

PROKARYOTIC GENOMES

PART 1

(OGGIONI)

The material of the second hours is **summarized** and highlighted in gray. None of it is going to be asked in the exam, except for the last lecture about antibiotics.

GENOME STRUCTURE

❖ CHARACTERISTICS OF BACTERIAL GENOME

Bacterial **genes do not have introns**, hence a bacterial gene is a sequence of nucleotides from the first codon to the stop codon.

- **SINGLE CHROMOSOME**, except: Leptospira e Burkholderia
- **HAPLOID CHROMOSOME** except: Neisseria gonorrhoeae
- **CIRCULAR CHROMOSOME** except: Streptomyces, Borrelia e Agrobacterium
- **CHROMOSOME WITH A SIZE OF 450.000 TO 9.000.000 bp** except endosymbionts as Carsonella ruddii (150K) or mitochondria (15K).
- **GENES WITHOUT INTRONS** except for rare group I introns, or mobile elements called group II introns
- **ONE GENE EVERY 1.000 BP** of course there are exceptions of genes that go from a distance of a few bases to over 10 kb
- **Genes do not code for polyproteins** except for cytochrome B and C1 in Bradyrhizobium
- **Genes do not overlap** except for the endosymbiont Carsonella ruddii

As we can see, the bacterial world is made of exceptions. There are a lot of rules, but each has an exception. In fact, **bacterial genomes are really different from each other** (more than a human genome compared to a plant one).

- while humans have different alleles (a child differs from their parents with respect to their alleles, not their genes), bacteria differ in genes.

As we have seen in a previous lecture, most bacteria have a chromosome but also have a few plasmids. **PLASMIDS** are not part of the chromosome of the bacterium, but are part of its genome. We define plasmids as small, **extrachromosomal DNA molecules within a cell that is physically separated from chromosomal DNA and can replicate independently**. We can also say that the genes contained in the chromosome are more essential than the ones contained in the plasmids, however this definition really depends on the environment, as some genes may be important in one environment and not important in another.

Many features about bacteria can be discovered by RNA sequencing, information that you wouldn't otherwise have from the genome. For example if certain genes are transcribed together or not.

❖ OPERONS

Bacterial genome is organised in genes, which are organised in **OPERONS**.

- ★ an operon is a unit of co-transcribed genes
- ★ a locus is a group of related genes

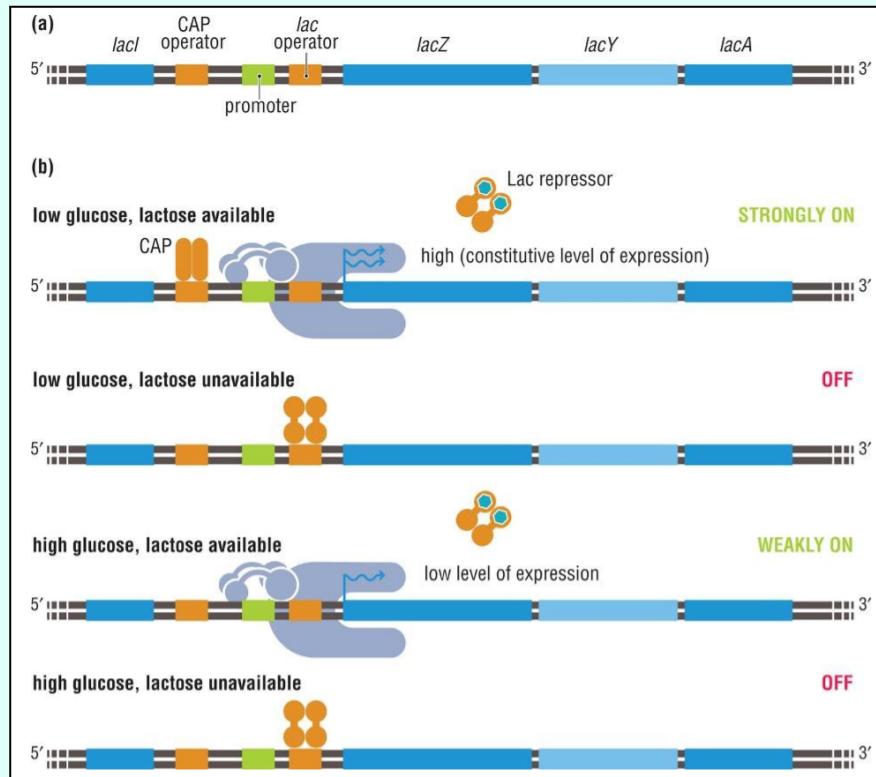
EXAMPLE: THE LAC OPERON.

lac operon contains genes (*lacZ*, *lacY*, *lacA*) that promote the **hydrolysis of lactose**.

- when **lactose levels are low**, the cell doesn't need any enzymes to hydrolyze it, and therefore the **transcription is off**. This happens because the **lacI repressor** is bound to the **lac operator**.¹ The lac operator **overlaps the -10 box of the promoter** of the lac operon, and therefore when a molecule is bound to the operator, there won't be space for RNAP to bind.
- the situation changes when **lactose levels are high**. In this case **allolactose** is present and acts as an inducer. Allolactose **binds to the lacI repressor**, which changes its conformation (we now say it's in *holo* form, in opposite to the *apo* form). This **conformational change** makes it so that lacI doesn't **bind to the operator anymore**, and the promoter sequence is therefore free. Now RNAP can bind to the promoter and begin transcription.

The functioning of lac operator is linked to the functioning of another operator, **CAP**, that is activated when glucose levels are low. **These two operons are linked as the cell doesn't need to hydrolyze lactose if glucose is already present, since it would be just spending energy.**

Hypothesizing that there is lactose present, when glucose levels are high the transcription of the lac operon isn't very high, while when glucose levels are low transcription of lac operon is strongly on.

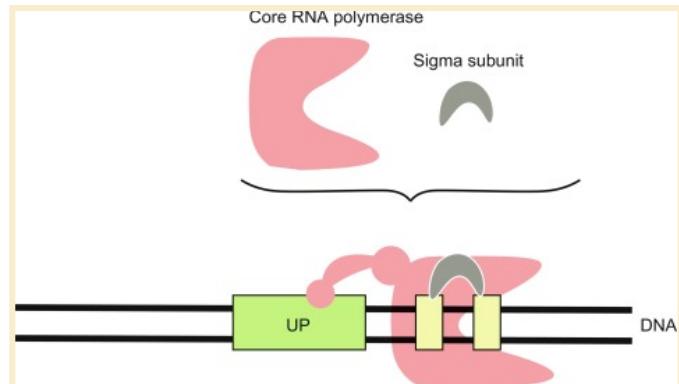


¹ In bacteria, **regulatory sequences** recognized by regulators are called operator sites. Operators are close or overlapping to the promoter, that is the sequence composed by the -10 and the -35 box that the RNAP reads.

❖ SIGMA FACTORS

Let's consider the example of *Bacillus subtilis* sporulation.

A sigma factor is a bacterial transcription initiation factor that **enables specific binding of RNA polymerase (RNAP) to gene promoters.**



More than one sigma factor exists per bacteria (they are called **alternative sigma factors**)

For example, *Bacillus subtilis* usually has a sigma that recognises fifty percent of the genes. However, as seen in the previous lecture, **when a bacteria cell is particularly dry it makes spores**, stopping every function. To do this *Bacillus subtilis* has to shut down every gene: the simplest approach to this is **simply changing the sigma factor in the RNA polymerase**. This new complex will not recognise the same old genes, allowing only sporulation since new genes are transcribed.

❖ OTHER IMPORTANT CONCEPTS

The bacterial genome is organised in functional units of related genes

- An **OPERON** is a unit of **co-transcribed genes**. Often the genes encode for enzymes which are metabolically related. An example of this is the lac operon, where three genes are transcribed at the same time. We can understand if genes are transcribed together from the previously mentioned **RNA sequencing**.
- A **LOCUS** is a group of **related genes**. It may include the operon and its regulator
- A **REGULON** is a group of co-regulated genes
- RNAseq data can indicate if genes are co-transcribed

1. The definition of species implies the creation of an offspring, as this concept was born in botany. However, bacteria do not sexually reproduce. And, furthermore, they are incredibly diverse. It is therefore wrong to define bacteria as a species, we define them as an **operational taxonomic unit** (OTU).
2. While the genome of a human cannot change, or, whenever change (ie cancer) the mutation cannot be passed on, the **genome of bacteria can change** thanks to plasmids.
3. **CORE GENOME**: includes all the genes present in all isolates² of a given species (or in the case of bacteria, operational taxonomic unit). In humans the core genome corresponds to the whole genome, instead in bacteria it corresponds to only about 70%.
 - a. If I have brown eyes and someone has blue eyes, it doesn't mean we have different genes. We just have different alleles. Instead bacteria have entirely different genes.
4. The genes that are not present in the core genome are named **ACCESSORY GENES**. We humans don't have them, but they constitute around 30% of bacteria's overall genome. It is also important to remember that the researcher decides the cut off.
5. The genome that comprehends accessory genes and core genes is named **PAN GENOME**.
 - a. we say a species has a **closed pan genome** when there are no new genes no matter how many individuals we sequence.
 - b. we say a species has an **open pan genome** when the more individuals you sequence the more different genes you find.

² a closer estimate would be 98% of all isolates because there can be sequencing errors.

❖ WHY DO WE NEED ALL OF THIS?

EXAMPLE: *Neisseria meningitidis*

Neisseria meningitidis causes meningitis in small children. As we can see from the picture, there are five different types of this kind of bacteria.



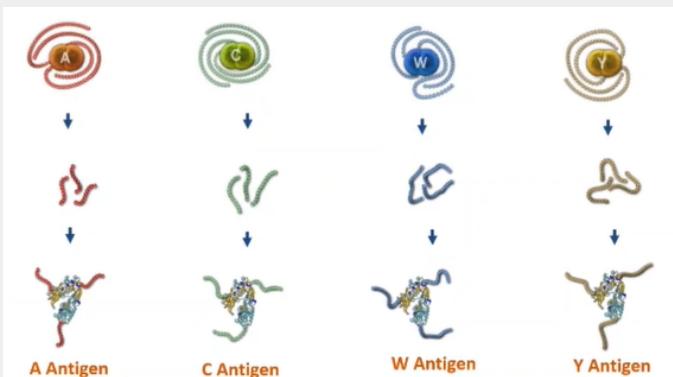
Neisseria meningitidis is surrounded by a capsular polysaccharide (A,B,C,W,Y). Bacteria with this coating don't dry out and are protected from the environment. This is because they are able to retain water on their surface.

→ This polymer of sugars, where each sugar has a distinct composition, keeps the bacteria wet.

Antibodies for this kind of bacteria have complementary binding with respect to the sugar coating.

Everything gets more complicated when vaccines come into the picture.

The most common way to create a vaccine (Pfizer) is to digest the bacteria, only leaving the sugar coating.



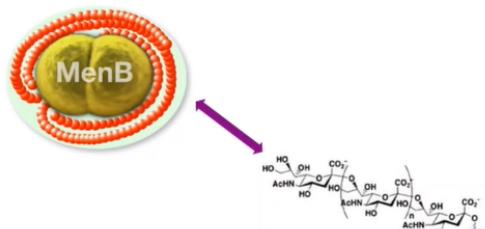
★ for A,C,W,Y

The sugar coating is then attached to a protein, forming an antigen that is injected into the body.

★ for B

Things get more complicated when trying to make a vaccine for meningitis B. This is because its sugar coating is similar to the sugar coatings we have in our organisms (self antigen): the fear is that the vaccine could trigger antibodies against our own cells causing autoimmune problems.

Meningococcus B Capsule Is a Self Antigen and Cannot Be Used for Vaccination

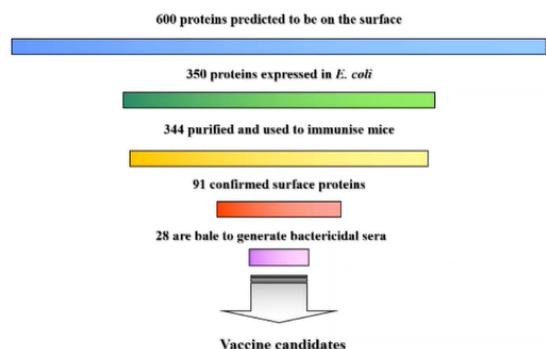


This is where **genomics** comes into place. Scientists sequenced the whole genome of meningitidis, and found what proteins were okay to attack.

However because of the fact that bacteria genomes change, making a vaccine is not as easy as sequencing the genome.

To know more: [Identification of Vaccine Candidates Against Serogroup B Meningococcus by Whole-Genome Sequencing](#)

The idea of reverse vaccinology



FUNDAMENTALS OF MUTATION

❖ INTRODUCTION TO MUTATIONS

Mutations are heritable changes in the genome of a cell/virus/organelle, usually altering either the nucleotide sequence or gene content

DNA repair systems repair most mutations, but some of them escape repair.

Mutations for the individuals are always bad (ie cancer), but for the species some mutations might be good.

Living organisms **evolve** because of **mutations**. If there weren't mutations, we wouldn't have no evolution and no life. However we cannot say that we mutate to evolve, rather that *we evolve because we mutate*.

- Genetic systems must make mistakes to evolve the species. Mutations happen by chance: of course they might be bad for the single organism, but they're still good overall for the species.

DEFINITIONS

- ★ **MUTATION**: heritable change in DNA.
- ★ **MUTANT**: an **organism** (be it a cell or a virus) that carries a mutation compared to the parental or wild type version of that organism/cell/ virus.
- ★ **REVERTANT**: restoration of gene function in a mutant by **reversal** of the original mutation.
- ★ **SUPPRESSOR MUTANT**: restoration of function of a mutated gene due to another **mutation** at a different location

In general, it is not the environment that detects the presence of mutations (although they might increment it), **mutations can happen spontaneously**. This was demonstrated during the **FLUCTUATION TEST BY DELBRUCK AND LURIA**.

The idea that **mutation occurs randomly** was unknown until recent times. It was demonstrated by the Luria-Delbrück fluctuation test.

We know that bacteria can grow in liquid and solid media. If bacteria are spread on semisolid media (+ 1.5% agar), single bacteria (or chains/clusters) **give rise to colonies**. By diluting a bacterial suspension and plating those dilutions on agar plates, we can count the bacteria present in the original suspension

The test involved the analysis of a *culture of E. coli bacteria*, that **can develop a mutation** so that they are resistant to the phage T1 infection.

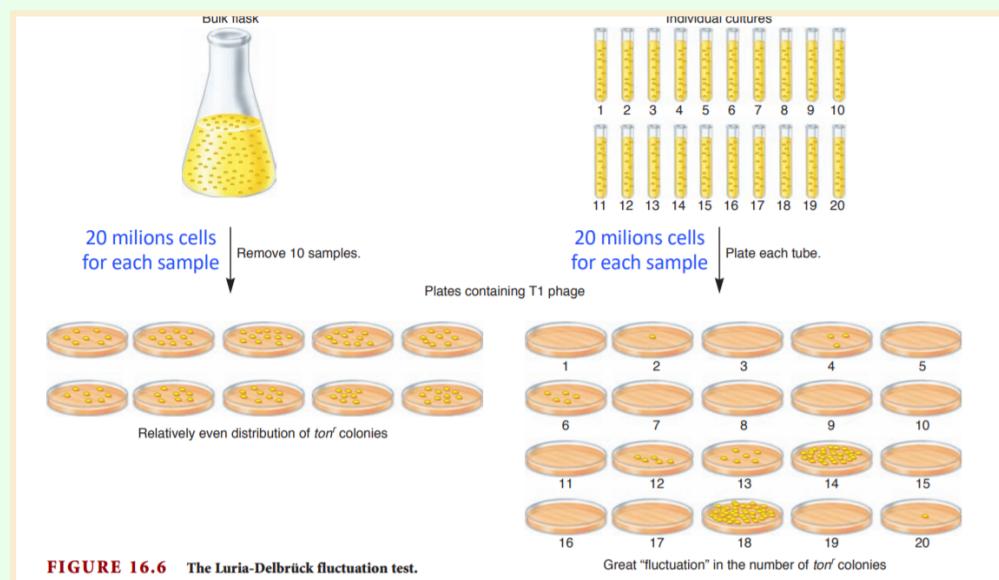


FIGURE 16.6 The Luria-Delbrück fluctuation test.

- **ONE CONTROLLED ENVIRONMENT:**

Scientists made a controlled experiment growing identical bacteria in the same environment and then taking various aliquots (same amount of bacteria in each aliquot) and putting them in different petri dishes with the **T1** phage: since it was a strictly controlled environment the hypothesis was that there would be no mutations, and so all the bacteria would die to T1.

However **some of them survived** and so, it was obvious that they had somehow mutated.

Since the environment was identical for all the identical bacteria, and only some of them survived, the **mutations must have occurred at random**.

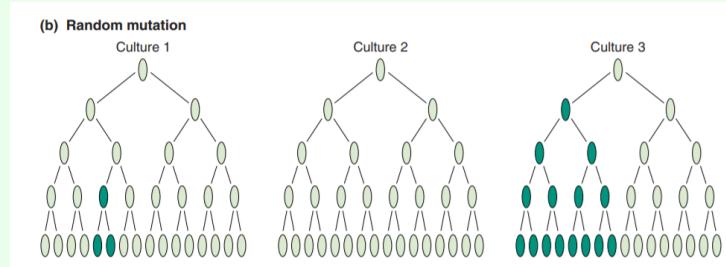
• DIFFERENT CONTROLLED ENVIRONMENTS:

In another experiment, bacteria were grown separately in different flasks and then put into different plates containing T1. In this case there was a **large fluctuation in the number of resistant bacteria**: some plates contained no surviving bacteria, some contained few and some contained a lot. This proved that the mutation took place independently: the bacteria did not **develop a resistance mutation** to the phage only after coming in contact with it, but **earlier**.

→ We say that the resistance mutation was developed before the environmental challenge.

This is because otherwise we would have seen the same number for every plate: there would be no reason for the mutants to have different numbers if they same quantity of bacteria met the same quantity of phage and developed *afterwards* the mutations

The different number of bacteria surviving meant that **mutations could happen at any time**. The sooner the mutation happened the more bacteria survived.



- random **mutations** can take place in different periods of cell division, at **any time or any place**.

❖ **TYPES OF MUTATIONS**

❖ **BASE CHANGES and LARGE CHANGES**

- **DELETIONS**, that are losses, big or small, of DNA
 - deletions have as nomenclature the letter delta (Δ): it means gene or genes has a deletion or is completely deleted.
- **INSERTIONS**, that are gains, big or small, of DNA (big or small)
 - Insertions are indicated with double colon (::) for known transposable elements
 - and are indicated with omega (Ω) for other insertions made in laboratory
- **INVERSIONS**, that are caused by losses of replicons
- **REARRANGEMENTS**, that are swapped sections of DNA.

Single base mutations can also be divided depending on what caused them:

- **DNA POLYMERASE ERRORS**: forces the wrong base into position
- **SPONTANEOUS MISPAIRING OF BASES**: for example bases that go from keto to Enol form or spontaneous deamination of bases (deaminated cytosine = uracil, pairs with adenine)
- **INDUCED DNA DAMAGE** (oxidative stress, UV light, chemicals) will also alter bases (eg. Thymine dimers) and cause mismatches
- **DNA repair enzyme mistakes**

Large mutations can also be divided depending on what caused them:

- **RECOMBINATION ERRORS**
- other types of **SPLICING EVENTS**: like transposons, IS elements, prophage insertions/deletions

❖ **PHENOTYPICAL CHANGES**

All kinds of mutations can also be divided depending on what phenotype they produce:

- **NEUTRAL MUTATION**: the change does not affect the amino acid encoded
- **MISSENSE MUTATION**: different amino acid in protein (might affect function)
- **NONSENSE MUTATION**: amino acid encoding a codon becomes a stop codon (TAA, TGA, TAG)

❖ TYPES OF MUTATIONS

The mutation rate can be determined by using the equation that will not be asked

$$\mu = [(r_2/N_2) - (r_1/N_1)] \times \ln(N_2/N_1) = (f_1 - f_2) \times \ln(N_2/N_1)$$

- r_1 is the observed number of mutants at time point 1,
- r_2 is the observed number of mutants at the next time point
- N_1 and N_2 are the numbers of cells at time points 1 and 2
- f_1 and f_2 are the mutant frequencies at points 1 and 2.

❖ FITNESS COST

Mutations that have the same phenotype don't necessarily have the same genotype.

Furthermore, mutations are usually damaging to a single individual, but some of them are necessary for the entire population to survive. This introduces the concept **FITNESS COST**.

WHAT IS FITNESS COST?

Let's say you have an antibiotic which targets an important biological pathway. Mutations that confer antibiotic resistance often involve modification of the target enzyme to prevent antibiotic binding. These mutations often make enzymes suboptimal compared to evolutionary optimized "wild-type" versions. This can reduce fitness, manifesting as decreased virulence, transmission, and growth rate. However, despite being less fit under normal growth conditions, this mutant can survive under conditions of antibiotic treatment. So this is a **TRADE OFF** also known as fitness cost.

Fitness costs can't be predicted but must be tested.

In *Salmonella enterica* mutations which confer resistance to streptomycin are two:

- *rpsL* Lys42Arg (lysine changes into an arginine) incurs no measurable cost
- *rpsL* Lys42Thr and Lys42Asn (lysine changes into a threonine or asparagine) (lysine changes into an arginine) incur a heavy fitness burden

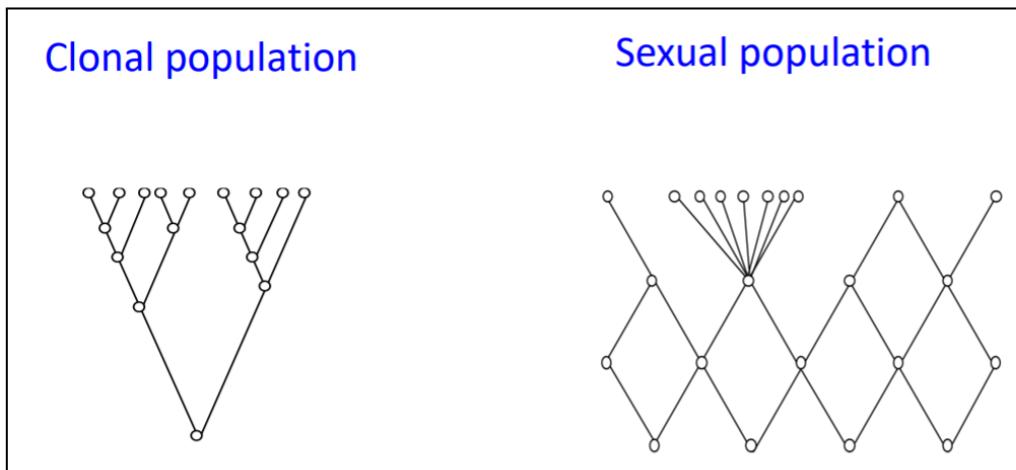
Some **mutations are easier to see than others**. This doesn't depend on whether a gene is easier to mutate (there's no such thing), but from the fact that:

1. **many mutations on the same gene** can induce the same phenotype (ex. streptomycin)
2. **many mutations on different genes** can induce the same phenotype (ex. isoniazid)

drug	gene	mutations per bacterium per cell division
rifampin	<i>rpoB</i>	3.32×10^{-9}
isoniazid	<i>katC, inhA, oxyR, ahpC, furA</i>	2.56×10^{-8}
streptomycin	<i>rrs, rpsL</i>	2.29×10^{-8}
ethambutol	<i>embB</i>	1.0×10^{-7}

❖ MUTATIONS IN EVOLUTION

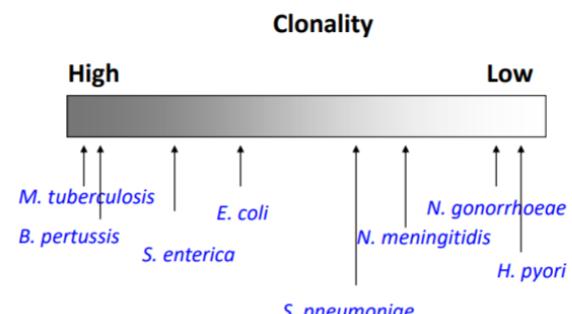
Bacteria can evolve differently, mainly in two ways:



- **CLONAL POPULATION:** a kind of population that **evolves only with mutations**. In this case there is no horizontal gene transfer. The distance between individuals is proportional to the **number of nucleotide changes**. For example I'm near a person with 6 mutated nucleotides with respect to my genome, and far away from a person who has 500 different nucleotides.
- **SEXUAL POPULATION:** a kind of population that evolves with **mutations** and with **horizontal gene transfer** (by transformation / conjugation / transduction). This makes it impossible to correlate evolutionary distance.¹

There are many kinds of bacteria:

- some species are **highly clonal** and only evolve through mutations, with no horizontal gene transfer. Examples of this can be *Mycobacterium tuberculosis*
- some species are **lowly clonal** because they exchange genes.



