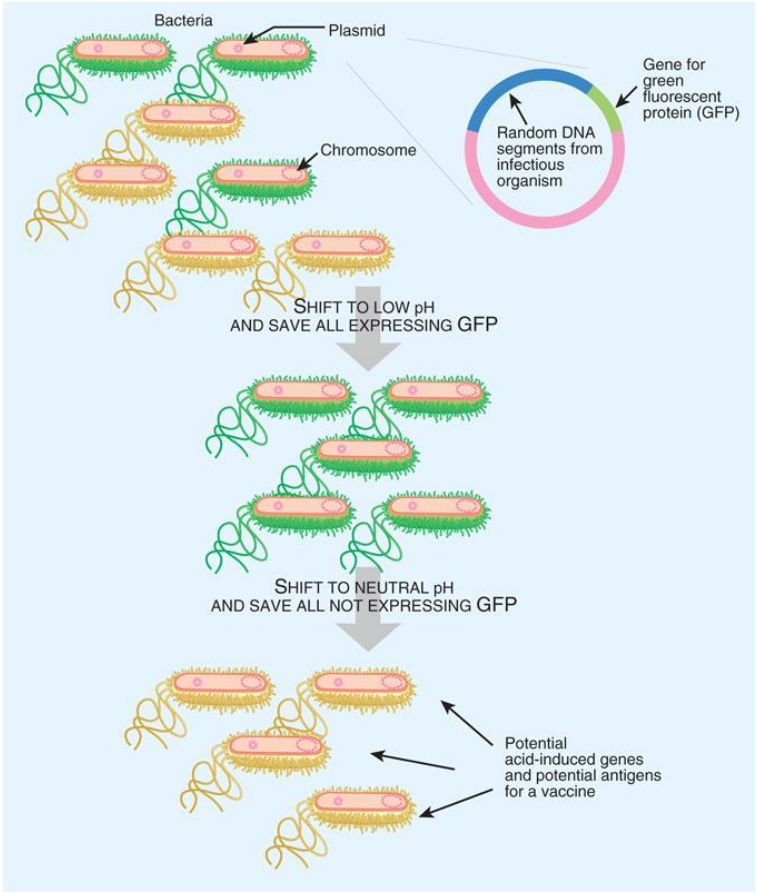
Adjuvant	Composition	Vaccines
Aluminum	One or more of the following: amorphous aluminum hydroxyphosphate sulfate (AAHS), aluminum hydroxide, aluminum phosphate, potassium aluminum sulfate (Alum)	Anthrax, DT, DTaP (Daptacel), DTaP (Infanrix), DTaP-IPV (Kinrix), DTaP-IPV (Quadracel), DTaP-HepB-IPV (Pediarix), DTaP –IPV/Hib (Pentacel), Hep A (Havrix), Hep A (Vaqta), Hep B (Engerix-B), Hep B (Recombivax), HepA/Hep B (Twinrix), HIB (PedvaxHIB), HPV (Gardasil 9), Japanese encephalitis (Ixiaro), MenB (Bexsero, Trumenba), Pneumococcal (Prevnar 13), Td (Tenivac), Td (Mass Biologics), Tdap (Adacel), Tdap (Boostrix)
AS04	Monophosphoryl lipid A (MPL) + aluminum salt	Cervarix
MF59	Oil in water emulsion composed of squalene	Fluad
AS01B	Monophosphoryl lipid A (MPL) and QS-21, a natural compound extracted from the Chilean soapbark tree, combined in a liposomal formulation	Shingrix
CpG 1018	Cytosine phosphoguanine (CpG), a synthetic form of DNA that mimics bacterial and viral genetic material	Heplisav-B
No adjuvant		ActHIB, chickenpox, live zoster (Zostavax), measles, mumps & rubella (MMR), meningococcal (Menactra, Menveo), rotavirus, seasonal influenza (except Fluad), single antigen polio (IPOL), yellow fever

