Lecture Supplements

Plant Bacteriology

Bacterial Genetics

Compiled by N. Hassanzadeh

Version 4.25

January 1, 2025

Website Address:
http://www.phytobacteriology.com

Contact address

- Department of Plant Protection, Faculty of Agricultural Sciences and Food Industries, Science & Research Branch, Islamic Azad University, Tehran-Iran.
- P.O. Box: 14155/775, Postal Code: 1477893855
- Branch website: www.srbiau.ac.ir
- e-mail addresses:
- hasanzadehr@srbiau.ac.ir
- hasanzadehr@yahoo.com

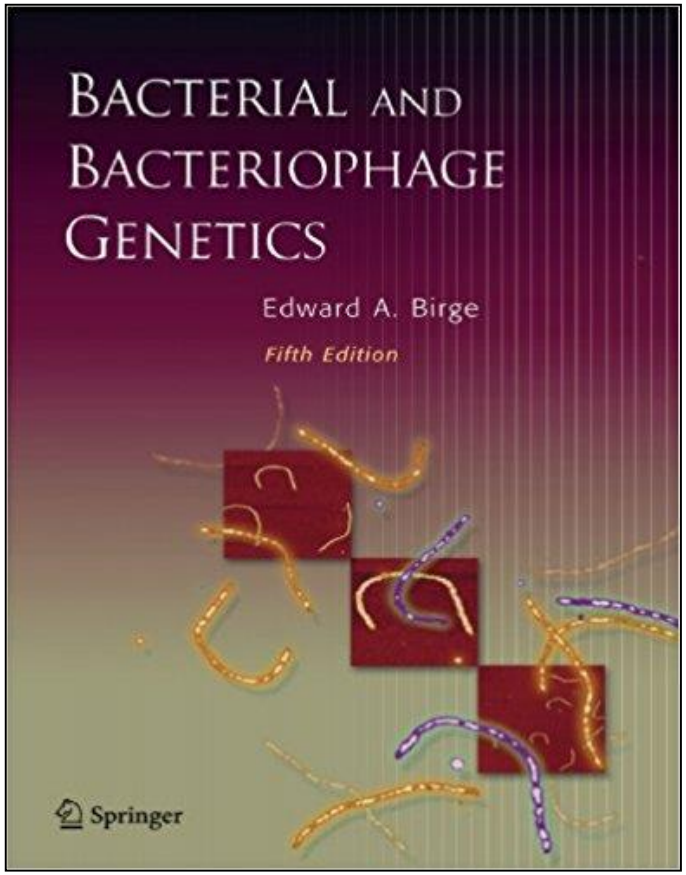
Table of Contents

Table of Contents

Table of Contents

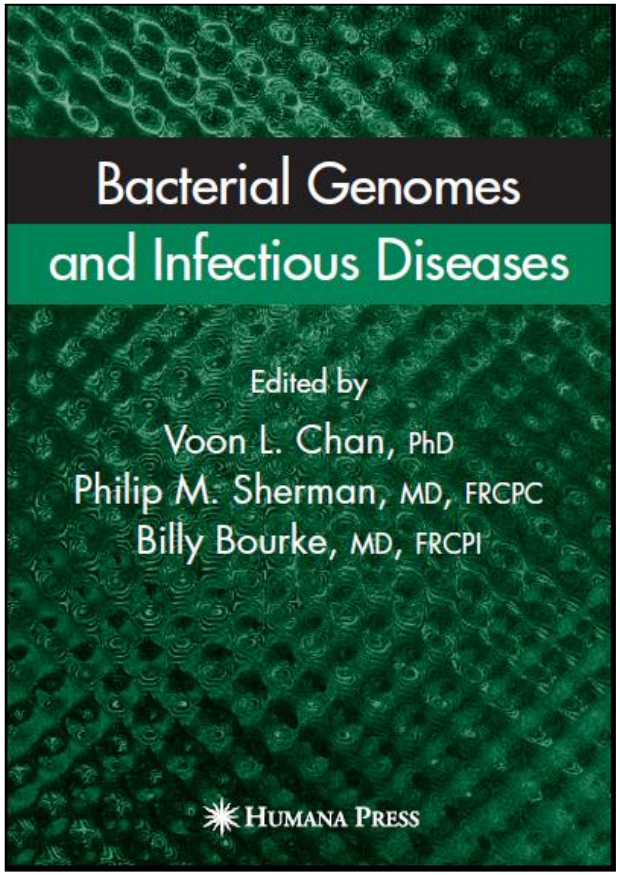
Bacterial and Bacteriophage Genetics

- **Bacterial and Bacteriophage Genetics**
- By Edward Birge
- Springer; Softcover reprint of hardcover 5th ed. 2006 edition (November 23, 2009)
- 578 pp.



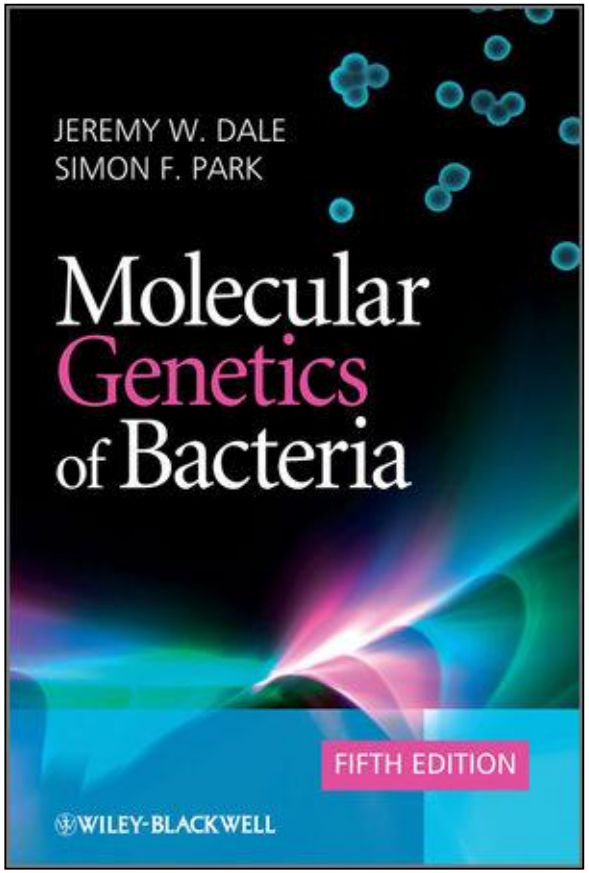
Bacterial Genomes and Infectious Diseases

- **Bacterial Genomes and Infectious Diseases**
- Voon L. Chan, Philip M. Sherman and Billy Bourke (eds.)
- 2006
- Humana Press Inc.
- 270 pp.



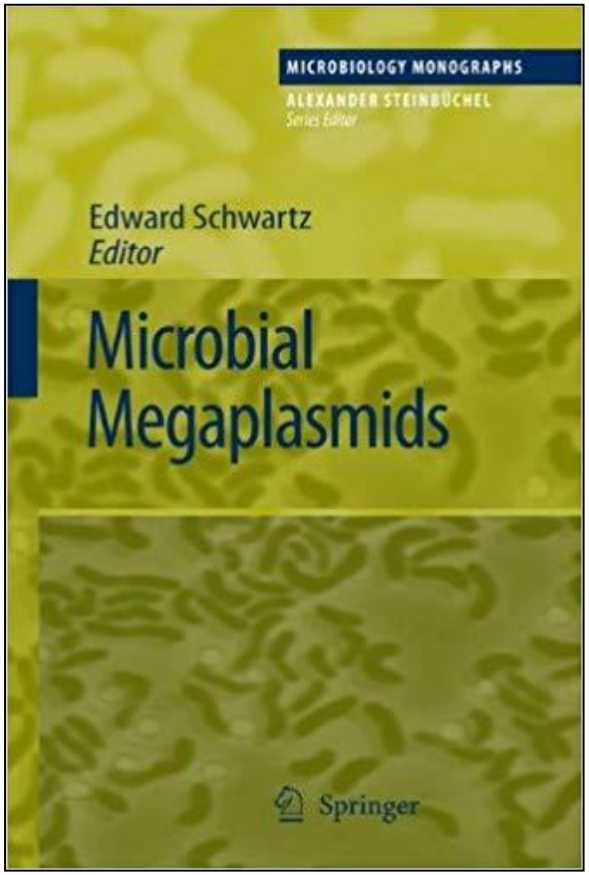
Molecular Genetics of Bacteria

- **Molecular Genetics of Bacteria**
- J. W. Dale and S. F. Park
- 2010
- John Wiley & Sons Ltd,
- Fifth Edition
- 400 pp.



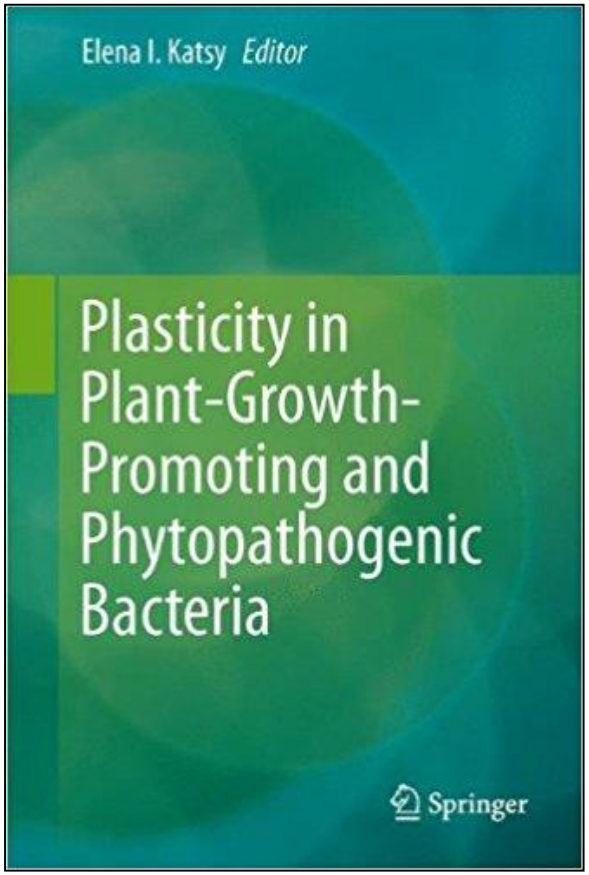
Microbial Megaplasmids

- **Microbial Megaplasmids**
- Edward Schwartz (Editor)
- Springer
- 2010
- 348 pp.



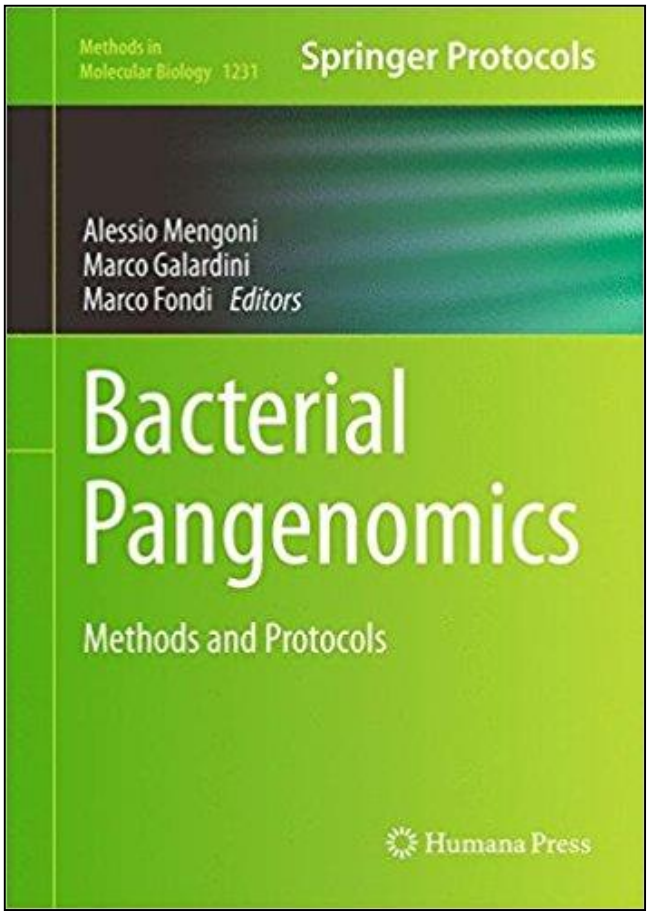
Plasticity in Plant-Growth-Promoting and Phytopathogenic Bacteria

- **Plasticity in Plant-Growth-Promoting and Phytopathogenic Bacteria.**
- by Elena I. Katsy
- Springer;
- 2014
- 208 pp.



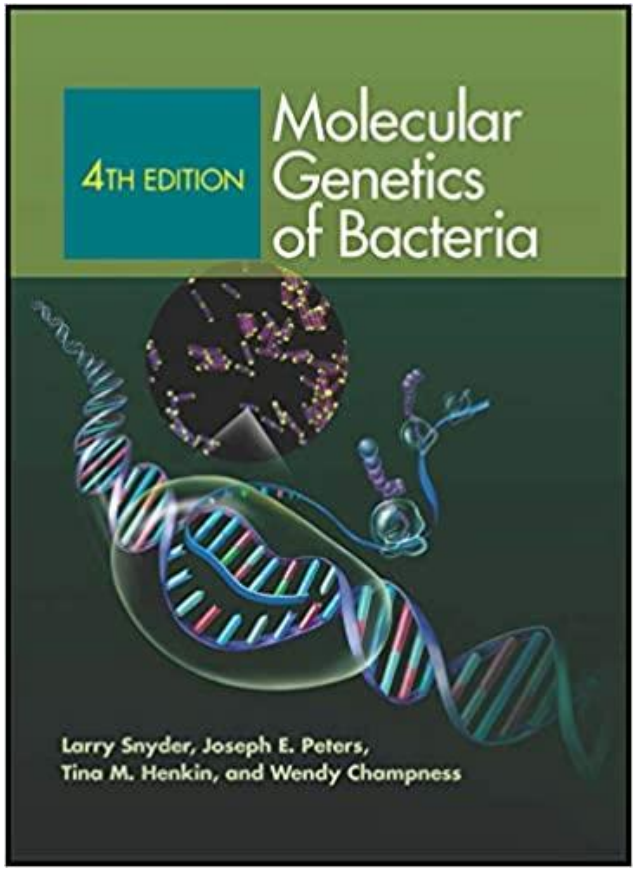
Bacterial Pangenomics: Methods and Protocols (Methods in Molecular Biology)

- **Bacterial Pangenomics: Methods and Protocols (Methods in Molecular Biology)**
- By Alessio Mengoni, Marco Galardini, Marco Fondi (Editors).
- Humana Press
- 2015
- 298 pp.



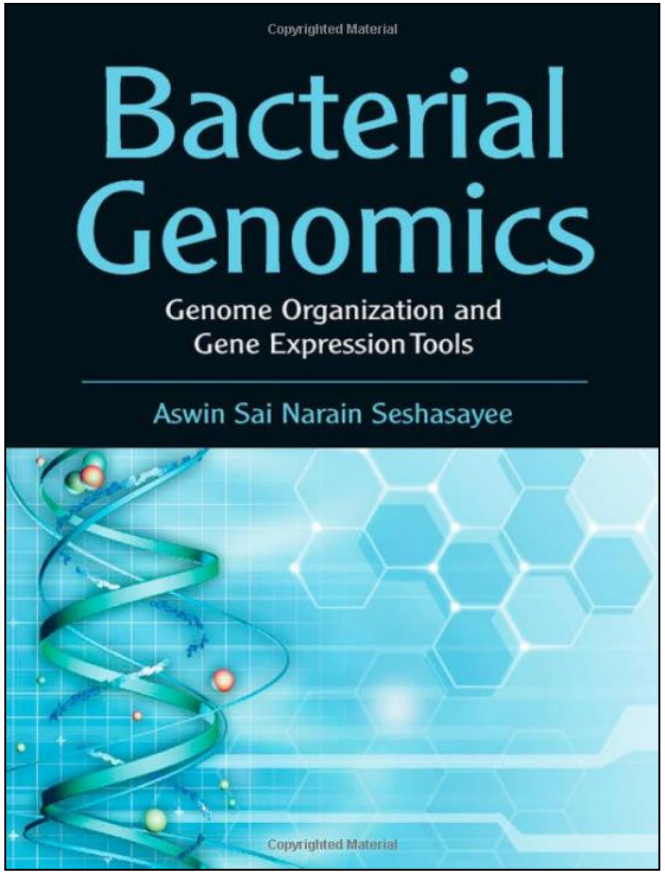
Molecular Genetics of Bacteria

- **Molecular Genetics of Bacteria**
- Larry Snyder, Joseph E. Peters, Tina M. Henkin, Wendy Champness
- ASM Press; 4 edition
- 2014
- 728 pp.



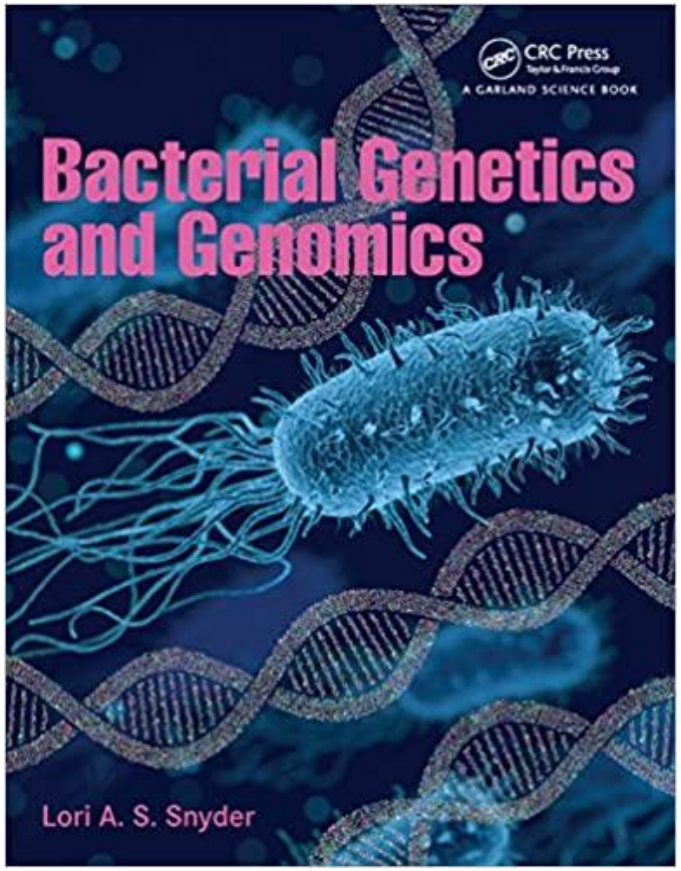
Bacterial Genomics: Genome Organization and Gene Expression Tools

- **Bacterial Genomics: Genome Organization and Gene Expression Tools**
- Aswin Sai Narain Seshasayee
- Cambridge University Press
- 2015
- 230 pp.



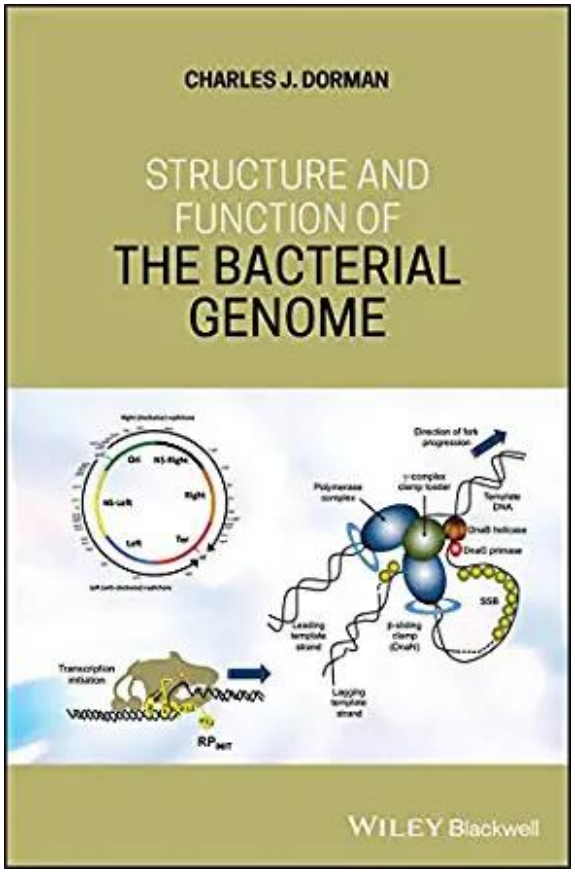
Bacterial Genetics and Genomics

- **Bacterial Genetics and Genomics**
- Lori Snyder
- Publisher: Garland Science
- 2020
- 413 pp.



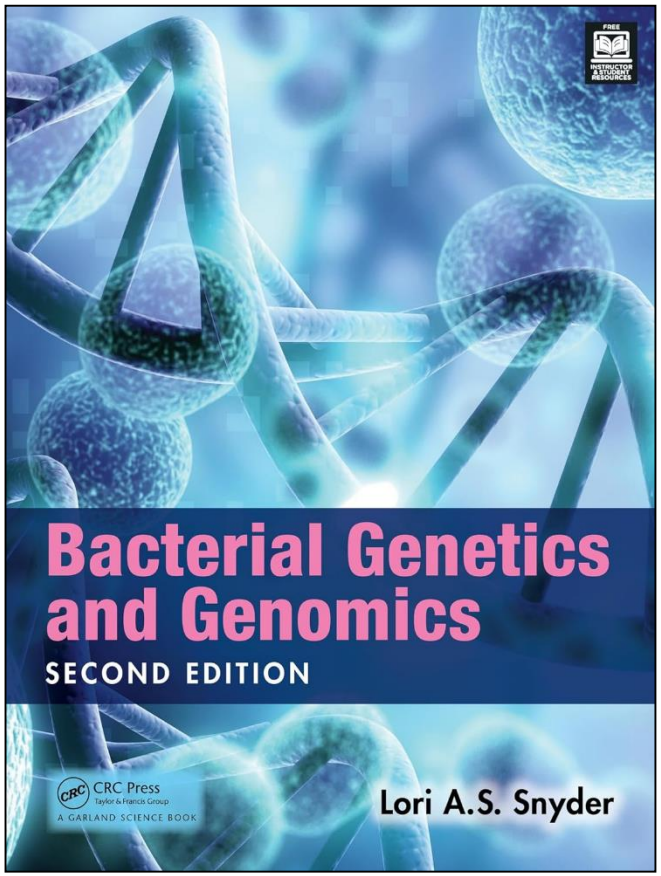
Structure and Function of the Bacterial Genome

- **Structure and Function of the Bacterial Genome**
- Charles J. Dorman
- Publisher: WILEY Blackwell
- 2020



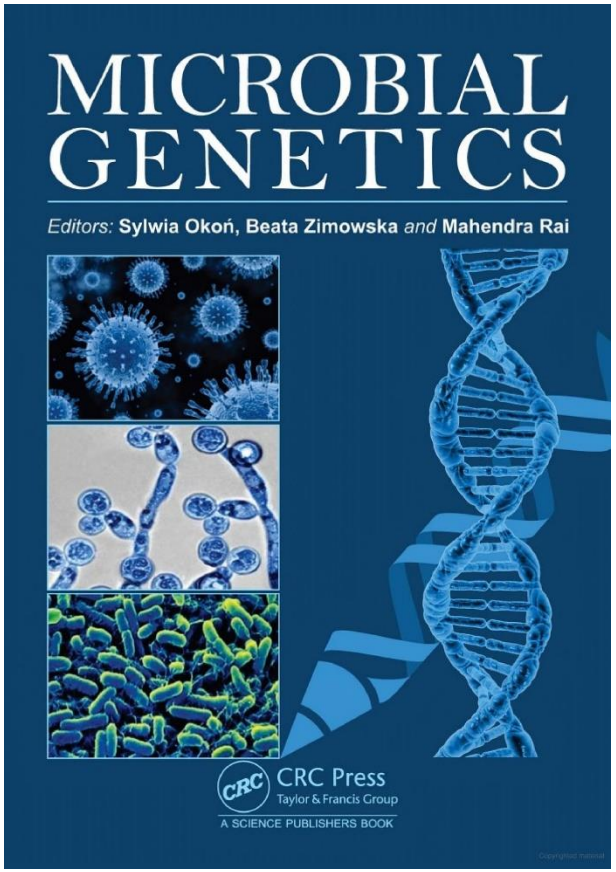
Bacterial Genetics and Genomics

- **Bacterial Genetics and Genomics**
- Lori Snyder and Lori A.S. Snyder
- Garland Science
- 2nd edition
- 2024



Microbial Genetics

- **Microbial Genetics**
- Sylwia Okoń, Beata Zimowska and Mahendra Rai
- CRC Press
- 2024
- 374 pages



PowerPoints, Monograph, PDF files, Lecture Notes

- Medini et al. 2008. Microbiology in the post-genomic era. Nature, 22 slides. (Presenter: Constantin Bode).
- El-Safey Mohmad El-Safey,2011. Introduction of bacterial taxonomy. 73 pages.
- Campbell, Lee Ann, L. Rajagopal and L. Goo. Biochemistry and Genetics of Pathogens and Hosts. PABIO 551. Course website: http://staff.washington.edu/lgoo/PABIO551.shtml. 94 slides.
- Microgen: Microbial Comparative Genomics. 2013. CSD2009-00006 17-12-2009/16-12-2014. 33 slides.
- Sojo,V, A. Pomiankowski and N. Lane.2014. A bioenergetic basis for membrane divergence in archaea and bacteria. PLoS Biol 12 (8): 1-12.
- Kogay, R., Y. I. Wolf and E. V. Koonin. 2024. Defence systems and horizontal gene transfer in bacteria. Environ Microbiol. 26:e16630.

Bacterial genome
DNA and plasmids

- Bacterial DNA can be located in two places:

1. Bacterial chromosome, located in the irregularly shaped region known as the nucleoid (DNA in the bacterial cell is generally confined to this central region);
2. extrachromosomal DNA molecule, located outside of the nucleoid region as circular or linear plasmids. Plasmids usually non-essential genes.