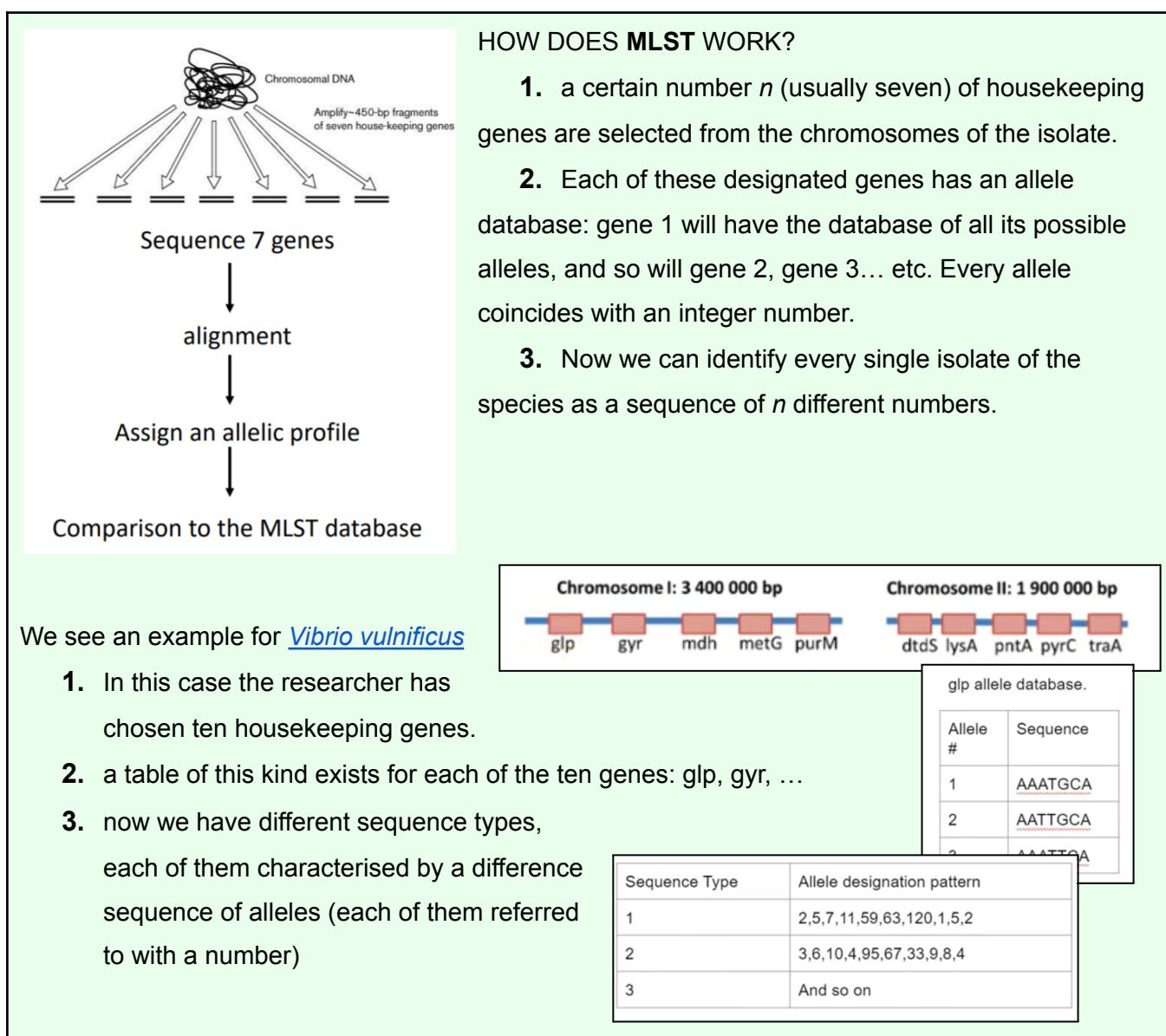
¹ Horizontal gene transfer is the **movement of genetic material between unicellular and/or multicellular organisms** other than by the ("vertical") transmission of DNA from parent to offspring (reproduction)

❖ MUTATIONS IN BACTERIA TYPING

Let's assume that someone has been infected with bacteria A in Italy and in France. Medical staff needs to know if it's the same variant of bacteria A in both places, or if there are different ones at the same time, or if it's a new bacteria.

It's very important to differ between bacteria species, however, **horizontal gene transfer makes it hard to understand how much things are related**. One way in which scientists have tried to overcome the horizontal transfer problem is MLST.

- **MULTILOCUS SEQUENCE TYPING**: is an unambiguous procedure for characterising isolates of bacterial species using the sequences of (usually) seven house-keeping genes.



Different sequences of different isolates can be compared: the more numbers they differ, the further apart they are genetically :

- in the example from earlier, if the variant of bacteria A in Italy and the variant in France have 7 numbers of difference, they are completely different variants. If all their numbers are equal, then its the same variant.

This means it is not important how alleles differ, only if they differ or not. This method avoids the complexity of horizontal gene transfer.

However it is very time consuming, especially because it is considering only seven loci.

Nowadays MLST is being substituted by core genome PCR.

CHROMOSOME PARTITIONING AND SPORULATION

❖ CELL DIVISION

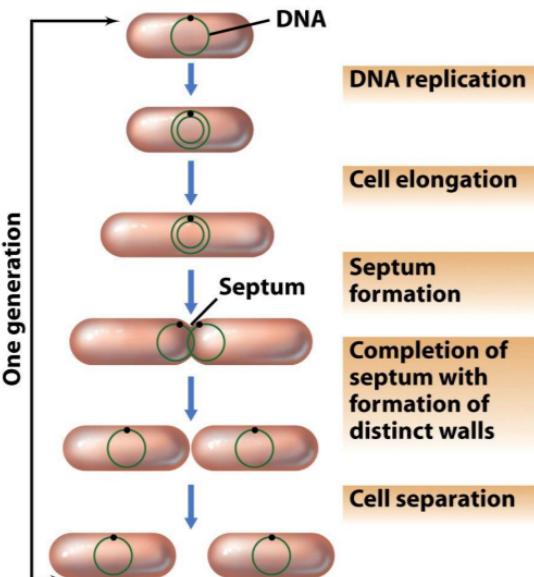
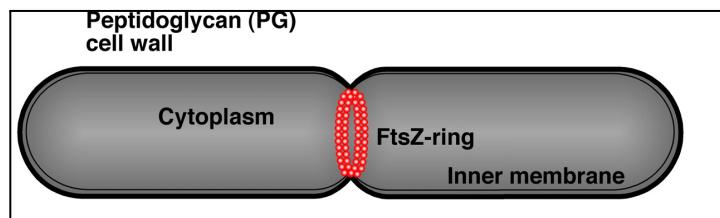


Figure 6-1 Brock Biology of Microorganisms 11/e
© 2006 Pearson Prentice Hall, Inc.

Cell division has been mostly studied in *Bacillus* since they have a longitudinal shape. Most bacteria have a rigid cell structure defined by the cell wall. Duplication of the genome precedes division, but both processes are linked. We see here a summary of cell division's most important steps:

1. DNA replication
2. Cell elongation, with the elongation also of the cytoplasm.
3. Septum formation and positioning: the septum is the new cell wall that forms between two daughter cells as a result of cell division. Building this cell wall is extremely tricky as there's an incredible amount of pressure (about 5atm).

4. Completion of septum with formation of distinct cell walls. The ring that closes down everything is called the **Z ring**, from the name of the protein that forms it.
5. Cell separation.



The most important things the cell must consider when replicating are:

- correct timing
- correct location of the chromosome and of the septum, as it is important not to cut it and lose information
- coordinated assembly and disassembly of the cell wall
- Z ring formation is important in the cell's division

Z RING FORMATION

The earliest stage of cell division in bacteria is the formation of a Z ring, composed of different proteins at the division site. This cellular process is managed by a large number of enzyme complexes which in a coordinated manner carry out processes. The many proteins that form the Z ring are present inside the cell from the beginning, and then assemble when the cell is dividing.

There are many different ways in which the bacterial Z ring can assemble, and it differs depending on the bacteria kind:

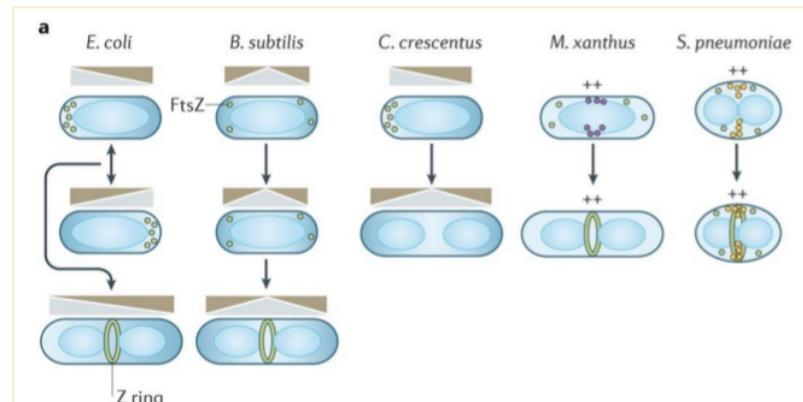
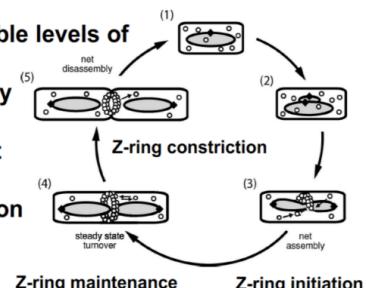
There are two main requirements for the formation of a Z ring:

- ★ reasonable levels of FtsZ
- ★ replication initiation

Requirements of FtsZ ring formation

1. Reasonable levels of ftsZ: necessary but not sufficient

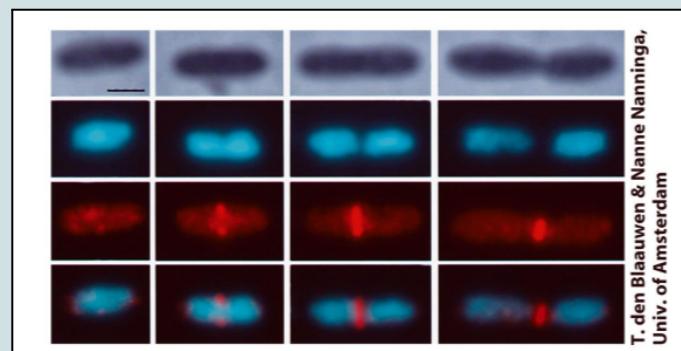
2. Replication initiation



One of the most important proteins for the Z ring is FtsZ, that defines the division plane in prokaryotes. The other proteins that compose the Z ring are mainly kinases and localis thanks to FtsZ. They start the formation of the peptidoglycan.

How do we study these proteins?

Proteins cannot be seen by normal eyes and not even by microscopes. However we can color proteins thanks to fluorescent dye: this is done by inserting a gene for a particular color inside the gene coding for the protein, so that the information for the color and the information for the protein are in the same mRNA. In this picture we can see that an intercalating agent has been used to color the chromosome in blue, and a red dye gene has been inserted to color the protein FtsZ.

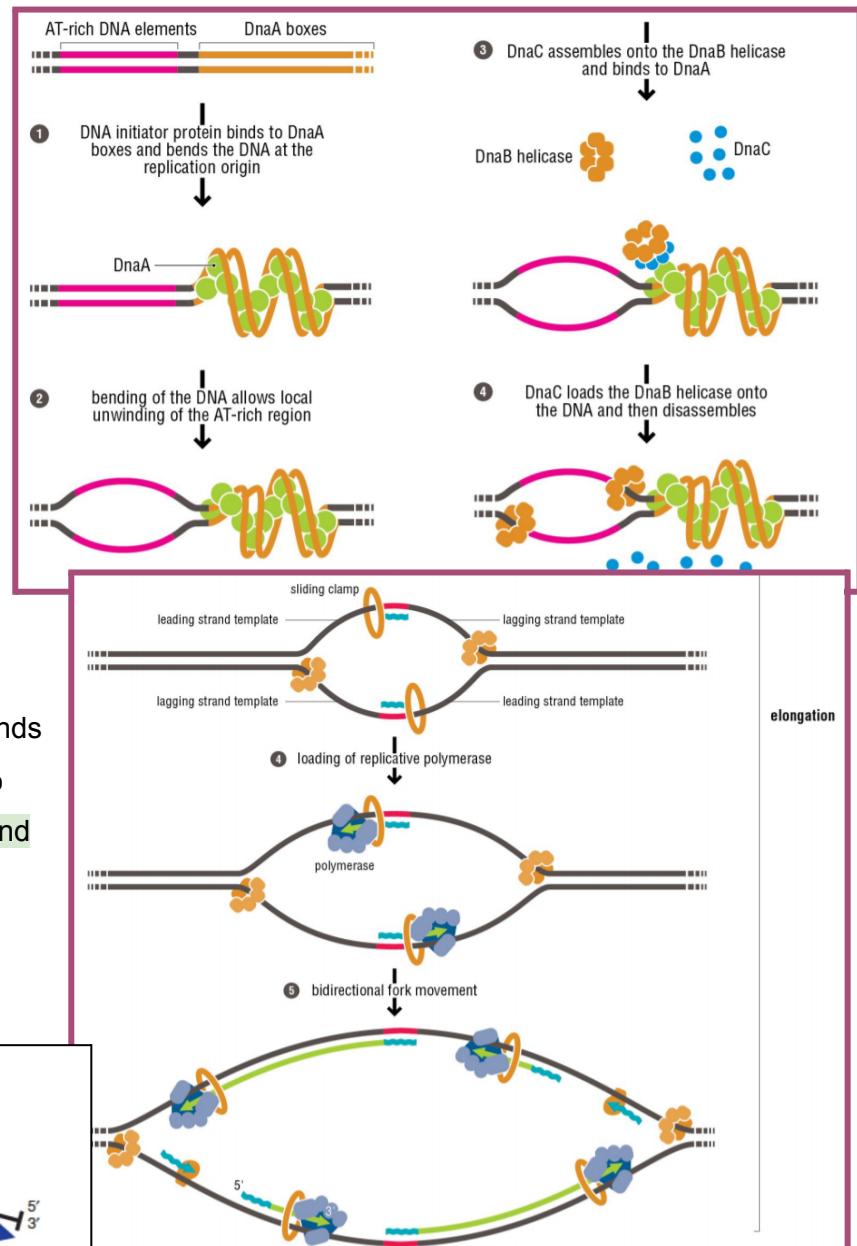
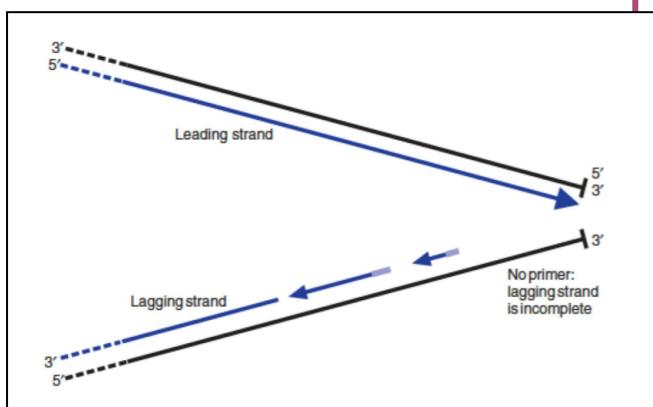


❖ CHROMOSOMAL REPLICATION

Let's now talk about the chromosomal replication, leaving FtsZ (that is connected to the cell wall structure) behind.

The chromosome has an origin of replication, that is a particular DNA sequence that is recognised by a DNA binding protein.

1. **DnaA** binding protein binds to 9 bp repeats at oriC
2. Initiation bubble
3. DNA helicase **DnaB** (with loader DnaC) unwinds DNA
4. DNA Primase generates a primer
5. **DNA Pol III** with **loaders** (clamp loader and slider clamp)
6. **DNA Pol III** replicates both strands at the same time, going into two different directions (lagging strand + leading strand)
7. Replication extends away using two replication forks.



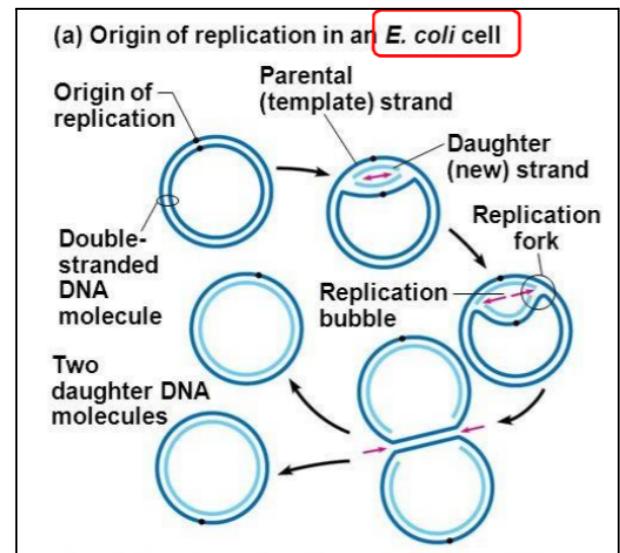
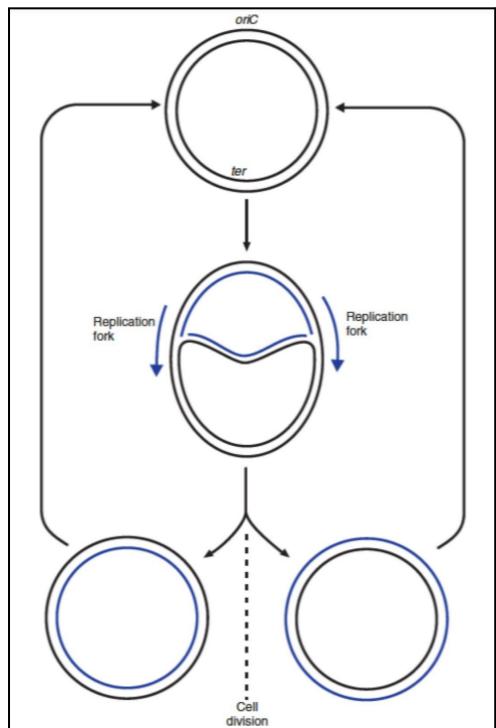
As replication is semiconservative in the two new strands we have:

- parental strand
- newly synthesized strand

As we can see, replication is not only semiconservative but it is also BIDIRECTIONAL.

Furthermore, replication in bacteria is done on a circular chromosome, which means the replication occurs as follows.

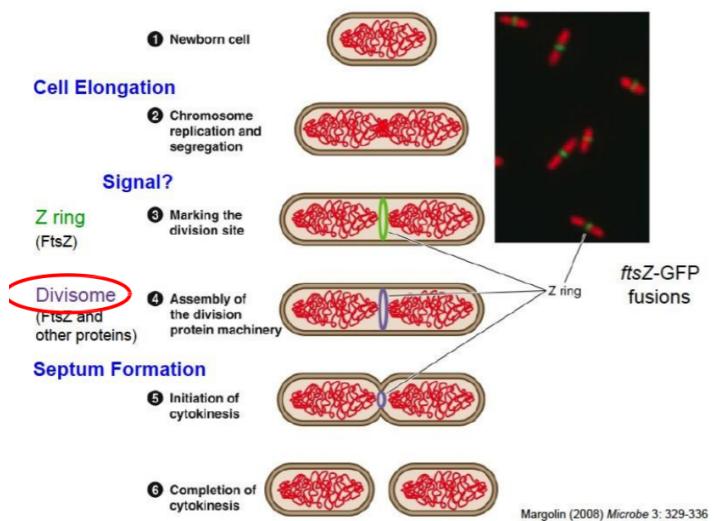
What occurs is defined as a theta (θ) structure, from the shape that DNA takes. Replication starts at *oriC* and continues to the termination site *ter*.



❖ CHROMOSOMAL REPLICATION AND CELL DIVISION MUST BE COORDINATED

The problem of bacteria is that as the chromosome is being duplicated there are many more proteins in the cell, which also acquires more material. The volume of the content of the cell increments and the cell is soon filled up. This means that replication and cell division must be coordinated.

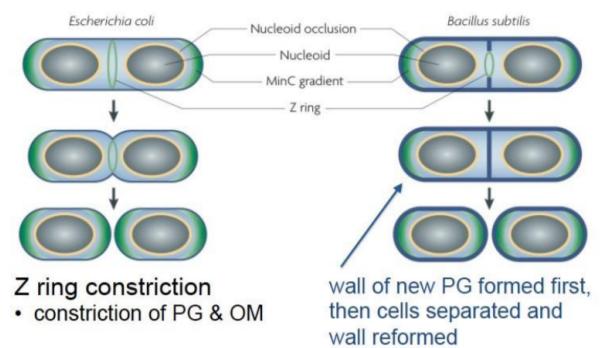
1. first the chromosome is replicated
2. the material is brought at opposite poles of the dividing cell
3. the cell must decide where to put the Z ring. The Z ring must be at the center of the cell, which coincides with the only place where there is no DNA, as it is going towards the sides.
4. the septum must be closed only when we are sure that the genetic material has been completely replicated



There are also differences between the division of various bacteria. Let's now consider the differences between *E. coli* and *B. subtilis*:

- In *B. subtilis* the new peptidoglycan wall forms first in the middle before the cell separates
- In *E. coli* we have first the almost complete cell separation before the peptidoglycan forms.

Differences between *E. coli* & *B. subtilis*



As we have already said, cell division depends on many more molecules than just FtsZ.

These proteins are mostly studied in *gram - E. coli*, and *gram + B. subtilis*.

→ “**FTS PROTEINS** (*filamentous temperature sensitive*). We have for example FtsZ, FtsA etc...

→ **MIN PROTEINS**, that are **negative spatial regulators**.

This means that their presence corresponds to the absence of other proteins, for example FtsZ, because they are occupying the space. They assemble at the cell poles so that the chromosome stays in the middle.

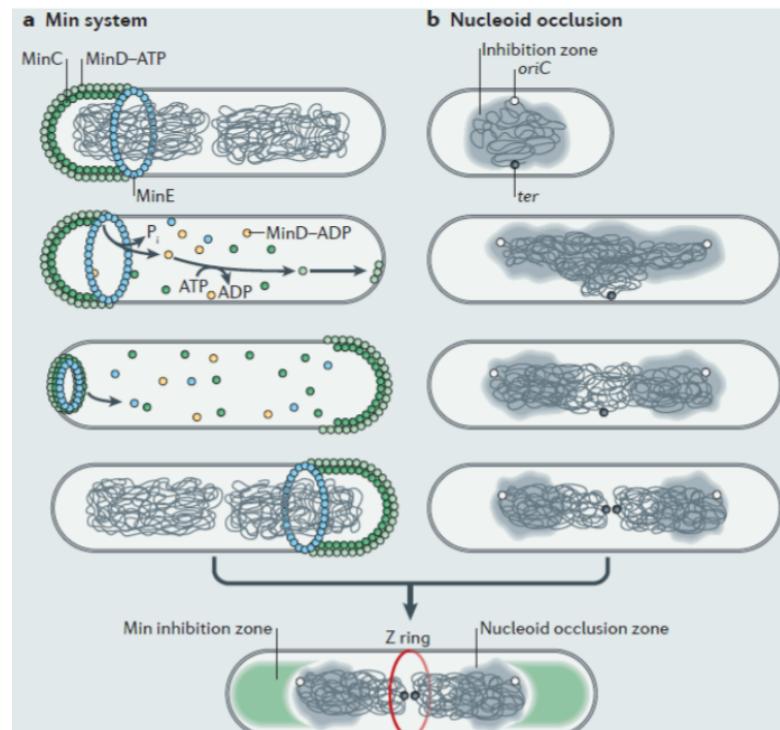
→ **NUCLEOID OCCLUSION PROTEINS**

are also negative spatial regulators.

They prevent the formation of a spetum over the nucleoids, because this would break the chromosomes.