Search for Suitable Antigens and Adjuvants

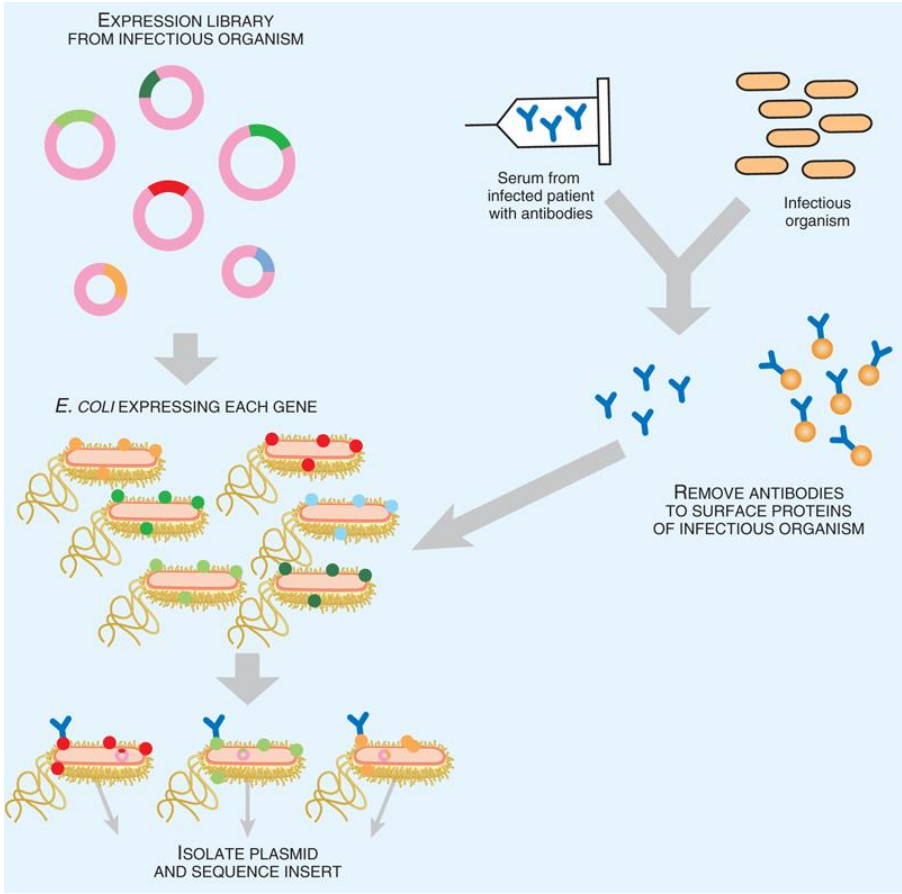
Reverse vaccinology



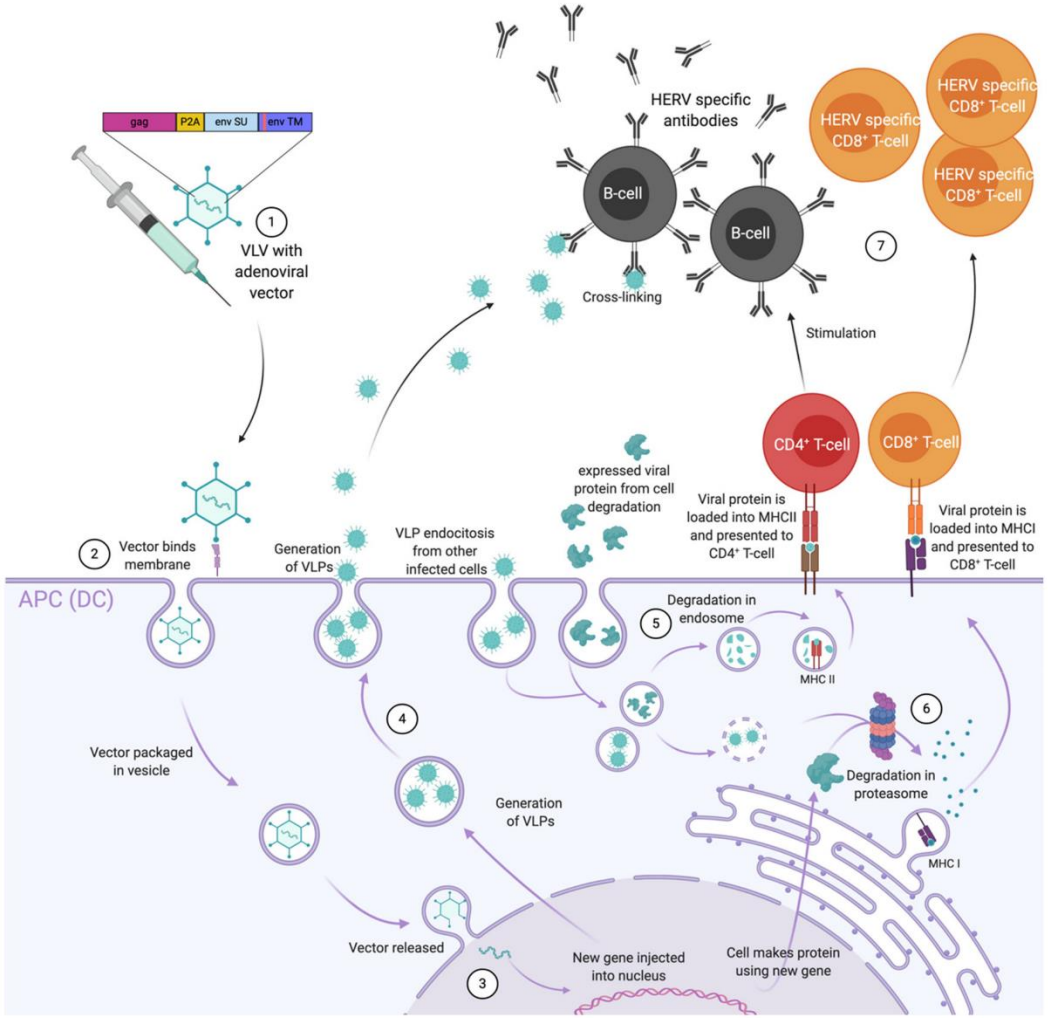
Differential fluorescence induction (DFI)