- Most bacteria have a genome that consists of a single DNA molecule (i.e., one chromosome) that is several million base pairs in size and is "circular".

The bacterial chromosome is one long, single molecule of double stranded, helical, supercoiled DNA.

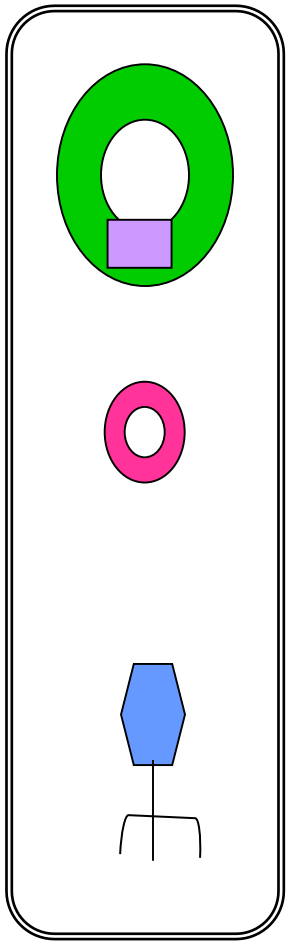
Genetic information in bacteria

Chromosome — Corresponds to the nucleoid

Plasmid — Extrachromosomal genetic material in the cytoplasm.

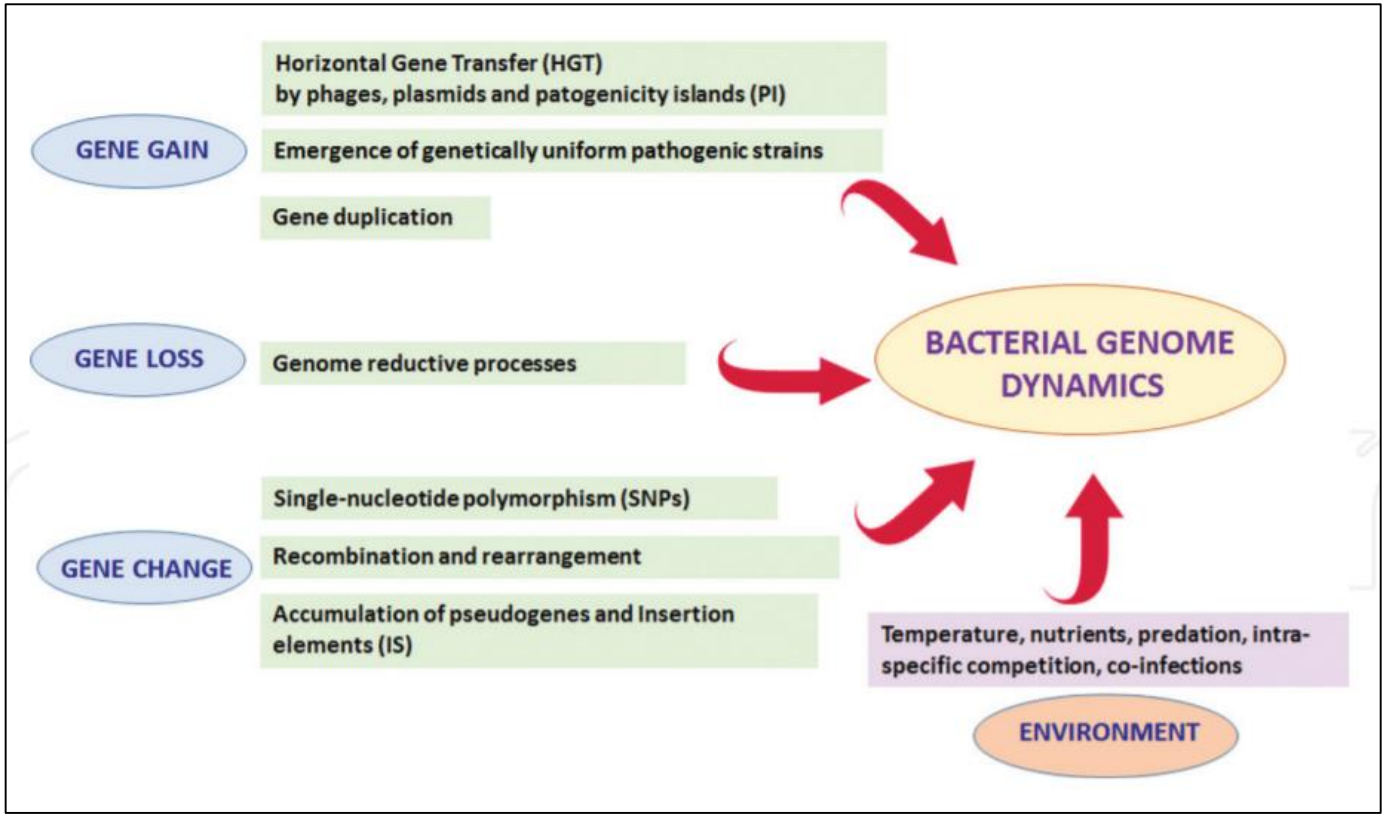
Replicate independently

Bacteriophage — Virus infecting bacteria



Genetic information in bacteria

Bacterial genome dynamics

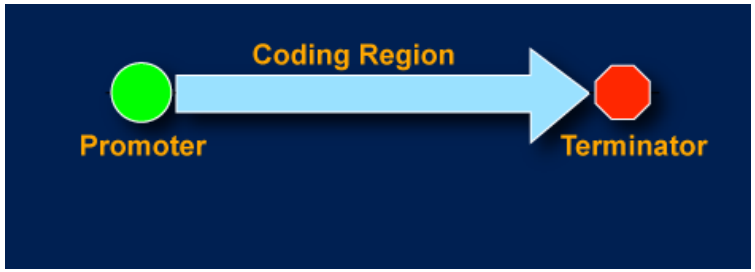


Gene gain, gene loss, gene change, and the environment are the main factors influencing on bacterial genome dynamics.

Gene
Definition

- A gene is one part of a genome.
- A genome is the collection of all the genes in a single organism.
- Each gene can be broken down into important parts:

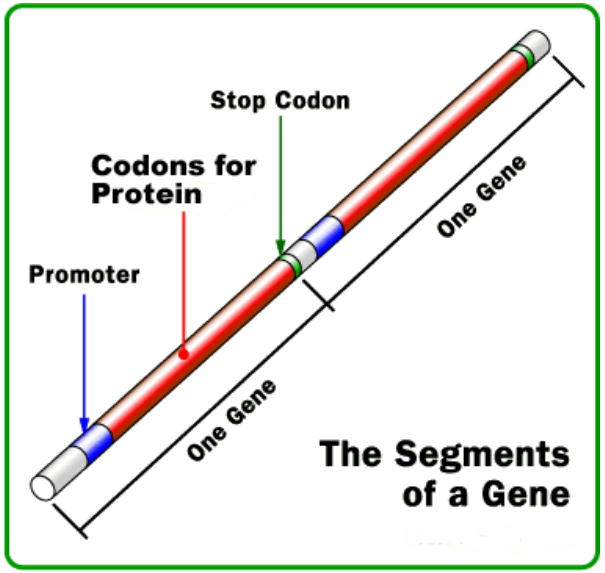
1. A promoter,
2. coding region, and
3. terminator.

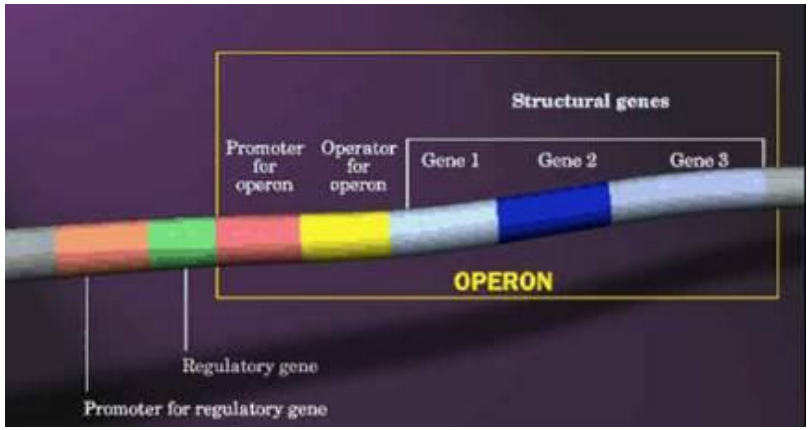


Coding sequence or CDS: portion of a gene's DNA or RNA that codes for protein. While the ORF may contain introns as well.

Gene
Promoter region

- Upstream region of a gene, which regulates when/where the gene is to be transcribed.
- Essentially the promoter region acts as the "on switch" for the gene.

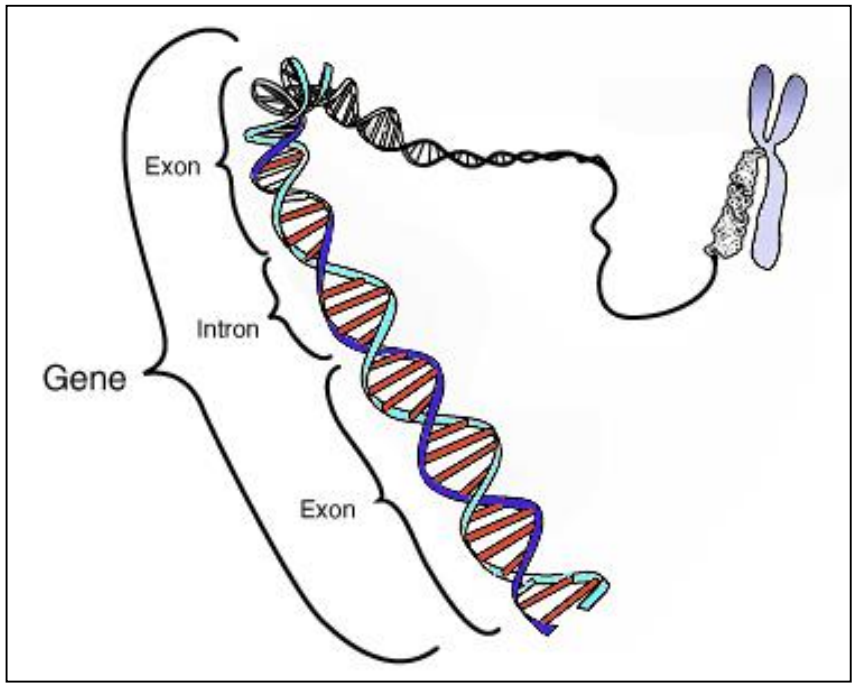




DNA structure

In bacteria exons (protein-coding regions) are present but not introns (non-coding sequences)

- Prokaryotes don't have introns - Genes in prokaryotes are continuous.
- Introns and exons are parts of genes.
- Exons code for proteins, whereas introns do not.
- Introns are regions often found in eukaryote but not prokaryote.
- In prokaryote only the exons encode the proteins.



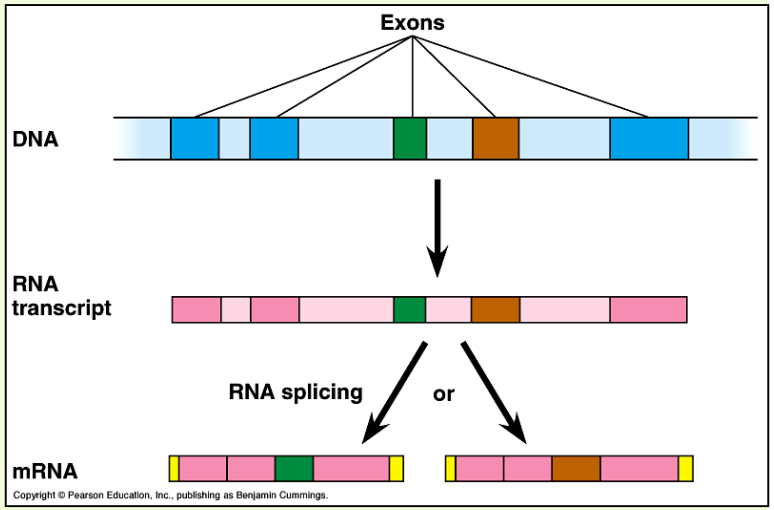
Representation of intron and exons within a simple gene containing a single intron.

Bacteria are tightly packed together; virtually all the DNA encodes proteins.

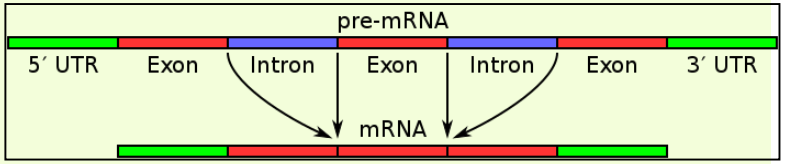
RNA splicing

Excision of the intron is accompanied by the precise ligation of the coding regions (exons)

- Sequences that are joined together in the final mature RNA (mRNA) after RNA splicing are exons.



- Simple illustration of an unspliced mRNA precursor, with two introns and three exons (top). After the introns have been removed via splicing, the mature mRNA sequence is ready for translation (bottom)

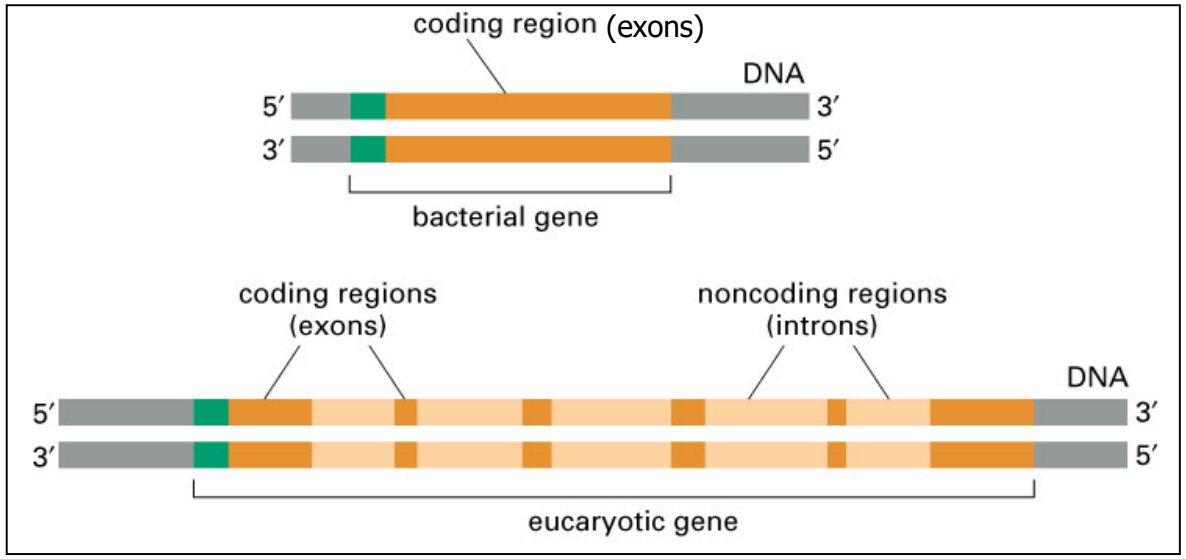


UTR = untranslated regions

Bacterial DNA

Bacterial genes don't have introns

- Intron (non-coding intervening sequence) and Exon (coding, or expressed, sequences) are found in eukaryotes.
- Exon (coding, or expressed, sequences) are found in bacteria.



Bjorkman,2011

Genetics

Prokaryotes vs Eukaryotes Genetics

Prokaryotes	Multicellular Eukaryotes
• Rapid responses to environment	• Stable internal environment
• Limited RNA processing	• Nutrient storage
• Coupled transcription & translation	• Extensive transcript processing
• Fast doubling times during exponential growth	• Relatively constant reproductive rate
• Small, compact genomes	• Few constraints on genome size
• Periodically-selected functions lost & regained through horizontal transfer	• New functions arise through gene duplication, junk recruitment
⇒ *Reductive evolution* ⇒ *Ancestral biochemistry replaced*	⇒ *Complexity, cell specialisation* ⇒ *Ancestral biochemistry conserved*

The size of the smallest of cells is constrained by the minimum amount of genetic material need to keep the cell going. In genetics, the term junk DNA refers to regions of DNA that are non-coding. Recruitment (hiring)junk DNA.

Genetics

Prokaryotes vs Eukaryotes Genetics

	Prokaryotes	Eukaryotes
DNA	DNA is naked	DNA bound to protein
	DNA is circular	DNA is linear
	Usually no introns	Usually has introns
Organelles	No nucleus	Has a nucleus
	No membrane-bound	Membrane-bound
	70S ribosomes	80S ribosomes
Reproduction	Binary fission	Mitosis and meiosis
	Single chromosome (haploid)	Chromosomes paired (diploid or more)
Average Size	Smaller (~1–5 μm)	Larger (~10–100 μm)

Genetics

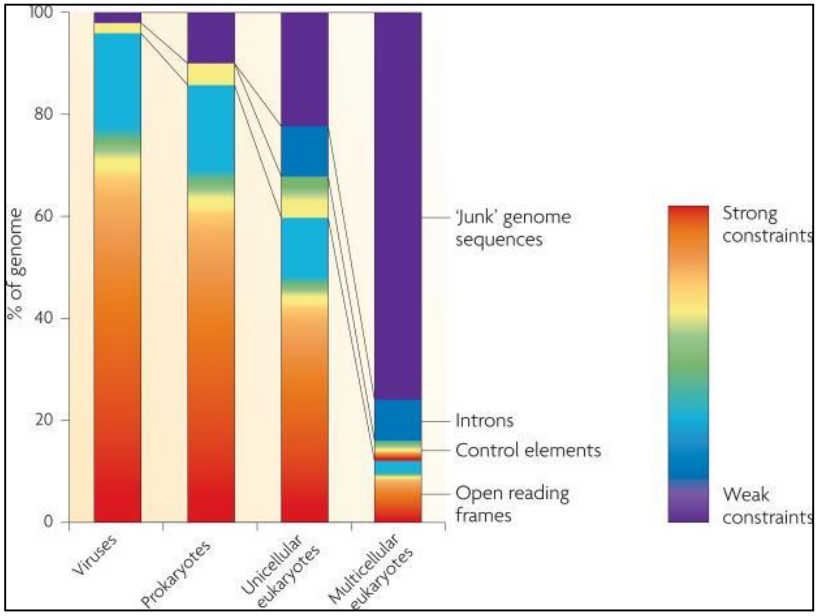
Prokaryotes vs Eukaryotes Genetics

	Bacteria	Archaea	Eukarya
Genome size	Most are 1-6 Mb		Most are 10-4,000 Mb, but few are much larger. Plants and animals have genomes greater than 100 Mb; humans have 3,000 Mb
Number of genes	1,500-5,700		5,000-40,000 genes. Unicellular fungi have about 5,000 genes. Multicellular eukaryotes can have up to at least 40,000 genes
Gene density	Higher than in eukaryotes		Lower than in prokaryotes (within eukaryotes, lower density is correlated with larger genomes)
Introns	None in protein-coding genes	Present in some genes	Unicellular eukaryotes: present, but prevalent only in some species. Multicellular eukaryotes: Present in most genes.
Other noncoding DNA	Very little		Can be large amounts: generally more repetitive noncoding DNA in multicellular eukaryotes

Constraints and plasticity in genome and molecular-phenome evolution

Approximate distribution of evolutionary constraints across genomes with different architectures

- Multiple constraints variously affect different parts of the genomes of diverse life forms.
- The constraints that shape the evolution of genomes and phenomes are complemented by the plasticity and robustness (hardy) of genome architecture, expression and regulation.



In eukaryotes few constraints on genome size.
In eukaryotes new function arise through duplication and junk DNA recruitment (hiring).
The term junk DNA refers to regions of DNA that are non-coding.
In prokaryotes the size of the smallest of cells is constrained by the minimum amount of genetic material need to keep the cell going.

Koonin and Wolf,2012;..

Gene Expression

Environmental influences on gene expression

- The expression of genes in an organism can be influenced by the environment, including the external world in which the organism is located or develops, as well as the organism's internal world, which includes such factors as its hormones and metabolism.
- One major internal environmental influence that affects gene expression is gender, as is the case with sex-influenced and sex-limited traits.
- Similarly, drugs, chemicals, temperature, and light are among the external environmental factors that can determine which genes are turned on and off, thereby influencing the way an organism develops and functions.

Lobo,2008

Genetics

Importance of study on microbial genetics

- Simple systems
  - studying genetic phenomena
- Useful tools
  - decipher the genetics mechanisms
- Molecular cloning
  1. isolate and duplicate specific genes from other organisms.
  2. genes are manipulated and placed in a microorganism where they can be induced to increase in numbers.

Genetics

Importance of study on microbial genetics

- Value in industry
  - Antibiotics
  - Increase yields and improve manufacturing processes
- Diseases
  - Understanding the genetics of disease-causing microorganisms
- Genetic transfer in prokaryotes
  - How genes can be transferred from one organism to another, even from one species to another.

DNA structure

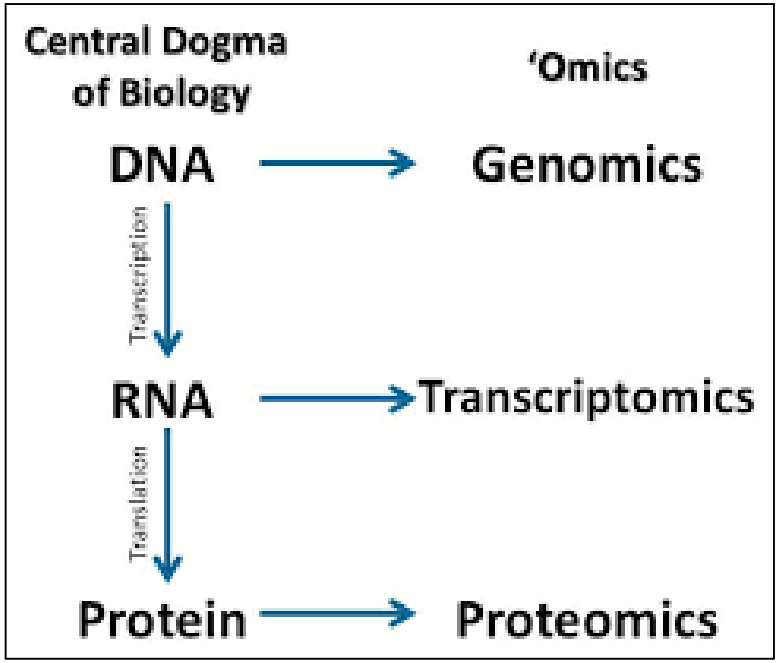
Prokaryotic vs. Eukaryotic genetics

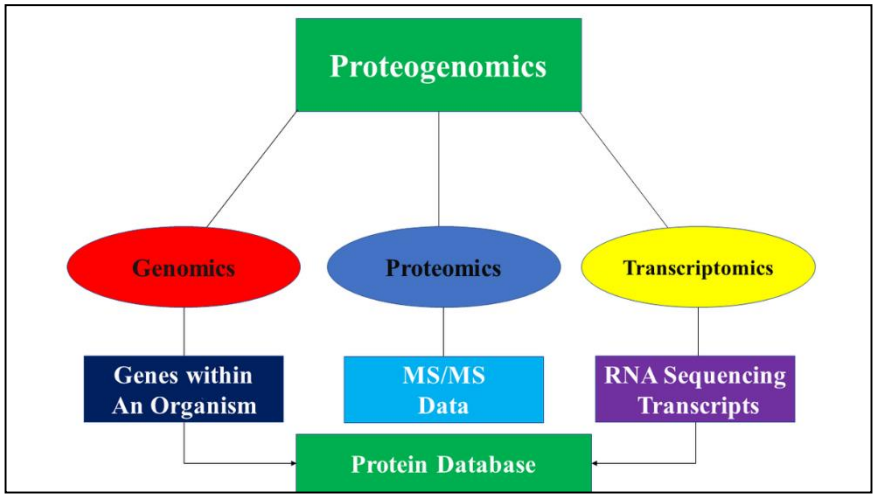
- Bacterial genetics are different.
- Prokaryote genes are grouped in operons.
- Prokaryotes have one type of RNA polymerase for all types of RNA.
- mRNA is not modified (no pre-mRNA),
- The existence of introns in prokaryotes is extremely rare.

DNA structure

Central Dogma of Biology: Protein Synthesis

Genomics, transcriptomics and proteomics





Transcriptomics is the study of transcriptomes, the complete set of RNA transcripts produced by the genome at any one time.

DNA structure
Prokaryotic vs. Eukaryotic genetics

- To initiate transcription in bacteria, sigma factors bind to RNA polymerases. RNA polymerases/sigma factors complex can then bind to promoter about 40 deoxyribonucleotide bases prior to the coding region of the gene.
- In prokaryotes, the newly synthesized mRNA is polycistronic (polygenic) (code for more than one polypeptide chain).
- In prokaryotes, transcription of a gene and translation of the resulting mRNA occur simultaneously. So many polysomes are found associated with an active gene.

During transcription, RNA polymerase makes a copy of a gene from the DNA to mRNA as needed.