These proteins are thought to be a back-up system if Min is absent. They bind to the chromosome but not near the terminus, which is found in the middle. Also

these proteins inhibit FtsZ polymerisation.



The Z-ring can form only mid-cell after the chromosomes have been separated.

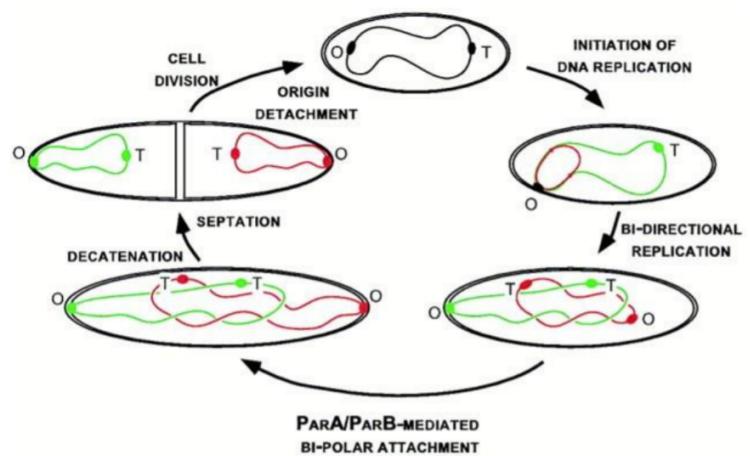
❖ PLASMID REPLICATION

The steps for plasmid partitioning inside the cell are three:

1. Starting from the duplication of the origins, replication of the plasmid begins.
2. plasmid DNA, which resides between the origin and terminus, is replicated, refolded, and partitioned separation of the newly replicated origins

- A pair of genes designated parA and parB are encoded by many low copy number plasmids and bacterial chromosomes. They work to ensure better pre divisional partitioning of the DNA molecule that encodes them. They do this by attaching the origins of the molecule that encodes for them to one of the cell's poles. This means that they also ensure the correct recombination of plasmids.¹
- Alternatively, parM and parR can be used. They create a system which makes the two copies of the plasmid push away from each other, until they are so far away the septum can be formed.

3. Thanks to these genes the concatenation of the two sister chromatids can finally happen.



¹ ParAB system is the most widespread in plasmids and is the only type present also in bacterial **chromosomes** [\[source\]](#)

❖ SPORES

Spores are dormant forms of vegetative bacteria.

- they are HEAT RESISTANT: if you heat bacteria up at 40° for 20 minutes they mostly die, however, most spores resist. This process is called pasteurization, and endospores are not killed by it.
- an example of spores are “anthrax” spores, that are extremely dangerous for lungs. They killed both by accident and as bioweapons
- as we have seen in chapter 2, spores are very important as they are a great way to study sigma factors: when spores are formed every single process in the cell must be interrupted, to do this they simply switch out the sigma factor used for RNA polymerase, so that the latter doesn't recognise any genes anymore, and doesn't bind to them. Different sigma factors recognise different promoter consensus sequences. Instead of the vegetative sigma factors we will then have the sigma factor that activates asymmetric cell division.

WHY ASYMMETRIC CELL DIVISION?

Asymmetric division gives rise to dissimilar progeny that follow different pathways of differentiation.

B. subtilis exhibits two modes of cell division:

- when growing vegetatively, the bacterium divides by forming a septum at the midcell (binary fission), which generates equal-sized progeny.
- upon entry into sporulation, the bacterium divides asymmetrically, switching the division site from the middle to a position close to one pole of the developing cell. In these two asymmetric daughter cells the sigma factors that are expressed are different, which means new genes will be produced.

Asymmetric division gives rise to unequal-sized progeny called the **forespore** (the small compartment) and the mother cell, which initially lie side-by-side in the sporangium. The **forespore ultimately becomes the spore**, whereas the mother cell is discarded by lysis when morphogenesis is complete.

The switch is brought about by a change in the location of the cytokinetic Z ring, which is composed of the tubulin-like protein FtsZ, from the cell middle to the poles during sporulation.

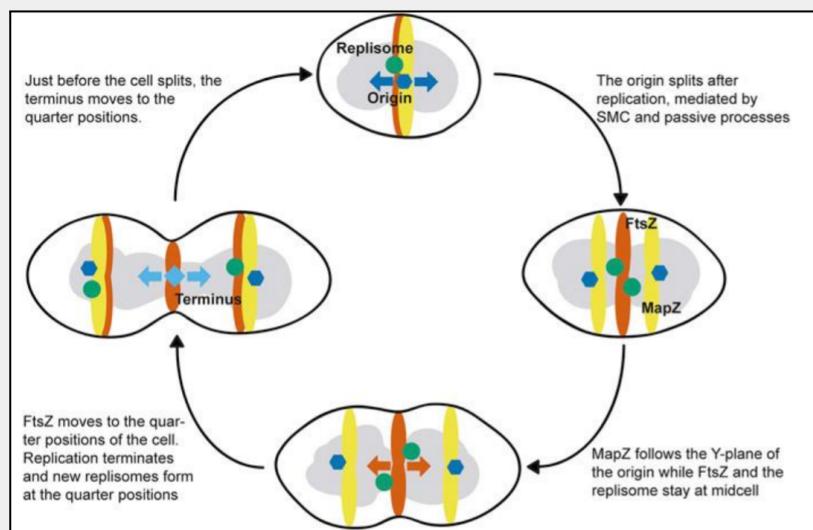
❖ ROUND CELL DIVISION

Everything that we have said until now only works if what is dividing are elongated cells, where the Z ring can be put in the middle.

But how do round cells divide?

Experiments were made on *S. pneumoniae*. This molecule is ovoidal shaped and has no Min and no nucleoid exclusion system. To find out what proteins were involved, scientists decided to tag different proteins, by fusing the genes that encoded for them with genes that encoded for fluorescent proteins.

1. Scientists initially thought that MapZ protein could be linked to the functioning of FtsZ. It is known that the MapZ protein follows the cell poles at constant distance during division and precedes FtsZ at the site of new cell division. Meanwhile, the FtsZ protein stays stably at mid-cell, until new division planes are built in the daughter cells.



However in mutants, it was discovered that the deletion of mapZ does not affect FtsZ localisation. If it is not MapZ which positions the division site, another system must be in place, so, new mutants were made.

Thanks to the superimposition of many images of targeted proteins at different times of replications, scientists found out where many proteins were, and thanks to mutants they realized what the proteins were doing.

Origins and termini were also colored thanks to particular DNA binding proteins that bind those sequences and express a color at the same time.

BACTERIAL CHROMOSOME AND ITS MANNER OF REPLICATION

The first progress in the study of bacterial DNA

replication was made by John Catrns.

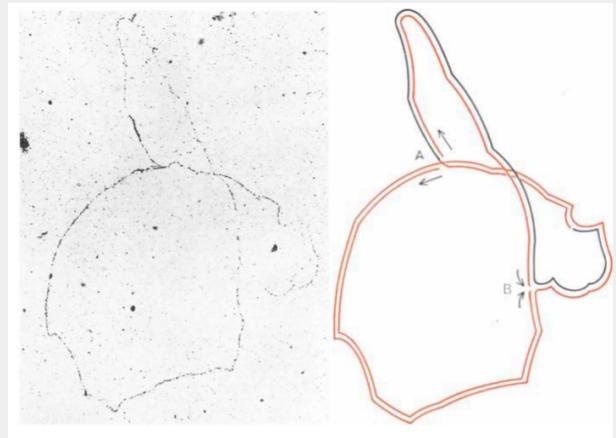
It was known that DNA is a double helix and that

replication was a semiconservative process.

In order to determine the form of replicating DNA, *E. coli* B3 and K12 Hfr were **labelled** for various periods with thymidine. Their **DNA was then extracted** gently and observed by autoradiography.

The results and conclusions can be summarized as follows.

1. The chromosome of *E. coli* consists of a **single piece of two-stranded DNA**, about 1mm long.
2. This DNA duplicates by forming **replication forks**. The new (daughter) limbs of the fork each contain one strand of new material and one strand of old material.
3. **Distal ends of the two daughter molecules appear to be joined during the period of replication.**



RECOMBINATION

❖ INTRODUCTION ON RECOMBINATION

DNA recombination involves the exchange of genetic material:

- either between multiple chromosomes
- between different regions of the same chromosome.

In EUKARYOTES:

- homologous recombination happens during meiosis or in post-replication repair (using sister chromatids);
- site specific recombination like mechanisms during retroviral DNA integration or homing of mobile introns

In PROKARYOTES:

- Homologous recombination during transformation
- Site specific recombination during integration of transposons and insertion sequences and bacteriophages.

❖ CLONING

Cloning genes means isolating them out of their original context and transferring them into a new context

Nowadays a very important field of biotechnology is **RECOMBINANT DNA TECHNOLOGY**, which consists in cutting and pasting DNA molecules to create new genes or constructs that did not exist. To cut DNA, scientists use restriction enzymes.

RESTRICTION ENZYMES, also called restriction endonuclease, a protein produced by bacteria that cleaves DNA at specific sites along the molecule. It usually recognises palindromic sequence.

As originally postulated by Arber, all restriction enzymes serve the purpose of defense against invading viruses. Bacteria protect their DNA by modifying their own recognition sequences, usually by adding methyl (CH_3) molecules to nucleotides in the recognition sequences and then relying on the restriction enzymes' capacity to recognize and **cleave only unmethylated recognition sequences**. Bacterial methylation is done as the DNA is replicated thanks to an enzyme called **methylase**.

There are different types of endonucleases:

- **type I**, which recognize specific DNA sequences but make their cut at seemingly random sites that can be as far as 1,000 base pairs away from the recognition site;
- **type II**, which recognize and cut directly within the recognition site
- **type III**, which recognize specific sequences but make their cut at a different specific location that is usually within about 25 base pairs of the recognition site

Examples of restriction enzymes can be *EcoRI* and *EcoRV*, where:

- *Eco* derives from *E. coli* that is the bacteria in which this enzyme was found
- *R* indicates the strain
- *I* indicates that it was the first (I) endonuclease found in this strain.

EcoRI binds to the palindromic sequence:

5' - GAATTC - 3'

3' - CTTAAG - 5'

and cuts between the G and the A. We get two strands:

5' - G 3'

5' - AATTC - 3'

3' - CTTAA - 5'

3' - G - 5'

As the two Gs are just hanging there, (we call them **SHORT SINGLE STRANDED OVERHANGS**) this type of cut is known as **STICKY ENDS**. They float around, but if they meet again, since they are complementary and have overhangs, they bind again.

EcoRV binds to the palindromic sequence:

5' - GATATC - 3'

3' - CTATAG - 5'

and cuts between the G and the A. We get two strands:

5' - GAT - 3'

3' - CTA - 5'

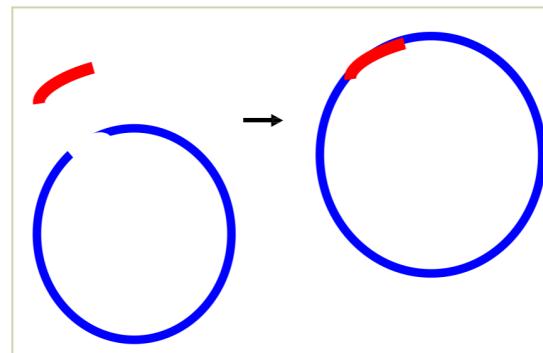
5' - ATC - 3'

3' - TAG - 5'

These are called **BLUNT ENDS**. As there are no overhangs, they do not tend to bind together again.

They are called *restriction enzymes* because they restrict the ability of a virus to infect the host.

In the biotechnological field restriction enzymes are used to cut open DNA. Then during a cloning process external DNA is inserted in the cell, as we can see from the picture on the right. Through recognition of overhangs (complementary ends) one 'cuts' and 'pastes' DNA fragments together. molecular cloning of DNA whereby DNA fragments are joined together to create recombinant DNA molecules, such as when a foreign DNA fragment is inserted into a plasmid.



Now what is left is to bind the strands of DNA together. This is done through DNA ligation.

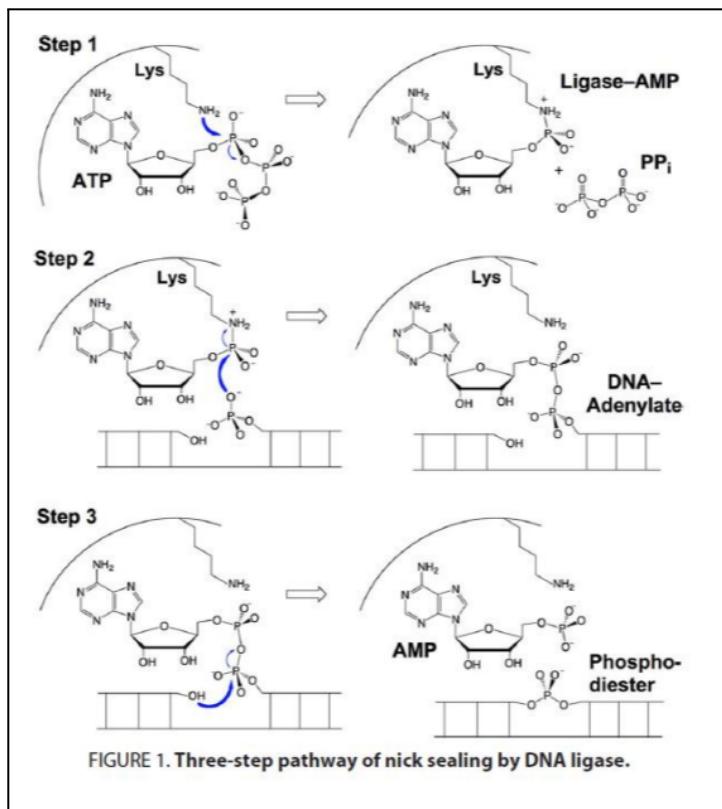
Ligation is the joining of two nucleic acid fragments through the action of an enzyme.

The ends of DNA fragments are joined together by the formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another.

Ligation in the laboratory is normally performed using T4 DNA ligase, however, procedures for ligation without the use of standard DNA ligase are also popular.

Two fragments of DNA may be joined together by DNA ligase which catalyzes the formation of a phosphodiester bond between the 3'-OH at one end of a strand of DNA and the 5'-phosphate group of another. In animals and bacteriophage, ATP is used as the energy source for the ligation.

1. The DNA ligase first reacts with ATP or NAD⁺, forming a ligase-AMP intermediate, with AMP linked to the amino group in the active site of the ligase
2. the AMP group is then transferred to the phosphate group at the 5' end of a DNA chain, forming a DNA-adenylate complex
3. phosphodiester bond between the two DNA ends is formed via the nucleophilic attack of the 3'-hydroxyl at the end of a DNA strand on the activated 5'-phosphoryl group



❖ GIBSON ASSEMBLY

Gibson assembly is a molecular cloning method which allows for the joining of multiple DNA fragments in a single, isothermal reaction.

1. It requires that the DNA fragments contain ~20-40 **base pair overlap with adjacent DNA fragments.**

→ something that can be done if the fragments don't already have overlaps is to reamplify one of the two fragments with a primer that has a tail complementary to the other fragment.

DNA fragments are mixed with a cocktail of three enzymes: exonuclease, DNA polymerase, and DNA ligase.

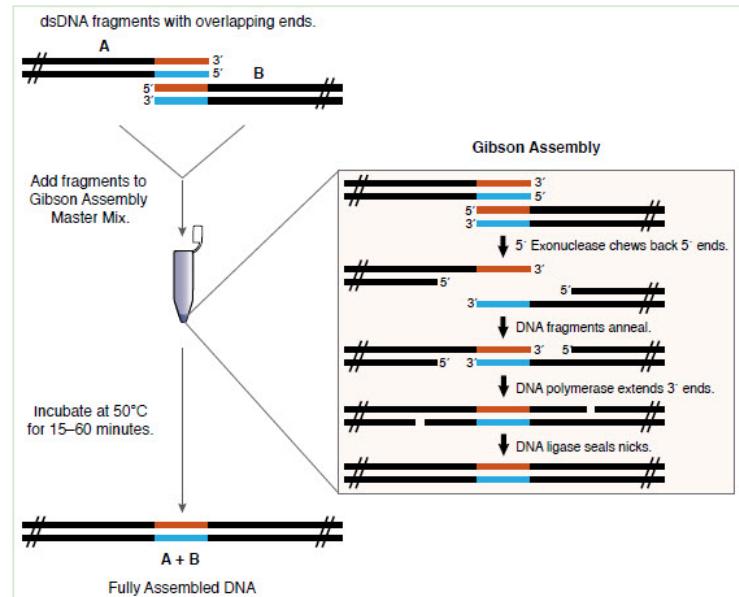
2. The exonuclease chews back DNA

from the 5' end, thus not inhibiting polymerase activity and allowing the reaction to occur in one single process.

3. The resulting single-stranded regions on adjacent DNA fragments can anneal.

4. The DNA polymerase incorporates nucleotides to fill in any gaps.

5. The DNA ligase covalently joins the DNA of adjacent segments, thereby removing any nicks in the DNA.



❖ LEGISLATION ON GMOS

We have talked about **DNA RECOMBINATION**, that gives rise to ***genetically modified organisms*** (GMO). To this particular field of science, certain laws are applied depending on the country.

Legislation of the contained use of genetically engineered microorganisms (Legislative decree 12 april 2001 n. 206.)

- The **executive of the company** (owner, CEO, rector) of the facility needs to notify the ministry about the facilities.
- The **user** (scientific responsible) is responsible to identify the risk and prepare the application for contained use (there is no authorization to use the GMO outside of the facility).

The biosafety classes are 1 to 4 (art. 75 of the legislative decree 626/94 (and successive modifications):

- class 1: contained use poses no risk to health (for example *E. coli* is class 1)
- class 2, class 3, class 4: contained use has increasing risk to health

❖ DEFINITIONS

- ★ **HORIZONTAL GENE TRANSFER:** movement of genes within a chromosome, between chromosomes and **between organisms** that are not one the offspring of the other.
- ★ **RECOMBINATION:** If two pieces of DNA sequence are identical or very similar, the two regions can then hybridize (base pair) with each other. Outcome can be the **swapping of similar DNA** between the two regions.

❖ RECOMBINATION

Genetic recombination (also known as genetic reshuffling) is the exchange of genetic material between different organisms which leads to production of offspring with combinations of traits that differ from those found in either parent

In general we can have **different kinds** of recombinations:

→ **HOMOLOGOUS RECOMBINATION**

→ **SITE SPECIFIC RECOMBINATION**: integrases from phages and elsewhere

The key difference between homologous recombination and site-specific recombination is that in homologous recombination, genetic material is exchanged between two identical molecules of double-stranded or single-stranded nucleic acids such as DNA or RNA, whereas in site-specific recombination, DNA strand exchange takes place between DNA segments that possess at least a certain degree of sequence homology but no extensive homology.

→ **TRANSPOSITION**: jumping genes, transposons, Retro-elements, IS elements, Phages

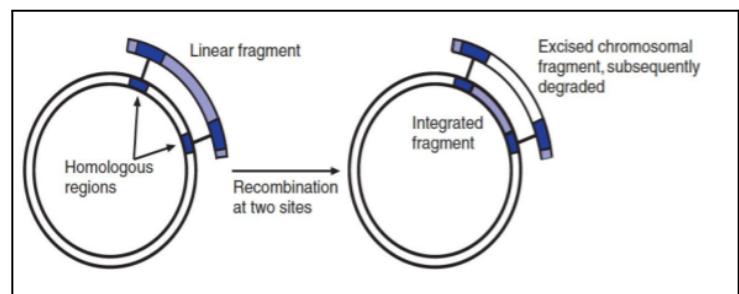
→ **IN VITRO RECOMBINATION**: recombinant DNA, gene cloning

Since we are in bacteria we have two different ways in which recombination can happen:

→ **RECOMBINATION BETWEEN A LINEAR DNA MOLECULE AND A CIRCULAR MOLECULE.**

The reaction requires at least two regions of homology.

Recombination at two sites leads to replacement of a portion of the circular molecule.

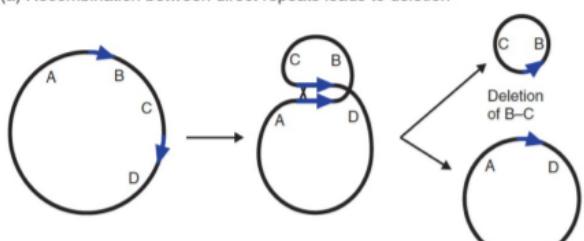


→ **RECOMBINATION ON THE SAME CHROMOSOME'S REPEATS**

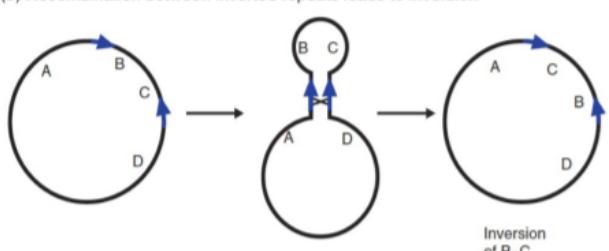
1. Recombination between **direct repeats** leads to deletions.

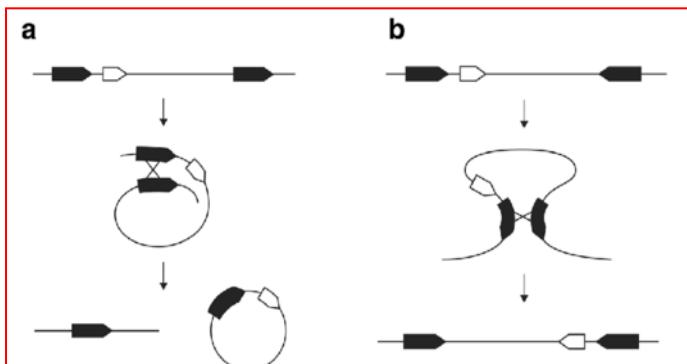
2. Recombination between **inverted repeats** leads to inversion. No information is deleted, but it is inverted.

(a) Recombination between direct repeats leads to deletion



(b) Recombination between inverted repeats leads to inversion





a) Direct repeat

5' ATG**CAGTG**CGCCATAACCATCA**CAGTG**TCGA 3'
3' TACG TCACGCGGTATTGGTAGTGT CACAGCT 5'

b) Inverted repeat

5' ATG**CAGTG**CGCCATAACCATCA**ACTG**TCGA 3'
3' TACG TCACGCGGTATTGGTAGT**GTGAC**AGCT 5'

WHY DOES RECOMBINATION TAKE PLACE?

- **DNA REPAIR (HOMOLOGOUS RECOMBINATION)** for double strand breaks, that can happen both during replication or because of external agents, like UV light.
- **GENERATION OF DIVERSITY**, acquisition of new genes: it is much easier for a bacteria to acquire external DNA through recombination rather than just wait for a mutation to happen.
- Preprogrammed part of the **LIFE STRATEGY** of selfish DNA elements like phage genomes, Plasmids, IS elements, etc.

Apart from this last point, recombination usually takes place in the core genome and is rarer in the accessory one. This is because of homology.

❖ HOMOLOGOUS REPAIR

For example if there is a nick in a DNA strand, when it is replicated it transforms into a **double strand break**. Such a situation causes what is known as a **collapsed replication fork** and is fixed by several pathways of [homologous recombination](#).