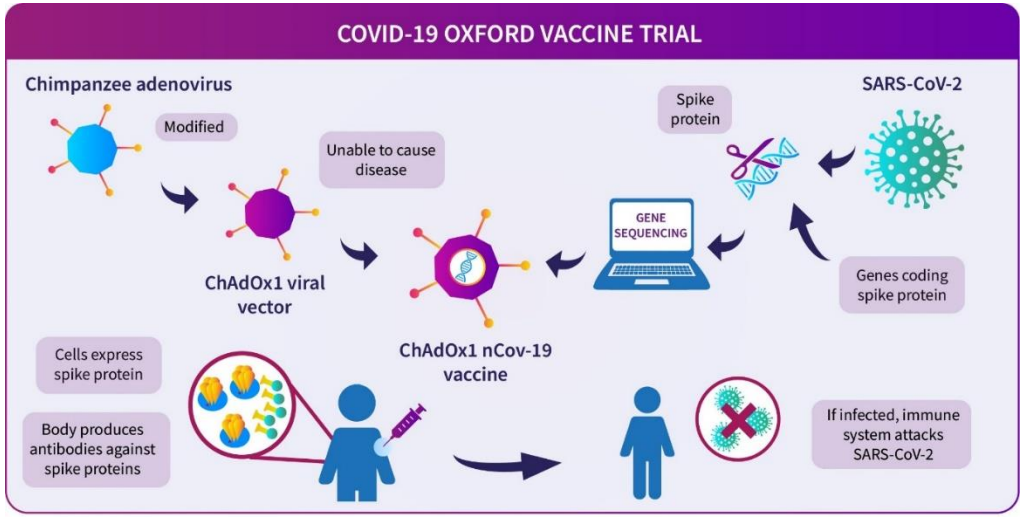
In-vivo induced antigen technology (IVIAT)