Transcription

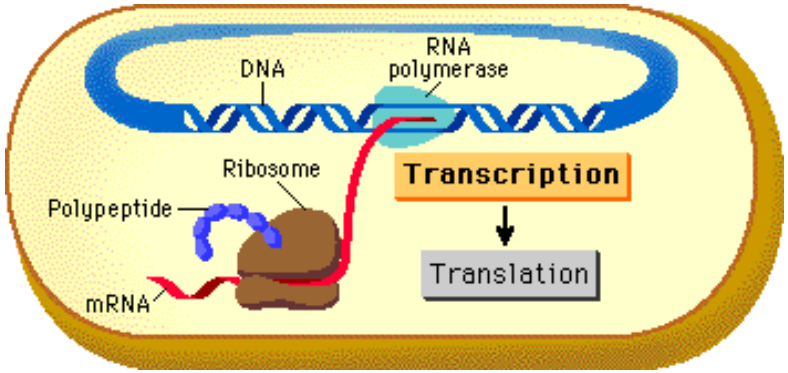
Prokaryotic vs. Eukaryotic transcription

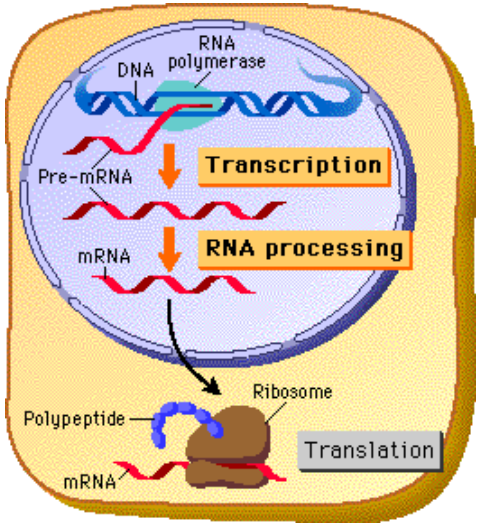
Prokaryotic organisms	Eukaryotic organisms
Lack nucleus	Contain nucleus
DNA is found in the cytoplasm	DNA is confined to the nuclear compartment
RNA transcription and protein formation occur almost simultaneously	RNA transcription occurs prior to protein formation, and it takes place in the nucleus. Translation of RNA to protein occurs in the cytoplasm.
Gene expression is regulated primarily at the transcriptional level	Gene expression is regulated at many levels (epigenetic, transcriptional, nuclear shuttling, post-transcriptional, translational, and post-translational)

Epigenetics is the study of mechanisms that switch genes on or off.
Nuclear shuttle (travel regularly between two or more places): Nuclear pore complexes are tiny channels where the exchange of substances between the cell nucleus and the cytoplasm takes place. RNA-binding protein that shuttles in and out of the nucleus.

Transcription

Prokaryotic vs. Eukaryotic transcription





In a prokaryotic cell, transcription and translation are coupled; that is, translation begins while the mRNA is still being synthesized.

This is because there is no nucleus to separate the processes of transcription and translation.

In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs in the cytoplasm.

This is because transcription occurs in the nucleus to produce a pre-mRNA molecule.

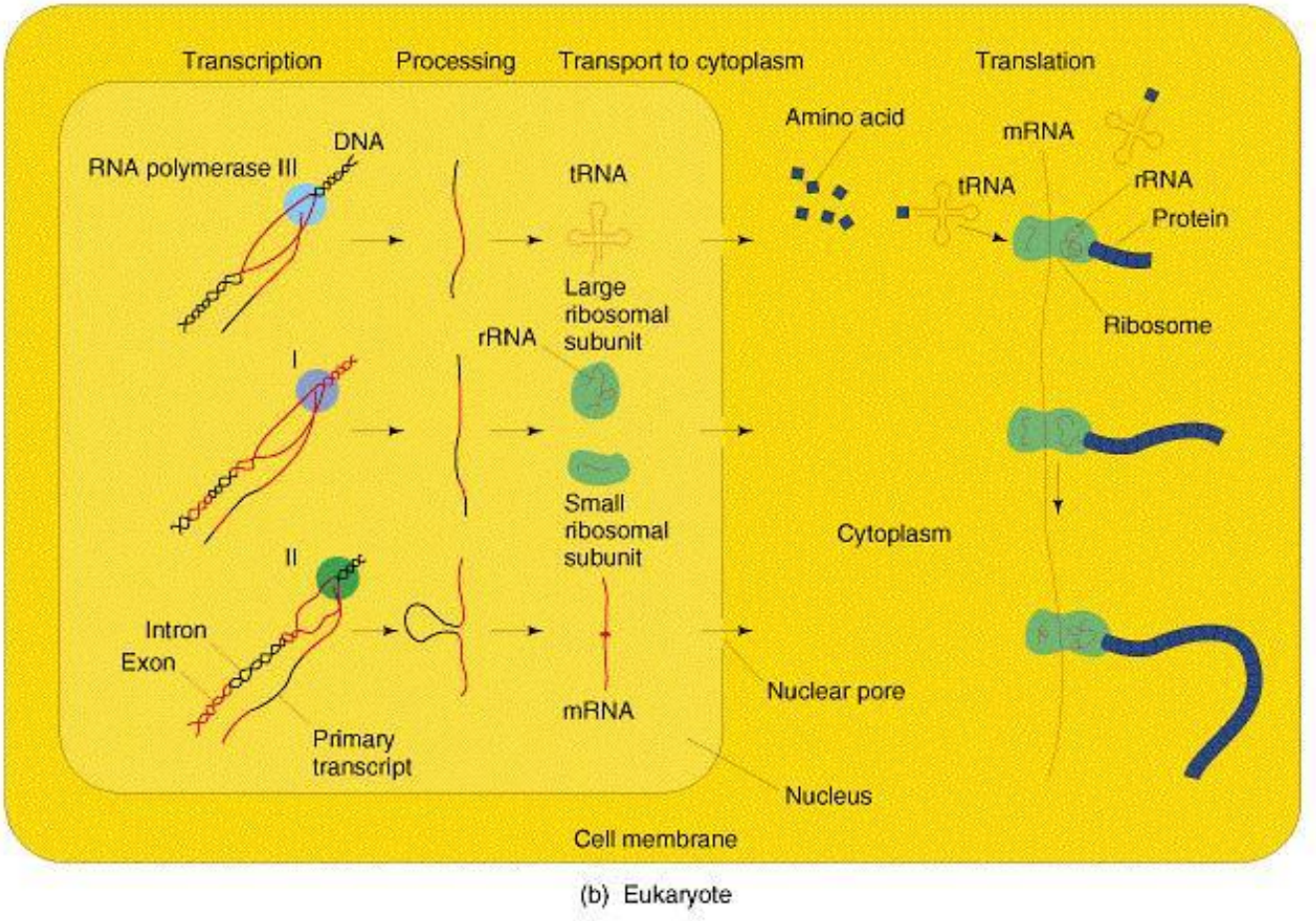
Pearson Education

Transcription

Prokaryotic vs. Eukaryotic transcription

In Eukaryotes, the process is a bit more complicated

Type of Polymerase	Product
RNA Polymerase I	rRNA
RNA Polymerase II	hnRNA (heterogeneous nuclear RNA)
RNA Polymerase III	tRNA



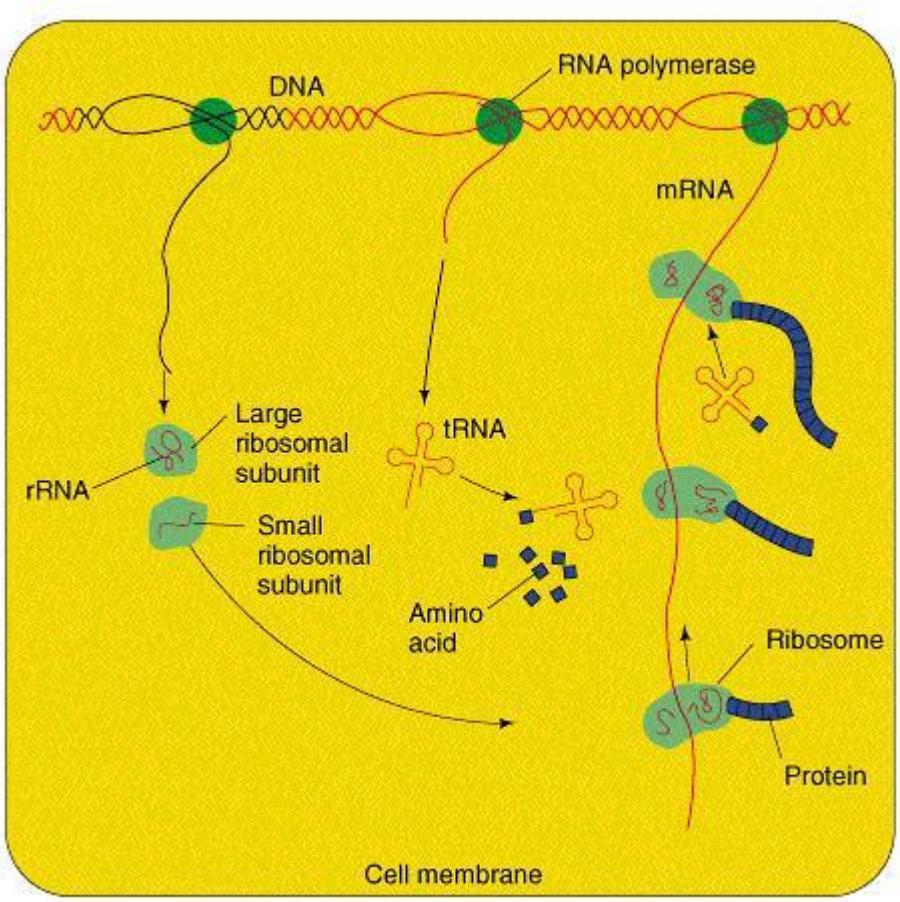
(b) Eukaryote

Biology 210 GENETICS,1998

Transcription

Prokaryotic vs. Eukaryotic transcription

In bacteria, transcription and translation take place at the same time…



(a) Prokaryote

Biology 210 GENETICS,1998

Biochemistry

Comparative biochemistry of archaea and bacteria

- Reconstructing the traits of the last universal common ancestor (LUCA) requires constraining the relationships between the three domains of life,

1. archaea,
2. bacteria, and
3. eukaryotes.

- Recent phylogenetic studies show that eukaryotes are secondarily derived.
- They (eukaryotes) are genomic chimeras (an organism containing a mixture of genetically different tissues), arising from an endosymbiosis between a bacterium and an archaeal host cell.

Sojo *et al.*,2014

Biochemistry

Comparative biochemistry of archaea and bacteria

- Archaea and bacteria share core biochemistry, including:

1. the genetic code,
2. transcription machinery, and
3. ribosomal translation.

Sojo *et al.*,2014

Genes Types

Patterns of gene expression

- **The seventeen important types of genes are:**
- House Keeping Genes
- Non-constitutive Genes
- Inducible Genes
- Repressible Genes
- Multi-genes (A group of genes that produces a specific phenotype or trait only when expressed together)
- Repeated Genes (e.g. ATTCG ATTCG ATTCG in which the sequence ATTCG is repeated three times)
- Single copy Genes (present in a genome in only one copy vs. satellite DNA)
- Pseudo-genes (dysfunctional genes lost their gene expression in the cell)
- Processed Genes (processed pseudogenes are genes which have been moved from their original location).
- Split Genes
- Transposons
- Overlapping Genes (genes having nucleotide sequences that may overlap in a way that involves control genes or structural genes).
- Structural Genes
- Regulator Genes
- Operator Genes
- Promoter Genes, and
- Terminator Genes.

BiologyDiscussion.com

Genes Types

Patterns of gene expression/regulation

- **Constitutive genes (housekeeping genes)**:
- A gene that is transcribed at a relatively constant level regardless of the cell environmental conditions.
- These are essentially unregulated and **encode proteins that are essential at all times**.
- always turned ON.
- e.g.
- 16s rRNA;
- tRNA genes.

The 16S rRNA is the RNA component of the small subunit (30S subunit) of the ribosome. The genes coding for it are referred to as 16S rRNA gene.

Genes Types

Patterns of gene expression/regulation

- **Facultative genes (regulated gene)**:
- Genes that are transcribed only when needed.
- Turned on/off in response to change in the cell's environment
- e.g., bacterial virulence genes.

- **Inducible genes**:
- Normally "off", switched on when needed.
- A gene whose expression is either responsive to environmental change or dependent on the position of the cell cycle
- e.g., inducible antibiotic resistance genes.

Genes Types

Patterns of gene expression/regulation

- **Repressible genes:**
  - normally on, but can be turned off when the end product is abundant
  - Common for anabolic (biosynthesis) genes

- **Terminative genes:**
  - In genetics, a transcription terminator is a section of nucleic acid sequence that marks the end of a gene or operon in genomic DNA during transcription.

Repetitive genes

Not all of the base sequences in DNA are translated

- The large size of eukaryotic genomes is that some genes are repeated many times.

1. Highly repetitive base sequences are not translated.
2. They consist of sequences of between 5 and 300 bases that may be repeated up to 10 000 times.
3. They constitute 5-45% of eukaryotic DNA.

- Single-copy genes or unique genes are translated and constitute a surprisingly small proportion of eukaryotic DNA.

Nonrepetitive genes

- Approximately one-third of the genome corresponds to highly repetitive sequences (in eukaryotic).
- Leaving an estimated 2×10^9 base pairs for:

1. functional genes,
2. pseudogenes, and
3. nonrepetitive spacer sequences.

- Whereas most prokaryotic genes are represented only once in the genome, many eukaryotic genes are present in multiple copies, called gene families.

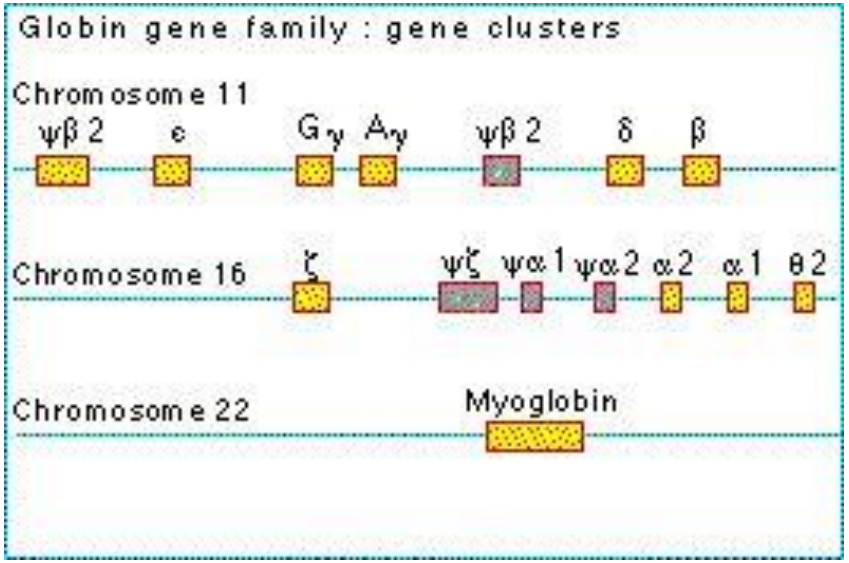
Pseudogenes are dysfunctional genes that have lost their gene expression in the cell or their ability to code protein. This a common phenomenon in eukaryotes and a few bacteria.

Gene family

Gene family vs gene cluster

What is a gene family?

- A gene family is a set of several similar genes, formed by duplication of a single original gene, that generally have similar biochemical functions.
- E.g.
- One such family are the genes for human haemoglobin subunits.
- The 10 genes are in two clusters on different chromosomes(16 and 22), called the α-globin and β-globin loci.

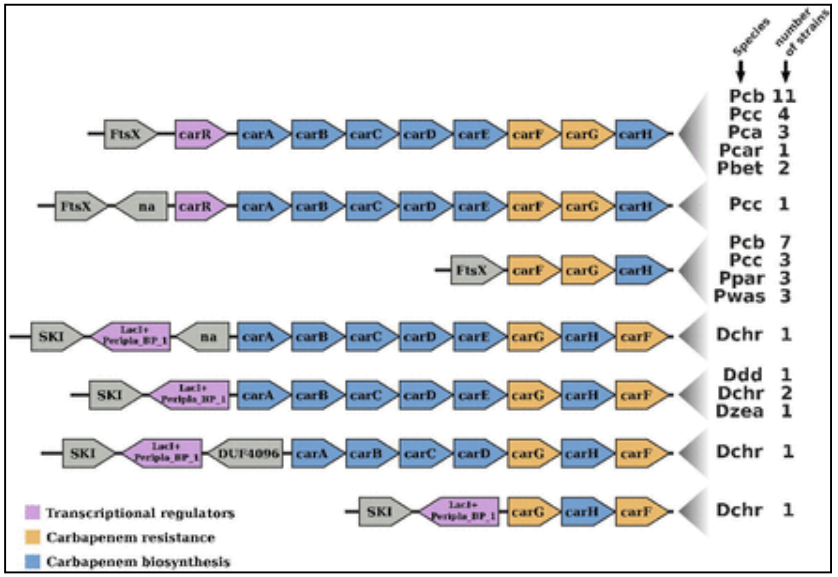


Gene family

Gene family vs gene cluster

What is a gene cluster?

- Gene clusters are the genetic building blocks of bacteria and archaea.
- Prokaryotic genomes are highly organized and the genes associated with a particular function often occur near each other (Fischbach *et al*.,2008).



***car* gene cluster in Soft-Rot-Enterobacteriacea (SRE):**
A cluster of nine carbapenem (car) genes has been identified on the chromosome of *P. carotovora*.
Carbapenem is a class of very effective antibiotic agents most commonly used for the treatment of severe bacterial infections.

Yutefar Shyntum *et al*.,2018

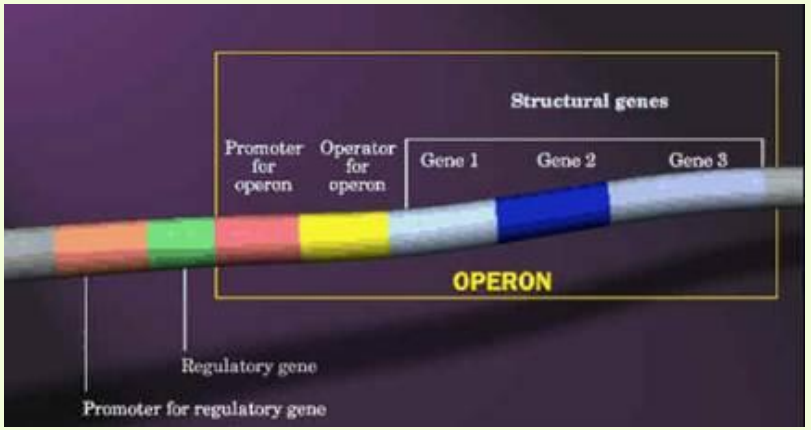
Gene family

Gene operon vs gene cluster

What is the difference between a gene cluster and an operon?

- A gene is one part of a genome. Each gene can be broken down into important parts:

1. A promoter,
2. coding region, and
3. terminator.



The coordinated transcription and translation of the lac operon structural genes is supported by a shared promoter, operator, and terminator.

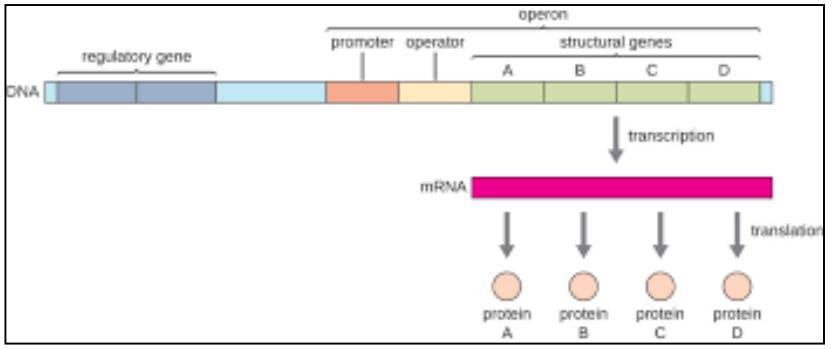
Nature Education,2014;..

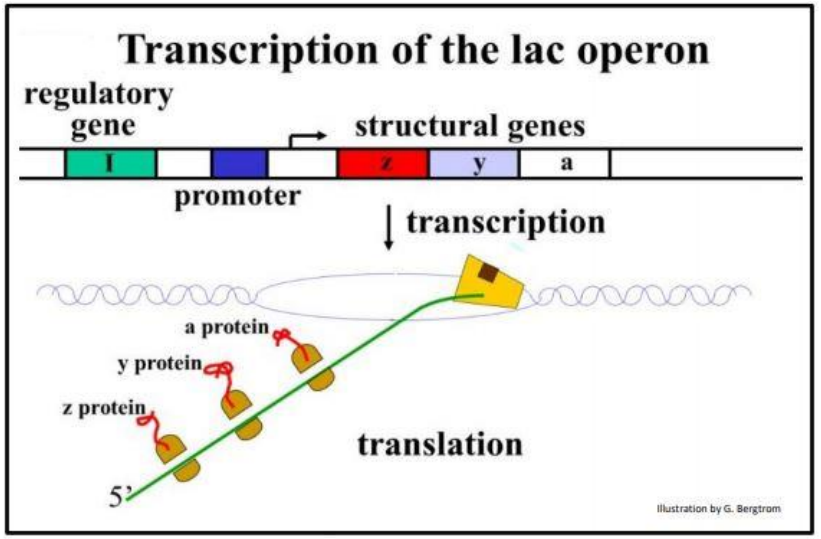
Gene family

Gene operon vs gene cluster

What is the difference between a gene cluster and an operon?

1. A gene cluster as a set of functionally related genes located in close physical proximity in a genome.
2. The term operon refers to a set of genes under common regulatory control, that are transcribed into a single mRNA and are all co-directional in orientation on the chromosome. E.g. lac operon. Three lactose metabolism genes (lacZ, lacY, and lacA) are organized together in a cluster called the lac operon.





Lumen waymaker; Biology LibreTexts

Gene family

A gene family is a group of genes that are related in structure and often in function

- A gene family is a set of homologous genes (from a few genes to several hundred genes) within one organism.
- A gene cluster is part of a gene family.
- A gene cluster is a group of two or more genes found within an organism's DNA that encode for similar polypeptides, or proteins, which collectively share a generalized function and are often located within a few thousand base pairs of each other.
- In bacteria, gene clusters may be similar to an operon in which all genes are controlled by a single promoter and operator.

Wikipedia,2022

Gene family

What is the significance of the gene family?

- Classifying individual genes into families helps researchers describe how genes are related to each other.
- Classification systems for genes continue to evolve as scientists learn more about:

1. the structure and function of genes, and
2. the relationships between them.
3. Additionally, gene families may provide clues for identifying genes that are involved in particular diseases.

Genetics Home Reference,2015

Gene family

A group of genes that are related in structure and often in function

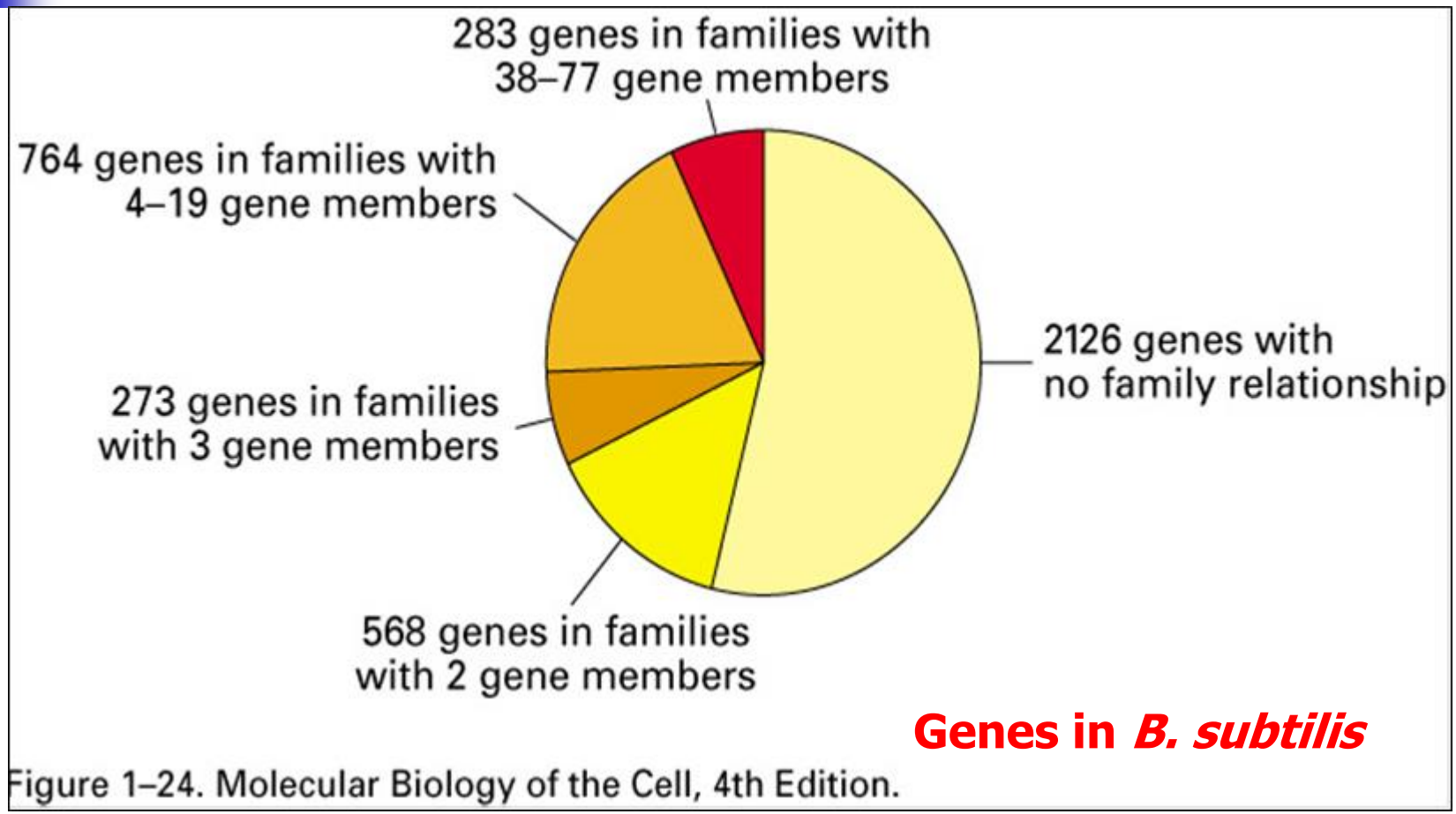
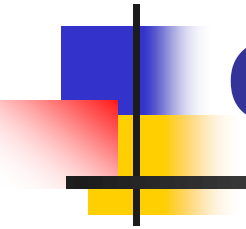


Figure 1–24. Molecular Biology of the Cell, 4th Edition.