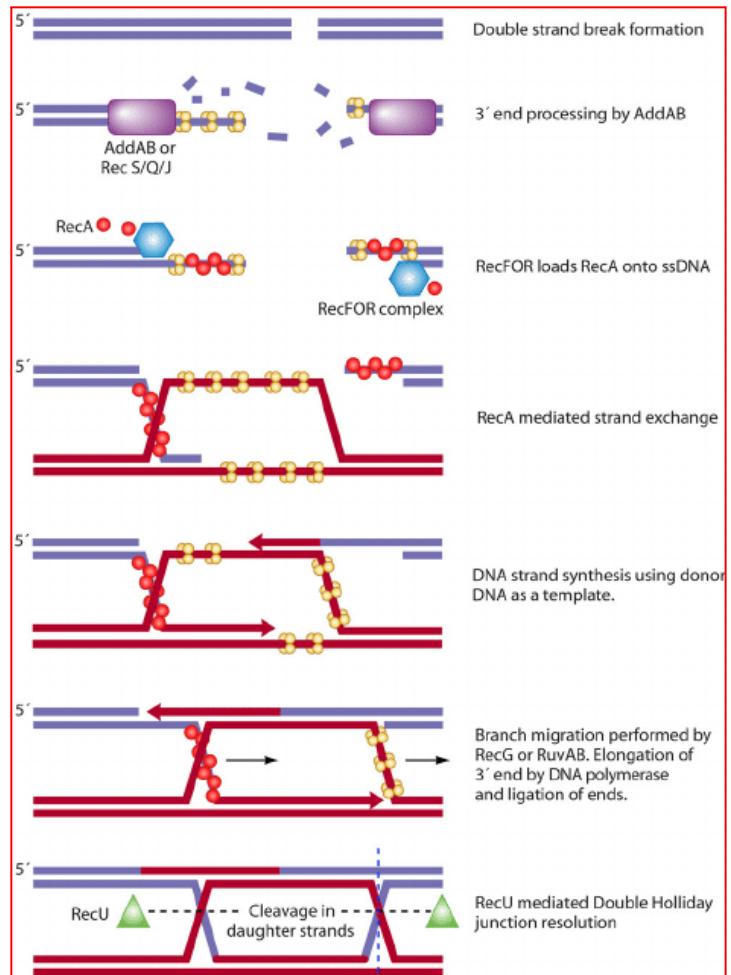
Generally, after a double strand break, in both **strands some nucleotides are removed** in order to create single stranded stretches. **RecA** proteins are then loaded onto the single-stranded DNA with the newly generated 3' end. The resulting RecA-coated nucleoprotein filament **searches out similar sequences of DNA on a homologous chromosome**. Upon finding such a sequence, the single-stranded nucleoprotein filament moves into the **homologous recipient DNA duplex** in a process called **strand invasion**.

In the strand invasion step that follows, an overhanging 3' end of the broken DNA molecule then "invades" a similar or identical DNA molecule that is not broken.

The invading 3' overhang causes one of the strands of the recipient DNA duplex to be displaced, to form a **D-loop**.

The resulting structure will be a **holliday junction** that can be resolved and the gaps will be filled thanks to ligasis.

This type of resolution produces only one type of recombinant (non-reciprocal).



HOMOLOGY IN PROKARYOTES.

Most prokaryotes are haploids, hence their chromosomes do not have homologs. This process however requires some sort of homology and so it can occur only if it occurs when bacteria are temporarily on a diploid condition:

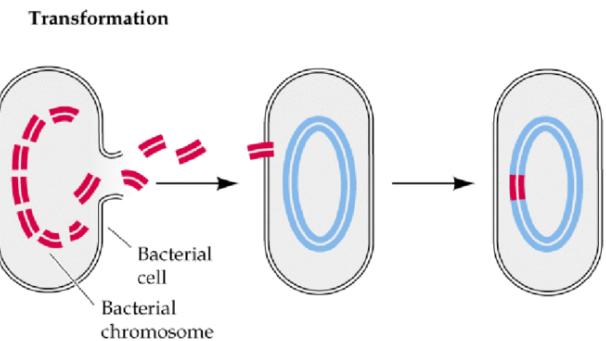
- just after replication
- homologous strands may also come from conjugation with other bacteria or through the process of transformation

Recombination requires homology of >20 bp (the more the better); and recombination events can be reversible (as long as the original molecules stay around)

❖ INSERTION OF LINEAR dsDNA INTO THE CHROMOSOME

Let's now try to put into context the recombination we have just talked about. Like we said it can happen during **transformation**, in three steps:

1. dsDNA upon cell entry during transformation is transformed into a single strand.
2. ssDNA is bound by single strand binding proteins and then RecA
3. RecA allows for homologous recombination.



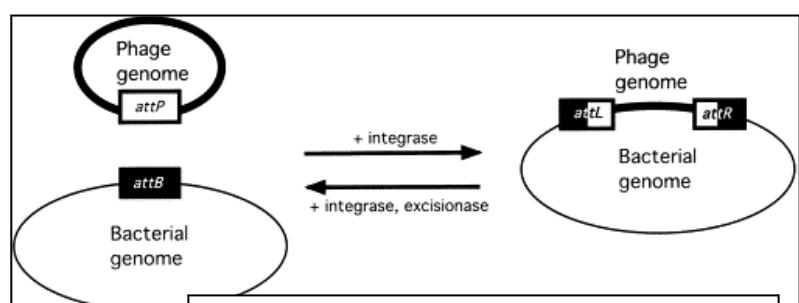
Permission pending from Sinaur Associates, Inc.

❖ INTEGRASES

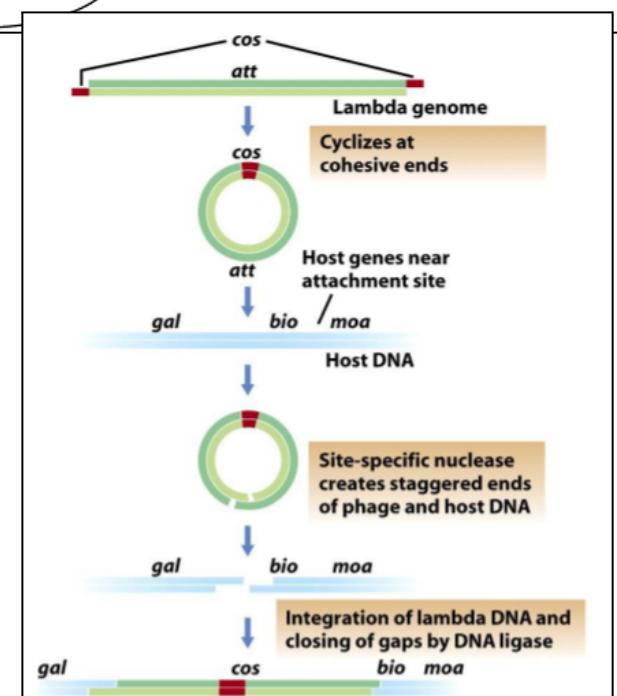
Phage integrases are enzymes that mediate unidirectional site-specific recombination between two DNA recognition sequences:

- the phage attachment site, **attP**,
- the bacterial attachment site, **attB**.

- These enzymes recognize specific DNA sequences ("att sites") and mediate recombination between these short sequences. These sequences are usually 15bp long.
- Genes encoding integrases are found in mobile elements: for example phages, but also chromosomal elements or plasmids



In this example we can see a lambda bacteriophage, our transposable element. It has two cohesive ends so it can close up in a circle or a concatemer. It then integrates into the bacteria (host DNA)



Streptococcus pneumoniae causes pneumonia

The polysaccharide capsule of *Streptococcus pneumoniae* is the dominant surface structure of the organism and plays a critical role in virulence; there are currently more than 90 different serotypes of capsules, each encoded by different genes. Our antibodies recognise the different sugar coatings. This gives rise to a new definition:

A serotype or serovar is a distinct variation within a species of bacteria or virus or among immune cells of different individuals.

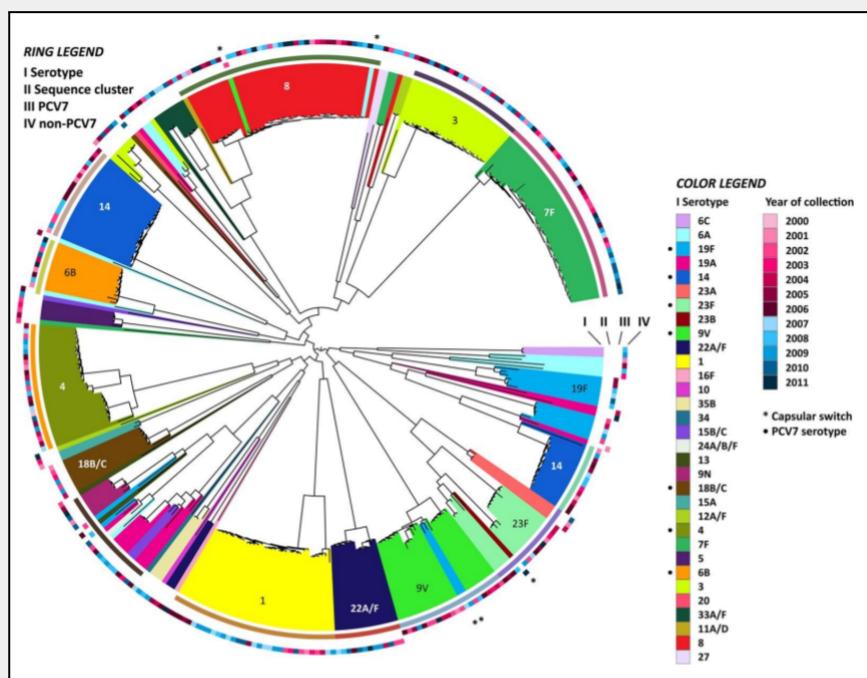
- **STRAIN** is a low level taxonomic rank, standing below species. A strain of a bacterial species represents interspecific diversity (diversity within the species). When colonies of a particular bacterial species are isolated on a petri plate, various **colonies** seen on the petri dish represent the different strains of that particular species of bacteria.
- A **SEROTYPE** is a sub-group of species, which are grouped according to their antigenic properties. Antigens are substances that are considered “foreign” to the host body. The host combats these antigens by eliciting an immune response, which involves the production of **antibodies**. These antibodies are found in serum ('sero-'), which is a part of our circulatory system.

Being affected by one serotype of pneumococcus (say, the blue one) doesn't mean you can't be affected by another serotype. It however usually means that you can't get infected by the same serotype again. This is because the **same serotype makes the body produce the same antigens**. So next time we already have antigens.

Furthermore, different pneumococcal serotypes are known to cause different clinical presentations of pneumococcal diseases.

The **capsule** is the target of current pneumococcal vaccines: antibodies to the capsule protect from infection.

Streptococcus pneumoniae vaccine was introduced in 2006: we can see in our phylogenetic tree when the serotypes of this bacteria were found: in blue we see when they were found after the vaccine, in red when they were found before the vaccine. Of course the vaccine didn't contain the serum for all 90 strains only for strains 4, 6B, 9V, 14, 18C, 19F, and 23F.

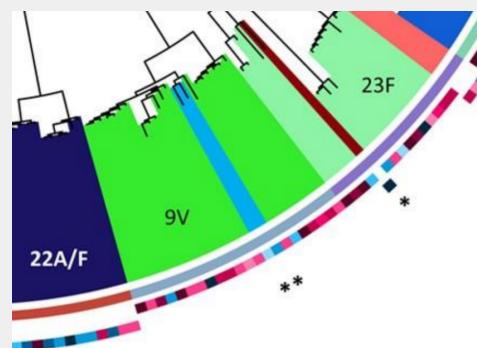


The inner bars each represent a pneumococcal strain (I) and are coloured by capsular serotype. The second circle displays the sequence clusters (II). The year of collection is marked by the colour (red: pre-PCV7, blue: post-PCV7) and intensity of coloration per year, for PCV7 vaccine serotypes (III) and non-vaccine serotypes (IV).

QUESTIONS

- Can you detect a **decrease in prevalence of (some of) the serotypes present in the vaccine?**

Yes, for example we can see that the presence of serotype 9V decreased after the vaccine: the blue is much less than the red. This however doesn't happen for the 4 serotypes. So it looks like that the vaccine worked better in some cases rather than others.



- Can you detect an increase in **prevalence of (some of) the serotypes not-present in the vaccine?**

Yes, for example in 1. This is probably because if some serotypes go, others become more spread.

- Some isolates are marked by an **asterisk on the outermost circle** – what happened there?

Change by chance of a gene of the capsid (mutation, which leads to the formation of a new type of capsid and therefore a new serotype. Of course the bacteria that mutate are only a few, but the vaccine kills all the rest, so the only ones that survive are the mutated ones (for example in 9V)

Helicobacter pylori is a gram-negative, microaerophilic, spiral (helical) bacterium usually found in the stomach.

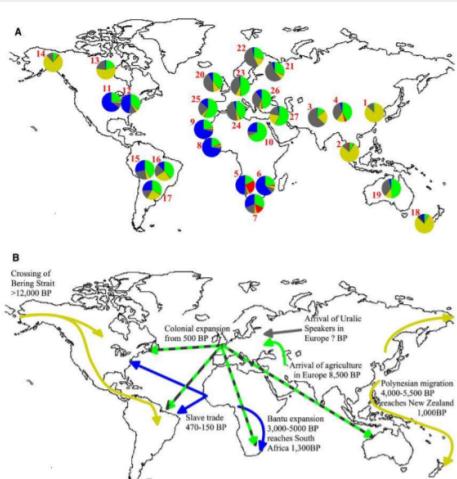
Helicobacter pylori, a chronic gastric pathogen of human beings, **can be divided into seven populations and subpopulations with distinct geographical distributions**. These modern populations derive their gene pools from ancestral populations that arose in Africa, Central Asia, and East Asia. **Subsequent spread can be attributed to human migratory fluxes** such as the prehistoric colonization of Polynesia and the Americas, the neolithic introduction of farming to Europe, the Bantu expansion within Africa, and the slave trade.

- individuals whose parents come from the same region have few mutations in comparison to the region's most common bacterial population. This corresponds to a small phylogenetic tree.
 - individuals whose parents come from different regions have few mutations in comparison to the region's most common bacterial population. This corresponds to a large phylogenetic tree.
- the more the populations have similar genes in *Helicobacter pylori*, the more those populations are near evolutionarily speaking.

Putative modern and ancient migrations of *H. pylori*.

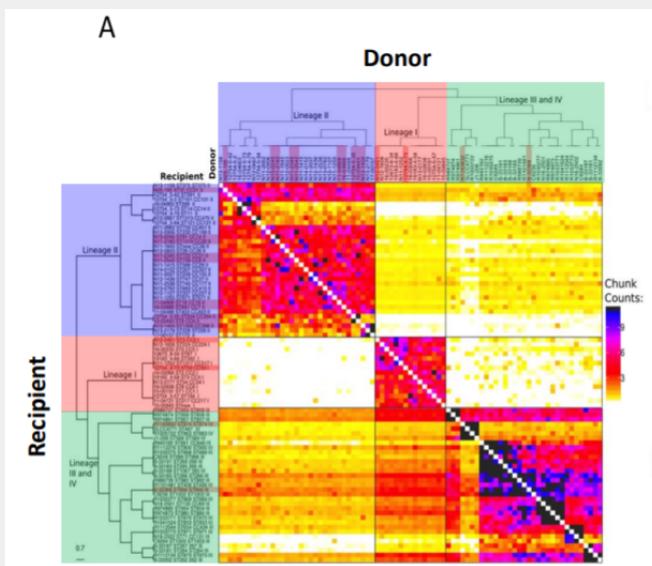
(A) Average proportion of ancestral nucleotides by source.

(B) Interpretation. Arrows indicate specific migrations of humans and *H. pylori* populations.



In *Listeria monocytogenes* the **DIRECTION OF DNA TRANSFER WAS ANALYSED**. This means that scientists tried to figure out if there are some lineages of *listeria monocytogenes* that donate or receive DNA more than others.

- To study the directionality of DNA transfer we used Chromopainter: three lineages are taken into consideration, lineage I (red), lineage II (blue), lineage III (green). The darker the correspondence table is, the more recombination there was between donor and receiver.
- Each lineage has the most chunks originating from the same lineage. This is because homologous recombination is the easiest and most common.
- Lineage III appears to be the best donor



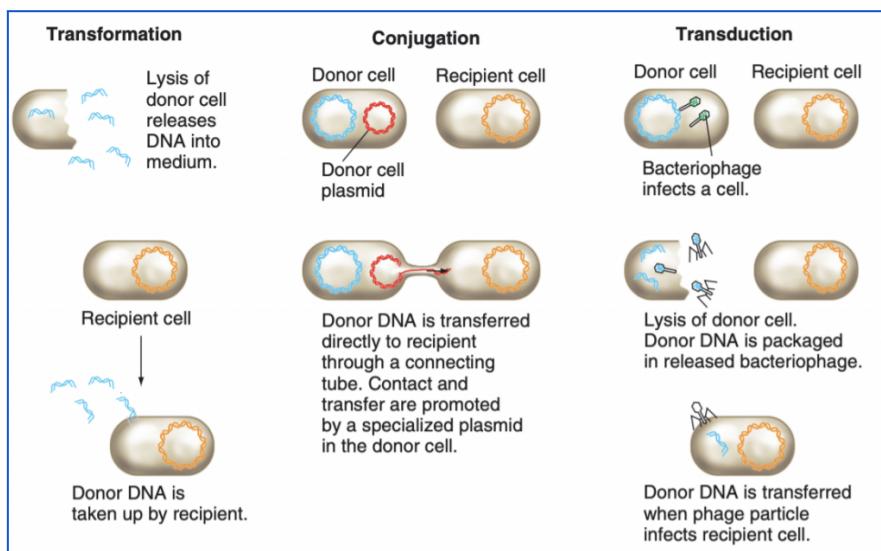
This means that some are better donors than others, but generally every lineage can recombine itself very well.

TRANSFORMATION AND TRANSDUCTION

★ HORIZONTAL GENE TRANSFER

horizontal gene transfer is the transfer of genetic material from another organism without being the offspring of that organism. **Cells are not related.**

Transfer of genetic information in bacteria can happen in three different ways:



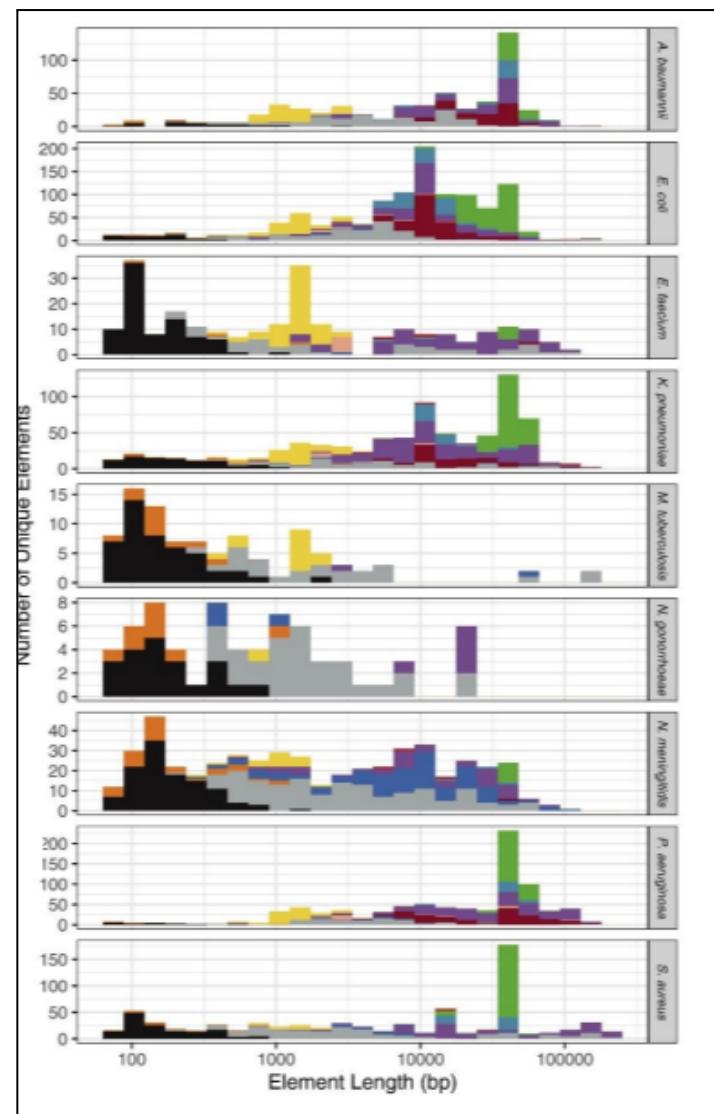
- **TRANSFORMATION:** is the **transfer of free DNA**, which includes any part of the chromosome including mobile elements. It happens because of the **lysis** of the donor cell, which releases all its DNA that is then taken up by the recipient cell.
- **CONJUGATION:** is a kind of **contact mediated transfer** thanks to a pilus
- **TRANSDUCTION:** is a kind of **phage mediated transfer**

Bacteria are very diverse as we have already said. This means that not all isolates have the same mobile elements.

What's on the right is an analysis of 1.200 -1.500 genomes of 9 species for integrated genetic elements, excluding plasmids.



For example in *S. aureus* the most common mobile element is **bacteriophages**. In *N. gonorrhoeae* there are no big elements such as phages, but a lot of small repeat elements (like the black one)



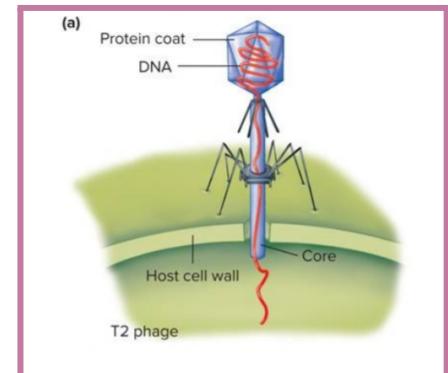
★ BACTERIOPHAGES

Viruses that infect bacteria are called **BACTERIOPHAGES** (phages).

- may be most abundant life-forms on Earth and have a significant impact on biosphere
- they are very important in the history of molecular biology, as they allow transduction.
- first genetic tool to map chromosomes and make clones.
- while with bacteria we can make a genetic tree established on 16rRNA, that is common in all core genomes; in viruses we can't do that because viruses are not cells. We cannot make a tree on viruses.

STRUCTURE OF A PHAGE:

- their genome can be composed of ssDNA, dsDNA, ssRNA
- and its contained inside a protein coat (CAPSID)



Phages can be:

- **VIRULENT phages**: infection results in death of cell (**LYTIC**)
- **TEMPERATE phages**: establish stable relationship with host cell (**LYSOGENY**)

Despite extensive gene exchanges, which generate diversity, and the absence of homology at the nucleotide and amino acid levels for most phage pairs, **some common virion structures exist in nature**. This raises the question of whether these structural similarities can be explained by divergent or convergent evolution.

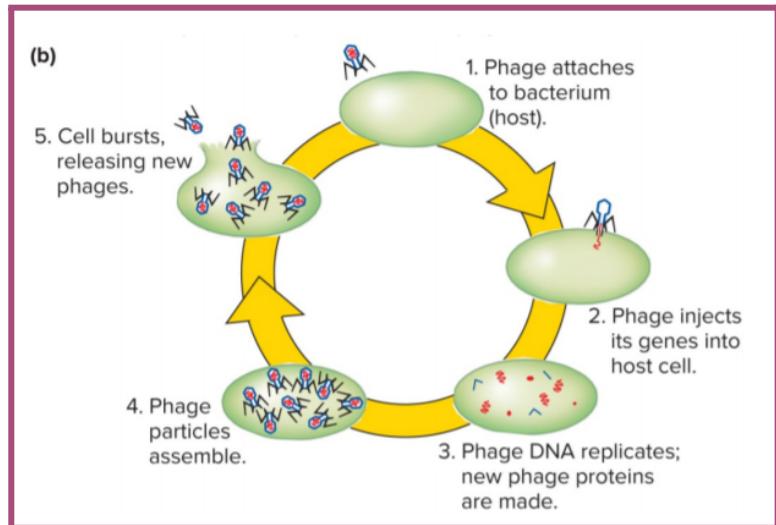
However as we have already said phylogenesis of viruses is impossible and **there cannot be a common ancestor**. This of course differs in cellular organisms.

- Phages have **very different sizes**, having a genome size of 15-30 kb.
- *Myoviridae* tend to be larger and have genome sizes of 80.000-150.000 kb.
- Some *jumbo* phages of genome sizes of >200.000 kb

We can conclude that phage phylogeny is traditionally illustrated with trees, but this representation fails to account for horizontal gene transfer events. A new proposed method uses networks to show the complexity of their evolutionary relationships and how phages are interconnected.

LYTIC CYCLE OF PHAGES

- a. Phages inject their genetic information into the bacteria.
- b. This viral DNA uses the molecular machinery of the bacteria to replicate its own DNA. It blocks all cellular gene transcriptions and replications, only allowing its own proteins and DNA to be active.
- c. Phages use the bacteria molecule to translate the viral DNA, encoding for the phage proteins by digesting the host DNA. Thanks to these proteins the capsid is formed.
- d. The new viral DNA will be identical to the initial phage. When lots of viral DNA and capsids have been produced, the cell bursts out and viruses spread.
- e. The virus goes on bursting more cells; one lytic cycle takes more or less 20 minutes.



There is a balance between phages and bacteria, which can develop immunity towards phages to survive them.