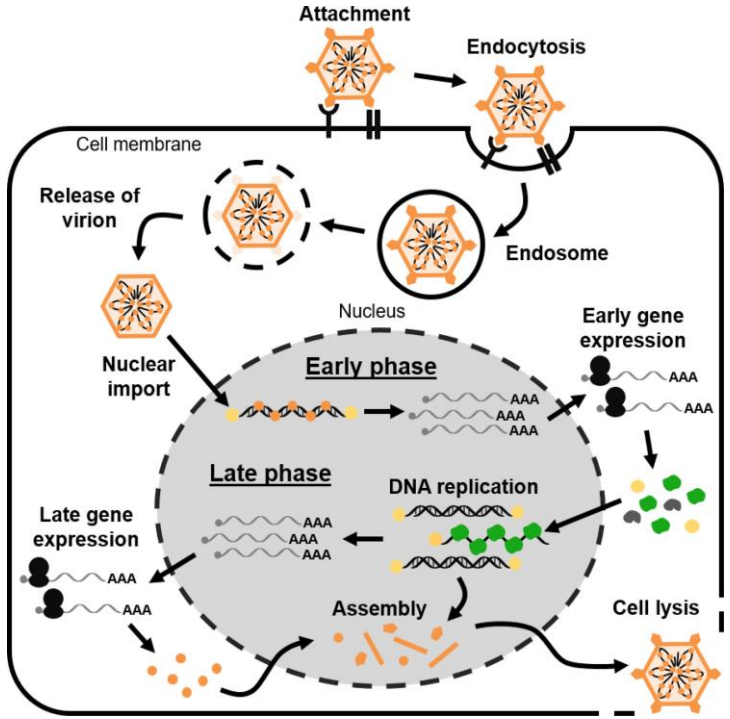
Adenovirus vaccines



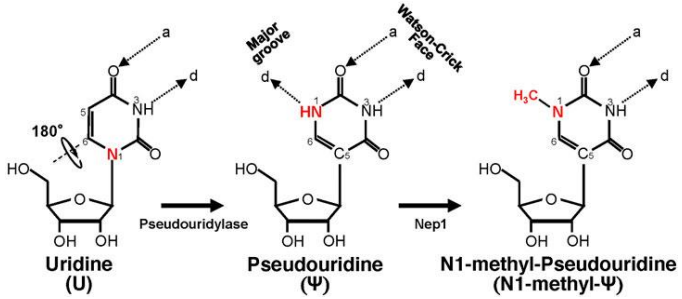
Bermejo et al, 2020

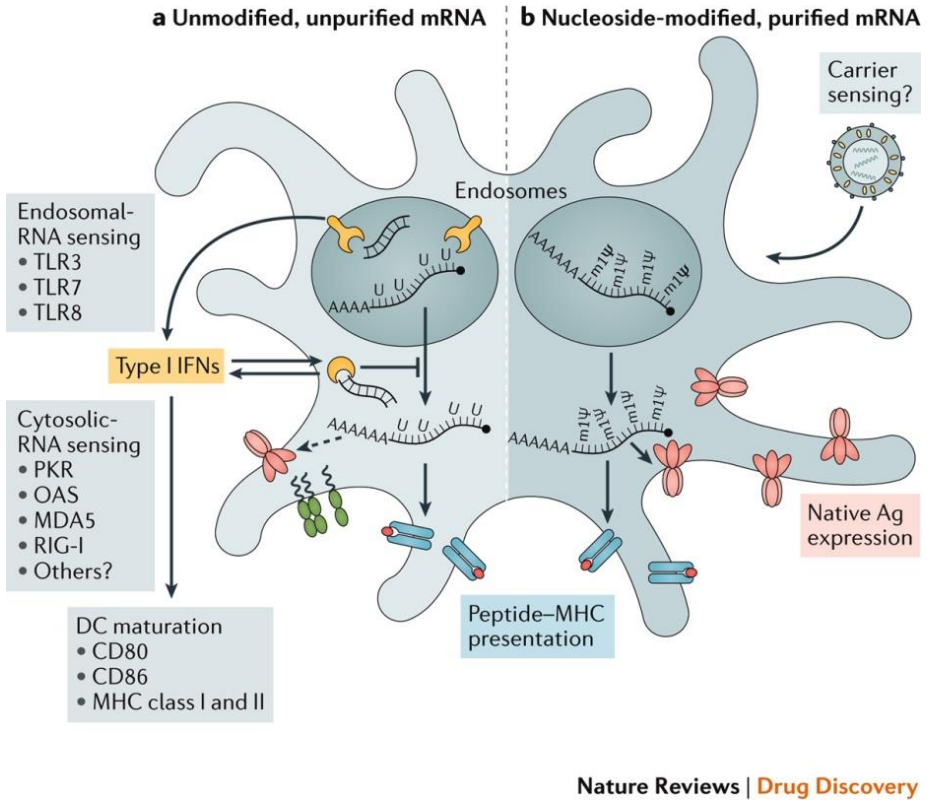


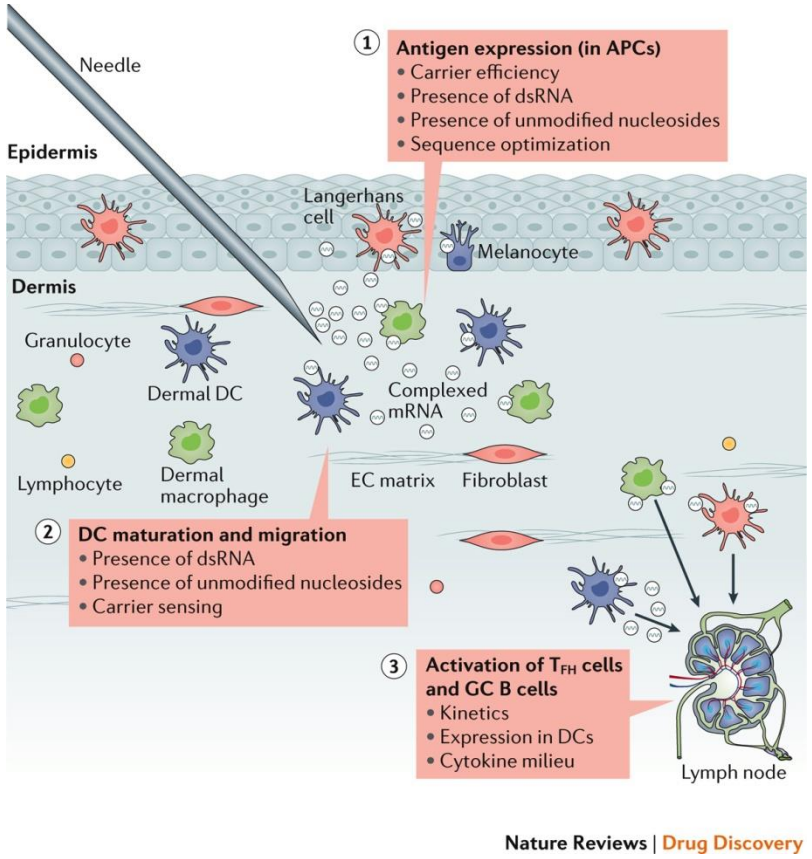