Campbell and co-workers



Genomes

Structure and diversity of microbial genomes

Genomes

Structure and diversity of microbial genomes

Bacteria and Archaea

- The structure of a microbial genome can be analyzed and compared with other genomes at various levels, including:

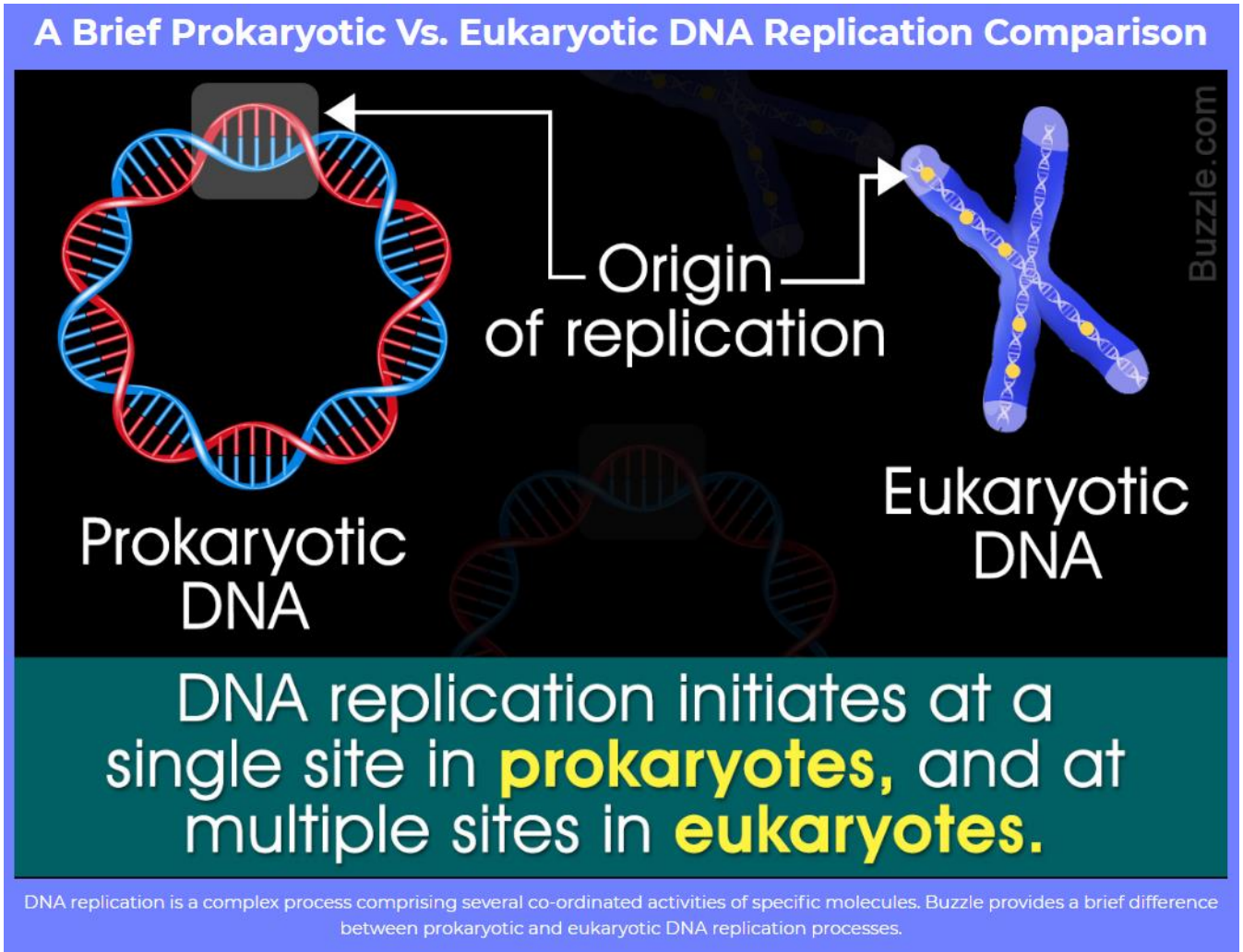
1. Guanine and cytosine (G+C) content,
2. Overall size, topology,
3. Organization of genes, and
4. The presence and abundance of accessory and mobile genetic elements.

Chan *et al.*,2006

Genomes

Origins of replication"*ori*"

Bacteria, Archaea and Eukaryotes



Genomes

Origins of replication "*ori*"

Bacteria, Archaea and Eukaryotes

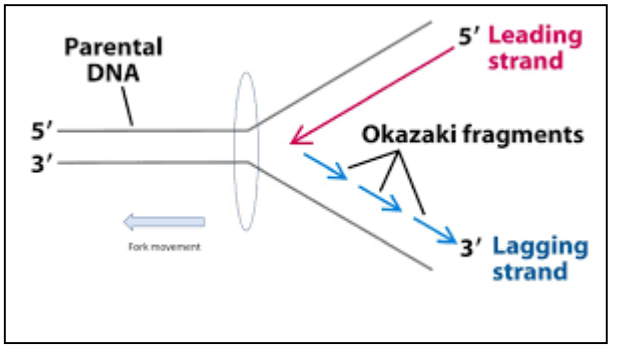
- There are also significant differences between prokaryotic and eukaryotic origins of replication:

1. Most bacteria have a single circular molecule of DNA, and typically only a single origin of replication per circular chromosome.
2. Most archaea have a single circular molecule of DNA, and several origins of replication along this circular chromosome.
3. Eukaryotes often have multiple origins of replication on each linear chromosome that initiate at different times (replication timing), with up to 100,000 present in a single human cell.

Genomes

Origins of replication "*ori*"

Prokaryotes and Eukaryotes



	Prokaryotic DNA Replication	Eukaryotic DNA Replication
Origin	**Prokaryotic DNA replication has a single origin.**	**Eukaryotic DNA replication has multiple origins.**
Time Taken	**Prokaryotic DNA replication takes 40 minutes.**	**Eukaryotic DNA replication may take up to 400 hours or more.**
Takes Place At	**Prokaryotic DNA replication occurs in the cytoplasm.**	**Eukaryotic DNA replication occurs in the nucleus.**
Rate of Base Pairs	**Prokaryotic DNA replication has a rate of 2000 base pairs per second addition.**	**Eukaryotic DNA replication has a rate of 100 base pairs per second.**
Nucleotide	**Prokaryotic DNA replication has 1000-2000 nucleotides long Okazaki fragments.**	**Eukaryotic DNA replication has 100-200 nucleotides long Okazaki fragments.**

Difference.Wiki

Genomes

Origins of replication"*ori*"

Bacteria, Archaea and Eukaryotes

- Origins of replication in all kingdoms share common characteristics such as high AT content (repeats of adenine and thymine are easier to separate because their base stacking interactions are not as strong as those of guanine and cytosine.
- Origins of replication are typically assigned names containing "*ori*".
- The *E. coli* replication origin is known as oriC.
- OriC consists of three A–T rich 13-mer repeats and four 9-mer repeats.

The plasmid *ori* sequences are similar to *oriC*, and are called *oriV* (origin of vegetative replication).

Genomes

Origins of replication"*ori*"

Bacteria, Archaea and Eukaryotes

- Bacterial chromosome replication is best understood in the well-studied bacteria *Escherichia coli* and *Bacillus subtilis*.
- Chromosome replication proceeds in three major stages: initiation, elongation and termination.
- This process is known as bidirectional replication.

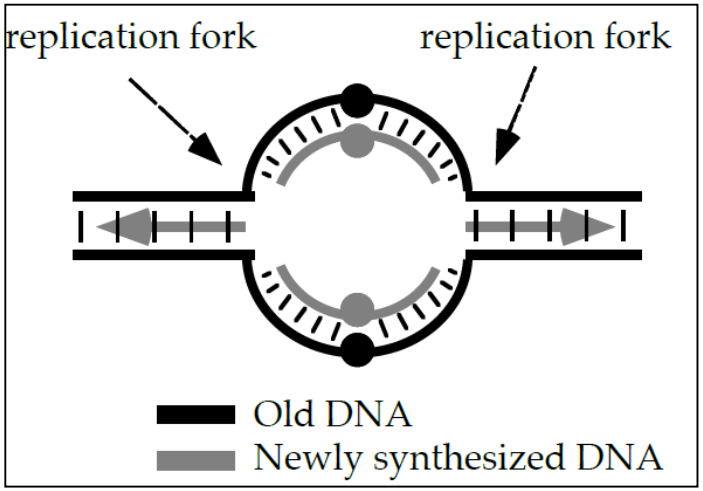
Note: In eukaryotic and prokaryotic cells, DNA replication is bidirectional.
Some plasmids and a few types of phage/viruses possess helicases that permit solely unidirectional replication.
See also the slide Nos. 244 onwards.

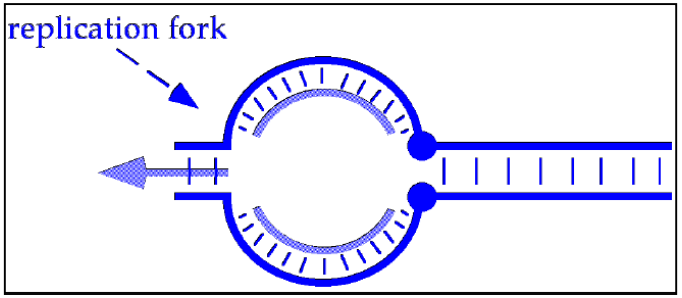
Wikipedia; DNA Replication 2/21/2013;..

Genomes

Origins of replication "*ori*"

Bacteria, Archaea and Eukaryotes





The difference is that the replication enzymes are moving in one direction away from the origin.

Bidirectional Replication

Each strand of the original DNA double helix is acting as a template for the synthesis of new DNA (gray).

Two independent sets of enzymes are moving away from a single origin in opposite directions (gray arrows) and synthesizing new DNA in their wake.

DNA Replication 2/21/2013

Unidirectional Replication

Each strand of the original double helix is still acting as a template for new DNA synthesis.

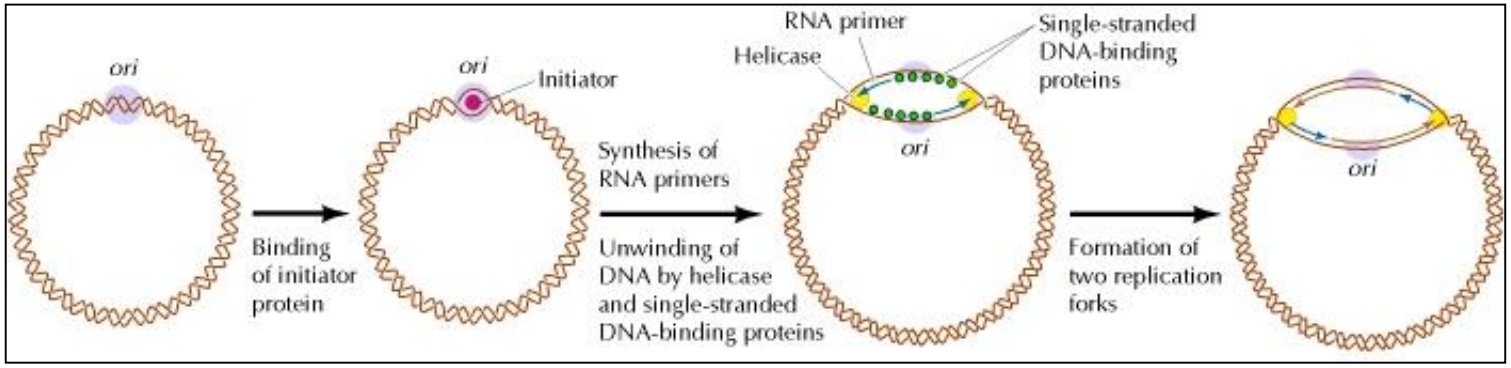
Unidirectional nucleic acid replication rapidly synthesize multiple copies of circular molecules of DNA or RNA, such as plasmids, the genomes of bacteriophages, and the circular RNA genome of viroids.

Genomes

Origins of replication"*ori*"

Bacteria

- The first event is the binding of an initiator protein to ori DNA, which leads to partial unwinding of the template.
- The DNA continues to unwind by the actions of helicase and single-stranded DNA-binding proteins, and RNA primers are synthesized by primase.
- The two replication forks formed at the origin then move in opposite directions along the circular DNA molecule.



Cooper,2000

Pangenome (or supra-genome)
Core genome and Accessory genome

- Definition:
- Pangenome: The bacterial pangenome was introduced in 2005.
- The pan-genome represents the complete gene pool of a species. It is made up of the core genome (genes shared by all strains) and the accessory genome (genes shared by some strains and not all).
- Core genome: The pool of genes common to all the studied genomes of a given species.
- Accessory genome: Not unique but not in the core genome.

Pangenome (or supra-genome)
Core genome and Accessory genome

- The pangenome is the entire gene set of all strains of a species.
- It includes both:

1. a single DNA molecule (i.e., one chromosome), and
2. plasmid DNA.

- The pan-genome of a bacterial species consists of:

1. a core genome which present in all strains, and
2. Flexible genomes genes present only in some strains of a species.

Plasmids are a unique feature of prokaryotes. A plasmid is a small piece of circular DNA which codes for additional genes that are not needed for normal growth or development. Plasmids are easily transferred between cells.

Pan-genome or supra-genome

A collection of all the DNA sequences that occur in a species

- Pan-genome: global gene repertoire of a bacterial species:

Core genome + Dispensable genome

1. Core genome: genes shared by all strains of the same species.
2. Dispensable genome: consisting of partially shared and strain-specific genes, i.e. genes present in some but not all of the same species.

Scientists continue to survey and catalogue the genomic variation across human populations and begin to assemble a human pan-genome. These efforts will increase our power to connect variation to human diversity, disease and beyond(Sherman & Salzberg,2020).

Super=able, over, plus, further

Campbell and co-workers;..

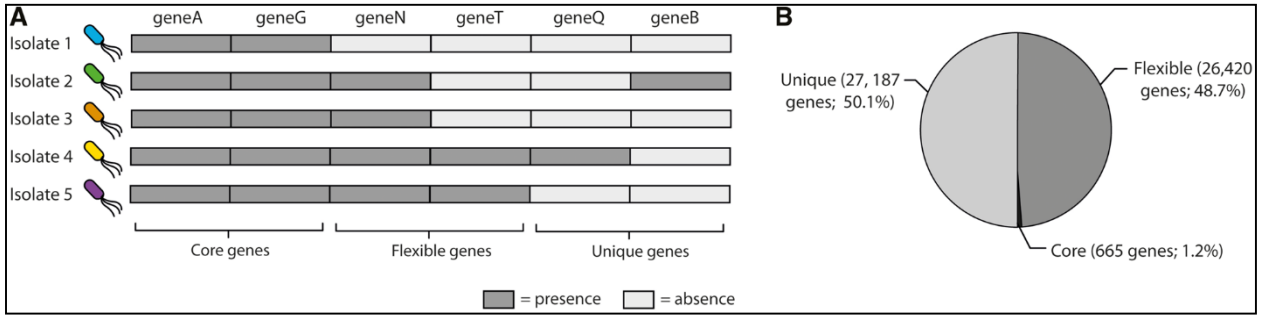
Pangenome
Core genome and Accessory genome

- **Core genome:** includes the genes present across all strains of a species.
- It typically includes housekeeping genes for cell envelope or regulatory functions.
- **Accessory genome:** (also: variable, flexible, dispensable genome) refers to genes not present in all strains of a species.
- These include genes present in two or more strains and genes unique to single strains.
- For example: genes for strain specific adaptation such as antibiotic resistance.

Pangenome
Core genome and Accessory genome

- As described by Tettelin *et al.*, 2005, a pan-genome is constituted by

1. core genes, that is, genes found in all strains of one species;
2. flexible genes, found in more than one strain but not all of them, and unique genes, found in one strain only.

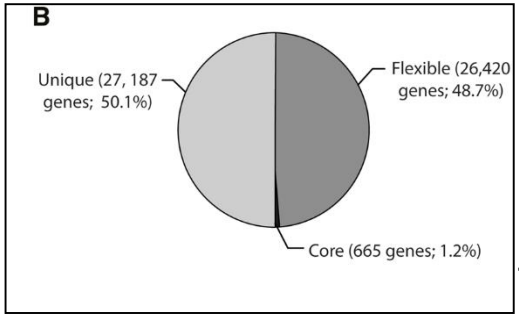


Freschi *et al.*,2019

Pangenome
Core genome and Accessory genome

- Pie chart showing the proportions of core, flexible and unique genes determined by SaturnV (https://github.com/ejfresch/saturnV; last accessed May 18, 2017).

1. Unique genes constitute 51% of the *P. aeruginosa* pan-genome,
2. flexible genes constitute 48% of it.
3. Core genes constitute only 1% of the pan-genome.



Freschi *et al.*,2019

Pangenome
Core genome and Accessory genome

- To analyze the *Corallococcus* pan-genome in detail, Roary was used (with a 90% BLASTp percentage identity cut-off) to cluster the genes encoding complete protein sequences into core (hard core and soft core) and accessory (shell and cloud) genomes(Page *et al.*,2015).
- hard core genes are found in >99% genomes,
- soft core genes are found in 95-99% of genomes,
- shell genes are found in 15-95%, while
- cloud genes are present in less than 15% of genomes.

Freschi *et al.*,2019

Pangenome
Core genome and Accessory genome

- When considering all bacteria, the majority of genome size variation is due to the gain and loss of accessory genes whose functions are thought to help bacteria cope with different niches or lifestyle.
- That bacteria with larger population sizes accommodate more accessory genes could reflect the fact that large populations likely span more diverse ecological conditions and require larger gene repertoires or that larger populations experience more competition, since many accessory genes are now known to be involved in bacterial warfare.
- Hence, accessory genes, and perhaps prophages, represent a diverse arsenal that allows bacteria to adapt to their ever-changing and competitive environments.

Bobay and Ochman,2017

Pangenome

Core genome and Accessory genome

The Fluidity of Prokaryote Genomes

- The realization of the high rates of gene content turnover has led to the paradigm of a core genome of genes present in all members of a taxonomic group and an accessory genome present in only a subset of members (with the total complement of genes in a taxon termed the pan genome.
- The extent to which related genomes differ in gene content varies for different species, with some having a relatively 'closed' genome (i.e., a core genome that is large compared to the accessory genome) and some species an 'open' or 'flexible' genome (i.e., a relatively small core genome and a large pan genome).

Vos *et al.*, 2015

Pangenome

Summary of the core and pan-genome

Core genome and Accessory genome

- The core genome and the pan-genome of the seven species were constructed according to their clustered orthologous proteins.
- Each species had distinctly different core-genome and pan-genome sizes and compositions, with no clear association found between the number of genomes calculated and the size of the pan-genome.
- For example, *S. aureus* had the second largest number of genomes (6282); however, it also displayed the smallest number of gene clusters constituting the pan-genome.
- On the other hand, the *E. coli* and *Shigella* spp. group with only 4401 genomes had the largest pan-genome (128,193 gene clusters).

Species	Average no. of genes	Core size	Percentage core (%)	Pan-genome size	Sum of branch length
Streptococcus pneumoniae	2076	1143	55.06	71,181	14.46
Staphylococcus aureus subsp. *aureus*	2668	2026	75.94	22,133	7.59
Salmonella enterica subsp. *enterica*	4606	3281	71.23	55,739	7.59
Escherichia coli (*Shigella* spp.)	4889	2608	53.34	128,193	9.73
Mycobacterium tuberculosis complex	4094	3343	81.66	33,490	4.12
Pseudomonas aeruginosa	6173	4464	72.31	49,744	5.69
Acinetobacter baumannii	3838	2355	61.36	29,845	5.08

Park *et al.*,2019

Pangenome

Summary of the core and pan-genome

Core genome and Accessory genome

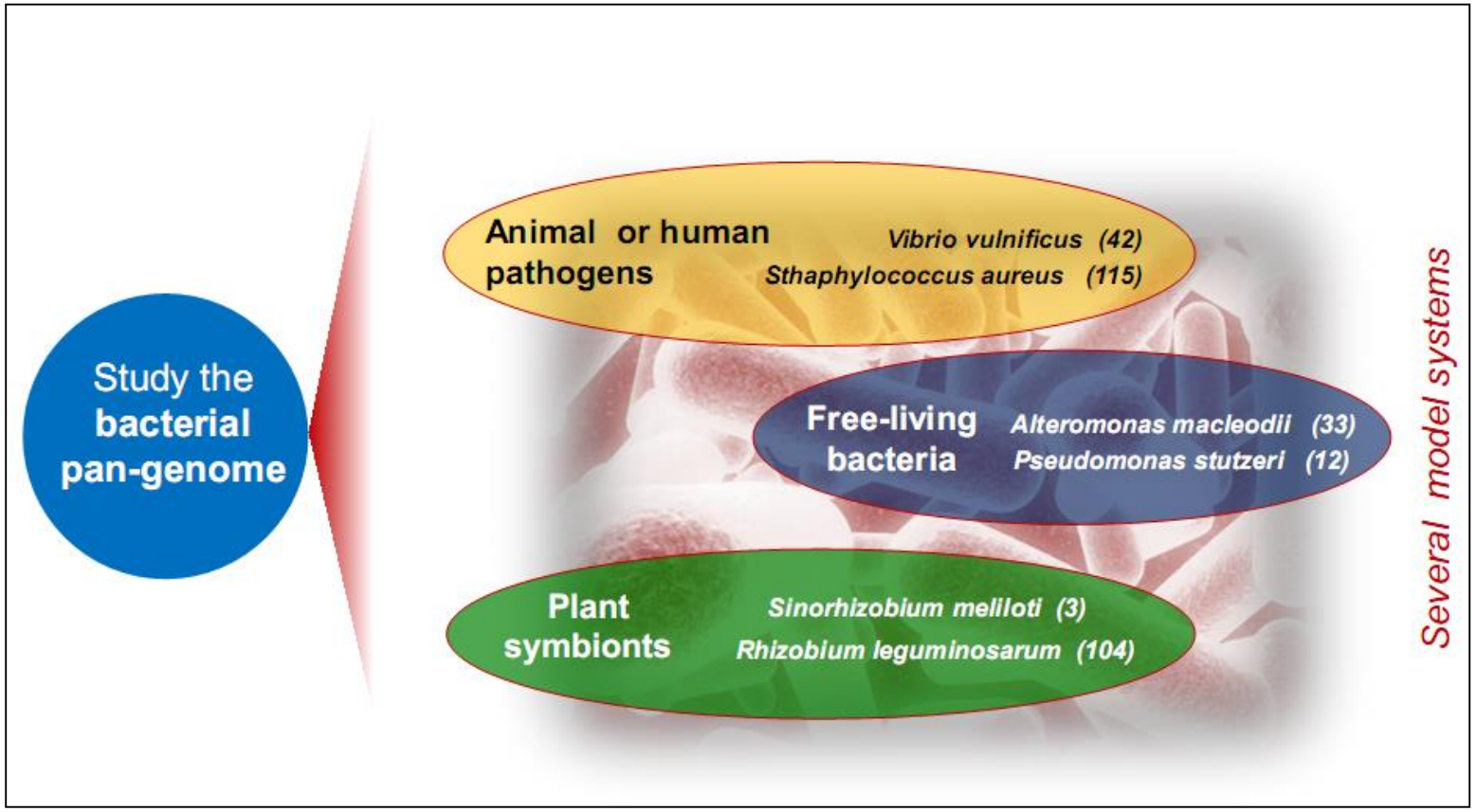
- Table also shows the relative proportion of genes belonging to the core genome as compared with the average number of genes (>50% in all seven species).
- A minimal relative proportion of the core genome was observed for *E. coli* (53%), although it had the largest pan-genome size.
- The second smallest relative proportion of the core genome was observed for *S. pneumoniae* (55%), but its difference with *E. coli* was very small.
- These were followed by *A. baumannii*, *S. enterica*, and *P. aeruginosa*. *M. tuberculosis* had an average of 4094 genes, with 81.66% of these belonging to the core genome.

Species	Average no. of genes	Core size	Percentage core (%)	Pan-genome size	Sum of branch length
Streptococcus pneumoniae	2076	1143	55.06	71,181	14.46
Staphylococcus aureus subsp. *aureus*	2668	2026	75.94	22,133	7.59
Salmonella enterica subsp. *enterica*	4606	3281	71.23	55,739	7.59
Escherichia coli (*Shigella* spp.)	4889	2608	53.34	128,193	9.73
Mycobacterium tuberculosis complex	4094	3343	81.66	33,490	4.12
Pseudomonas aeruginosa	6173	4464	72.31	49,744	5.69
Acinetobacter baumannii	3838	2355	61.36	29,845	5.08

Park *et al.*,2019

Pan-genome and Core genome

Pan-genome (supra-genome) describes the full complement of genes in a species



Microgen,2013

Pan-genome and Core genome

Update on the *E. coli* pan genome

- Genome size ranges from 4.6 to 5.6 Mb, with:

1. A core genome of about 1700 genes, and
2. A pan-genome of over 16,000 genes.

- The *E. coli* chromosome composed of ca 4.63 x 10^6 base pairs (bp) codes for 4288 proteins.

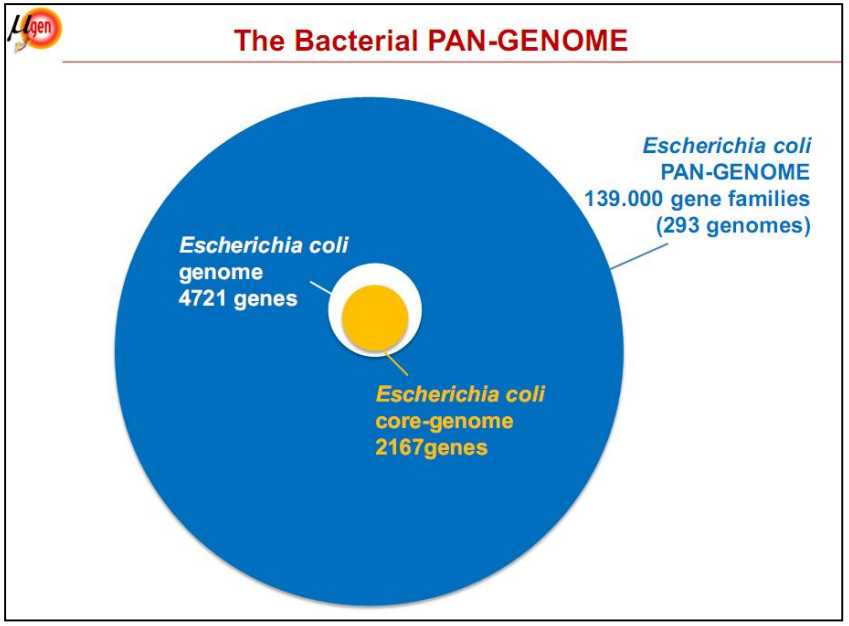
Rasko *et al.*,2008; Waleed Khalid

The Bacterial Pan-Genome

Update on the *E. coli* pan genome

- Relevance of the PAN-GENOME model:

1. Evolution of Prokaryote
2. Ecology
3. Pathogenicity
4. Biotechnology.



Microgen,2013

The Bacterial Pan Genome

Advantages and disadvantages

- Advantages:
  - Simple gene structure
  - Small genomes (0.5 to 10 million bp)
  - No introns
  - Genes are called Open Reading Frames (ORFs)
  - High coding density (>90%)
- Disadvantages:
  - Some genes overlap (nested)
  - Some genes are quite short (<60 bp)

The Bacterial Pan Genome

Example ORF in bacterial DNA

- The region of the nucleotide sequences from the start codon (ATG) to the stop codon (TAA, TGA, TAA) is called the Open Reading frame.