★ TRANSDUCTION

Transduction, which is the phage mediated transfer, can be divided into two types:

→ SPECIALISED TRANSDUCTION:

when during excision from the chromosome a phage erroneously excises some **neighbouring chromosomal DNA**, that is then incorporated into the capsid.

Through experiments we can foresee which chromosomal genes will be replicated, as:

- ◆ in a genome genes are in a particular order;
- ◆ phages erroneously excise always the same parts.

→ GENERALISED TRANSDUCTION: when during packaging **random genome fragments** are incorporated into a capsid

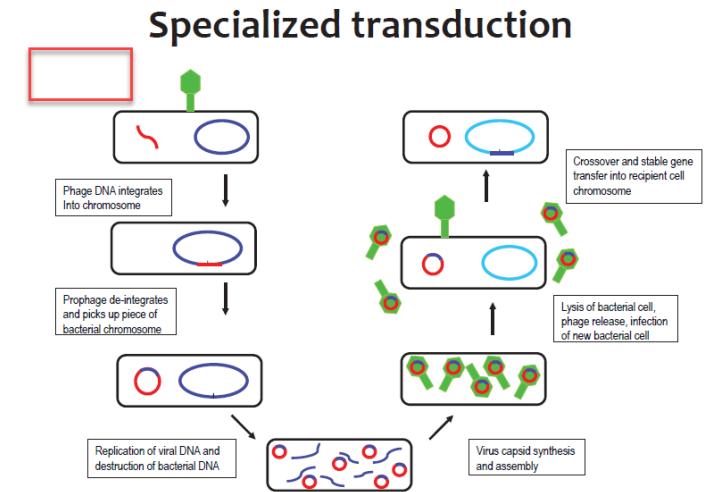
★ PHAGE THERAPY

Sometimes bacteria cause illnesses in the human body. To eliminate bacteria we usually use antibiotics, but another possibility is to **use phages that are specific for the bacteria causing that particular disease**.

Phages have been used (mainly in Russia) as therapeutics for treatment of human infections. However this is all just experimental as there are:

- problems with the fact that the phages should be highly specific not to damage beneficial bacteria
- some issues with resistance

Some phages are already used in the industry, for example to combat listeria infections.

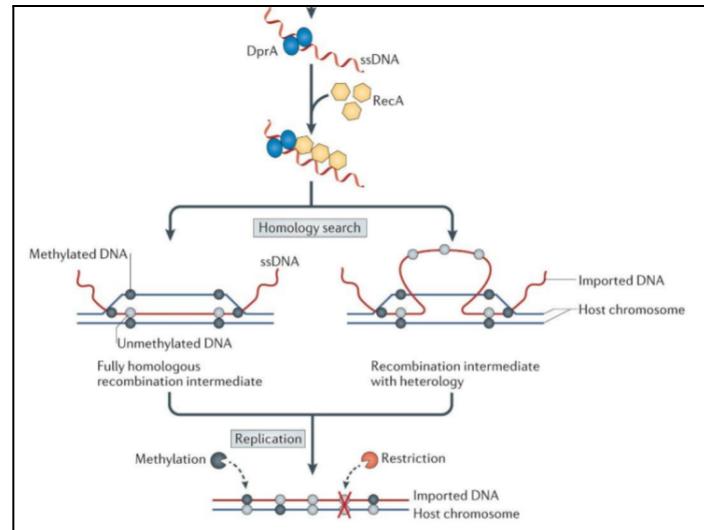
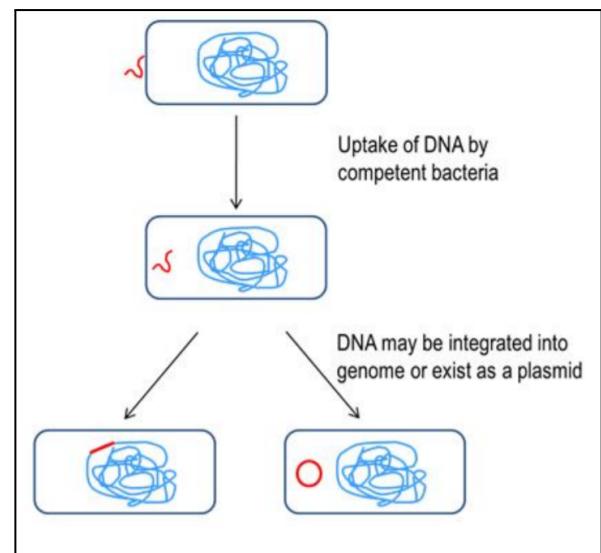


Courtesy of M. Mulks (MSU)

★ TRANSFORMATION

Transformation is the **entry of foreign DNA** into a **recipient cell** and is characterized by the **RECOMBINATION** between the foreign DNA and its **homologous region in the recipient chromosome**. In particular, at the beginning of transformation the donor cell breaks, releasing its DNA into the medium. Afterwards, the DNA is taken up by the recipient cell.

1. The DNA-uptake machinery comprises a transformation pilus, which consists mainly of subunits and **captures exogenous double stranded DNA**.
2. A nuclease receives DNA and degrades one DNA strand to generate single-stranded DNA (**ssDNA**).
3. Internalized ssDNA is presumably bound by DprA (DNA processing protein A), which recruits the recombinase **RecA**.
4. RecA polymerizes on ssDNA and **promotes a homology search along chromosomal DNA**, followed by strand exchange. If it finds homology then it will integrate into the bacterial DNA. However we can also bring completely different genes using **flanking homology**.
5. The methylation and restriction activities of the restriction-modification (R-M) can kill transformants.



Even if the DNA is taken up by every bacteria cell, this doesn't mean that the DNA will be integrated into the recipient cell: only "**COMPETENT**" bacterial cells may integrate **foreign DNA**. If the donor DNA is not integrated into the chromosomal genome, it will exist as a plasmid.

Bacteria become competent only in specific moments.

HOW DOES THIS WORK?

The **comCDE** operon encodes a peptide **CSP**, that is exported by a dedicated export protein **ComAB**. The more bacteria are present, the higher the CSP concentration in the medium (quorum sensing).

At a certain concentration transcription of both the **comAB** and **comCDE** genes is enhanced, so the quantity of CSP becomes even more (this is a positive feedback loop).

- When this critical concentration is reached, CSP interacts with the ComD receptor protein of the neighboring cells and activates its cognate response regulator, ComE, through autophosphorylation. **ComE** activates also transcription of **comX**, that is a **Sigma factor**. It recognises specific sequences in promoter sequences and drives expression of a whole panel of competence specific genes.

Gene expression time course after stimulation of bacteria with competence peptide changes, as we can see from these two graphs. At first none are active, then some start to be active early and then are turned off, while others are first off and then are turned on later.

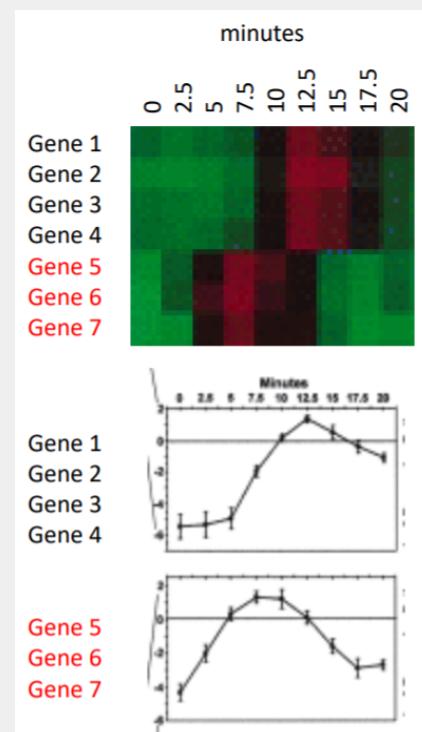
- green is off
- red is on

This means that many new genes are regulated by the competence system and there are distinct sets of early, delayed and late genes.

Again the whole process was studied making mutants to explain the function of every gene: for example in mutants that had **comX** deletions were not able to continue competence changes.

It was found that the genes involved in the initial feedback loop were not controlled by the **comX** sigma factor, which instead controls the later genes.

Obviously the sigma factor has promoter specificity: the specificity is in the -10 sequence.



Sometimes bacteria transform big fragments that end up recombining, however this is very difficult to see.

→ objective: **see recombination**

He used two bacteria that were the same species but phylogenetically quite afar, with 16000 SNPs. When aligning the two genomes, we have the transfer of not just a few nucleotides, but tons of SNPs. This means that it was easy to identify which parts were from the donor and which parts were from the receiver.



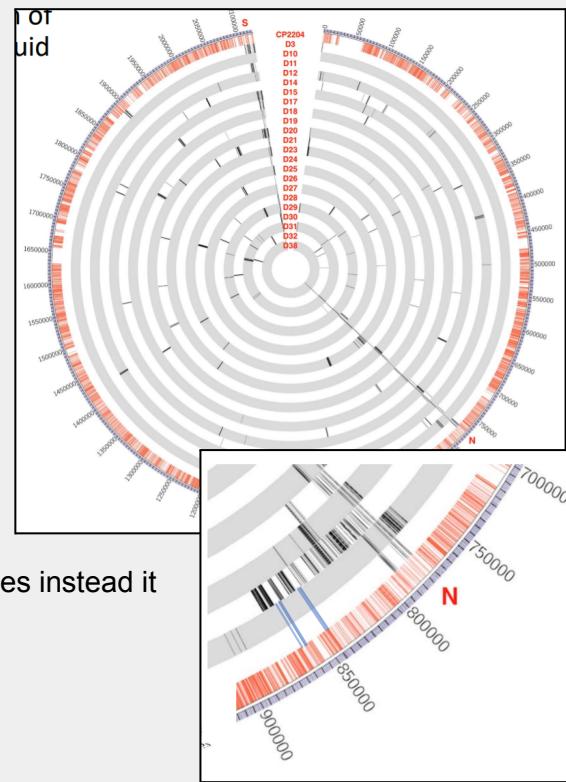
In this picture, red outer circle SNPs distinguish donor and recipient.

Each inside circle is the genome sequence of a transformant.

Black bars in circles are the donor SNPs detected in the **transformant**.

Not all SNPs are transferred in one region of recombination. The most likely explanation is that **larger DNA fragments align and recombine but some breaks may occur during this process.**

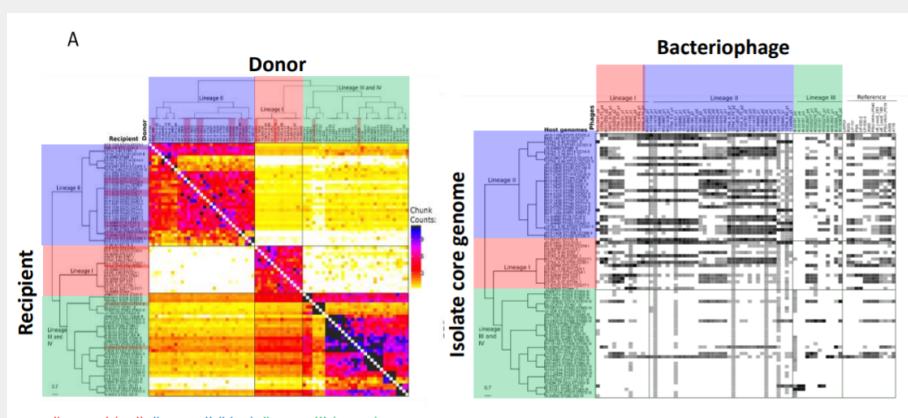
→ Sometimes the transferred DNA is a lot, sometimes instead it is not much. It honestly depends.



But what is the mechanism for this horizontal gene transfer?

Let's think again about the DIRECTIONALITY OF GENE TRANSFER. As we have already said between the same lineages there are bigger regions of homology, and so recombination is more frequent. In the picture below only core genomes are considered, and the experiment is done on Listeria.

★ the question was how does listeria exchange DNA?



→ Transformation: possible but never proved to be present in *listeria*

→ Transduction: possible both generalised and specialised transduction could transfer core genome genes

→ Conjugation: import of whole genetic elements should not impact on the core genome, so in this case it isn't considered.

Scientists' reasoning was that if Listeria exchanges DNA through transduction, we should be able to see remnants of the phages in some of the bacteria. Genomes of all listeria strains were collected.

- in the first graph we consider DNA that is recombinanting. We can see that similar lineages recombine better because of homology.
- in the second graph we are analysing the presence of the same phages. We can see for example that phages of the blue lineage are present in blue lineage as well as red. This means that phages do not go around like the DNA in the first graph does. We don't know why phages do this, but it is enough to tell us that DNA and viruses do not move around in the same way.

This means that transduction is not responsible for ***Listeria* gene transfer**, and it is probably **transformation instead**.

Mycobacterium tuberculosis instead evolves only through mutations.

PLASMIDS AND CONJUGATION

★ PLASMIDS

A plasmid is a small, extrachromosomal DNA molecule within a cell that is physically separated from chromosomal DNA and can replicate independently.

- they are made of dsDNA and are arranged like covalently closed circles. However some linear plasmids also exist.
- they have a very small size; 1.5 - 150 Kb ; (2 - 200 genes) but some can also be very big.
- they are very widespread; found in all bacterial

WHAT'S THE DIFFERENCE BETWEEN A SECOND CHROMOSOME AND A PLASMID?

A chromosome contains key information that is necessary for the survival of the cell.

Plasmids instead contain only information that pertains to the accessory genome.

Since we are learning that in microbiology there are always exceptions, it is important to remember that even if most bacteria have only one chromosome, some plasmids have begun to contain essential information, and thus we can consider those bacteria as having two chromosomes.

Plasmids are non-essential for host viability¹, but they are able to confer different phenotypes:

- provide selective advantage;
 - antibiotic resistance (R-factors)
 - heavy metal tolerance (Mer)
 - ability to degrade toxic hydrocarbons (TOL)
 - bacteriocin production (Col)
- determinants of pathogenicity;
 - haemolysin production (Hly)
 - entero-toxin production (Ent)
 - plant Crown-gall tumours (Ti)

Plasmids are also important as vectors for Genetic Engineering, since if constructed in vitro they can have useful attributes.

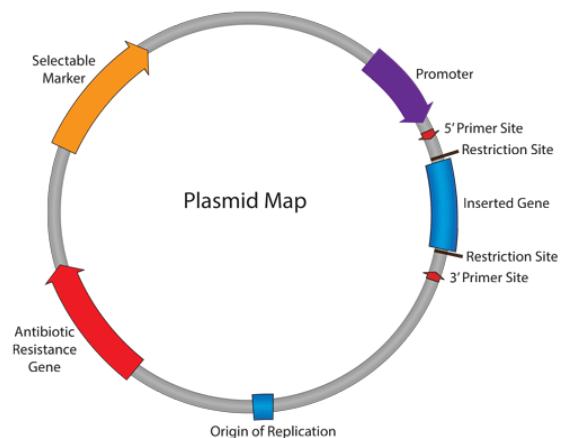
¹ of course if we are in an environment with penicillin and a plasmids has a penicillin resistance, it is fundamental for the bacteria to survive. However this regards isolates and not the entire species.

WHAT EXACTLY IS ANTIBIOTIC RESISTANCE?

Of every antibiotic, there is a maximum dosage we can take before experiencing any side effects. This maximum dosage corresponds to a certain concentration of antibiotic in our body. If the bacterium is still alive at this maximum concentration, we call it resistance.

Plasmids have their own genes to replicate to ensure that each daughter cell still contains the plasmids.

- plasmids have systems involved in replication
- plasmids also have to control their copy number



CONJUGATION

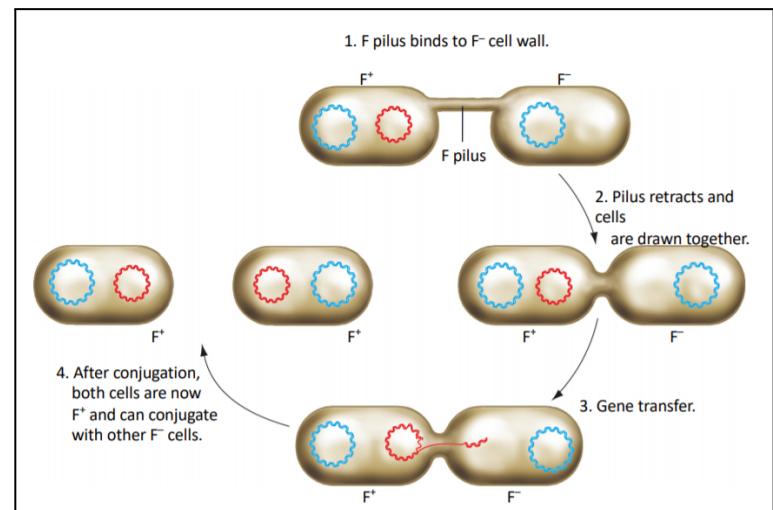
Conjugation requires **two cells**: a donor cell and a recipient cell. The donor cell contains a F(fertility) plasmid (we will call it **F⁺**), while the recipient cell doesn't carry a F plasmid (we will call it **F⁻**).

The fertility plasmid contains an **origin of transfer** that is cut by an enzyme (endonuclease): it's the origin by which the origin on the transfer of this plasmid will start.

When the donor and the recipient cell are near each other, a **F pilus**, a cytoplasmic bridge, is constructed to make the contact between the two cells possible.

The F pilus then retracts, and the cells are drawn together. **Gene transfer starts**,

single DNA strand breaks in the plasmid, going towards the recipient cell. While this happens, **DNA replication** will take place in both the recipient and the donor cell, making it so that on both sides at the end there will be a double stranded plasmid. This is called a **ROLLING CIRCLE MECHANISM**. Like this, the gene is transferred to F⁻, that becomes F⁺. After conjugation, **both cells are now F⁺** and contain the same double stranded copy: they can conjugate with other F⁻ cells.



Not all plasmids replicate: we can differentiate plasmids into two categories:

- ❖ **CONJUGATIVE PLASMIDS** are generally large (50 - 100Kb) and contain a unique origin of transfer replication (oriT). They carry genes encoding “transfer functions”. Their ssDNA transferred by “rolling circle” replication from oriT, as we can see above.
- ❖ **NON-CONJUGATIVE PLASMIDS** carry transfer origin (oriT/mob) but no transfer (Tra) genes. They are transferred by co-resident conjugative plasmid.

HFR CELLS.

Episomes are F **plasmids** that can **integrate** into the bacterial genome of the **DONOR** cell. This is possible due to the fact that there can be sequences on the F plasmid and on the chromosomal DNA that are homologous. We call these sequences IS (insertion sequences), and the **homology** will lead to a **single crossing over** event.

A **high frequency recombination** (HFR) cell, that is by definition a bacterium with a **conjugative plasmid** (for example, the F-factor) integrated into its chromosomal DNA.

HFR CELLS WERE USED TO MAP THE FIRST GENOMES.

All Hfr donor cells carry the F factor at the same site and in the same orientation on their chromosomes, and transfer always initiates from the same spot in the F factor. Thus, the time a gene first enters the recipient cell reflects the distance of that gene from the origin of transfer. Lysing the donor cell at different times, it is possible to realise where/when each gene is.

PLASMID TYPING.

Plasmid typing is the identification and classification of plasmids.

- plasmids cannot be typed by resistance genes or cargo genes, as these are mobile elements. Therefore plasmids are typed thanks to their origin of replication and how they divide during cell replication.

Many bacteria contain more than one type of plasmid. However, These types of plasmid coexist stably in the bacterial cell and equally divide as the cell separates. Even after many cell generations, they may remain in it.

- If another plasmid is present in the cell with a **different kind of control**, the cell is still able to separate the plasmids.
- If another plasmid is present in the cell with the **same kind of control**, the cell won't be able to divide the plasmids in the daughter cells correctly. This means that one of the two types of plasmids will probably be lost. We say that similar plasmids are therefore **INCOMPATIBLE**.

Plasmid incompatibility refers to the inability of two plasmids to coexist stably over a number of generations in the same bacterial cell line. Generally, closely related plasmids tend to be incompatible, while distantly related plasmids tend to be compatible.

Incompatibility is a manifestation of the relatedness of plasmids that share common replication controls, which is the same origin of replication. Incompatibility was defined as the inability of two related plasmids to be propagated stably in the same cell line; thus, only compatible plasmids can be rescued in transconjugants.

→ Plasmids are so typed into incompatibility groups.

- ★ Plasmids are part of the accessory genome.
- ★ PCR-Based Replicon Typing (PBRT) method was proposed in 2005 to detect plasmid content. Typing is not really used anymore as sequencing exists.
- ★ pMLST is tricky because there are too many types of plasmids, and there is no core genome.
- ★ few cells can survive without plasmids. This is because plasmids provide resistance to antibiotics or other methods of survival, so if there are no plasmids, the cell dies.
- ★ most plasmids have accessory segments for more than one thing, protecting the host in different ways.
- ★ Plasmids carry accessory genes between different strains, species. Some plasmids have a wider range of host cells than others.
- ★ Plasmids are very different from one another, as they can be considered almost like mosaics, comprehending IS elements, Tn transposons, gene cassettes, integrons. Some elements are however, sometimes, in common. Still, plasmids can't be put on a phylogenetic tree.

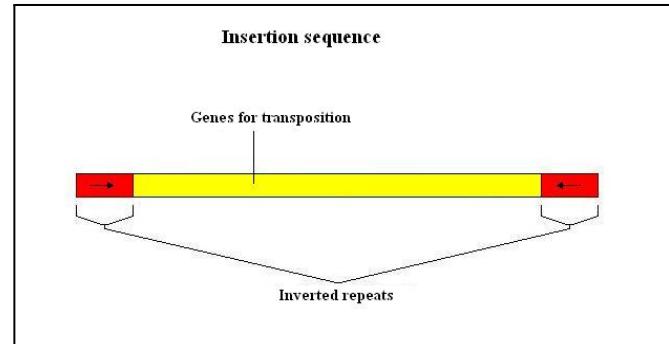
★ OTHER MOBILE ELEMENTS

Chromosomal elements and phages are able to be integrated in the bacterial chromosome (or in other cells) and also be excised.

★ INSERTION ELEMENT

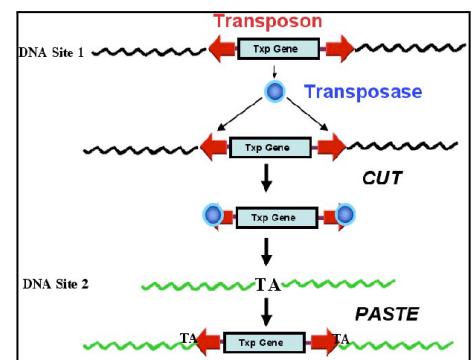
(also known as an IS, an insertion sequence element, or an IS element) is a short DNA sequence that acts as a simple transposable element. Insertion sequences have two major characteristics: they are small relative to other transposable elements (generally around 700 to 2500 bp in length) and only code for proteins

implicated in the transposition activity



★ TRANSPOSONS

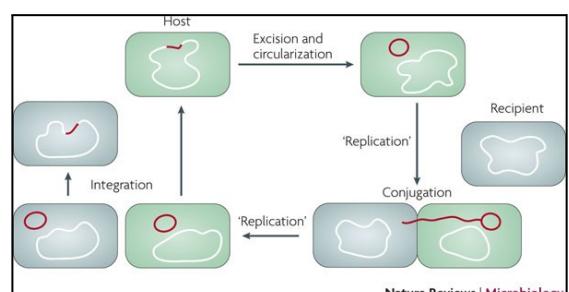
are DNA sequences that can change position within a genome, sometimes creating or reversing mutations and altering the cell's genetic identity and genome size. They also carry accessory genes such as antibiotic resistance genes



★ INTEGRATIVE AND CONJUGATIVE ELEMENT

(ICEs) are located primarily on the chromosome, but have the ability to excise themselves from the genome and transfer to recipient cells via bacterial conjugation. They possess two kinds of genes:

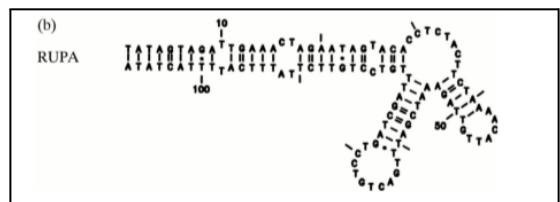
- core genes
- cargo genes, that are not usually related to the ICE life cycle and that confer phenotypes to host cells:
 - ★ antibiotic resistance
 - ★ catabolism
 - ★ restriction systems
 - ★ virulence genes



Unlike plasmids, these elements don't have to self regulate their own replication, as they propagate during chromosomal replication and cell division.
but have to consider their excision instead.