https://sputnikvaccine.com/about-vaccine/



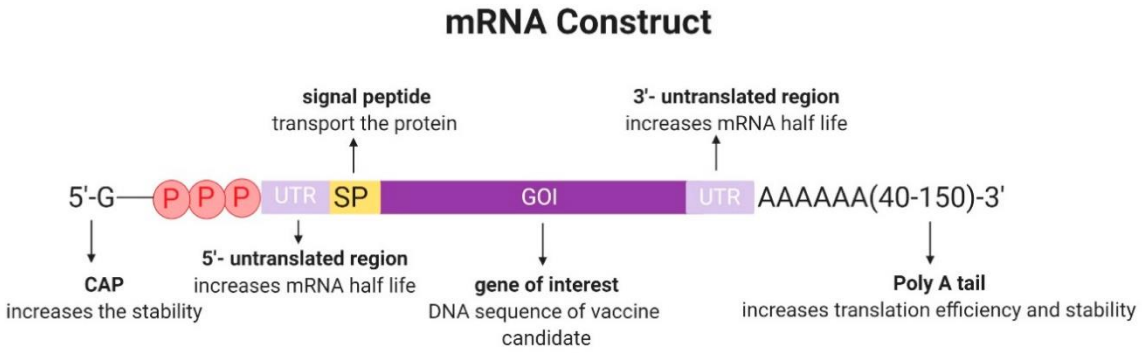
mRNA vaccines

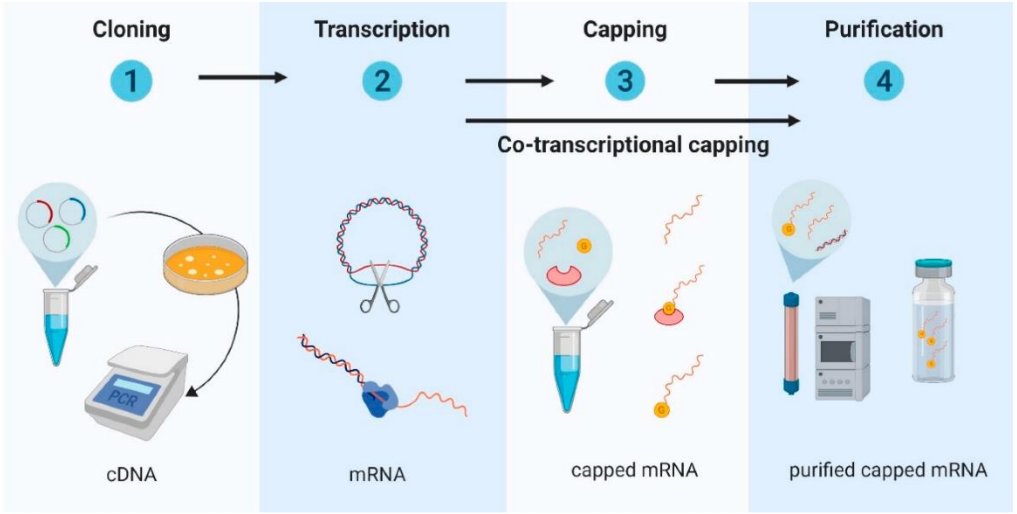






mRNA vaccines





Versteeg et al, 2019