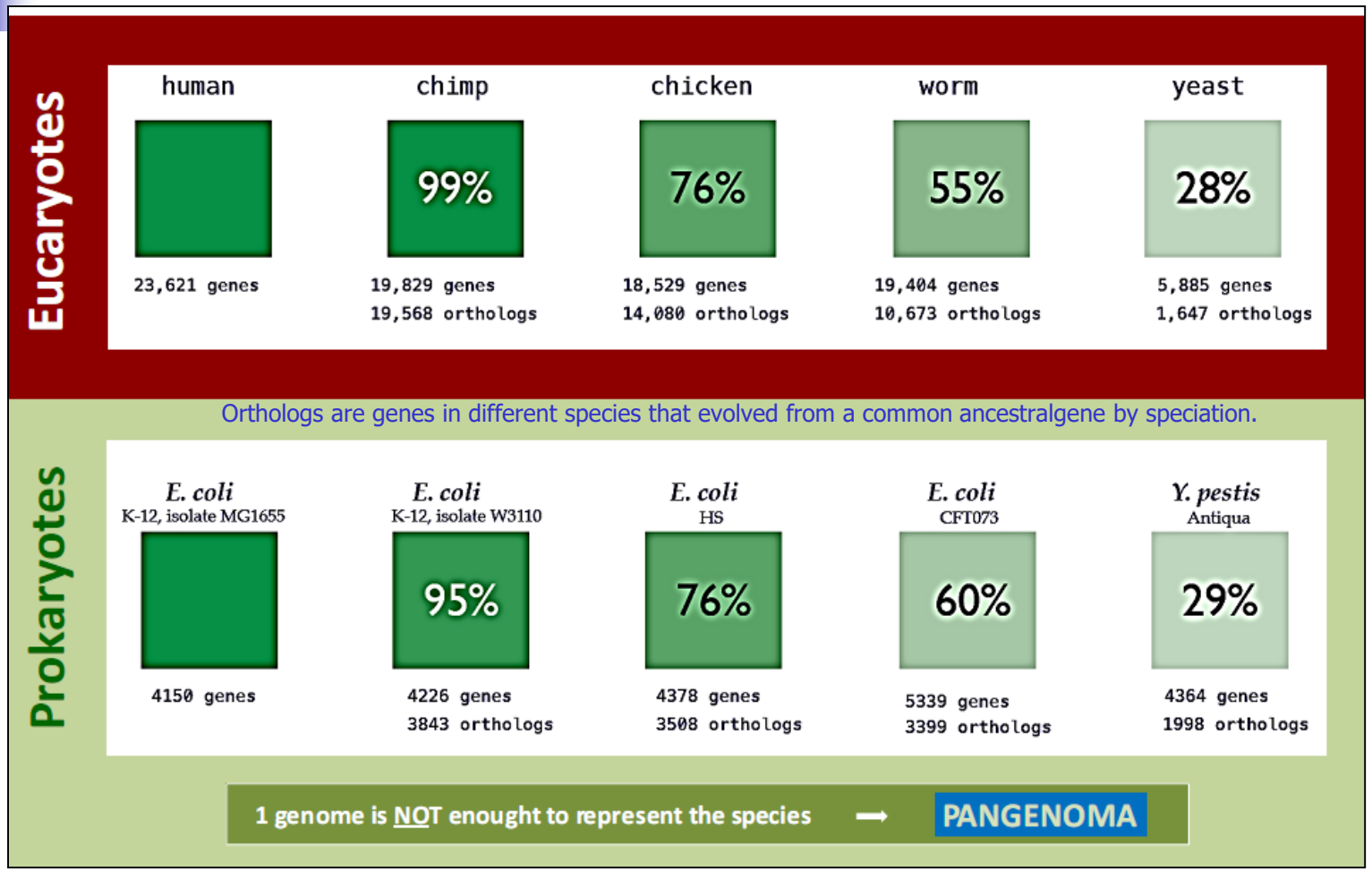
```
 5'                                                    3'
  atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa

1 atg ccc aag ctg aat agc gta gag ggg ttt tca tca ttt gag gac gat gta taa
   M   P   K   L   N   S   V   E   G   F   S   S   F   E   D   D   V   *
2  tgc cca agc tga ata gcg tag agg ggt ttt cat cat ttg agg acg atg tat
    C   P   S   *   I   A   *   R   G   F   H   H   L   R   T   M   Y
3   gcc caa gct gaa tag cgt aga ggg gtt ttc atc att tga gga cga tgt ata
     A   Q   A   E   *   R   R   G   V   F   I   I   *   G   R   C   I
```

In the standard genetic code, there are three different stop codons:
in RNA: UAG, UAA, UGA
in DNA: TAG, TAA, TGA

The Bacterial Pan-Genome

Intra-species variation

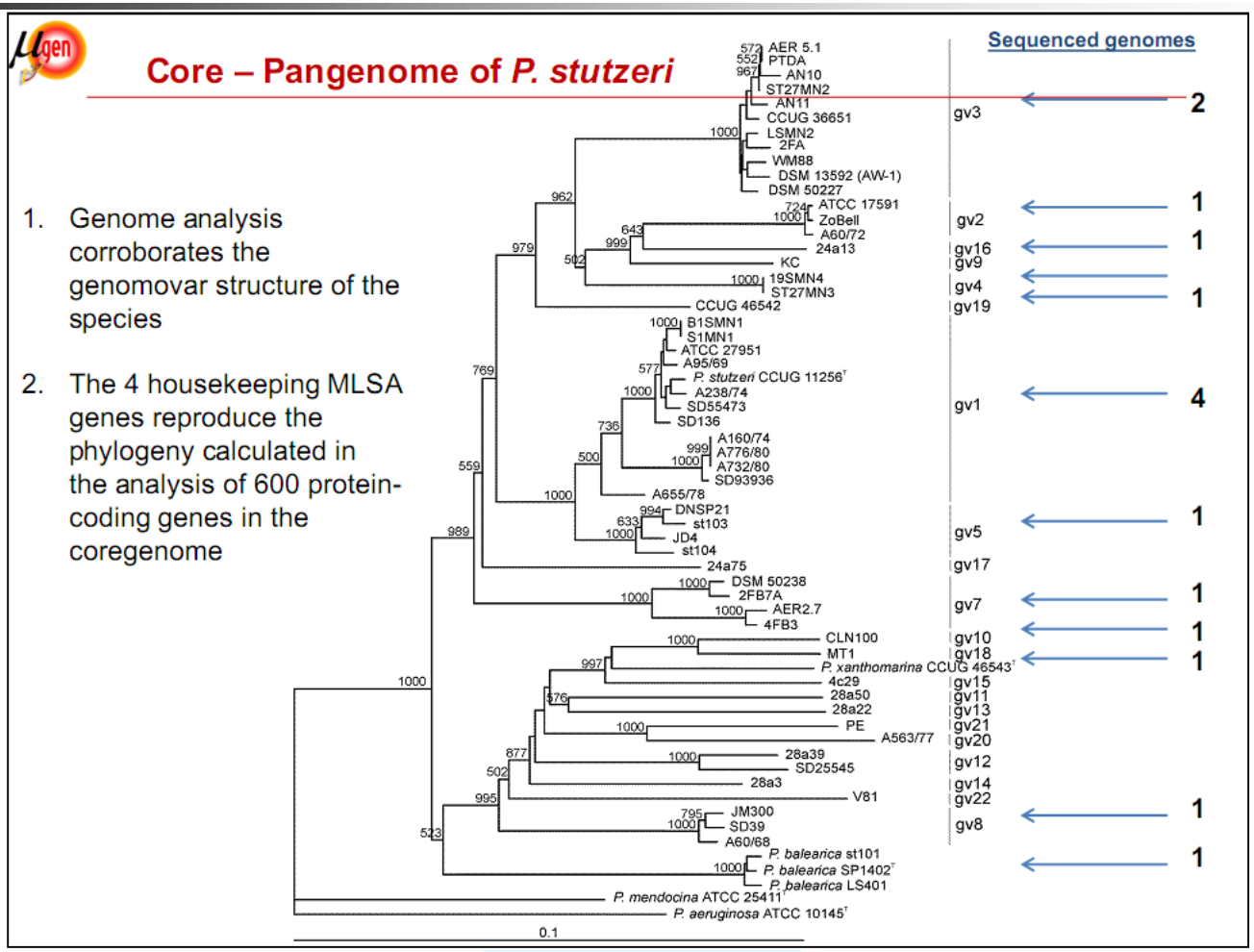
1 genome is NOT enough to represent the species



Microgen,2013

Pan-genome

Pan-genome of *Pseudomonas stutzeri*



Microgen,2013

Genome sequencing
Accessory genes

- The pan-genome of a species is composed of core genes present in all strains and dispensable genes that provide a selective advantage under specific conditions.
- Movement of these dispensable genes between species, genera and kingdoms is known as horizontal gene transfer (HGT).
- There are three primary mechanisms of HGT in bacteria.

1. Transformation: uptake of naked DNA from the environment by naturally competent cells.
2. Transduction: transfer of bacterial DNA between cells using bacteriophages as vectors.
3. Conjugation: intimate cell-to-cell contact with transfer of single-stranded DNA by a type-IV-like secretion system.

Blakely,2015

Genome sequencing
Accessory genes

- Mobile genetic elements (mobilome) are:
- Plasmids,
- Transposons,
- Insertion sequences,
- Integrons,
- Prophages,
- Genomic islands, and
- Pathogenicity islands.
- These are part of the accessory genes, which can have a significant influence on the phenotype and biology of the organism.
- These mobile elements facilitate interspecies and intraspecies genetic exchange.
- They play an important role in the pathogenicity of bacteria, and are a major contributor to species diversity.

Chan *et al.*,2006

Accessory genes

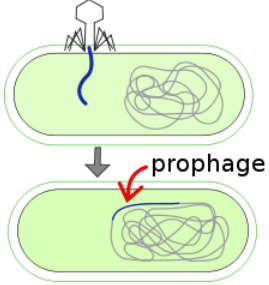
Mobile genetic elements involved in HGT in bacteria

HGT, horizontal gene transfer

Mobile element	General features	Examples	Phenotype of specific element
Gene cassette	Mobility of 'simple genes' Immediate constitutive expression Extremely broad host range	IntI1 IntI4	Antibiotic resistance (simple) Gene 'maturation' in *Vibrio*
Transposon	Mobility of genes and operons Loci screening for best performance Very broad host range	Tn21 Tn4651	Antibiotic resistance (complex) Toluene degradation
Plasmid	Mobility of operons and regulons Immediate regulated expression Very broad host range	pNRG234a pNL1 pCD	Nodulation plasmid of *Rhizobium* Degradation of aromatic compounds Virulence determinants of *Yersinia*
Bacteriophage	Mobility of regulons 'Maturation' in chromosomes (via lysogeny) Narrow host range	CTXϕ	Virulence phage of *Vibrio cholerae*

In cell biology and genetics, a regulon is a collection of genes or operons under regulation by the same regulatory protein.

de la Cruz and Davies,2000



Integron and prophage

Capture and spread of genes by site-specific recombination

- **Prophage:**
- Bacteriophage DNA are integrated into the genome of the bacteria and produce prophages, which are silent within bacterium and don't disrupt the bacterial cell.
- Prophages are important agents of horizontal gene transfer.
- Prophages can constitute as much as 10–20% of a bacterium's genome.

Waites,2000; Wikipedia,2015, Casjens,2003

Integron and prophage

Capture and spread of genes by site-specific recombination

- **Integrons (vectors for antibiotic resistance):**
- Integrons are mobile DNA elements with the ability to capture genes (gene cassettes), notably those encoding antibiotic resistance, by site-specific recombination.

1. Integrons may be found as part of mobile genetic elements such as plasmids and transposons.
2. Integrons can also be found in chromosomes.

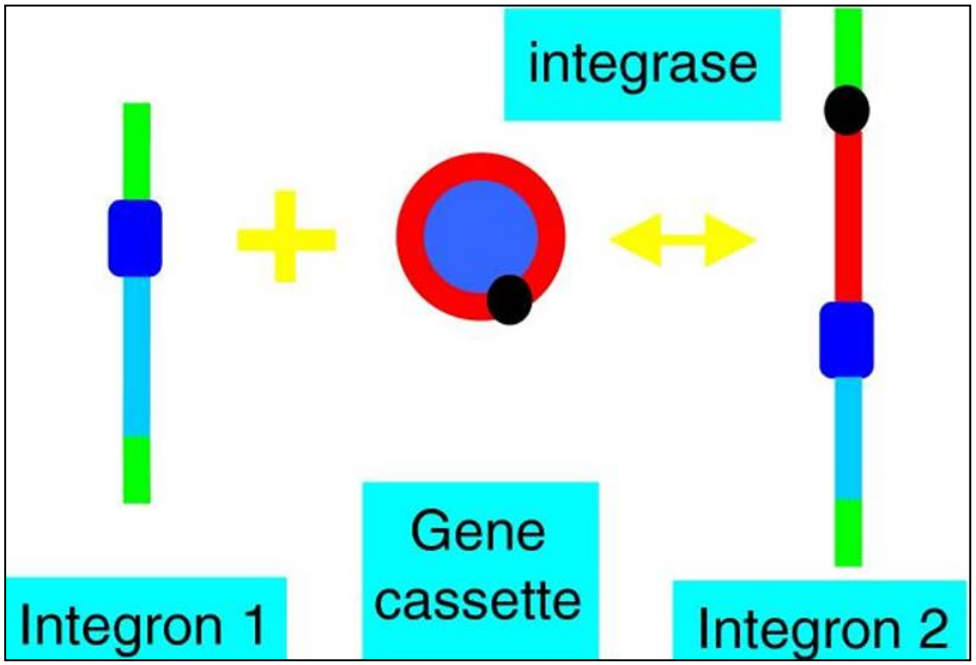
- Integrons are important agents of horizontal gene transfer.

Waites,2000; Wikipedia,2015

Integrons

An acquisition tool for capturing antibiotic resistance genes by pathogenic bacteria

- Cartoon of gene cassette capture by a bacterial integrons (mobile DNA elements).



Bennett,2008

Simple rule-based gene finding
Gene/ORF prediction

- Genes are called Open Reading Frames (ORFs).
- The genetic code can be expressed as either RNA codons or DNA codons.
- Gene finding in organism specially prokaryotes starts form searching for an open reading frames (ORF).
- ORF is usually predicted based on DNA sequence and not proven to be transcribed.
- Computer analysis of DNA sequence can predict the existence of genes based on ORFs.

Pangenome tools

- Roary: Fast tool for extracting complete pangenomes, core gene sets, or differences between reference genomes
- PanOCT: considers both gene homology and conserved gene neighborhoods
- OrthoMCL:- extracting the core genomes, etc.
- LS-BSR: rapid comparison of the genetic content of large numbers of genomes.
- PanPhlAn: pangenome based detection of gene compositions of strains in environmental WGS sample.

see more tools: http://omictools.com/pangenomics-category

Bacterial genetics

- **Includes two sections:**
- **Part I - Bacterial Genetics**
- **Part II- Recombinant DNA technology**

Part I
Bacterial Genetics

1. Genetic Elements Structures
2. Genetic Elements Functions

1. Genetic Elements in Prokaryotes

- Chromosome
- Plasmids
- Bacteriophages: temperate/virulent
- Transposable elements:

I. Simple transposons (Insertion sequence, IS)
II. Complex transposons (Transposons,Tn)
III. Bacteriophage Mu

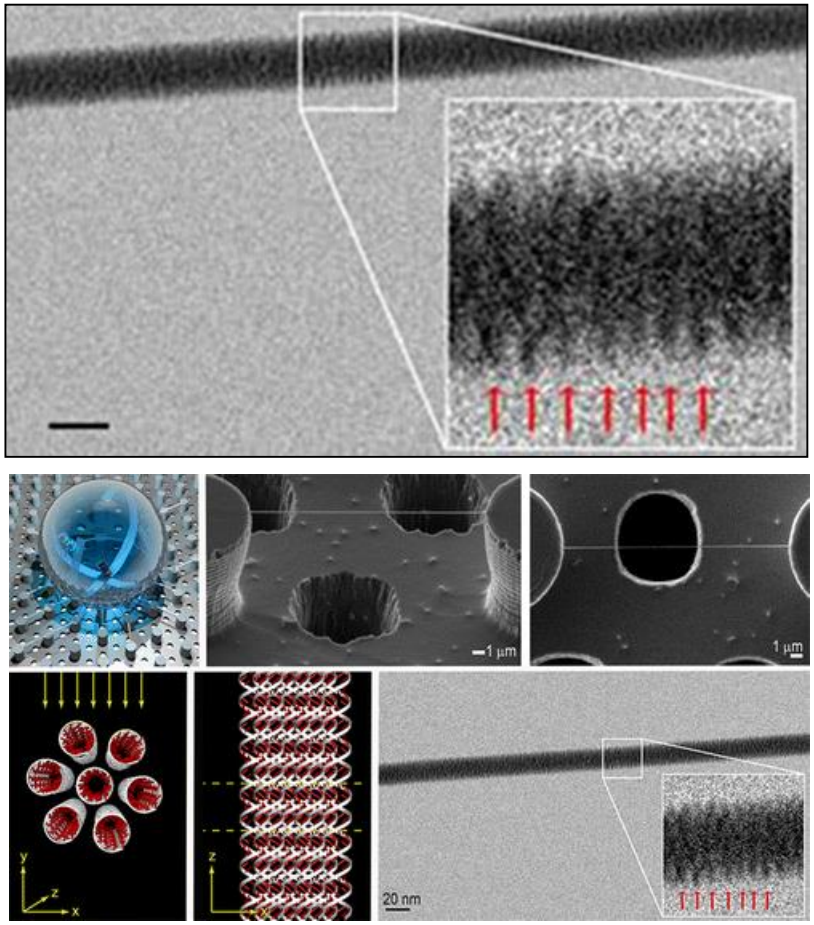
1. Chromosome

- **Chromosomes consist chiefly of proteins and DNA.**
- **Genes occupy precise locations on the chromosome.**

Structure of DNA

World's very first true photograph that shows the double helix structure of DNA

- Enzo di Fabrizio and his colleagues at the University of Genoa, Italy developed a new technique of imaging the double helix structure of DNA molecules.
- These images were obtained by combining a novel sample preparation method based on super hydrophobic DNA molecules self-aggregation process with transmission electron microscopy (TEM).



Gentile *et al.*,2012

Bacterial chromosome

- Prokaryotic cells (bacteria) contain their chromosome as circular DNA.
- The bacterial DNA is packaged in loops back and forth.
- The bundled DNA is called the nucleoid.
- It concentrates the DNA in part of the cell, but it is not separated by a nuclear membrane (as in eukaryotes).

Bacterial chromosome

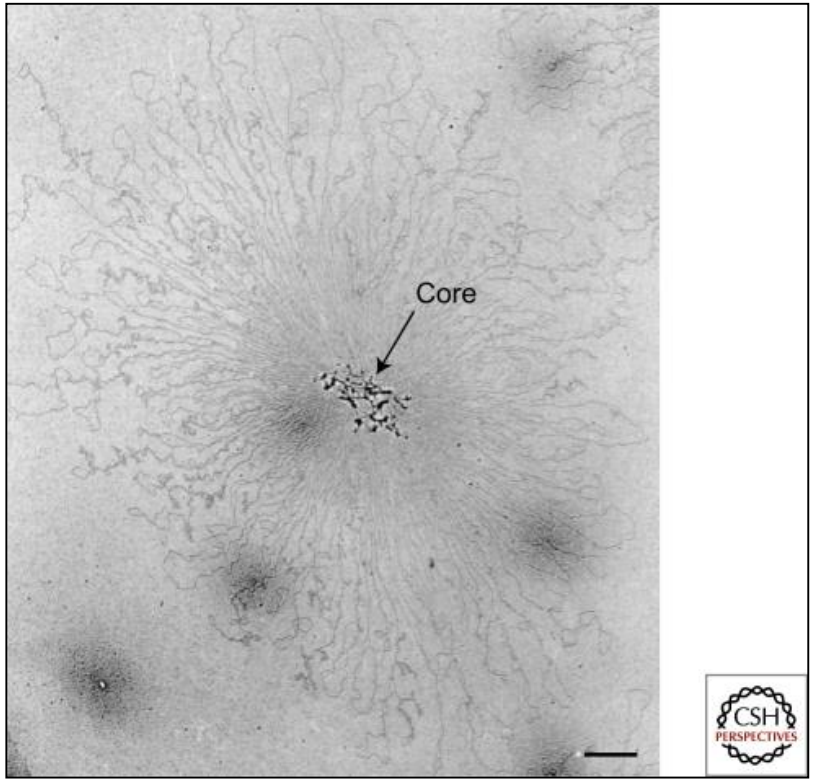
How does the packaging of a bacterial chromosome differ from eukaryotic chromosome packaging?

- Bacterial DNA is packaged into a single chromosome that is a continuous loop.
- In both cases, the DNA must be folded or coiled so that it will fit into the cell.
- In bacteria the compaction involves DNA binding proteins that help to form initial loops followed by coiling of DNA.
- Nucleoid means nucleus-like and also known as the nuclear region, nuclear body or chromatin body.
- In eukaryotes the DNA is wrapped around groups of histone proteins (collectively called chromatin).

The bacterial chromosome

Bacterial chromosomes are generally ~1000 times longer than the cells in which they reside

- Electron micrograph of isolated membrane-free chromosomes from *E. coli*.
- The central core, from which several tens of loops(domains of supercoiling) radiate out from this core, is sensitive to RNAse.
- Bar=1 μm.

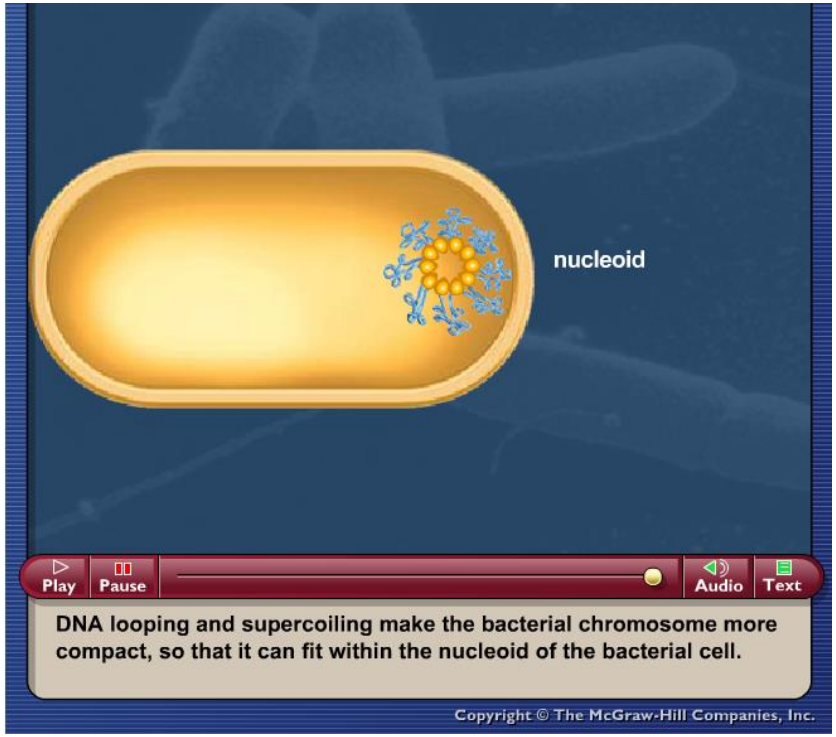


Reprinted from Kavenoff and Bowen,1976

The bacterial chromosome

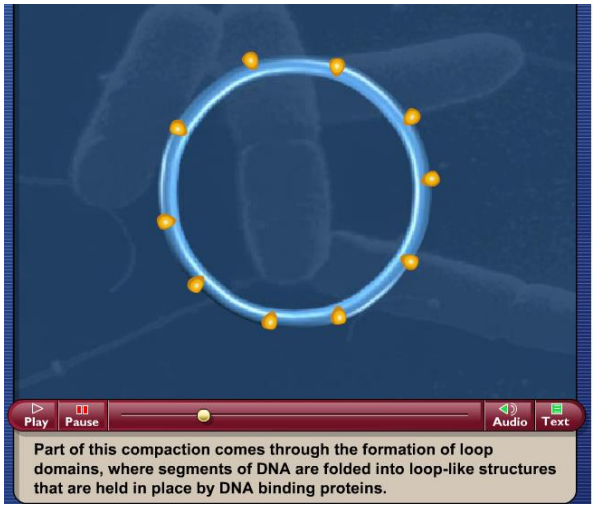
Bacterial chromosomes are generally ~1000 times longer than the cells in which they reside

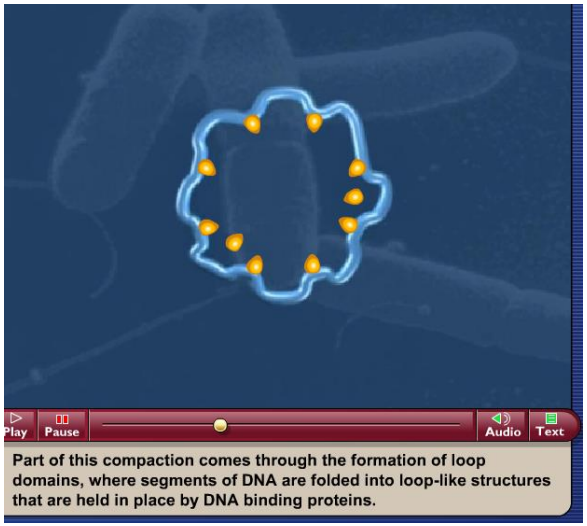
- DNA looping and supercoiling make the bacterial chromosome more compact, so that it can fit within the nucleoid of the bacterial cell.
- The nucleoid is a region of cytoplasm where the chromosomal DNA is located.
- It is not a membrane bound nucleus, but simply an area of the cytoplasm where the strands of DNA are found.