MINIATURE INVERTED REPEAT

TRANSPOSABLE ELEMENTS (MITEs) are a group of non-autonomous Class II transposable elements (DNA sequences). Being non-autonomous, MITEs cannot code for their own transposase. They're just formed by an inverted repeat that forms a hairpin and therefore can be excised.



TRANSPOSEASE is an enzyme that binds to the end of a transposon and catalyses its movement to another part of the genome by a cut and paste mechanism or a replicative transposition mechanism.

Chromosomal mobile genetic elements can be also divided by size:

- Large mobile elements (5-100 kb) are vehicles for genes and horizontal gene transfer
 - Transposons and insertion sequences (0.7-2 kb) can mobilise nearby genes
 - Small mobile elements (50-100 bp) may contribute to genome plasticity like IS elements and MITEs

GENOME PARTITIONING

Rhizobium Etli is one of the many soil-living bacteria able to live in conditions of nitrogen limitation due to its distinctive ability to settle onto root nodules of legumes. Like other *rhizobia*, it is characterized as aerobic, gram-negative, and able to form symbiotic relationships with legumes.

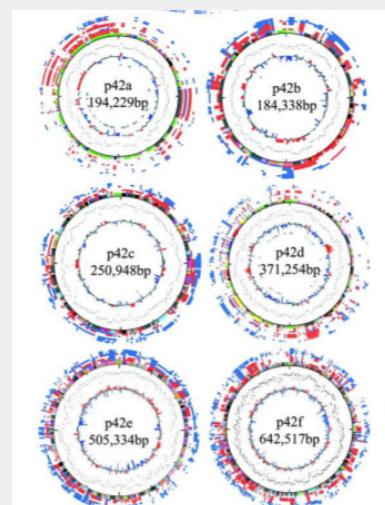
If we look at the bacteria *Rhizobium etli*, it has a genome of 6,530,000 base pairs.

- in the chromosome 4.381,500 base pairs are present
 - so the remaining 2.000.000 base pairs are on plasmids

This means that an average 30% of the genome is on plasmids.

Two of the plasmids, p42a and p42d, are abnormal as it contains a lower GC value of 58% as compared to the other four plasmids at 61.5%. Also, the complete genome sequence reveals identical or more than 100 nucleotide repeats which are located in plasmid p42a and p42d.

- ★ chromosome of *Rhizobium etli* is similar in related species
 - ★ In some plasmids there is homology between related species, but in other cases there is no homology whatsoever.



In classical genetics, **SYNTENY** describes the physical co-localization of genetic loci on the same chromosome within an individual or species.

Extensive synteny at the nucleotide level was found among the *Rhizobium etli* chromosome and other related species. There is however no synteny in plasmids.

FUNCTIONAL GENOMICS

★ OMIC RESEARCH

The goal of *omics* sciences is to take a biochemical/physiological picture of a system under defined conditions

FUNCTIONAL GENOMICS is the study of **how genes and intergenic regions of the genome contribute to different biological processes**. A researcher in this field typically studies genes or regions on a “genome-wide” scale (i.e. all or multiple genes/regions at the same time), with the hope of narrowing them down to a list of candidate genes or regions to analyse in more detail.

→ **COMPARATIVE GENOMICS** is a field of biological research in which researchers use a variety of tools to **compare the complete genome sequences of different species**. By carefully comparing characteristics that define various organisms, researchers can pinpoint regions of similarity and difference. Identifying DNA sequences that have been "conserved" - that is, preserved in many different organisms over millions of years - is an important step toward understanding the genome itself. It pinpoints genes that are essential to life and highlights genomic signals that control gene function across many species.

➤ Comparative and functional Genomics can be considered intimate partners

We also have many other “omics”:

<u>GENOMICS</u>	Genomics is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes
<u>TRANSCRIPTOMICS</u>	Transcriptomics is the study of the complete set of RNA transcripts produced in one cell or in a population of cells under specific environmental conditions.
<u>PROTEOMICS</u>	Proteomics is the large-scale study of proteins. The proteome is the entire set of proteins produced or modified by an organism or system. It tells you when mRNA is transcribed and how stable the mRNA is.
<u>METABOLOMICS</u>	Metabolomics is the scientific study of chemical processes involving metabolites, the small molecule substrates, intermediates and products of cell metabolism .

other 'omics'	
	System
• Molecular	Phenomics
Interactomics	Organomics
Regulomics	Pathogenomics
Drugomics	Physiomics
Lipidomics	
Glycomics	
Phosphorylomics	
Resistomics	Environome (<i>Ecology</i>)
	<ul style="list-style-type: none"> • Metagenome • Metaproteome - Etc.etcetc....

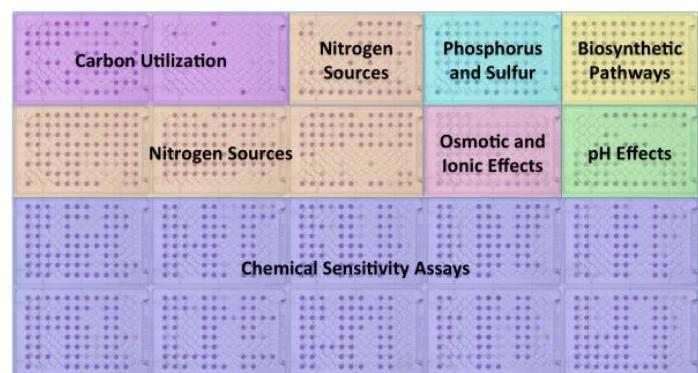
Omics sciences:

- give an overview
- typically do NOT answer a specific question
- Provides data to generate a hypothesis of what is going on
- All 'omics approaches have bias in the methodologies
- Remember humans are bias in choosing the questions and approaches

★ FUNCTIONAL SCREENING

Functional genomics enables the discovery of **gene function and their involvement in biochemical, cellular, and physiological pathways.**

- we could put the same type of bacteria in different wells, and in each well we change the environmental conditions (like presence or lack of certain carbohydrates). The bacteria that survive have a specific phenotype that suffices for the lack of that precise carbohydrate.



Phenotypic microarray technology is growth of microorganisms in **multi-well plate each with a different test component in each well, enabling a screen of the phenotypic characteristics of the test culture.**

Phenotype MicroArray technology enables researchers to evaluate nearly 2000 phenotypes of a microbial cell in a single experiment. Through comprehensive and precise quantitation of phenotypes, researchers are able to obtain an unbiased perspective of the effect on cells of genetic differences, environmental change, and exposure to drugs and chemicals. You can:

- Correlate genotypes with phenotypes
- Determine a cell's metabolic and chemical sensitivity properties
- Discover new targets for antimicrobial compounds
- Optimize cell lines and culture conditions in bioprocess development
- Characterize cell phenotypes for taxonomic or epidemiological studies

★ SYSTEMS BIOLOGY

Systems biology is the study of an organism viewed as an integrated and interacting network of genes, proteins and biochemical reactions which give rise to life.

- system biology brings together a lot of different fields to figure out how they interact;
- networks organise and integrate information at different levels to create biologically meaningful models;
- networks formulate hypotheses about biological functions and provide temporal and spatial insights into dynamical changes.
- genomics approaches are key contributions to systems biology: while experimentalists provide data and system biologists analyze models, genetics have to interpret the data to give models.

★ BIOREMEDIALTION

Bioremediation consists in using biological systems to treat contaminated sites

- for example imagine is there is a petrol leak in the ocean and some bacteria were able to digest it.

This is possible because bacteria perform alkane degradation in the environment if engineered. This process is extremely hard though, there are many different oils and also bacteria behave differently in different environments.

Deepwater Horizon was a drilling rig. In April 2010 it exploded because of a leak. The problem was that the explosion was at the bottom of the structure, and so the oil was at the bottom of the ocean. Scientists analysed the level of bacteria in the contaminated water at that particularly high pressure, and realised that certain bacteria exist, in those conditions, that are able to degrade oils.

The dispersed hydrocarbon plume stimulated deep-sea indigenous γ -Proteobacteria that are closely related to known petroleum degraders. Hydrocarbon-degrading genes coincided with the concentration of various oil contaminants.

PEP group translocation, also known as the phosphotransferase system or **PTS**, is a distinct method used by bacteria for **sugar uptake where the source of energy is from phosphoenolpyruvate** (PEP): it couples the transport of a sugar to its phosphorylation.

ATP binding cassette (ABC) transporters constitute a ubiquitous superfamily of integral membrane proteins that are responsible for the **ATP powered translocation** of many substrates across membranes.

Studies were done in regards to the utilization of sugars by bacteria.

FIRST, HOW DO YOU TEST FOR SUGAR UTILISATION?

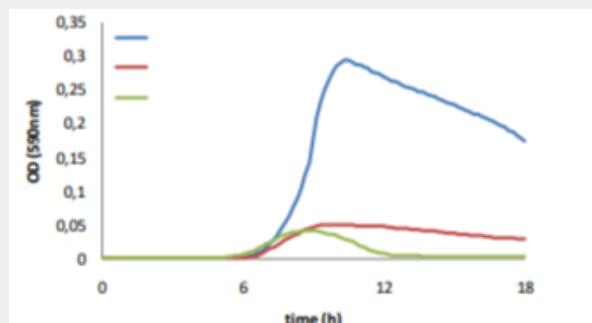
1. Recover bacteria from a frozen stock
2. Plate by isolation on the appropriate agar medium (semi solid medium) and incubate at 37 °C for 16h (over night)
3. Collect few colonies and inoculate in liquid medium.
4. Positive growth is evidenced by turbidity of the medium
5. Turbidity of the medium (optical density, OD) is measured by a spectrophotometer at 590 nm

WHAT DOES THE GRAPH MEAN?

Exp 1: three different bacteria are inoculated with galactose as sole carbon source. The data show that only the blue bacterium is able to grow and therefore able to metabolise galactose

Exp 2: the same bacterium is inoculated with galactose (blue), glucose (red) and sucrose (green). The data show that this bacterium can grow only on galactose and therefore is only able to metabolise galactose

Exp 3: A wild type bacterium (blue) and two mutants (green and red) are inoculated in a medium with galactose. The data show that both mutants are unable to grow in galactose which indicates that both mutated genes are involved in galactose metabolism



THE CONCLUSION OF THE EXPERIMENT:

- Phenotypic and genomic characterisation of carbohydrate uptake systems in *Staphylococcus epidermidis*
- identification of variation in an ABC sugar transporter which is missing in 4 isolates
- Identification of a sorbitol PTS system not previously described in *S. epidermidis*

HOW TO MAKE A MUTANT

1. Amplify by PCR segments upstream and downstream of the gene(s) to be deleted (blue and green)
2. Join in vitro the two fragments to the DNA encoding for a gene conferring antibiotic resistance (red). Clone the construct into a integration vector for your bacterium
3. Recover bacteria from a frozen stock, grow and wash them to remove any salts. Add DNA to your cells and electroporate.
4. Plate the bacteria on agar plates which contain the appropriate antibiotic. Depending on the mechanism, by which you can select for integration of your plasmid (a temperature sensitive replicon), you may need to transfer bacteria to a different temperature.
5. Isolate representative transformants and check if the genetic construct in the transformant clone is ok
6. Test the phenotype by growing in media with only one sugar as carbon source and record culture turbidity

SWEET YOGURT is made by fermentation of milk by bacteria. Starter cultures contain both *Streptococcus thermophilus* and *Lactobacillus delbrueckii*. **Bacteria take up lactose as a carbon source and ferment it to lactic acid.**

Since GMOs cannot be sold, natural mutations of these bacteria were isolated to make yogurt; for example:

- Wild type: Both bacteria in yogurt import lactose and use the glucose-part of lactose to make energy through the glycolysis and produce lactic acid and make an acidic yogurt
- Mutations of the kinase GK1 in *Streptococcus* does not suffice as glucose is re-imported by the PTS transporter..

COMPARATIVE GENOMICS

★ ACCESSORY AND CORE GENES

The goal of *omics* sciences is to take a biochemical/physiological picture of a system under defined conditions

COMPARATIVE GENOMICS is a field of biological research in which researchers use a variety of tools to **compare the complete genome sequences of different species.**

The major principle of comparative genomics is that common features of two organisms will often be encoded within the DNA that is evolutionarily conserved between them. Therefore, comparative genomic approaches start with making some form of alignment of genome sequences and looking for orthologous sequences (sequences that share a common ancestry) in the aligned genomes and checking to what extent those sequences are conserved

When its genome was first sequenced, scientists thought that alone would have been a revolutionary discovery for antibiotics and such. However, as more genomes were sequenced, they soon realised that genomes of apparently similar species are in reality very different.

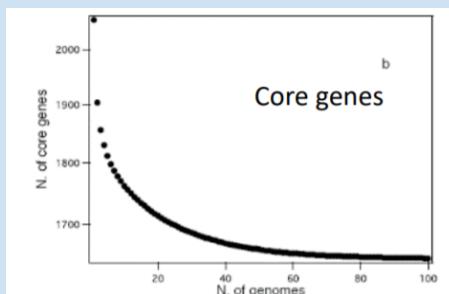
→ there was about a 20% difference in various bacterial genomes.

44 genomes of different species (or rather, OTU) were analysed, and scientists realised that **the more genomes of different species you consider, the less genes in common you find.**

→ remember the concept of **CORE GENOME**, that is the defined as “all genes present in all isolates”

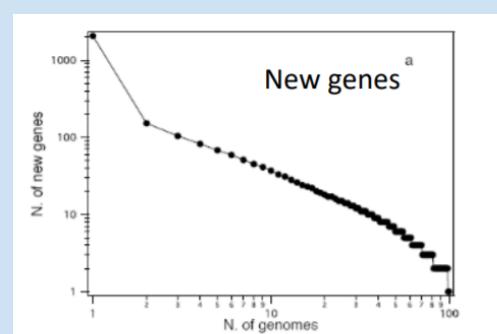
This graph shows that the more genomes you analyse, the smaller the core genome becomes.

→ this is of course valid only for bacteria, and not for eukaryotes.



However, **the more genomes you sequence the more new genes you find**. Even if the graph hypothesises that after a while you stop finding new genes, this was proved inaccurate.

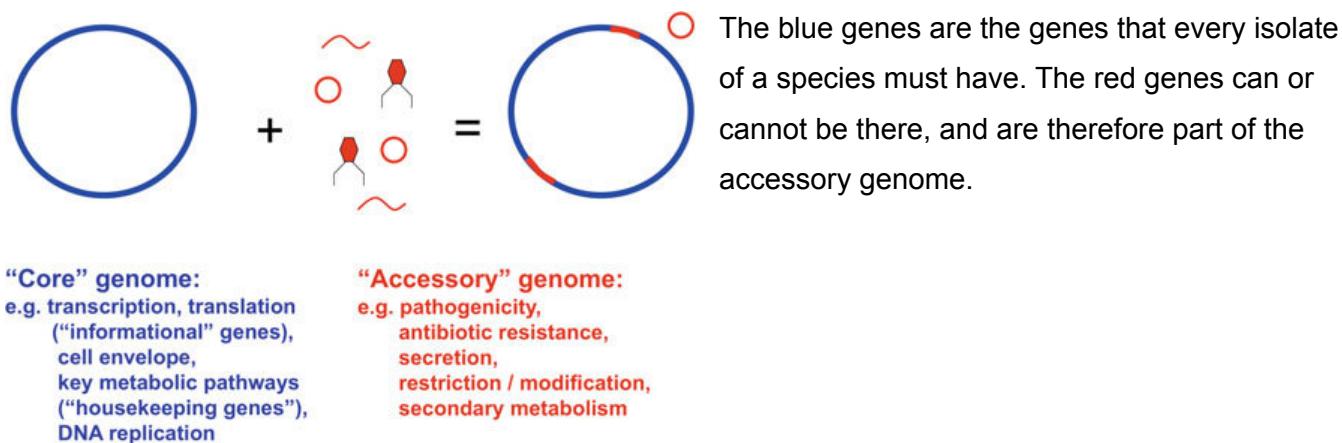
→ **ACCESSORY GENES** are not present in all isolates. If acquired recently from other species.



It is however true that the more you sequence the genome of some species, the number of new genes each time goes down. This means that the species as it is does less recombination than before, and we talk about a **CLOSED PAN GENOME**.

For other species instead, in new isolates the number of accessory genes doesn't diminish as much, which means recombination is still active. In this case we talk about a **CLOSED PAN GENOME**.

- the **core genes are needed for life**, we cannot live without them.
- however there are **essential genes also in the accessory genome**.
 - something that's really important when deciding if a gene is viable is quantity: if there are two genes that are necessary for an iron pump, the loss of one doesn't imply death, but the loss of two does.
 - furthermore some genes are necessary for the formation of a disease and so they're almost "essential", however they're not in common to the whole species.



★ ACCESSORY AND CORE GENES

A **PHYLOGENETIC TREE** is a **branching diagram or a tree showing the evolutionary relationships** among various biological species or other entities based upon similarities and differences in their physical or genetic characteristics.

→ phylogenetic trees are based on core genomes, as accessory genes are too different between species.

Different species of bacteria are very different, even closely related ones. So much that it is hard to establish borders between species.

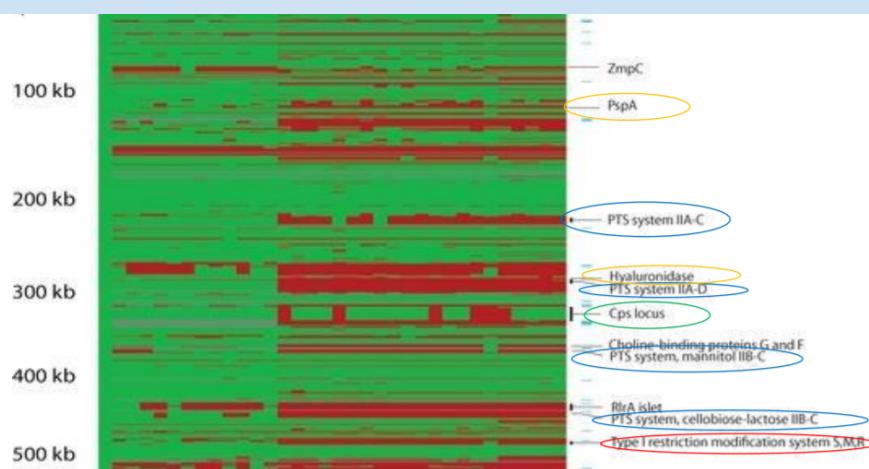
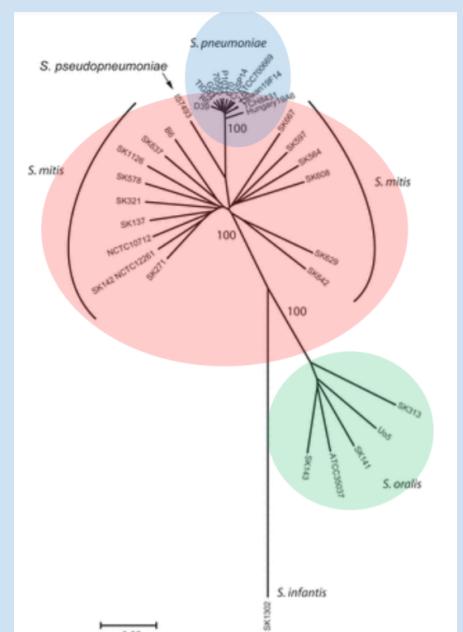
In this case let's analyse *S. mitis*, a really common bacteria in the mouth. *S. pneumoniae* looks like a subtype of *S. mitis*, as it is quite similar to it and totally different from *S. oralis*.

However the functions of *S. mitis* and *S. pneumoniae* are completely different: one is harmless while the other is pathogenic.

On the other hand, *S. mitis* and *S. oralis* are very afar on a phylogenetic tree but they have similar functions.

- the comparative genomic analysis showed that pathogenic genes are only in *S. pneumoniae*, and absent from the related non-virulent oral commensal *S. mitis*.

These are, supposedly, the proteins that make the disease.



Some “obvious” virulence genes were identified, as they were present only in *S. pneumoniae* and not in other bacterial species.

→ green is when a gene is present, red when it's missing

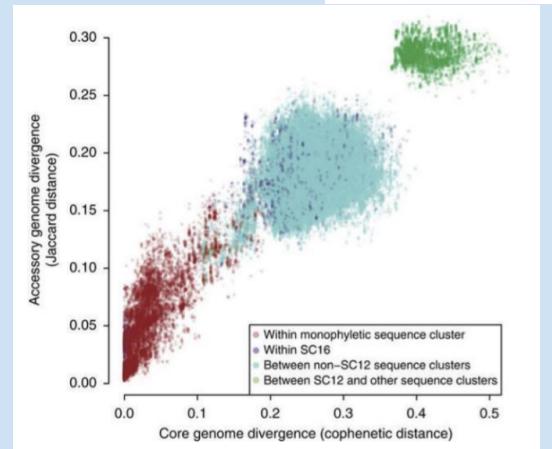
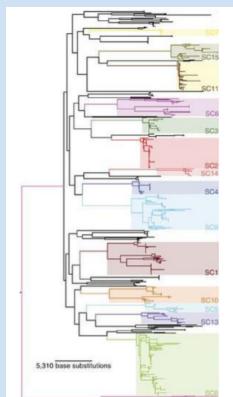
Many of these genes are connected to the transportation of sugar (we can see genes involved in the PTS system), however no specific connection was found.

Many experiments were done in regards to the accessory genome.

1. Scientists made a phylogenetic tree based on the core genome SNPs; defining core genomes that group together as **SEQUENCE CLUSTERS SC**. This tree in particular therefore reflects the DIVERSITY OF THE CORE GENOME.
2. Species that had similar core genomes (and therefore were from the same cluster) also had similar accessory genomes; **accessory genome diversity reflects core genome diversity**. We can see that different isolates from the red, green and blue clusters were considered.
3. Then different clusters were compared to one another: the green cluster to the red and so on.. different clusters of course were different.

This means that the **accessory genome doesn't move**

around randomly. We don't exactly know why. Maybe it's the fact that similar genomes recombine more easily as they have bigger homology regions, however we are not completely sure. There are also elements that are much more mobile.



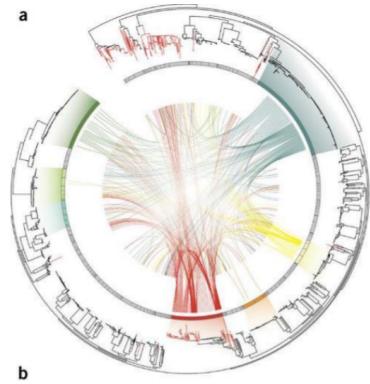
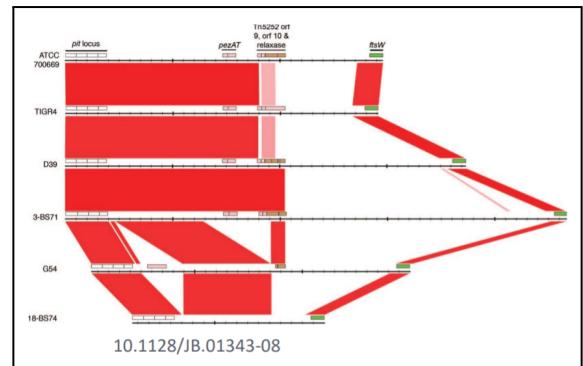
GENOMIC ISLANDS are chromosomal segments imported by horizontal gene transfer and often differ by their GC content

PATHOGENICITY ISLANDS are considered to be a subclass of genomic islands that are acquired by horizontal gene transfer. They encode various virulence factors and are normally absent from non-pathogenic strains of the same or closely related species.

27-kb **pathogenicity island of *Streptococcus pneumoniae***, termed pneumococcal pathogenicity island 1 (**PPI1**), contains iron uptake locus piaABCD, **required for full virulence in mice**. Data indicate that there is striking variation in gene content and structure of the 3' region of PPI1 among strains and that this region includes at least one virulence determinant. Gene variation within horizontally acquired DNA such as that of PPI1 may be one factor modulating differences in virulence among strains.

As we can see the accessory genes (green) are all different, while the core genomes (red) are more similar: the more one variates, the more also the other variates.

All accessory genes however must have a purpose, otherwise they would have been eliminated by evolution. Accessory genes must give some kind of advantage in natural selection, otherwise they wouldn't be there.



A **genome-wide association study (GWAS)**, is an observational study of a genome-wide set of genetic variants in different individuals/strains to see if any variant is associated with a trait.