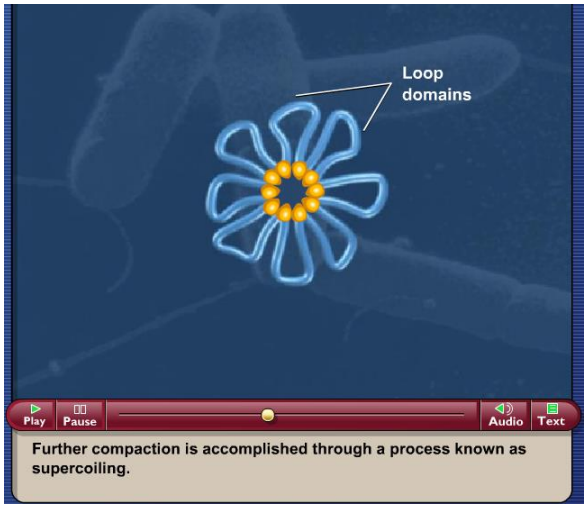


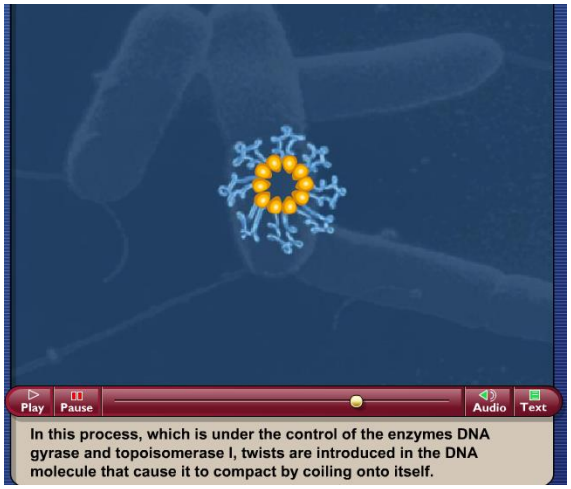
The bacterial chromosome

Bacterial chromosome compaction process









The bacterial chromosome

Bacterial genes

- A sequence of bases along a DNA strand that codes for the production of a protein is known as a gene.

  Or

- A segment of the DNA that codes for a specific polypeptide is known as a structural gene.

The bacterial chromosome
DNA/RNAs

- **DNA (deoxyribonucleic acid):**
- Is a double-stranded helix of nucleotides which carries the genetic information of a cell.
- DNA typically serves the permanent information-storage role while RNA serves a temporary information-storage role.

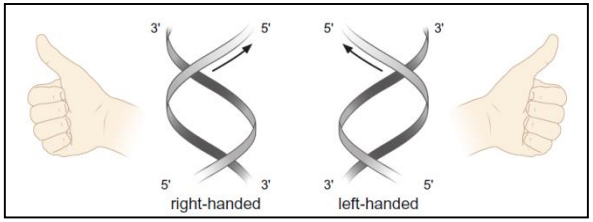
The bacterial chromosome
DNA/RNAs

- **RNA (ribonucleic acid):**
- Unlike DNA it is more often found in nature as a single-strand folded unto itself, rather than a paired double-strand.
- Is an information encoded strand of nucleotides, similar to DNA, but with a slightly different chemical structure.
- There are three main forms of RNA, each a slightly different function:
  - mRNA (messenger RNA)
  - tRNA (transfer RNA)
  - rRNA (ribosomal RNA)

The bacterial chromosome
RNAs

- Three types of RNA:
  - mRNA: messenger RNA
    - serve as temporary copies of the information found in DNA. Contains 3 bases (codon)
  - rRNA: ribosomal RNA
    - serve as structural components of protein-making structures known as ribosomes. Comprises the 70 S ribosome.
  - tRNA: transfer RNA
    - Transfers amino acids to ribosomes for protein synthesis;
    - Contains the anticodon (3 base sequence that is complimentary to codon on mRNA).

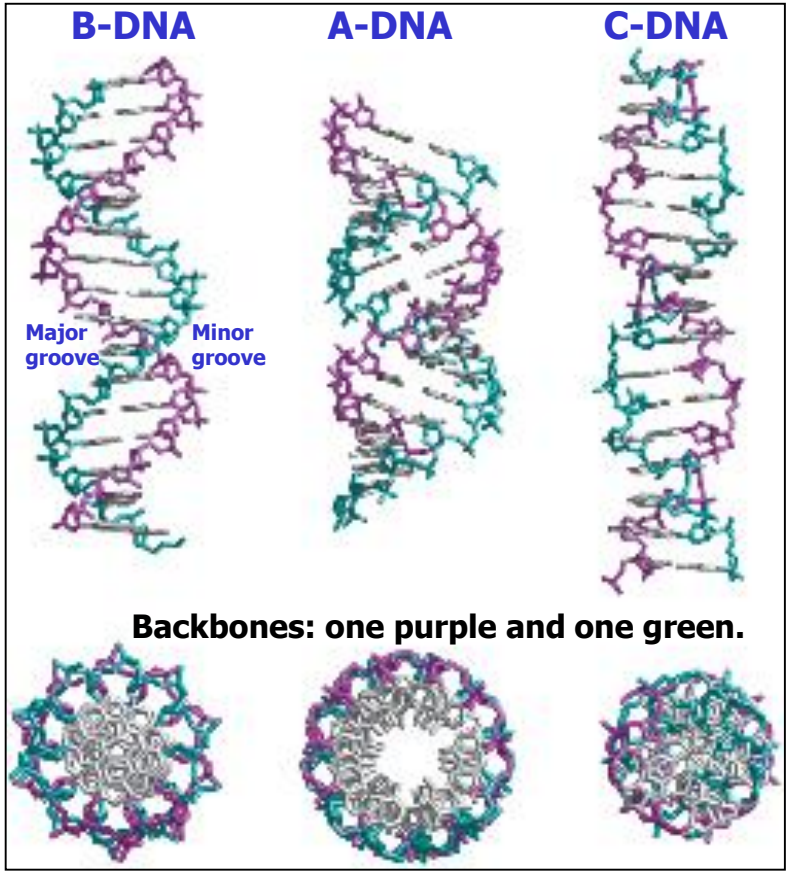
Johnson,2003;..



DNA

Major and minor grooves of DNA

1. At least three DNA conformations are believed to be found in nature:
2. A-DNA (right-handed double helix).
3. B-DNA (right-handed double helix).
4. Z-DNA (left-handed helix and has a zigzag (Z) appearance).

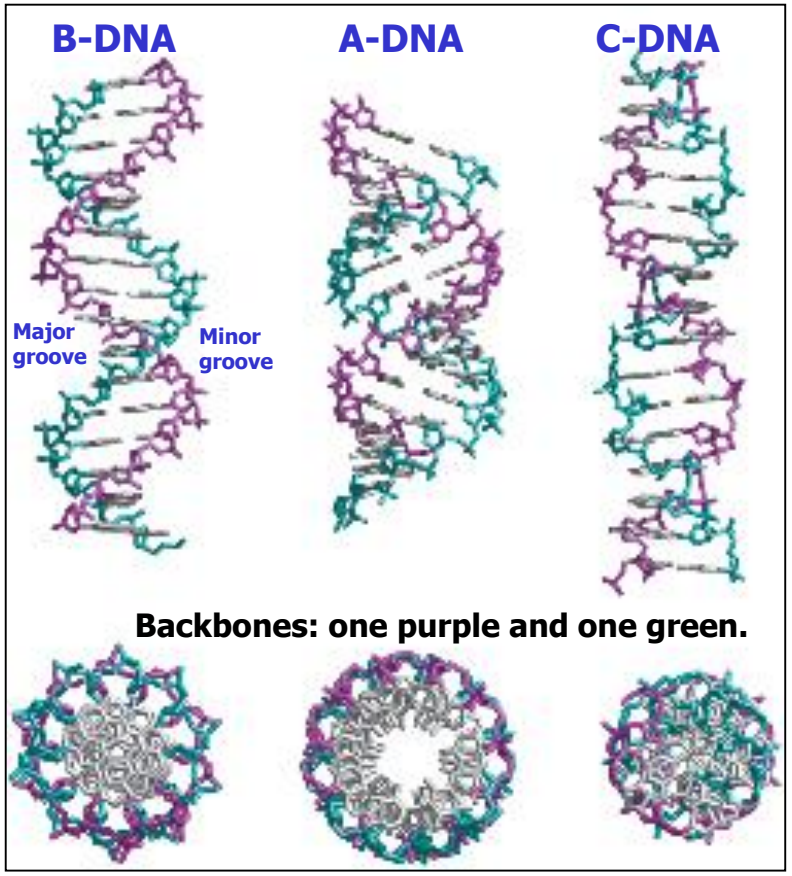


Z DNA is a variant of B DNA, with slightly different configuration.
B-DNA is driven into the A-DNA when under dehydrating conditions.

DNA

Major and minor grooves of DNA

1. A-DNA:
   - major groove is narrower and much deeper
   - minor groove is broader and shallower.
2. B-DNA:
   - major groove is wider than the minor groove.
3. Z-DNA:
   - The major and minor grooves, unlike A- and B-DNA, show little difference in width.



DNA conformations

A comparison of the structural properties of A, B, and Z DNAs as derived from single-crystal X-ray analysis

Parameter	A-DNA	B-DNA	Z-DNA
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per turn	11	10	12
Diameter of helix (nm)	2.3	2.0	1.8
Stability	less stable	More stable	Unstable
Found in:	Bacteria to eukaryotes	Most organisms including bacteria	Bacteria to eukaryotes except *E. coli*

- A-DNA occurs when DNA is dehydrated.
- A-DNA helix being less stable than the B-DNA.
- A-DNA also plays a role in some processes that do not involve RNA. For example, in sporulating bacteria, there is a protein which can bind to DNA in the B-conformation and induce a change to the A-DNA helix.
- The Z-helix is narrower than the A-and B.
- In some eukaryotic genomes, 10% or more of the genome contains sequences capable of forming Z-DNA. Most human genes have been found to have Z-DNA-forming sequences near the transcription start site.

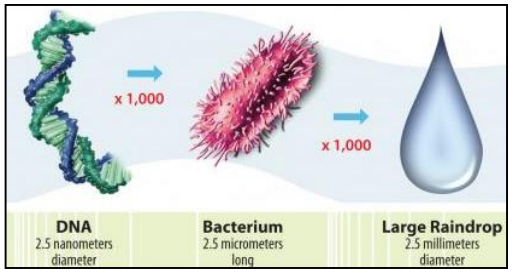
DNA conformations
DNA persistence length

- DNA in solution does not take a rigid structure but is continually changing conformation due to thermal vibration and collisions with water molecules, which makes classical measures of rigidity impossible.
- Hence, the bending stiffness of DNA is measured by the persistence length, defined as:
- The length of DNA over which the time-averaged orientation of the polymer becomes uncorrelated by a factor of *e.*
- In an aqueous solution, the average persistence length is 46-50 nm or 140-150 base pairs (the diameter of DNA is 2 nm).
- This makes DNA a moderately stiff molecule.

Wikipedia,2017

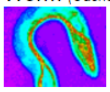
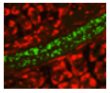
DNA conformations

Size of DNA



At least 46,831 genes are protein-coding genes in the human genome

- Genome size does not correlate with the number of genes.
- The size of bacterial chromosomes ranges from 0.6 Mbp to over 10 Mbp.
- Mammalian cell nucleus ~10 micron diameter.
- A strand of human DNA is 2.5 nanometers in diameter.
- A human hair is about 60,000 nanometers in diameter, and a DNA molecule is between 2-12 nanometers wide.

Species	Size of genome	Number of genes
Human 	2.9 billion base pairs	30,000
Fruit fly *(Drosophila melanogaster)* 	120 million base pairs	13,601
Baker's yeast *(Saccharomyces cerevisiae)* 	12 million base pairs	6, 275
Worm *(Caenorhabditis elegans)* 	97 million base pairs	19,000
E. coli 	4.1 million base pairs	4,800
Arabidopsis *(Arabidopsis thaliana)* 	125 million base pairs	25,000

The human genome contains approximately 3 billion of these base pairs, which reside in the 23 pairs of chromosomes within the nucleus of all our cells. Each chromosome contains hundreds to thousands of genes, which carry the instructions for making proteins.

Genome size

Comparison of genome size in different organisms

Organism type	Organism	Genome size (base pairs)		Approx. no. of genes	Note
Virus	*Pandoravirus salinus*	2,470,000	2.47Mb		Largest known viral genome.
Virus	Megavirus	1,259,197	1.3Mb		Until 2013 the largest known viral genome.
Virus	Bacteriophage MS2	3,569	3.5kb		First sequenced RNA-genome
Plant	*Paris japonica* (Japanese-native, pale-petal)	150,000,000,000	150Gb		Largest plant genome known
Plant	Genlisea tuberosa	61,000,000	61Mb		Smallest recorded flowering plant genome, 2014.
Plant	Arabidopsis thaliana	135,000,000	135 Mb	27,655	First plant genome sequenced, December 2000.
Nematode	Pratylenchus coffeae	20,000,000	20Mb		Smallest animal genome known
Mammal	Homo sapiens	3,289,000,000	3.3Gb	20000	Homo sapiens estimated genome size 3.2 billion bp Initial sequencing and analysis of the human genome
Mammal	Bonobo	3,286,640,000	3.3Gb	20000	Pan paniscus estimated genome size 3.29 billion bp
Insect	Drosophila melanogaster (fruit fly)	175,000,000	175Mb	13600	Size variation based on strain (175-180Mb; standard y w strain is 175Mb)
Insect	Solenopsis invicta (fire ant)	480,000,000	480Mb	16569	
Fungus-yeast	Saccharomyces cerevisiae	12,100,000	12.1Mb	6294	First eukaryotic genome sequenced, 1996
Fungus	Aspergillus nidulans	30,000,000	30Mb	9541	
Bacterium – cyanobacterium	*Prochlorococcus* spp.	1,700,000	1.7Mb	1884	Smallest known cyanobacterium genome
Bacterium	*Haemophilus influenzae*	1,830,000	1.8Mb		First genome of a living organism sequenced, July 1995
Bacterium	*Escherichia coli*	4,600,000	4.6Mb	4288	

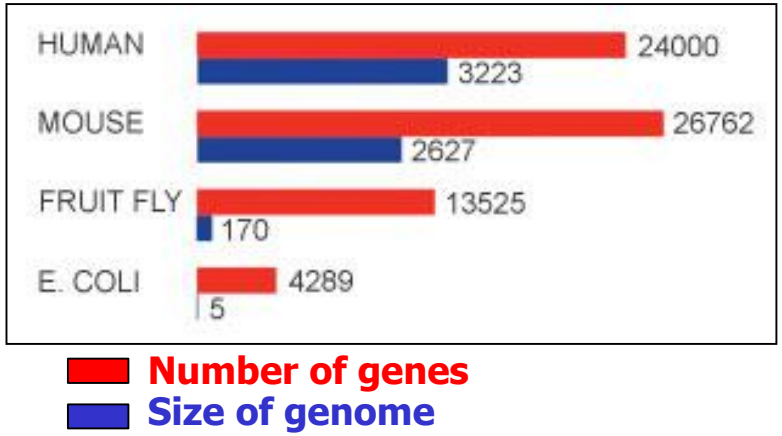
Wikipedia,2017

Genome size

Comparison of genome size in four different organisms

Genome comparison chart

- Genome size is the total number of base pairs in human, mouse, fruit fly and *E. coli* cells.
- While the number of genes in an organism's DNA (red bars) varies from species to species (numbers at right), it is not always proportional to genome size (blue bars, in millions of base pairs).
- Note how many genes a fruit fly can squeeze out of its relatively small genome.



Note: new total of human protein-coding genes comes to at least 46,831(Saey,2018)

Nature Education,2009

Genome size

Comparison of genome size in some organisms

Human, mice, pig and chimpanzees

- Have a look genome size and # of genes:
- 3.2Gbp/41,000 for humans,
- 2,8 Gbp/ 35,000 for pigs,
- 3.3 Gbp/31,000 for chimpanzee.
- Humans and chimpanzees shared a common ancestor ~5-7 million years ago (Mya). The difference between the two genomes is actually not ~1%, but ~4%-comprising ~35 million single nucleotide differences and ~90 Mb of insertions and deletions.
- Humans, chimpanzees, gorillas, orangutans are mammals belong to family Hominidae. Pig is a mammal of the Suidae family.

Genome size

Comparison of genome size in some organisms

Human, mice, pig and chimpanzees

- Genetic similarity between human and mice is approximately 99%. But still they are different.
- On the other hand, humans and pigs are both composed of animal cells, but a number as high as 98% still seems too high. It's almost the same number as the similarity between humans and chimpanzees.
- We can conclude:

1. These similarities do not reflect similarity on phenotype (outer look) but it reflects the resemblance in Genetic makeup of the genome.
2. Genetic similarity does not respond to strong ancestral relations.

Genome size

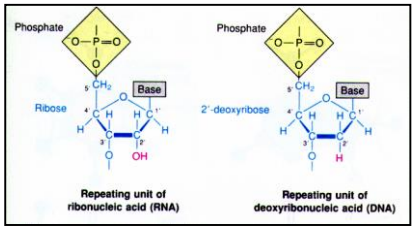
Comparison of genome size in some organisms

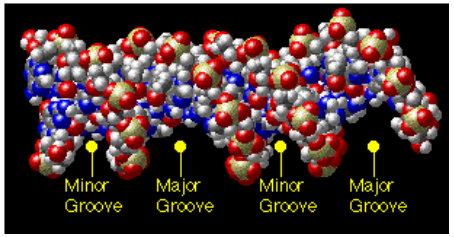
Human, mice, pig and chimpanzees

- The other reasons:

1. The reason behind this is evolutionary change in expression of various genes and also some set of genes are conserved and express themselves while others are silent in expression.
2. 98% identity corresponds to ~60,000,000 differences in a 3 Gbp (3 billion base pairs) genome.

- Even if only 1% of those have an effect, that easily accounts for the existing differences.
- And keep in mind that changes in the DNA do not necessarily carry over into the protein sequence.
- 99% similarity in the DNA level can result in anywhere between 97 and 100% similarity on the protein level.



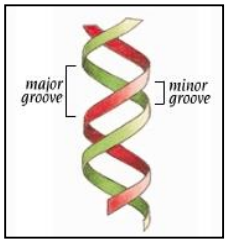


DNA

Major and minor grooves of DNA

- B-DNA or common DNA: The most common form of DNA found in most organisms including bacteria.
- B-DNA present at neutral pH and physiological salt concentrations.
- B-DNA, is right-handed double helix with about 10-10.5 base pairs per turn.
- This translates into about 20-21 nucleotides per turn.
- The major groove is approximately 50% wider than the minor.
- Thus, many proteins which bind to B-DNA do so through the wider major groove.

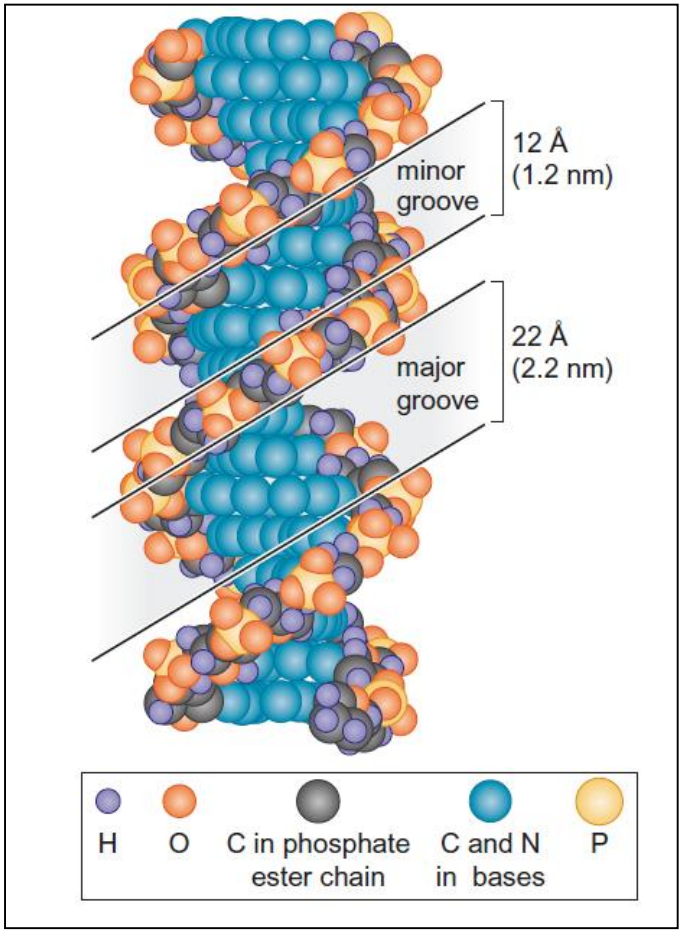
Watson_6;..



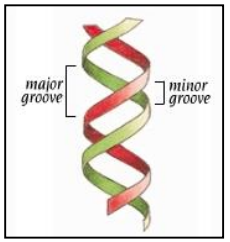
DNA

Major and minor grooves of DNA

- Twin helical strands form the DNA backbone.
- The major groove occurs where the backbones are far apart,
- The minor groove occurs where they are close together.
- The grooves twist around the molecule on opposite sides.
- The major groove is approximately 50% wider than the minor.
- The major groove, is 22 Å wide and the minor groove, is 12 Å wide.



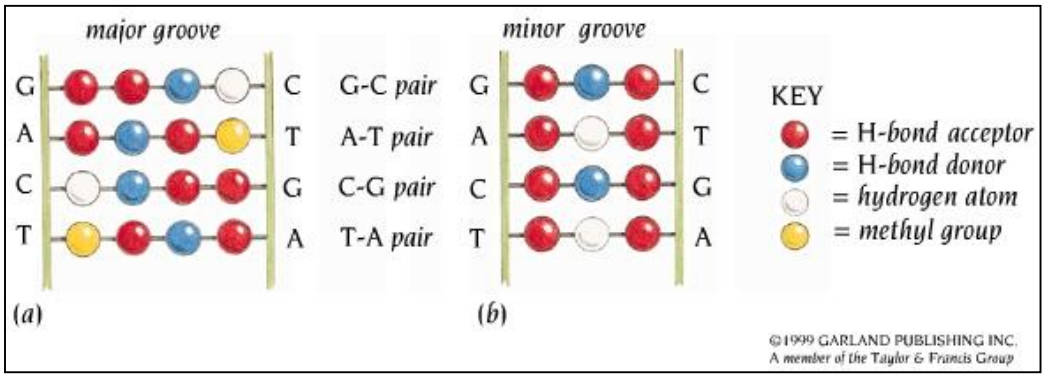
1 angstrom is a unit of length equal to 10^{-10} m or 0.1 nm.



DNA

Major and minor grooves of DNA

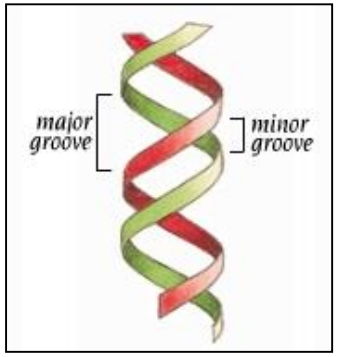
- The major groove has twice the information content of the minor groove.
- A-T base pairs can be differentiated from G-C pairs via either groove.
- But minor groove could not discriminate between A-T and T-A base pairs, or G-C and C-G.
- Hence the major groove has a four-symbol code, whereas the minor groove has only a two-symbol code, and only half information content.



DNA-binding proteins

DNA-binding domains

Structural DNA-binding motif

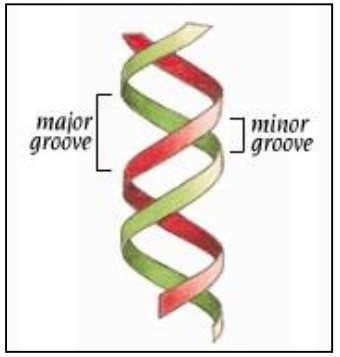


- The term domain in biochemistry and molecular biology is usually used to refer to a part of a protein that has a conserved structure and function, is similar between related proteins, and can generally exist or function on its own if it was separated from the rest of the protein.
- A DNA-binding domain is a protein structure that has a high affinity for DNA, and so binds to it when the two molecules are in the same vicinity.
- An example is a type of protein called a transcription factor, which binds to DNA in the nucleus of the cell (in Eukaryotes, bacteria and Archaea don't have nuclei - the DNA just floats free) and causes changes in the transcription of genes.

DNA-binding proteins

DNA-binding domains

Structural DNA-binding motif



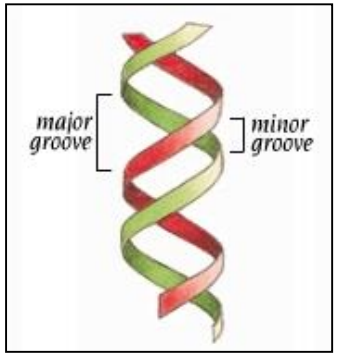
- Domain and motif are indeed related, but I would think that domain indicates the region in a protein (or DNA) involved in the function (be it DNA binding, catalytic etc for protein/transcription factor binding etc for nucleic acids), while motif is used to describe the particular sequence of building blocks in that domain/region.
- Motif is used to indicate:

1. the particular amino acids that one typically finds in the protein domain,
2. but also to indicate the particular nucleotides in a DNA/RNA strand.

DNA-binding proteins

DNA-binding domains

Structural DNA-binding motif



- Each structural motif contains features that are highly conserved among many organisms.

1. Some proteins bind DNA in its major groove (e.g. the zinc finger).
2. Some other in the minor groove (e.g. HMG-I(Y)–DNA complex), and
3. Some need to bind to both (e.g. the Leucine Zipper Motif).

The major groove, being wider than the minor groove, can accommodate larger structural motifs.

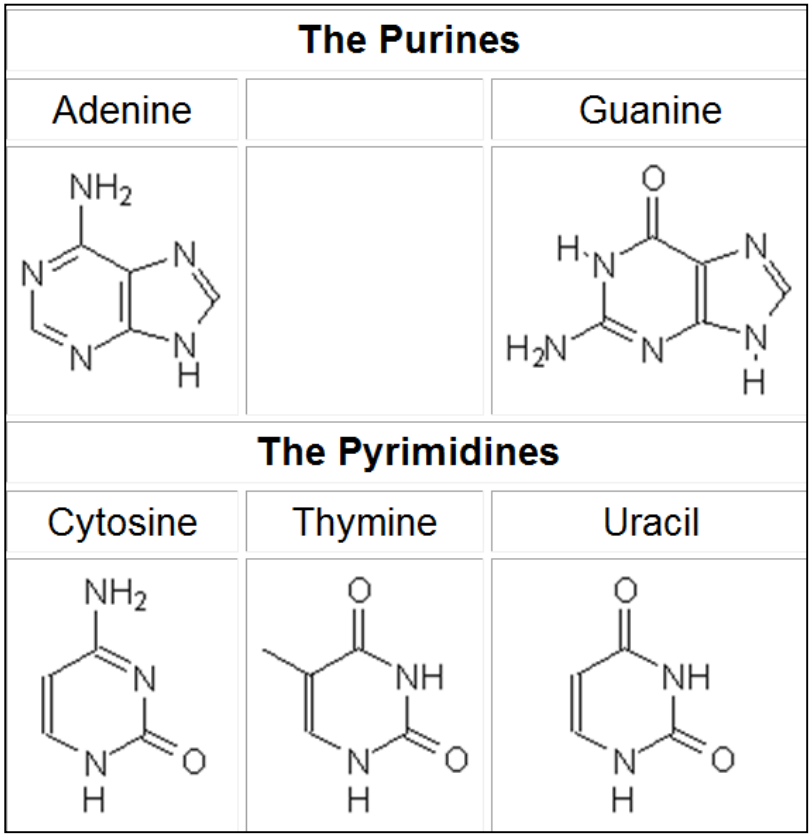
Atlas Genet Cytogenet Oncol Haematol, 2003

DNA vs. RNA

DNA bases

Purines 2 carbon rings, pyrimidines only have one

- A and G are double-ringed larger molecules (called purines);
- C and T are single-ringed smaller molecules (called pyrimidines).
- Only two purines and three pyrimidines occur widely in nucleic acids.



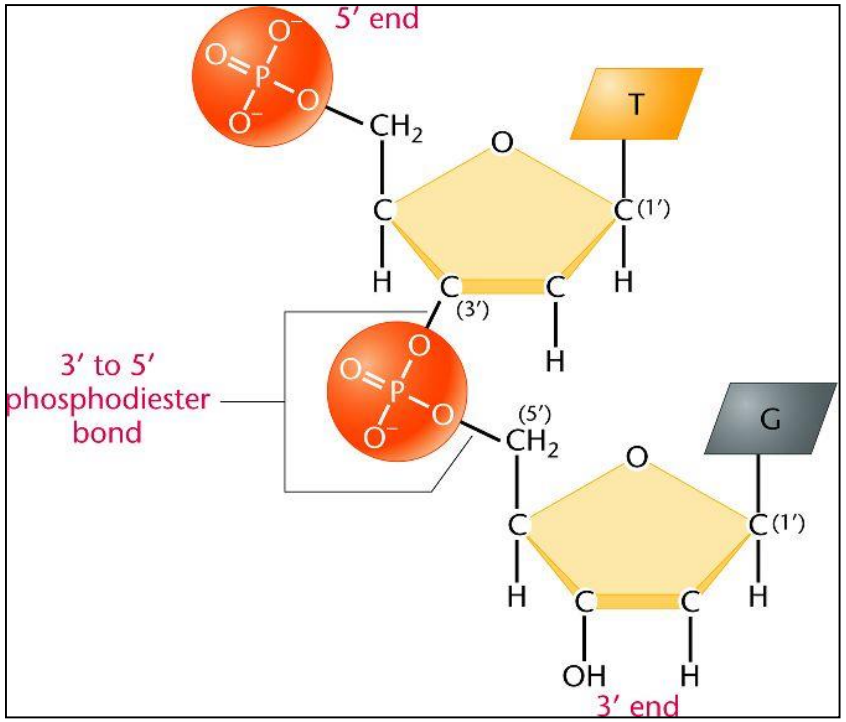
University of Maine,2008

A dinucleotide

Units of bacterial DNA/RNA

Dinucleotides form from a phosphodiester link between 2 mononucleotides

- Linkage of two nucleotides by the formation of a 3'-5' phosphodiester bond, producing a dinucleotide.
- Multiple phosphodiester bonds form a polynucleotide chain.



DNA/RNA structure
Upstream/Downstream

- DNA and RNA are synthesized (replicated) in the 5′ to 3′ direction.
- **In an RNA:**
- Anything towards the 5' end of a reference point is upstream of that point. Downstream is toward the 3' end.
- **In DNA:**
- The situation is a bit more complicated.
- Due to the anti-parallel nature of DNA, the 3' end of the template strand is upstream of the gene and the 5' end is downstream.

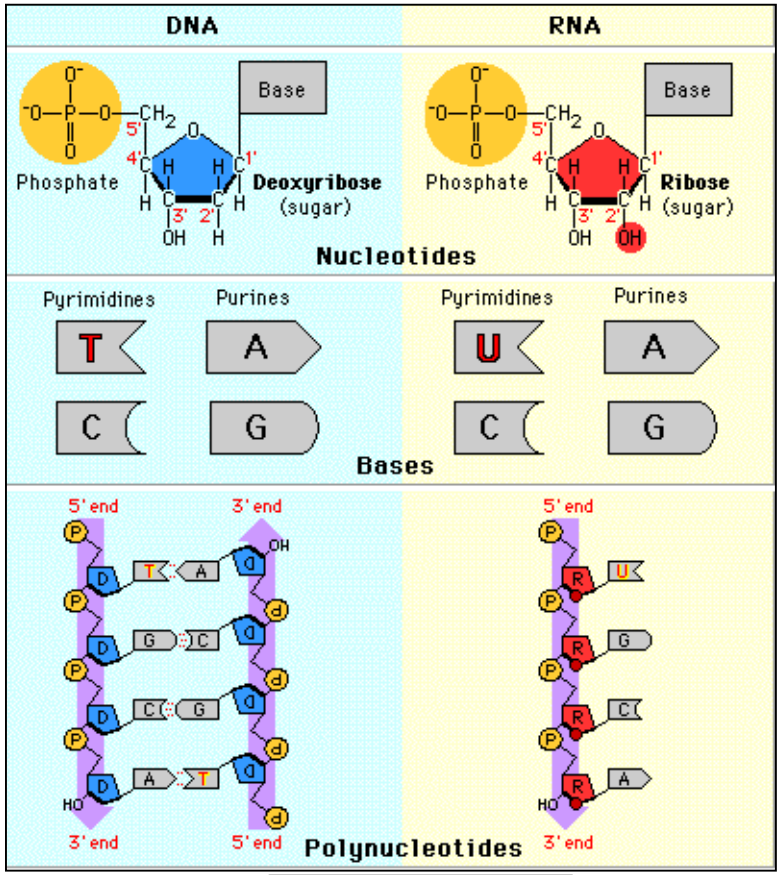
A ribosome moves along an mRNA from 5′ to 3′.
mRNA is also made in the 5' to 3' direction.

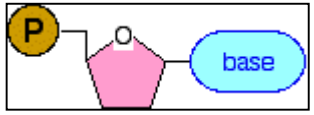
The bacterial chromosome
DNA vs. RNA

- The 2 DNA strands are antiparallel [5'--->3' pairs to 3'<---5'].
- The combination of a base and a pentose is called a nucleoside.
- Nucleotides consist of:

1. A five-carbon sugar ribose (RNA) or deoxyribose (DNA);
2. A nitrogenous base(ring structure),and
3. One or more phosphates.

- Inosine is a purine nucleoside containing the base hypoxanthine and the sugar ribose, which occurs in transfer RNAs.





Watson & Crick Model,1953

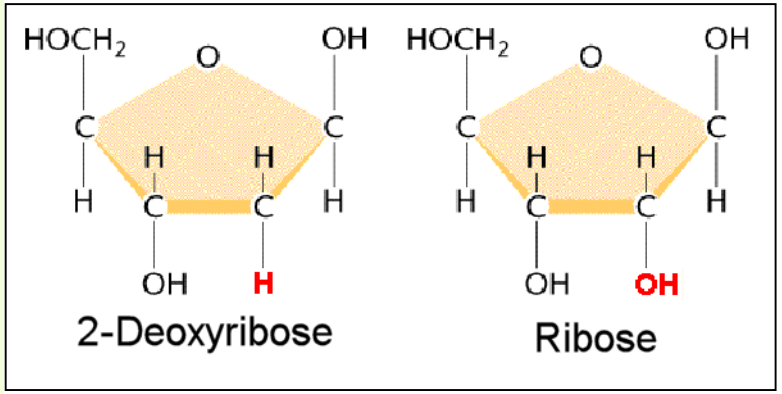
DNA vs. RNA
Pentose sugars

- Both ribose and deoxyribose are pentose sugars.

1. Ribose (in RNA only);
2. 2-deoxyribose (in DNA only).

- One oxygen distinguishing them:

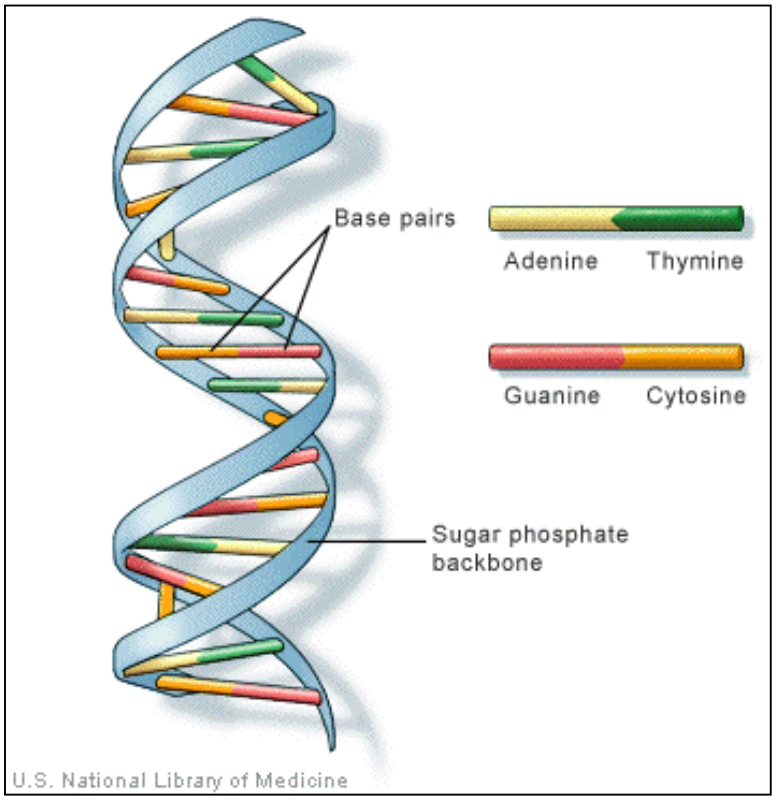
1. a hydrogen (H) bonds in deoxyribose and
2. a hydroxyl (OH) bonds in ribose.



The nucleotides in RNA are ribonucleotides - that is, they contain the sugar ribose (hence the name ribonucleic acid).

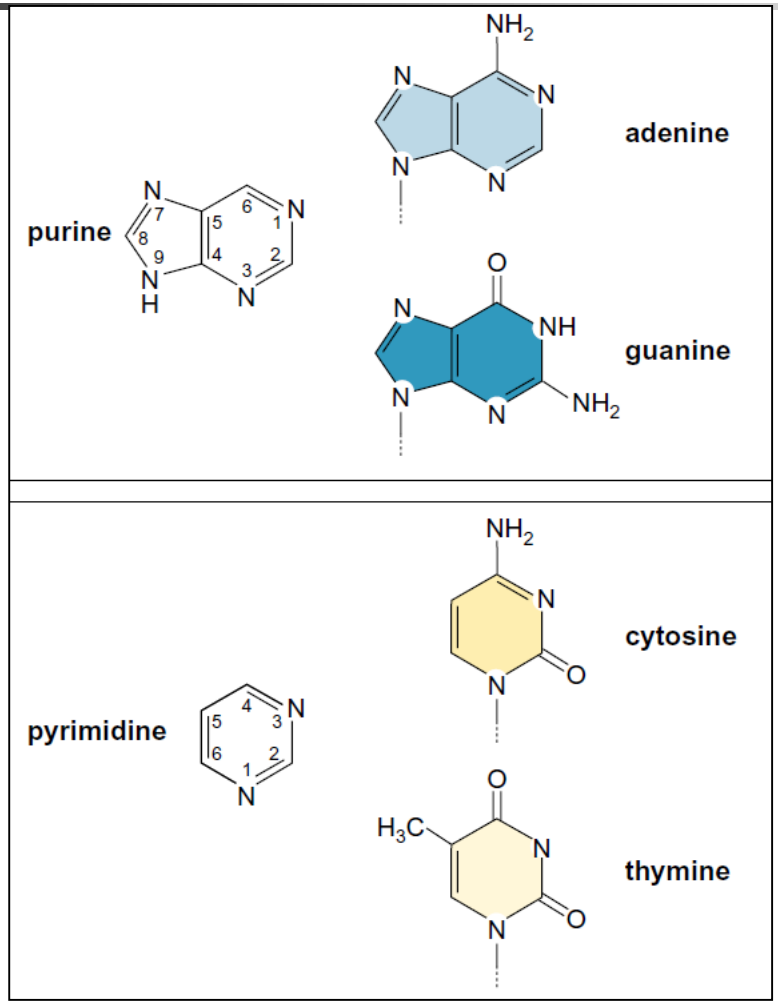
Nucleotide
DNA bases

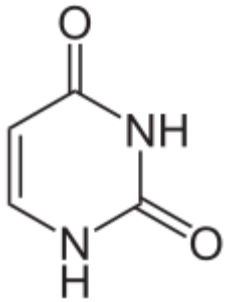
- DNA bases pair up with each other, A with T and C with G, to form units called base pairs.
- Each base is also attached to a sugar molecule and a phosphate molecule.
- Together, a base, sugar, and phosphate are called a nucleotide.
- Nucleotides are arranged in two long strands that form a spiral called a double helix.



Genetics Home Reference,2015

DNA/RNA Base structures



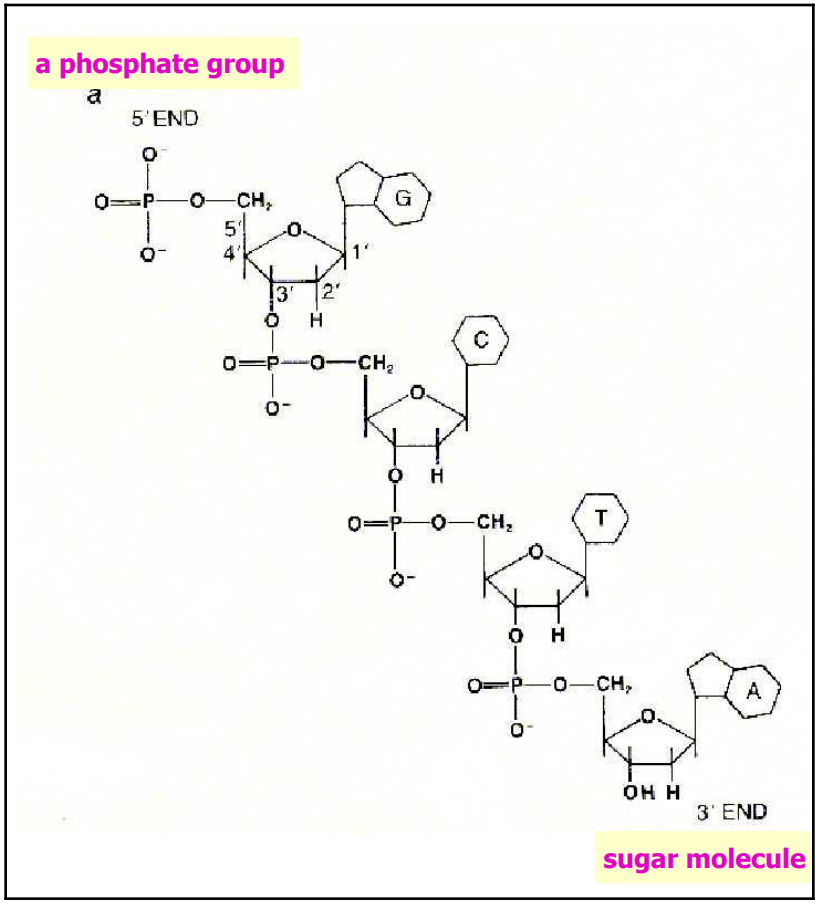


The chemical structure of uracil (RNA).

DNA structure

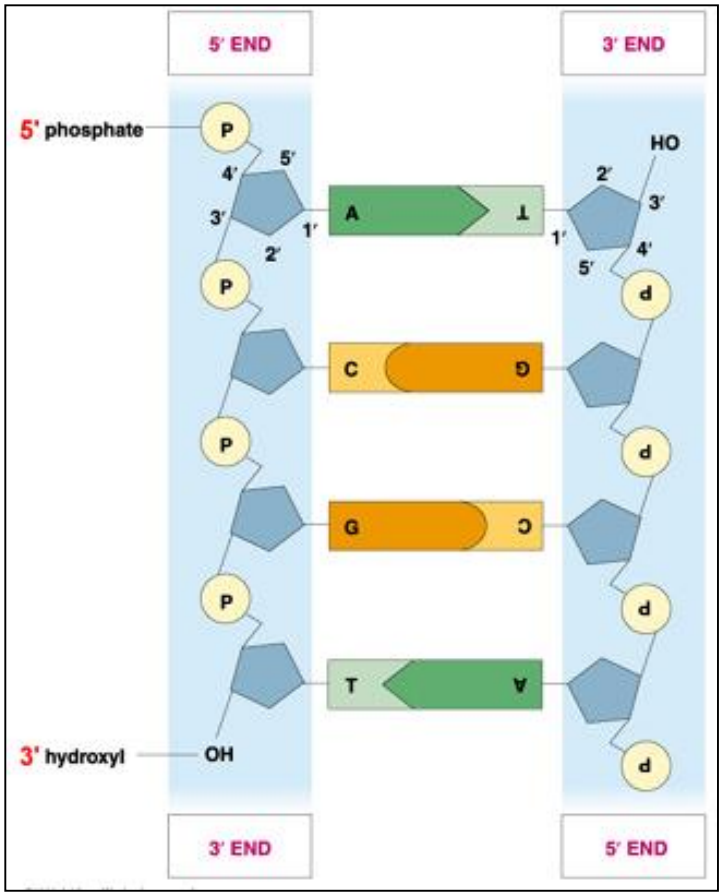
- All DNA strands are read from the 5' to the 3' end where:

1. the 5' end terminates in a phosphate group and
2. the 3' end terminates in a sugar molecule.



DNA Structure

In a DNA double helix, the strands run in opposite directions to permit base pairing between them, which is essential for replication or transcription of the encoded information

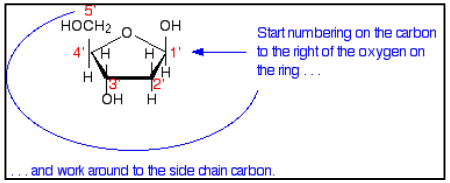


downstream

upstream

Complementary DNA (cDNA)

DNA/RNA structure
Upstream/Downstream

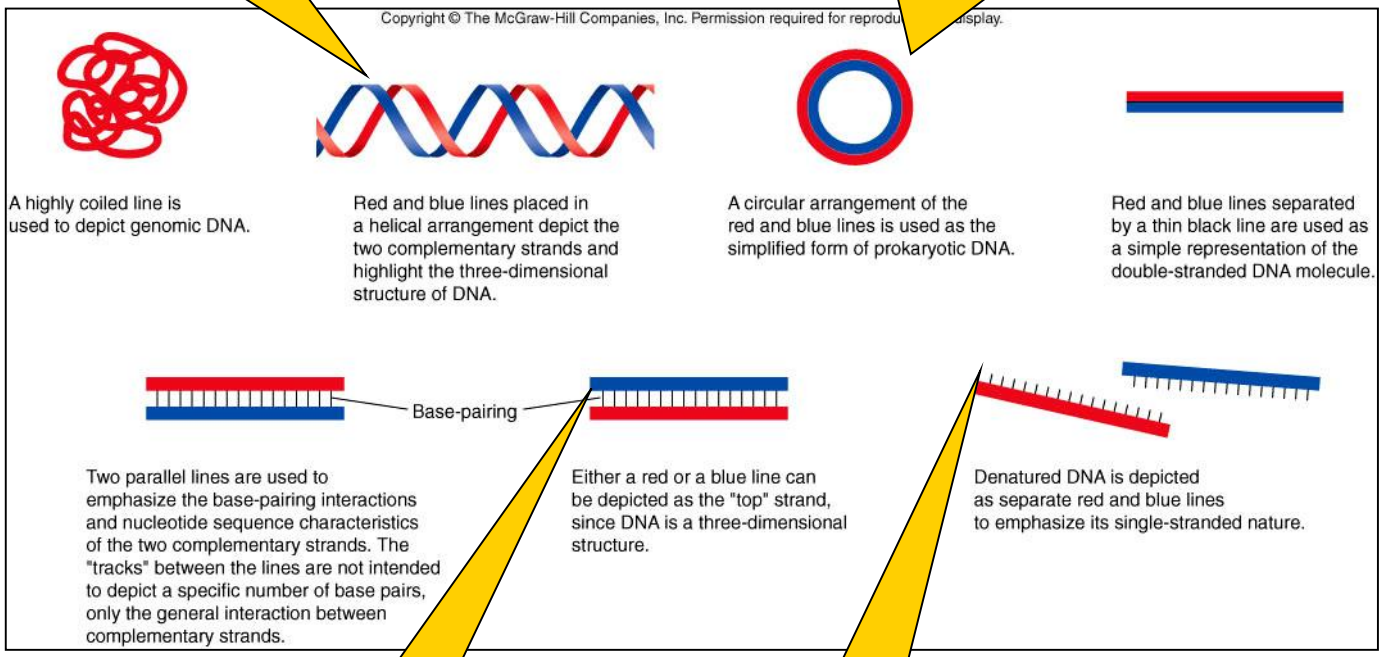


- Each strand of DNA or RNA has a 5' (pronounced five prime) end and a 3' end, so named for the carbon position on the deoxyribose (or ribose) ring.
- In DNA, the sugar is 2-deoxyribose, and in RNA, it is ribose.
- DNA and RNA are synthesized (replicated) in the 5’ to 3’ direction.
- In an RNA,
- Anything towards the 5' end of a reference point is upstream of that point. Downstream is toward the 3' end.
- In DNA,
- The situation is a bit more complicated.
- Due to the anti-parallel nature of DNA, the 3' end of the template strand is upstream of the gene and the 5' end is downstream.

Representations of DNA Structures

Double helix

Closed-circular DNA

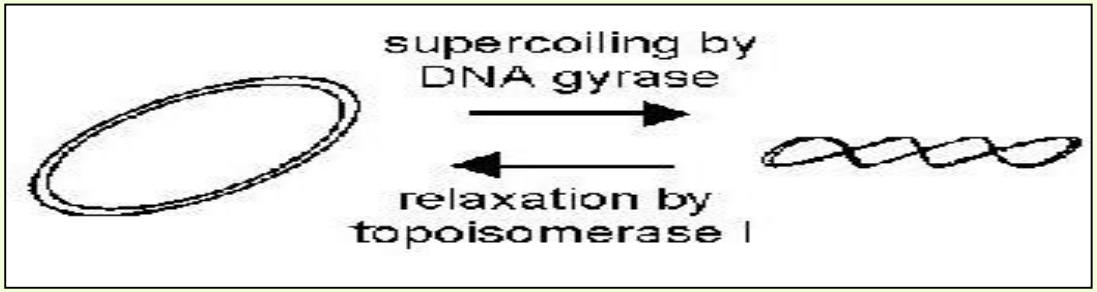


dsDNA

ssDNA

Topology of DNA

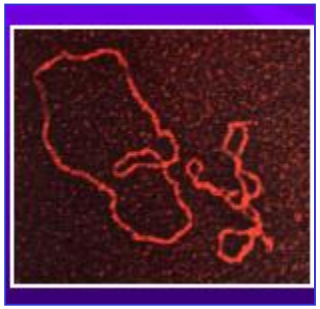
- The coiling (or wrapping) of duplex DNA around its own axis is called supercoiling.
- Supercoiling aids in the condensing of DNA into structures that can fit into the dimensions of bacteria.
- The circular chromosome is negatively supercoiled (underwound) by DNA gyrase (topoisomerase II).
- Supercoils can be relaxed by the enzyme topoisomerase I by nicking the DNA.



- Positive supercoils (twists in the opposite direction of naturally occurring negative supercoils).