1. Single nucleotide polymorphisms (SNPs) are mapped throughout the genome by array or sequencing
2. If a SNP is more frequent in subject/strains with the disease/phenotype, the variant is said to be associated with the disease/phenotype

Experiment done sequences 3.085 isolates of *Streptococcus Pneumoniae*, which were taken in a closed population.

- **GENOMIC RECOMBINATION** hotspots show remarkable consistency **between lineages**, indicating common selective pressures acting at certain loci, **particularly those associated with antibiotic resistance**. This allows isolates to stay alive.
- suppose that there's antibiotic consumption, and a particular gene is against it. Then the gene will spread because it survives. In the following page we bring an example about penicillin, which is an antibiotic.
- Temporal changes in antibiotic consumption are reflected in changes in recombination trends, demonstrating rapid spread of resistance when selective pressure is high.

PENICILLIN EXAMPLE

The antibiotic penicillin inhibits bacterial surface enzymes involved in peptidoglycan (cell-wall) synthesis and modification, such as transpeptidase or bifunctional transpeptidases.

Penicillins inhibit the transpeptidases active site by homology to the natural substrate.

- Genes may mutate to encode for penicillin resistant PBPs. These mutant genes may be exchanged by horizontal gene transfer (transformation)

EPIGENOMICS

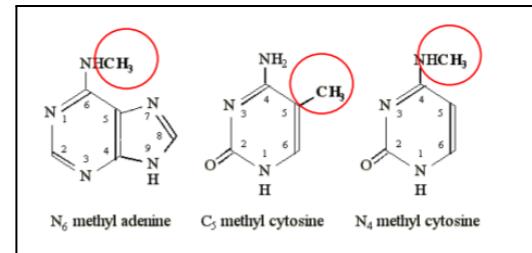
Epigenomics is **the study of all epigenetic modifications** on the genetic material of a cell.

- we define an epigenetic modification as an **hereditary changes which do not affect the primary sequence of the DNA**

Two of the most characterized epigenetic modifications are DNA methylation and histone modification

Different bases are methylated in bacteria and eukaryotes:

- bacterial DNA contains **N6-methyladenine (m6A)** and **N4-methylcytosine (m4C)**
- mammalian DNA contain 5-methylcytosine (m5C)



★ METHYLATION: RESTRICTION MODIFICATION SYSTEMS

Inside a prokaryote, the restriction enzymes selectively cut up foreign DNA in a process called restriction digestion; meanwhile, host DNA is protected by a modification enzyme (a methyltransferase) that modifies the prokaryotic DNA and blocks cleavage. Together, these two processes form the **restriction modification system**.

Restriction modification systems are therefore generally composed of:

- a **METHYLTRANSFERASE**, which protects the own genome by methylating it.
- an **ENDONUCLEASE**, that cleaves foreign DNA.

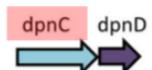
RMS are generally part of the **accessory genome**

- ★ endonucleases can be usually passed on thanks to horizontal gene transfer.
- ★ RMS composed of 1 to 3 genes/enzymes which allow for recognition, methylation (methyltransferase) and restriction (endonuclease)
- ★ The own chromosome is generally protected by methylating the same site which is cleaved
- ★ In many species RMS genes are clearly associated with certain lineages.

Let's now analyze a few restriction systems. They are called *Dpns*, where Dpn stands for *Diplococcus pneumoniae*. The numbers I, II, III etc... specify the number of enzymes found in bacteria.

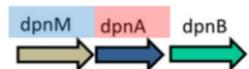
- **DpnI** cleaves fully **adenomethylated** sites and **hemi-adenomethylated** sites

- there is no gene that encodes for methyltransferases
- **dpnC** leads to cleaved GATC when it is methylated.



This is to **restrict foreign DNA that has been methylated**, while the cell itself hasn't been methylated at all.

- **DpnII** cleave sites **lack adenomethylation**. In this system we have three co transcribed genes:
 - **dpnM** encodes for the methyltransferase that attaches the methyl group to the adenines of GATC sequences. In this case we talk about **Dam methyltransferases**, (Dam= *DNA adenine methylase*). The sequence goes from GATC to GmATC. Both sides of the double helix are methylated, so in the next replication cycle only the parent strand will have methyl groups.
 - **dpnA** encodes for a restriction enzyme that, if it sees unmethylated GATC, it cuts it.



This is a way to **restrict incoming foren DNA that has not been methylated by the cell**.

- **DpnIII** cleave sites **lack**. In this system we have three co transcribed genes:
 - **m.dpnIII** encodes for the methyltransferase that attaches the methyl group to the citosins of GATC sequences
 - **r.dpnIII** encodes for a restriction enzyme that, if it sees unmethylated GATC, it cuts it.



This is a way to **restrict incoming foren DNA that has not been methylated by the cell** but has a different methylated base.

There are four categories of restriction modification systems: type I, type II, type III and type IV, all with restriction enzyme activity and a methylase activity (except for type IV that has no methylase activity).

- **TYPE I** systems are the most complex ,they can both cleave and methylate DNA. However the **cleavage often occurs a considerable distance from the recognition site**.
- **TYPE II** systems are the simplest and the most prevalent. Instead of working as a complex, the methyltransferase and endonuclease are encoded as two separate proteins and act independently. **Cleavage occurs at a defined position close to or within the recognition sequence**.
- **TYPE III** systems form a complex of modification and cleavage. The M protein, however, can methylate on its own. Methylation also only occurs on one strand of the DNA unlike most other known mechanisms.
- **TYPE IV** systems are not true RM systems because they only contain a restriction enzyme and not a methylase. Unlike the other types, type IV restriction enzymes recognize and cut only modified DNA.

★ METHYLATION: MISMATCH REPAIR.

Methyltransferases are not only present in restriction modification systems.

Methylation is not only useful because it protects bacteria's own genome from accidental restriction, but also because it allows correct mismatch repair.

- ❖ During replication the newly synthesised DNA strand is not methylated, as to allow error correction. If there are no errors, **Dam (DNA adenine methylase)** methylates the newly synthesised DNA filament.

BUT WHAT IF THERE ARE ERRORS?

Normal DNA polymerase usually makes mistakes when synthesizing the new DNA strand: the most common error made by DNA pol is the **incorporation of the wrong nucleotide (base mismatch)**. However thanks to the activity of **proofreading of DNA polymerase**, DNA pol is often able to understand where there has been a mismatch, then correcting the wrong nucleotide with the right one by going back.

However, sometimes the proofreading of DNA pol doesn't work and there keeps being a mismatch in the DNA sequence. Therefore there is a second mechanism: **mismatch repair mechanism (MMR)**.

- This mechanism is able to recognise the presence of a mismatch and make a cut in the phosphodiester bond to remove a small DNA strand containing the mismatched base. Then the DNA pol will resynthesise a new strand of DNA thanks to the template strand.

Since there is a mismatch between two bases, we **wouldn't know on which strand the error occurred and what to repair**. Like us, the enzyme of the MMR mechanism should only be able to see the distortion of the double helix caused by the mismatch and not understand which of the two strands is wrong. There would therefore be an only 50% chance of correcting the actual wrong base.

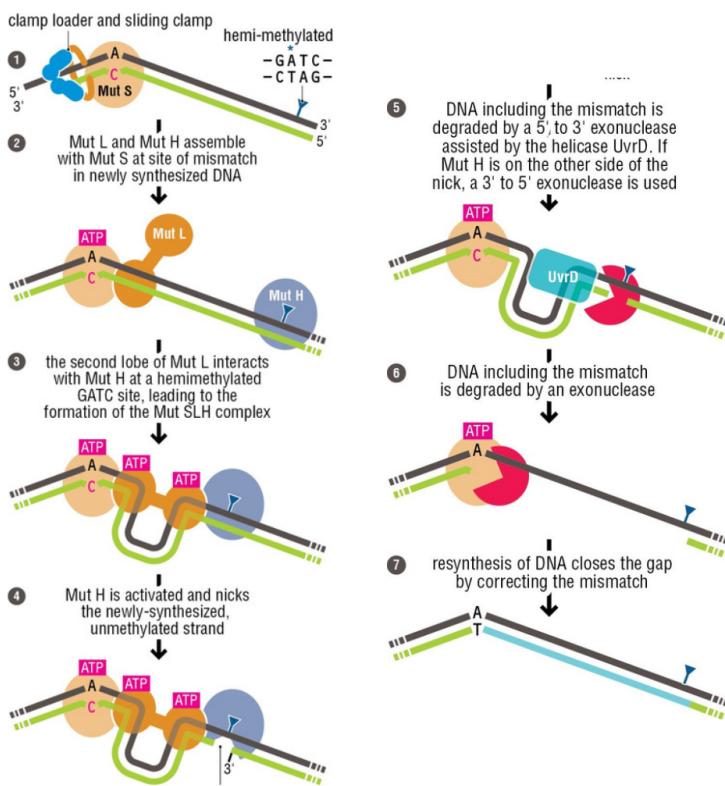
However, at least in E.coli, there is a system that recognises which one of the two strands is brand new and which one is the template strand. **The new strand will of course contain the error.**

In bacteria the detection of the wrong base depends on the **detection of methylated bases**.

- the new strand is not methylated, while the old strand is.

The proteins of the MMR system therefore will recognise the mismatch and the **hemimethylated** (non methylated) **site** nearby, making a **nick** (break of the phosphodiester bond) in the strand that is not methylated and therefore contains the error. Then the exonuclease will cut off the DNA that contains the error and DNA pol will create a new strand.

MOLECULAR PROCESSES OF MMR.



★ **MutS:** is found behind the replication fork to scan for errors. It recognizes and binds mismatched base pairs

★ **MutH:** nicks the newly synthesized strand recognizing hemi- methylated GATC sites

★ **MutL:** recruits helicase II to cleave the unmethylated strand.

★ METHYLATION INFLUENCES TRANSCRIPTION

EPIGENETIC CONTROL OF PHASE VARIATION OF THE PAP-PILUS LOCUS

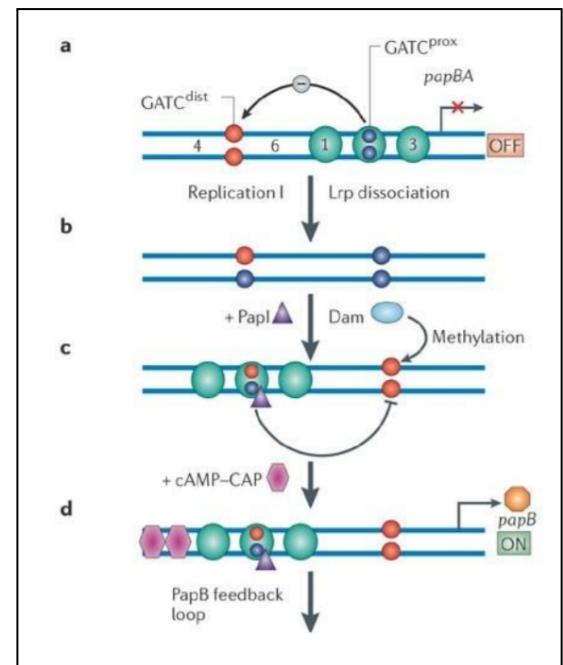
Upstreams of the genes in bacteria are DNA binding motifs that are recognised by regulators and RNA polymerase to regulate transcription.

PAP PILI are fimbrial appendages found on the surface of many *Escherichia coli* bacteria. The P fimbriae is considered to be one of the most important virulence factor. Its expression is under the control of methylation-dependent transcriptional regulation, for which the leucine-responsive regulatory protein (Lrp) is essential.

- the regulatory protein **Lrp** (leucine-responsive regulatory protein) has six binding sites on the DNA upstream of a few genes. Some of these sites are Dam sites (the two colons), which means they can be bound to DNA adenine methylases (Dam).
 - orange ones are **methylated** by Dam
 - blue ones are **not methylated**

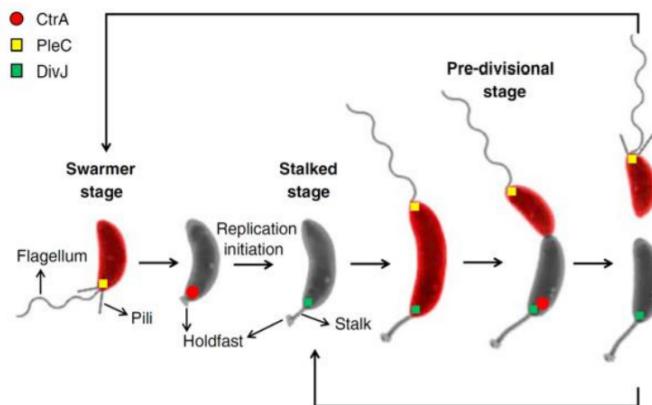
To bind to the DNA chain however **the regulator Lrp needs an unmethylated Dam site.**

This means that if the Dam comes first, the site will be methylated and the Lrp won't bind in the Dam sites (2,5) but in other sites (4,6,1,3). Otherwise the Lrp will bind first and the Dam won't be able to bind.



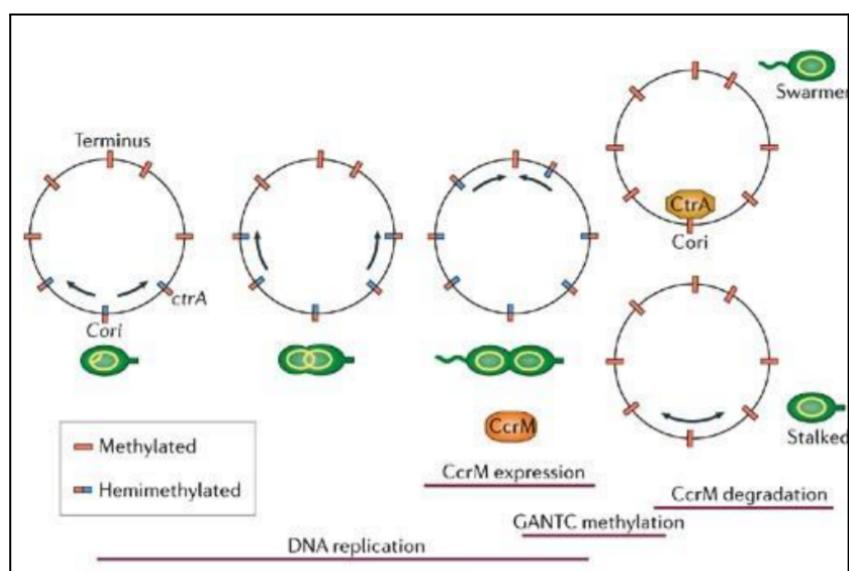
★ METHYLATION INFLUENCES CELL CYCLE

In *Caulobacter crescentus* the **cellular differentiation** is part of its normal cell division cycle. This is actually pretty rare in bacteria, but we've seen another example in *Bacillus Aptilus* which makes spores. In particular *Caulobacter crescentus* is usually in a swarmer form, however in some cases it can take a stalker form and then divide.



- the **CcrM methylase** (cell cycle regulated methyltransferase, which recognises 5'-GANTC-3' and inserts a methyl group on the adenine) influences initiation of replication and decision of generation of a replication stalker cell or a non-replicating swarmer cell
 - ***ccrM*** expression is **repressed by CcrM methylation in front of the gene**, which means it is **expressed only when hemimethylated**.
 - these genes if near the origin are expressed earlier in the cell cycle
 - if they're at the end near the termination site they are expressed at the end
- as it is the only moment they are present in an unmethylated way in the cell: this creates a gradient of hemi-methylated DNA which may have a **clock function**. They are time dependent.

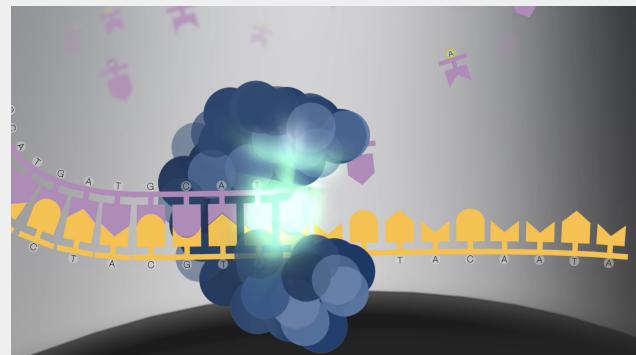
It is produced at the end of the replication cycle when Ccrm recognition sites are hemimethylated, rapidly methylating the DNA.



SINGLE MOLECULE REAL TIME (SMRT SEQUENCING).

SMRT is a technology that shows in real time the addition of nucleotides by a DNA polymerase that is immobilised at the bottom of each chamber.

1. a single DNA molecule is isolated from either bacteria, vertebrates, etc...
2. a single DNA polymerase enzyme is affixed at the bottom of a chamber
3. single nucleotides are also found in the chamber. Each of the four DNA bases is attached to one of four different fluorescent dyes.
4. When a nucleotide is incorporated by the DNA polymerase, the fluorescent tag is cleaved off.
A detector detects the fluorescent signal of the nucleotide incorporation, and the base call is made according to the corresponding fluorescence of the dye



In SMRT sequencing, DNA polymerases catalyze the incorporation of fluorescently labeled nucleotides into complementary nucleic acid strands. The arrival times and durations of the resulting fluorescence pulses yield information about polymerase kinetics and **allow direct detection of modified nucleotides in the DNA template**, including N6-methyladenine, 5-methylcytosine and 5-hydroxymethylcytosine.

- some sequences in the DNA of bacteria are methylated.

PHASE VARIATIONS are present when in the same species some individuals have a gene turned off, and some have a gene turned on. Since this is just about the activity of a gene we don't say this is a mutation, rather, a phase variation.

This happens for example in pili-like we have already studied, where the on and off of a gene is determined by the methylation of other genes.

the research can be summed up in this way:

1. Bacteria have **DNA methylation systems**
2. Methylation systems serve as **defence from bacteriophages**
3. **Methylation has impact on gene expression and phenotypes**

METHYLATION HAS IMPACT ON GENE EXPRESSION AND PHENOTYPES 1

Methylome SMRT								
name	Type	ORF ^a	Specificity ^b	Variable domains	Modified Base	Number of sites ^a	Predicted n. of sites ^a	Strand specificit ^c
DpnI	Type II	SPD_1630-1	5'-GATC-3' 3'-CTAG-5'		-	7164	7324	-2 % -
DpnII	Type II	SPG1732-4 ^d	5'-GATC-3' 3'-CTAG-5'		m6A	7164	7324	-2 % -
SpnI	Type II	SPD_1259-60	5'-TCTAGA-3' 3'-AGATCT-5'		m6A	644	438	47 % -
SpnII	Type II	SPD_1079-80	5'-TCGGG-3' 3'-AGCTG-5'		m6A	1509	1454	4 % -
SpnIIIA	Type I	SPD_0450-5	5'-CRAAN ⁿ CTG-3' 3'-GYTTN ⁿ GAC-5'	V1.1, V2.1	m6A	720	438	64 % 66%
SpnIIIB	Type I	SPD_0450-5	5'-CRAAN ⁿ TTC-3' 3'-GYTTN ⁿ AG-5'	V1.1, V2.2	m6A	1029	665	55 % 64%
SpnIIIC	Type I	SPD_0450-5	5'-GAC ⁿ TTT-3' 3'-GTGN ⁿ AG-5'	V1.2, V2.2	m6A	641	876	-27 % 66%
SpnIID	Type I	SPD_0450-5	5'-GAC ⁿ CTG-3' 3'-GTGN ⁿ GAC-5'	V1.2, V2.1	m6A	428	577	-26 % 67%
SpnIIIE	Type I	SPD_0450-5	5'-CRAAN ⁿ CCT-3' 3'-GYTTN ⁿ GAA-5'	V1.1, V2.3	m6A	1028	665	55 % 63%
SpnIIIF	Type I	SPD_0450-5	5'-GAC ⁿ CTT-3' 3'-GTGN ⁿ GAA-5'	V1.2, V2.3	m6A	796	876	-9 % 64%
SpnIVA	Type I	SPD_0782-4 ^e	5'-GAYN ⁿ TATC-3' 3'-GTYN ⁿ ATAG-5'	V1.1, V2.1	m6A	723	665	9 % 42%

Another study was conducted to study more the relationship between methylation and phenotypes.

Six variants of the same gene, with different methylations, were analysed via SMRT.

Differences between the phenotypes were noticed:

- SpnIIIA clones show opaque and SpnIIIB transparent colony morphology
- SpnIIIB clones are better in colonising mice
- SpnIIIA clones are better in starting bacteraemia
- SpnIIIB clones have less capsule locus expression

This shows that this is in fact an epigenetic change. Other experiments were also made.

METHYLATION HAS IMPACT ON GENE EXPRESSION AND PHENOTYPES 2

How can we demonstrate that methylation really changes gene expression?

- in some cases the **methylation is in the promoter** or in other binding sites, it is **blocking site specific binding** of something else. Like in the papillus.
- but can we really say that methylation itself has an effect on gene expression?

To experiment with this, we take the promoter regions and clone them in front of a reporter (in this case the **luciferase gene**, that can express light if in contact with ATP and oxygen)

- if the promoter was off, light wasn't present
- if the promoter was on, light was present.

Some bacteria could be methylated, others couldn't. The experiment was carried out on both types to understand if methylation influences the presence of light.

1. Recombinant bacteria which **methylate different sites** didn't really make light. This means that the promoter wasn't really active.
2. Recombinant bacteria which **methylate in front of the promoter** made a lot of light. This means that the promoter was active.
3. Recombinant bacteria which **do not methylate made no light**. This means that the promoter wasn't active.

So methylation does influence gene expression. We don't understand why, but it does.

ANTIBIOTIC RESISTANCE

★ ANTIBIOTICS IN GENERAL

Most antibiotics are designed against a broad range of bacteria.

- This is very important because, if for example you have pneumonia, you don't know exactly which bacteria is causing it. Of course diagnostic methods could be used, but they would take days, so having an antibiotic that works against different bacteria who all cause pneumonia is better. This means that we want a medication that works in general against pneumonia, rather than an antibiotic that works only on one of the many bacteria that can cause pneumonia.

An antibiotic needs therefore two main things:

1. to **kill all of the bacteria** that can cause a certain illness
2. to **not kill the human host.**

We say that the drug target has to:

- have **SELECTIVITY**: as the drug has to be **specific to bacterial cells**
- **TARGET ESSENTIALS** for bacterial replication and/or survival. Drugs therefore attack core genome structures:
 - cell wall, which is good because humans don't have that
 - RNA polymerase
 - etc...

This is done in order to control mechanisms common to all/most bacterial cells (we say that the antibiotic has a **broad spectrum**)

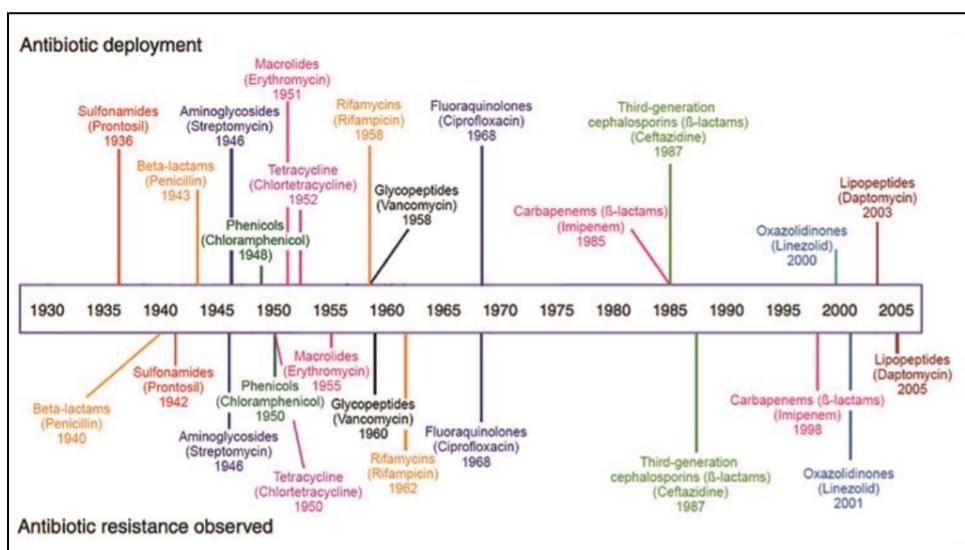
- Currently **DRUG RESISTANCE** is the main factor reducing the activity spectrum of drugs (**narrow spectrum**). Whenever we add a damaging agent to a population, we select all those individuals that are, by chance, immune to it. This means that **the use of antimicrobial drugs selects for drug resistance.**

We can also say that:

- ★ Most antimicrobial agents are active only against bacteria; smaller numbers of antifungal, antiviral, antiprotozoal agents are available. This is because it's easier to kill something that's completely different to us than something that's very similar and has more in common.
- ★ Antimicrobial agents used in clinical medicine that do not affect spores of bacteria or fungi, or latent viruses.

★ In our body we have many different kinds of bacteria that are actually very useful to us. When we take antibiotics those bacteria suffer but they don't all die. This is because drugs only kill bacteria that are actively replicating. The only bacteria that are always replicating inside our body are those that are pathogenic, because they're out of control. Instead in our body, normal bacteria do not usually replicate. This is because otherwise we would get incredibly large amounts in a small time. Bacteria in our body are in a stationary state. Of course some of them might die while taking antibiotics, but we never completely run out of them.

If we have a population that is mostly susceptible to an antibiotic and die because of it, there is a chance that some of those bacteria mutated and are not susceptible. This means that if we use the antibiotic, later we will have left only bacteria that are resistant to the antibiotic. This is because everything else died and the other ones reproduced. Thus, drug resistance develops.



As we can see in the timeline, for every antibiotic that was invented a few years later bacteria that were immune to it were found.

Another problem is that the number of antibiotics is decreasing. This is because since bacteria keep developing resistance, industries are less prone to invest in them.

- industries are more likely to invest in a cure for cancer, that will last for 40+ years, than antibiotic that will only last a few.

★ ANTIBIOTICS AND GENOMICS

Great expectation was placed into drug discovery using genomics: the objective was to understand what enzymes were used in which pathway. In particular any enzymes that were necessary for bacterial life but not human life could be used.

Unfortunately genome based drug discovery was not very successful as practical aspects as drug penetration and lack of bactericidal activity terminated most discovery projects.

★ DEFINITIONS

CLINICAL RESISTANCE AND CLINICAL BREAKPOINTS:

A clinical breakpoint is a **chosen concentration (mg/L)** of an antibiotic which defines whether a species of bacteria is susceptible or resistant to the antibiotic

CLINICALLY SUSCEPTIBLE (S):

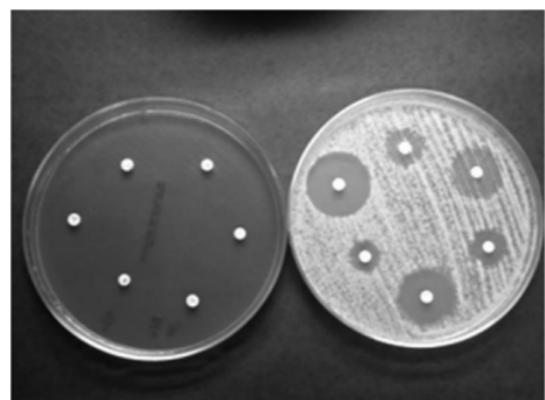
- ★ a micro-organism is defined as **susceptible** by a level of antimicrobial activity associated with a high likelihood of therapeutic **success**
- ★ a microorganism is categorized as susceptible (S) by applying the appropriate breakpoint in a defined phenotypic test system

CLINICALLY RESISTANT (R):

- ★ a micro-organism is defined as **resistant** by a level of antimicrobial activity associated with a high likelihood of therapeutic **failure**
- ★ a microorganism is categorized as resistant (R) by applying the appropriate breakpoint in a defined phenotypic test system

A way to understand if bacteria is susceptible or not is the so called disc diffusion susceptibility testing.

- Seed bacteria on agar plate
- Overlay disc with antibiotic
- Measure diameter of inhibition zone where bacteria did not grow.
- **Compare to tables with breakpoints**
- Report susceptible or resistant phenotype (S, R)



However something that's really important to mention is that not all bacteria of the same strain have the same clinical breakpoint, as some of them mutate. We therefore need to talk about: **MICROBIOLOGICAL RESISTANCE AND EPIDEMIOLOGICAL CUT-OFF VALUES**

WILD TYPE (WT):

- ❖ a micro-organism is defined as wild type (WT) for a species by the **absence of acquired and mutational resistance mechanisms** to the drug in question.
- ❖ a microorganism is categorized as wild type (WT) for a species by applying the appropriate cut-off value in a defined phenotypic test system.

NON-WILD TYPE (NWT):

- ❖ a micro-organism is defined as non-wild type (NWT) for a species by the presence of an **acquired or mutational resistance mechanism** to the drug in question.
- ❖ a microorganism is categorized as non-wild type (NWT) for a species by applying the appropriate cut-off value in a defined phenotypic test system.

★ HOW DO WE TEST ANTIBIOTICS?

MINIMUM INHIBITORY CONCENTRATIONS (MICs) are defined as the lowest concentration of an antimicrobial that will inhibit the VISIBLE growth of a microorganism.

MICs are used to evaluate the antimicrobial efficacy of various compounds by measuring the effect of decreasing concentrations of antibiotic/antiseptic over a defined period in terms of inhibition of microbial population growth. MIC depends on the microorganism, the affected human being (in vivo only), and the antibiotic itself.

Suppose that we want to know the minimal inhibitory concentration (MIC) of an antibiotic.

1. Incubate bacteria in 2-fold dilutions of drugS
2. Incubate for 24h
3. a turbid well (turbid because there was growth) indicates resistance and so we need to increase the inhibitory concentration.
4. a transparent well indicates growth inhibition. Being transparent doesn't mean there is no bacteria.

MINIMUM BACTERICIDAL CONCENTRATIONS (MBCs) as the lowest concentration of antimicrobial that will kill an organism after subculture on to antibiotic-free media.

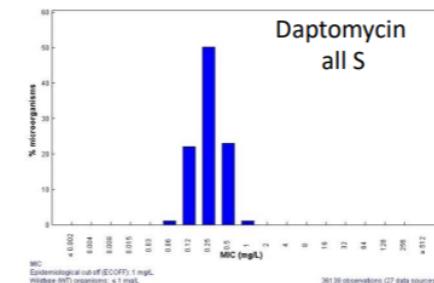
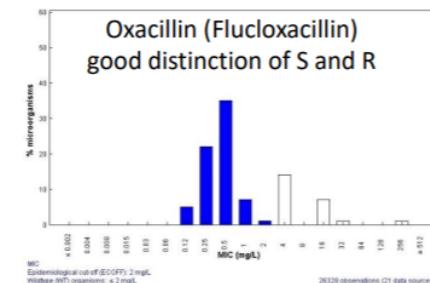
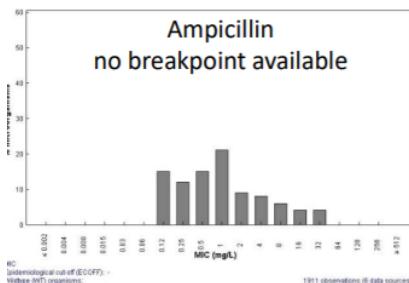
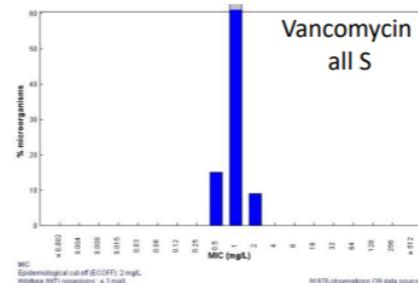
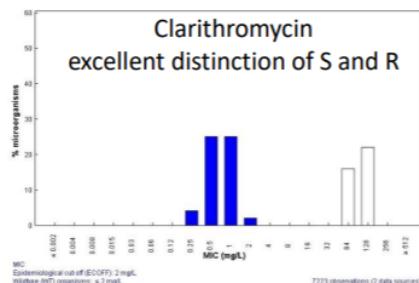
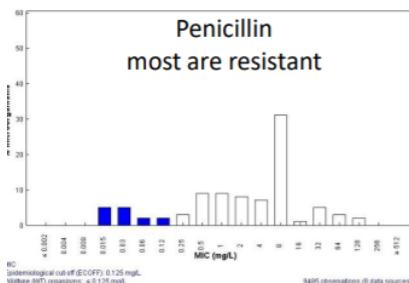
Suppose that we want to know the MBC of an agent:

1. Plate on solid media 10 microliter of medium from transparent wells
2. Incubate 24h
3. Absence of growth indicates bactericidal effect

While **MIC** is the lowest concentration of an antibacterial agent necessary to inhibit **visible growth**, minimum bactericidal concentration (MBC) is the minimum concentration of an antibacterial agent that results in **bacterial death**. The closer the MIC is to the MBC, the more bactericidal the compound.

Let's now have a look at the following graphs of MIC distributions for *Staphylococcus aureus*.

on the X axis we find the level of MIC, while on the Y axis we find the percentage of microorganisms.



- susceptible bacteria are in blue
- resistant ones are white.

As we can see there are some drugs (*daptomycin*, *vancomycin*) to which all *Staphylococcus aureus* are susceptible to. However, they are very expensive and not very used, as they would select for only resistant strains in no time.

This means that other antibiotics are usually used: for example in the case of *oxacillin* and *clarithromycin*, there are bacteria that are both susceptible and resistant to them.

★ ANTIMICROBIAL DRUG RESISTANCE (AMR)

Antimicrobial drug resistance is, as we said, the resistance that bacteria have against a particular drug.

- any use of a “selective agent” like an antibiotic, will select for resistance against that agent: the more i use an antibiotic the more i kill all the bacteria that are susceptible to it, and the more only the resistant bacteria remain.

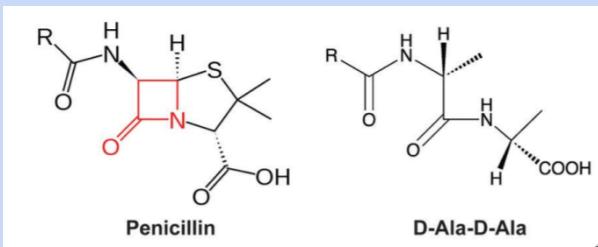
This doesn't work as well for humans because we are much less: there are more bacteria in a test tube than humans in the world, so it's easier to look at their mutations.

Resistance can be both natural and acquired:

- **NATURAL RESISTANCE**: for example some bacteria do not have a cell wall, so an antibiotic that targets the cell wall won't work by principle.
- **ACQUIRED RESISTANCE**: means that an antibiotic that first worked against a bacteria now doesn't work anymore, because all resistant bacteria were selected by it. This is what happens in the hospital environment. For example:
 - the **drug target might have mutated**: as “target” we mean the element of the bacterium that the drug is targeting in particular: it could be a cell wall, a particular enzyme etc.. if these particular things mutate, the antibiotic might not be able to bind and therefore to degrade particular molecules.
 - **new gene** imported by horizontal gene transfer.
 - **TARGET MODIFICATION**
 - **DRUG INACTIVATION**
 - **EFFLUX**
 - **ALTERNATIVE TARGET**

EXAMPLE: drug target mutates.

- Let's now consider *Streptococcus pneumoniae* and the antibiotic *penicillin*.



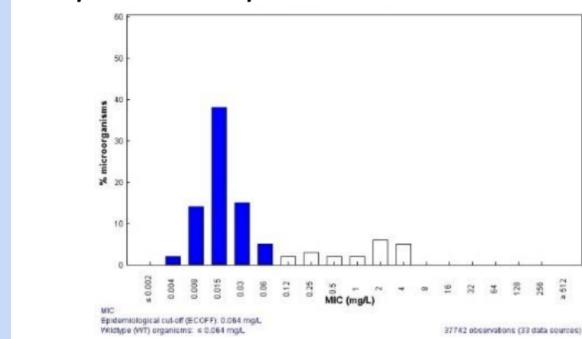
PENICILLIN targets the cell wall to kill the bacteria: it inhibits its target protein by mimicking D-alanine-alanine dipeptide in the peptidoglycan.

When a bacterium divides in the presence of penicillin, it cannot fill in the “holes” left in its cell wall. Without a full cell wall to support the bacterium, it “pops” from the turgor pressure.

However, all three pneumococcal core-genome genes encoding the cell wall biosynthesis enzymes PBP1a, PBP2b and PBP2x can **have missense mutations conferring decreased susceptibility to penicillin**. This means that some of the bacteria can develop a resistance due to mutations.

Of course then the mutated genes can then be transformed by horizontal gene transfer, however they originate from mutations: no new genes are build.

Streptococcus pneumoniae - Penicill



EXAMPLE: NEW GENES ARE FORMED (alternative target)

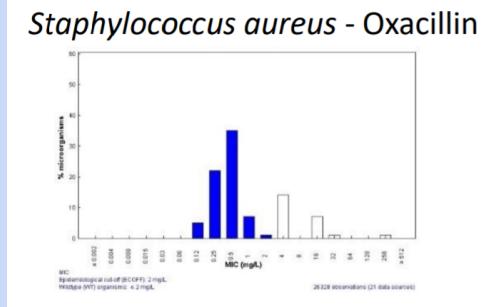
Notice that there is a difference between acquiring a mutated gene via horizontal gene transfer and getting a completely **new different gene obtained through recombination**.

2. Let's now consider *Staphylococcus aureus* and the antibiotic *oxacillin / methicillin*.

Staphylococcus aureus is normally susceptible to methicillin, however, some bacteria have developed a resistance for it.

In the **resistant bacteria**, the staphylococcal chromosomal cassette element **SCC** carries the **mecA gene**. This **new gene**

is resistant to antibiotics such as methicillin, penicillin and other penicillin-like antibiotics. It encodes for a new protein binding methicillin, and therefore **providing an alternative target**.



mecA is on staphylococcal cassette chromosome *mec*, it is a **mobile gene element** from which the gene can undergo horizontal gene transfer and insert itself into the host species, which can be any species in the *Staphylococcus* genus. Proper insertion of the *mecA* complex into the host genome requires the **recombinases**. (**recombinases are enzymes that catalyse site-specific recombination events in DNA**)

Therefore the **formation of the *mecA* gene is due to recombination**, and then the gene is transferred via horizontal gene transfer.

- SCC elements were discovered as they are the genetic platform for mobilisation of the *mecA* gene. conferring methicillin resistance. The SCC elements are excised by the site specific recombinases. The SCCmec and joint SCCmec-cop elements are mobilizable as circular intermediates can be detected. Apparently the SCoop does not excise independently even if carrying a *ccrC* recombinase. Experiments done by quantitative PCR detection of the attachment site (*attS*) formed during circle formation or chromosomal attachment site (*attB*) formation

Detection of *mecA* is usually done through PCR.

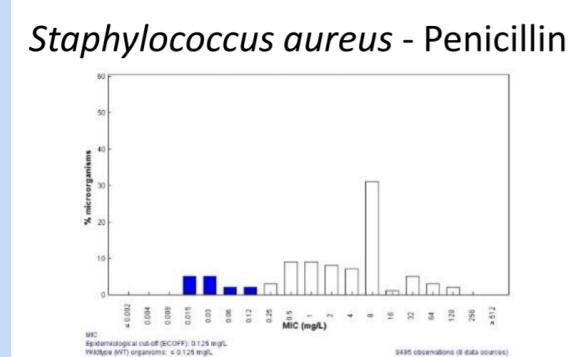
EXAMPLE: NEW GENES ARE FORMED (drug inactivation)

3. Let's now consider *Staphylococcus aureus* and the antibiotic *penicillin*.

The blaZ gene encodes a **beta-lactamase**, which is an enzyme conferring high level resistance to penicillin, as it **cuts the antibiotic**. Beta-lactamases take their name from beta-lactam antibiotics that prevent the synthesis of the cell wall.

- there are many beta-lactamases are there are many beta-lactam antibiotics. Resistance plasmids encoding β -lactamase vary widely in origin.

However, they always possess a common or closely related DNA sequence of about 3×10^6 daltons, even when no other sequences are shared. This striking uniformity is attributable to the fact that the β -lactamase gene (bla) resides within a 3.2-megadalton DNA sequence, termed transposon A (TnA), which is capable of translocating at high frequency from one replicon to another.



Resistance genes are often on mobile elements which in turn are located on plasmids (which are again mobile).

- A **TRANSPOSON** is a DNA sequence that can change its position within a genome, sometimes creating or reversing mutations and altering the cell's genetic identity and genome size
- **INTEGRONS** allow bacteria to adapt and evolve rapidly through the stockpiling and expression of new genes. These genes are embedded in a specific genetic structure called gene cassette that generally carries one promoterless open reading frame (ORF) together with a recombination site (attC). Integron cassettes are incorporated to the attI site of the integron platform by site-specific recombination reactions mediated by the integrase. Integrons are on transposons that are on plasmids: we have a mobile element *on* a mobile element.

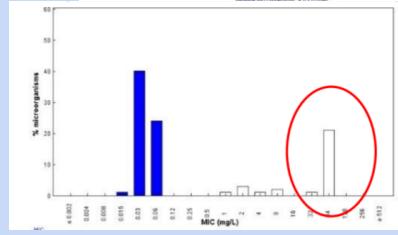
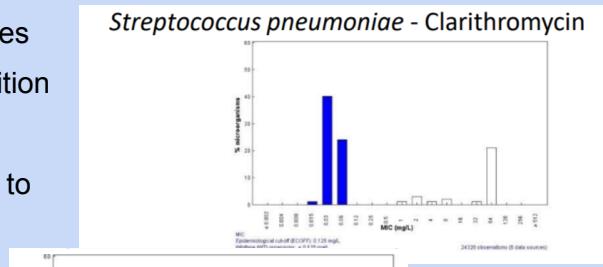
EXAMPLE: NEW GENES ARE FORMED (target modification)

4. Let's now consider *Streptococcus pneumoniae* and the antibiotic *Clarithromycin*.

Usually *Streptococcus pneumoniae* is susceptible to Clarithromycin, which is an antibiotic that attacks the ribosome of bacteria.

However, a plasmid encoded **ErmC methylase** methylates the 23S rRNA. The modification of the target though addition of a methyl group prevents the binding of the antibiotic.

This methylase therefore confers **high level resistance** to macrolides



EXAMPLE: NEW GENES ARE FORMED (efflux)

Efflux pumps are capable of moving a variety of different toxic compounds out of cells, such as antibiotics, heavy metals, organic pollutants,

5. Let's still consider *Streptococcus pneumoniae* and the antibiotic *Clarithromycin*.

The **MefA efflux system** encoded by a phage like element confers low level macrolide resistance. As we can see from the graph the efflux is not as efficient al blocking completely the activity of the antibiotic, but it does give selective advantage.

OBSERVATION: Bacterial efflux transporters are classified into five major superfamilies, based on their amino acid sequence and the energy source used to export their substrates:

- The major facilitator superfamily (**MFS**) that also includes the MefA system. The major facilitator superfamily (MFS) are membrane proteins which are expressed ubiquitously in all kingdoms of life for the import or export of target substrates.
- The ATP-binding cassette superfamily (**ABC**): ABC transporters often consist of multiple subunits, one or two of which are transmembrane proteins and one or two of which are membrane-associated AAA ATPases. The ATPase subunits utilize the energy of adenosine triphosphate (ATP) binding and hydrolysis to provide the energy needed for the translocation of substrates across membranes, either for uptake or for export of the substrate

Transporters can have strict **substrate specificity** or transport multiple substrates. Furthermore, genes for multi-drug efflux transporters are often mobilised on plasmids. Core genome efflux systems can be upregulated by mutation of promoter regions.

WORKSHEET

Aminoglycosides block the bacterial ribosome by targeting ribosomal 16S rRNA.

Aminoglycoside antibiotics naturally concentrate in the kidney and middle ear.

- there exists a maternally inherited disease: aminoglycoside-induced hearing loss.
- amongst mutations that underlie deafness, an A1555G base substitution in the small (12S) rRNA gene of mtDNA is of particular interest as a main cause of antibiotic-induced deafness. This mutation increases sensitivity to aminoglycoside ototoxicity, but has been reported also in deaf individuals who have not been exposed to these antibiotics.

An important thing to remember is that the **DNA of our mitochondria comes from a bacterium**.

This means that when making an antibiotic we need to make sure the mitochondria isn't affected.

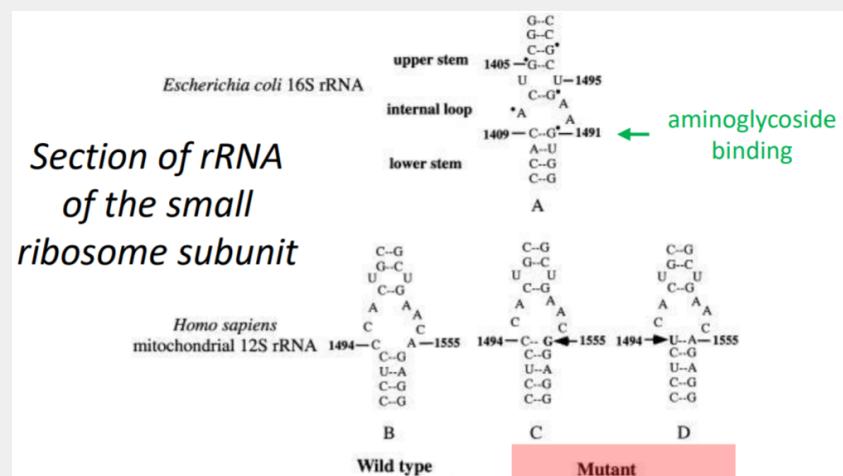
1. Why do individuals with an A1555G 12S rRNA mutation experience hearing loss?

The mutation from A to G allows the double helix to form a hydrogen bond between the C and the G. This particular bond allows the aminoglycoside to bind.

2. Would you consider this as a side effect?

No. It is not the aminoglycoside's fault if the mitochondria mutates. It binding the mutated mitochondria causes hearing loss, but this doesn't depend on the antibiotic itself. Also these antibiotics are not given systemically.

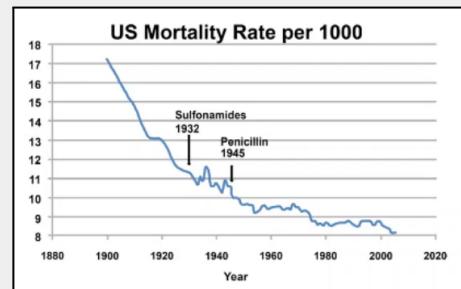
- an actually side effect would be people being allergic to penicillin.



Online tools identify readily resistance associated genes. Some of the genes listed are genes which clearly confer resistance to a drug. Many hits are genes which can confer resistance and are mutated. Mutations in the promoter can upregulate efflux pumps. Mutations in the coding sequence of drug targets can confer resistance. This means that data have to be evaluated critically

THE IMPACT OF ANTIMICROBIAL DRUG RESISTANCE: DISCUSSION.

Antibiotics are super important of course for battling infections. However as this mortality rate graph shows us, infectious decreases have been decreasing long before the invention of antibiotics. This is probably due to hygienic conditions and other living improvements.



There are some bacteria that however are pan drug resistant, which means they are resistant to every antibiotic.

Grouping	Organism or family	Resistance
1: Critical	<i>Acinetobacter baumannii</i> <i>Pseudomonas aeruginosa</i> <i>Enterobacteriaceae</i>	carbapenems carbapenems carbapenems, third-generation cephalosporins
2: High	<i>Enterococcus faecium</i> <i>Staphylococcus aureus</i> <i>Helicobacter pylori</i> <i>Campylobacter</i> spp. <i>Salmonella</i> spp. <i>Neisseria gonorrhoeae</i>	vancomycin methicillin, vancomycin clarithromycin fluoroquinolone fluoroquinolone Third-generation cephalosporins, fluoroquinolone
3: Medium	<i>Streptococcus pneumoniae</i> <i>Haemophilus influenzae</i> <i>Shigella</i> spp.	penicillin ampicillin fluoroquinolone

WHO made a priority list to combat antimicrobial drug resistance (AMR).

AMR depends also on the country, as not every country uses the same antibiotics with the same conditions. Italy's AMR is high, which is bad.

A report said that without new antibiotics to tackle resistance, millions of lives around the world could be at risk each year from infections by 2050. **Which means that antibiotic resistance is getting worse.**

→ governments are trying to do something about this giving

However some threats are exaggerated as:

1. as we just said **antibiotics did not cure alone infectious diseases**
2. some **numbers might be over the top**, and it is difficult to compare things with the past, as conditions are very different.
3. also it is difficult for bacteria alone to kill an individual: for people who have good health this won't happen. So a worsening in AMR would mainly interest only the **weakest part of society**.
4. there are **ways to go against antibiotic resistance**. For example hospitals in leicester prevented the spread of methicillin resistance by making every adult patient disinfect themselves.