2. Plasmids

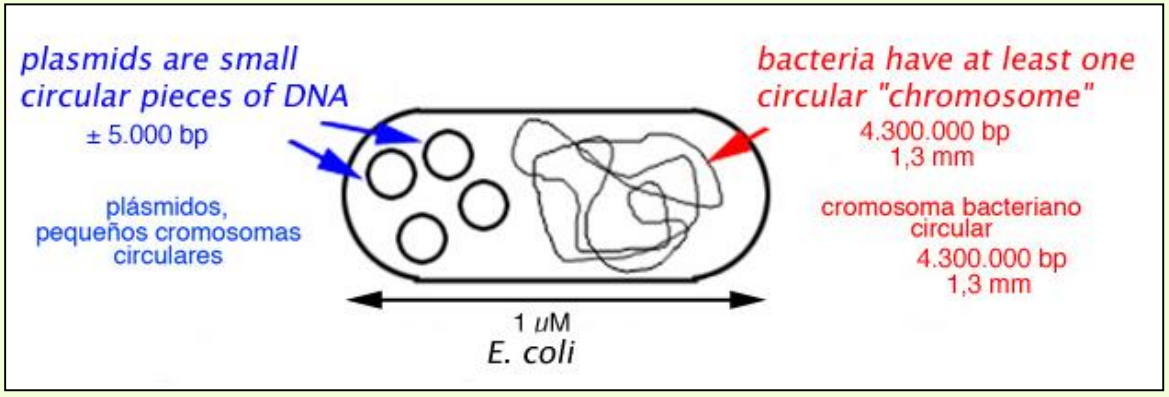
Plasmids are only valuable if the genes on them can be expressed and if the plasmid can be replicated.



How do bacteria store genetic information?

- Genetic information in bacteria is stored in the sequence of DNA in two forms, that is bacterial:
- Large plasmids are usually present in one to two copies per cell, whereas small ones may be present in 10, 40, or 100 copies.
- Within the bacterium:

1. The chromosome is long thin closed line DNA;
2. Plasmids are small to medium size circles.



Wikimedia commons; Waleed Khalid

Chromosome and plasmids
In prokaryotic and eukaryotic cells

- Many bacteria have circular plasmids and chromosomes.
- Linear plasmids and chromosomes were unknown in prokaryotes until recently but have now been found in:

1. spirochaetes,
2. Gram-positive bacteria, and
3. Gram-negative bacteria.

Hinnebusch and Tilly,1993

Chromosome and plasmids
In prokaryotic and eukaryotic cells

- Some bacteria have linear chromosomes and linear plasmids. E.g.
- *Streptomyces* have linear plasmids.
- *Agrobacterium* species contain has two chromosomes:

1. one linear, and
2. one circular.

	Bacteria	Achaea	Eukaryote
Plasmids	➢ **very common** ➢ **circular, but some linear**	**very common**	**fungi and higher plants**

Esser *et al.*,1986; Bauman,2008;..

The bacterial plasmids

Small circular mini-chromosomes with their own origin of replication

1. Found in most bacteria (Gram-ve and Gram+ve),
2. Number of lower eukaryotes (yeasts and fungi), and
3. Even in the mitochondria of some plants.
4. Ori is the DNA sequence that signals for the origin of replication, sometimes referred to simply as *origin*.
5. *E. coli* has on his chromosome a oriC for replication of its genome
6. In *E. coli*, *ori* is some 250 nucleotides in length for the chromosomal origin (*oriC*).
7. The plasmid *ori* sequences are similar to *oriC*.

Plasmids can drive gene expression in a wide variety of organisms, including plants, worms, mice and even cultured human cells. Plasmid may also be used for gene transfer into human cells as potential treatment in gene therapy so that it may express the protein that is lacking in the cells.
Animal cells don't typically carry their own plasmids.

The bacterial plasmids

Advantage and disadvantages of circular and linear plasmids

1. Circular plasmids has more efficiency than the linear one.
2. Circular DNA is digested more slowly than linear DNA within bacterial cells.

- For transformation of foreign DNA:
- *E. coli* prefer circular plasmid, while
- *Bacillus subtilis* get higher transformation efficiency with linearized plasmid.
- Linear DNA is actively taken up by the cells.
- One strand is degraded and the other is recombined into the chromosome.

In Gram-ve *E. coli*, DNA passes intact through a weakened cell envelope; closed circular plasmid reaches the cytoplasm in a state ready to begin replication, which is why it transforms better.

Plasmid vs. chromosome

How can we distinguish bacterial chromosomes from plasmids?

- **Plasmids** carry genes that are not needed under all growth conditions - that is, plasmids are dispensible under certain growth conditions.
- For example, if a plasmid encodes antibiotic resistance:

1. It may be needed when cells are exposed to that antibiotic;
2. But not needed when cells are growing in the absence of the antibiotic.

- **Chromosomes** carry genes that are essential for growth of the organism under all conditions.

Copy number is also a key parameter. The copy number of the plasmids within a cell is often not the same as chromosome.

Features of bacterial chromosome

Location	➢ **Within nucleoid region** ➢ **not surrounded by nuclear envelope**
Number	**1 chromosome each cell**
Size	***E. coli* 4640 kbp (4.6 million base pairs)**
Component	➢ **Single, double stranded, circular DNA (some linear)** ➢ **Also contains RNA and proteins that take part in DNA replication, transcription and regulation of gene expression**
Information	**Contain genes essential for cellular functions**

Features of bacterial plasmids

Location	**Plasmids are found in cytoplasm and never integrate with the bacterial chromosome**
Number	**From 1 to several**
Size	➢ **Much smaller than chromosomes** ➢ **Most are between 1,000 to 25,000 base pairs**
Component	**Single, double stranded, circular DNA (some linear)**
Information	➢ **Contains drug resistant genes as well as heavy metal resistant genes.** ➢ **Not essential for growth and metabolism of bacteria**

The bacterial plasmids

Cryptic plasmids

- Plasmids are not essential to their bacterial host.

1. Plasmids can be easily gained or lost by a bacterium,
2. Plasmids can be transferred between bacteria as a form of horizental gene transfer.

- Plasmids have been identified in a large number of bacterial genera.
- Some bacterial species harbor plasmids with no known functions (cryptic plasmids) which have been identified as small circular molecules present in the bacterial DNA.

Plasmid size

- In general, bacterial plasmids can be classified into two groups on the basis of:

1. the number of genes, and
2. functions they carry.

- Plasmids vary in size.
- Smallest plasmid is only 846 bp long and contains only one gene.
- Some plasmids carry more than 500 genes.
- The largest plasmid (megaplasmids) carries 1,674 genes. E.g. 2.1-Mb megaplasmid in the case of *R. solanacearum*.

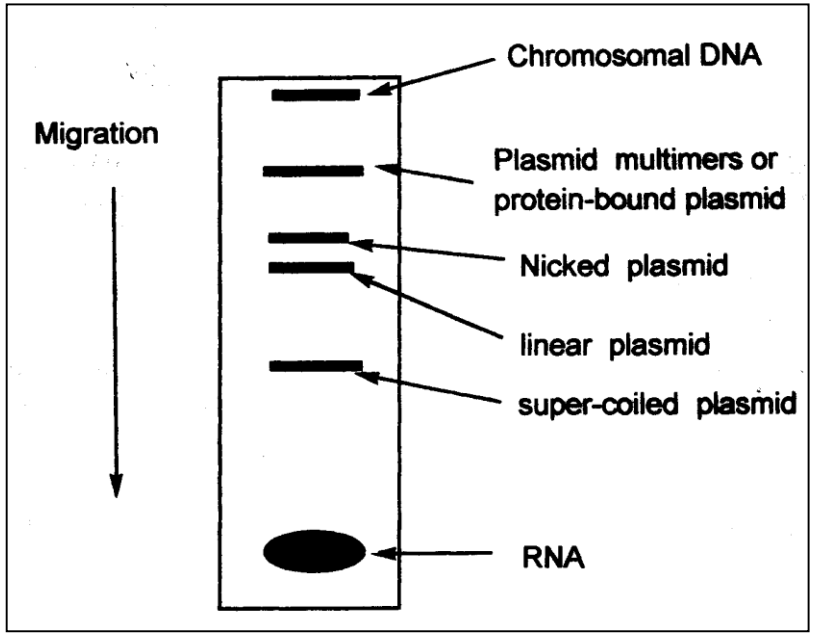
Rao,2012;..

Agarose Gel profile

Migration of Various DNA Species in Our System

A schematic diagram of an agarose gel separation

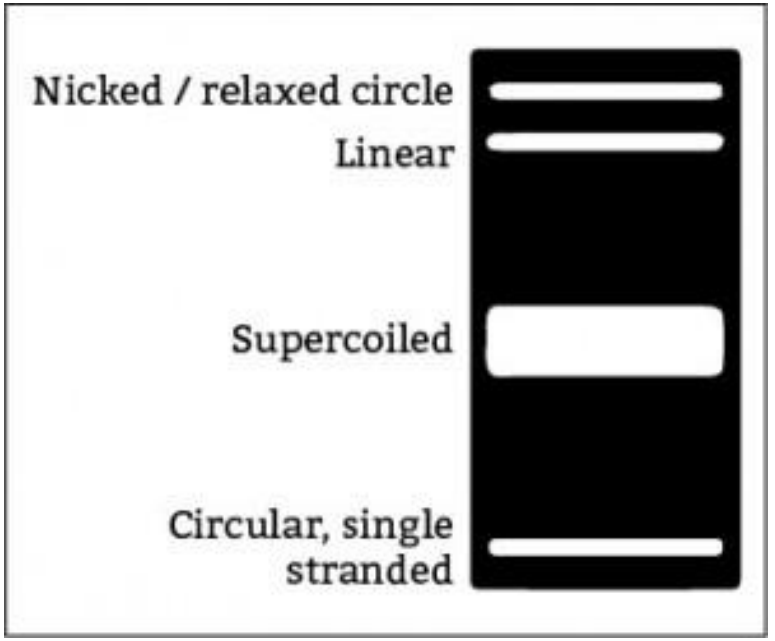
1. Supercoiled DNA is the most compact form of DNA and will consequently migrate more rapidly than other forms of DNA of the same or larger molecular weight.
2. Nicked relaxed Plasmid DNA has an intermediate mobility under our electrophoretic conditions.
3. Chromosomal DNA is large and will migrate the least distance.



Plasmid size

Supercoils and relaxed forms can be seen under the electron microscope, in electrophoresis or after centrifugation

- In bacteria, plamids are found in the supercoil form.
- Topoisomerases help to pass from one form to the other.
- When plasmid DNA is isolated and run on an agarose gel, you may observe 2, 3 or even 4 or more bands.
- The supercoiled form migrate quicky in eletrophoresis.
- Linear DNA generally migrates between the nicked/relaxed circle and the supercoiled forms.



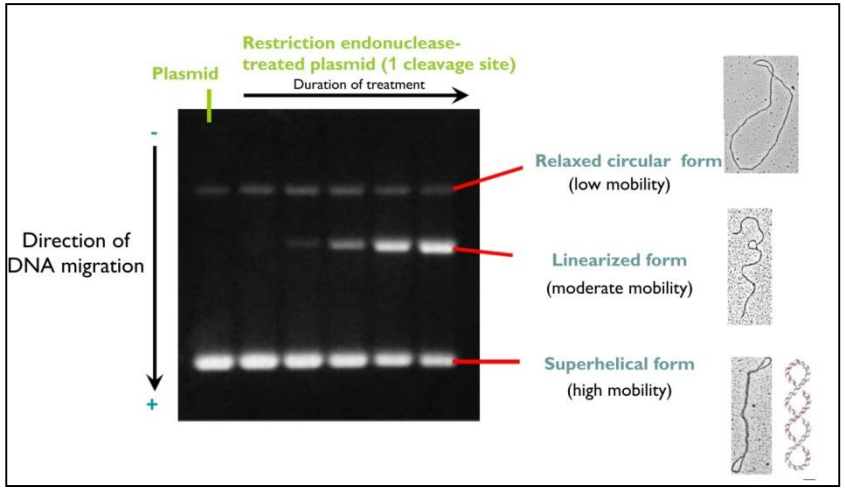
Tirabassi,2014

Plasmid size and shape

Mobility of DNA molecules

Gel electrophoretic image of plasmid DNA

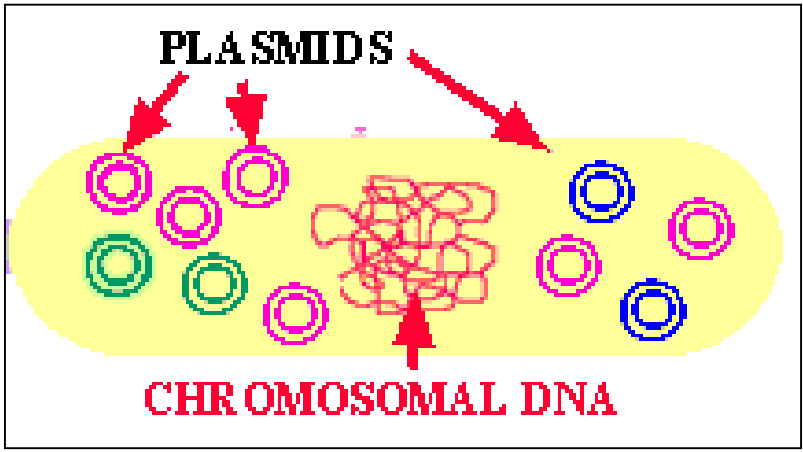
- Besides their size, the electrophoretic mobility of DNA molecules is also significantly affected by their shape.
- Circular plasmid DNA has a compact structure, and its hydrodynamic size is much smaller—and its electrophoretic mobility is therefore greater—than that of linear DNA molecules of the same size, as the latter form a freely moving entropic chain.



Hegyi *et al.*,2013

Number of plasmids

- A bacterial cell may contain:
1. no plasmids, one plasmid or
2. many copies of a plasmid.
- The number of copies of the plasmid would depend of its size and the characteristics of its replication origin.



Some bacteria can survive without plasmids. e.g *Brucella*, *Rickettsia*, and some endosymbionts. Some other may carry megaplasmids that acts more than little chromosomes. e.g. *R. solanacearum*.

Numbers of plasmids

Copy number of plasmids

- Copy number refers to the number of copies of a plasmid present in a cell.

1. Larger plasmids are present in smaller numbers (1-2), and
2. Small plasmids may be found in high copy numbers (~40).

Common feature of all plasmids

Independently plasmid replication

Replicon

- One common feature of all plasmids is a specific sequence of nucleotides termed an origin of replication (ori).
- Ori is the DNA sequence that signals for the origin of replication.
- This sequence, together with other regulatory sequences, is referred to as a replicon.
- The replicon allows a plasmid to replicate within a bacterial cell independently of bacterial chromosome.
- Replicons controls replication in the cell.

Common feature of all plasmids

Independently plasmid replication

Replicon

- Plasmids can function in the absence of ongoing protein synthesis = absence of host (bacterial cell) replication.
- If you inhibit protein synthesis with antibiotics (chloramphenicol, spectinomycin), these antibiotics bind to ribosomes and hence, shut down protein synthesis.
- Since DNA(plasmid) replication requires polymerase (long half life), DNA synthesis continues while cell growth does not.

Common feature of all plasmids
Replicon

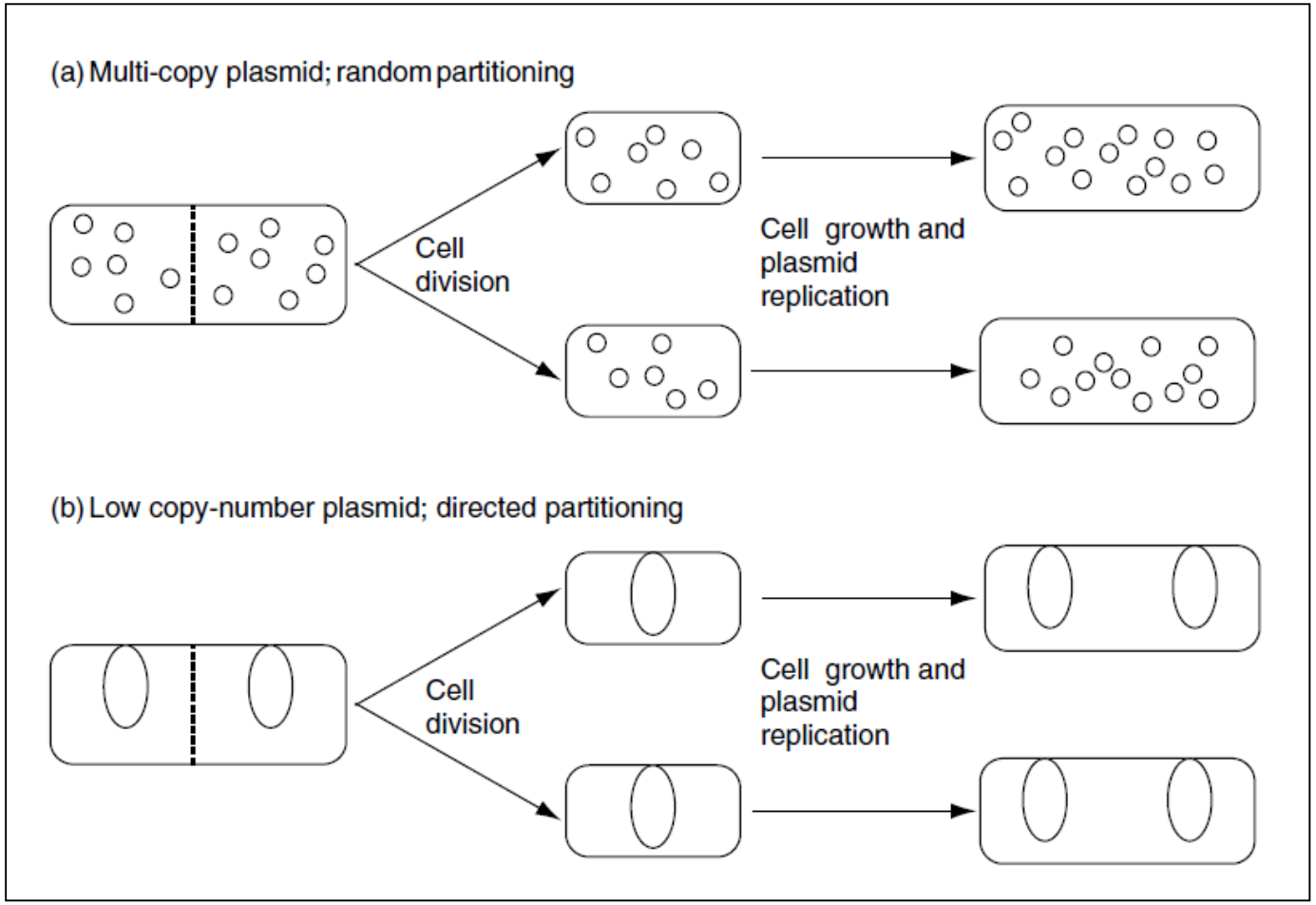
- In *E. coli*, *ori* is some 250 nucleotides in length for the chromosomal origin (*oriC*).
- The plasmid *ori* sequences are similar to *oriC*.

1. If the plasmid makes many copies of itself per cell, it is termed a "relaxed" plasmid. Plasmids such as pMB1 and ColE1 (multicopy plasmids) replicate in a "relaxed" fashion.
2. If it maintains itself in fewer numbers within the cell it is termed a "stringent" plasmid.

Note: Bacteria have a single origin for replication(*ori*).
The *E. coli* replication origin is known as oriC.
Eukaryotes have multiple replicons, each with an ori.

Partitioning of plasmids at cell division

Relaxed and stringent plasmids



Dale and Park,2004

Bacteria with or without plasmids

Bacteria can survive without plasmids

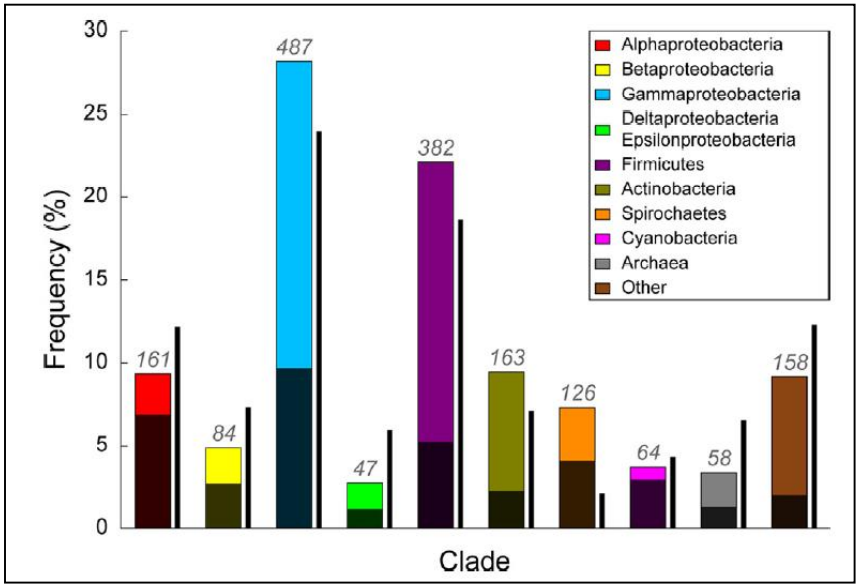
- Plasmids under natural circumstances is unavailable.

1. Absence of plasmids doesn't kill bacterium,
2. But their presence provides additional benefits to the bacterial cell.

Rao,2012

Plasmid database with respect to prokaryotic phyla

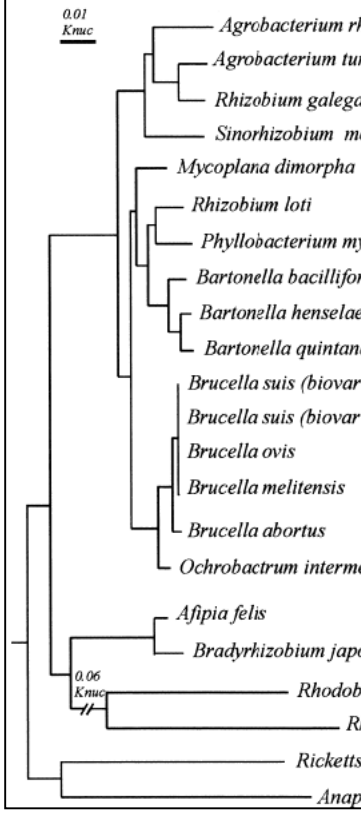
- Distribution of the 1,730 GenBank plasmids by taxon:
- By early 2009 there were 1,730 complete plasmid sequences.
- Bars are topped by the numbers of plasmids for the clades.
- The thin bars represent the proportion of plasmids of each clade among the plasmids sequenced with the prokaryotic host.



Smillie *et al.*,2010;..

Bacteria with or without plasmids

Bacteria can survive without plasmids



	Plasmid size in kb	Major chromo-some in kb	Minor chromo-some in kb	G+C %	Preferred habitat and host
Agrobacterium rhizogenes	200-600	4000	2700[a]	62	Soil and pericellular pathogen of dicotyledones[b]
Agrobacterium tumefaciens	200-500	3000	2100[c]	62	Soil and pericellular pathogen of dicotyledones[b]
Rhizobium galegae	400-600	5850	1200[a]	63	Soil and intracellular symbiont of legumes
Sinorhizobium meliloti	500-140	3400	1700[d]	63	Soil and intracellular symbiont of legumes
Mycoplana dimorpha	100-300	3200	none	64	Soil free living
Rhizobium loti	400-600	5500	1200[a]	62	Soil and intracellular symbiont of legumes
Phyllobacterium myrsinacearum	250-700[e]	5330	none	60	Soil and peri- and intracellular pathogen of sugar beets
Bartonella bacilliformis	none	1540	none	40	Peri- and intracellular pathogen of humans and sand flies
Bartonella henselae	none	1400	none	41	Intracellular pathogen of cats, humans and lice
Bartonella quintana	none	1700	none	39	Intracellular pathogen of humans and lice
Brucella suis (biovar 3)	none	3100	none	57	Intracellular pathogen of suines
Brucella suis (biovar 2 and 4)	none	1850	1350	57	Intracellular pathogen of suines and reindeer
Brucella ovis	none	2100	1150	57	Intracellular pathogen of ovines
Brucella melitensis	none	2100	1150	57	Intracellular pathogen of caprines
Brucella abortus	none	2100	1150	57	Intracellular pathogen of bovines
Ochrobactrum intermedium	50-200	2700	1900	59	Soil and opportunistic of humans and nematodes
Afipia felis	40-100	Unknown	Unknown	63	Opportunistic of cats and humans
Bradyrhizobium japonicum	200-800[e]	8700	none	65	Soil and intracellular symbiont of legumes
Rhodobacter sphaeroides	40-200	3046	914	65	Water free living photoheterotroph
Rhodobacter capsulatus	100-200	3800	none	66	Water free living photoheterotroph
Rickettsia rickettsii	none	1300	none	33	Intracellular pathogen of rodents and ticks
Anaplasma marginale	none	1200	none	51	Intracellular pathogen of rodents and ticks

Moreno,1998

Bacteria with or without plasmids

General features of the *Xanthomonas oryzae* pv.*oryzae* genome

No plasmid was detected in the course of genome assembly

- The assembled sequence was consistent with a single, 4 941 439 bp, circular chromosome.
- No autonomous plasmids were apparent.

Length (bp)	4 941 439
G + C content (%)	63.7
Protein coding genes	
With function assigned	3340
Conserved hypothetical	1151
Hypothetical	146
Total	4637
Transfer RNA	54
Ribosomal RNA operons	2
Plasmids	0
Insertion sequence element (IS)	207

Lee *et al.*,2005

Bacterial plasmids
Transferable plasmids

- Transferability is the ability of certain plasmids to move from one bacterial cell to another.
- Not all plasmids are transferable. Most very large plasmids are nonmobilizable.
- Half of the plasmids are nonmobilizable and that half of the remaining plasmids are conjugative.
- Only those plasmids that have genes for self transfer can be transferred.
- E.g.
- Transferable antibiotic resistance plasmids;
- F-type and R-type plasmids.

Smillie *et al.*,2010;..

Bacterial plasmids
Transferable plasmids

- Plasmids are passed on to other bacteria through two means:

1. For most plasmid types, copies in the cytoplasm are passed on to daughter cells during binary fission.
2. Other types of plasmids, however, form a tubelike structure at the surface called a pilus that passes copies of the plasmid to other bacteria during conjugation, a process by which bacteria exchange genetic information.

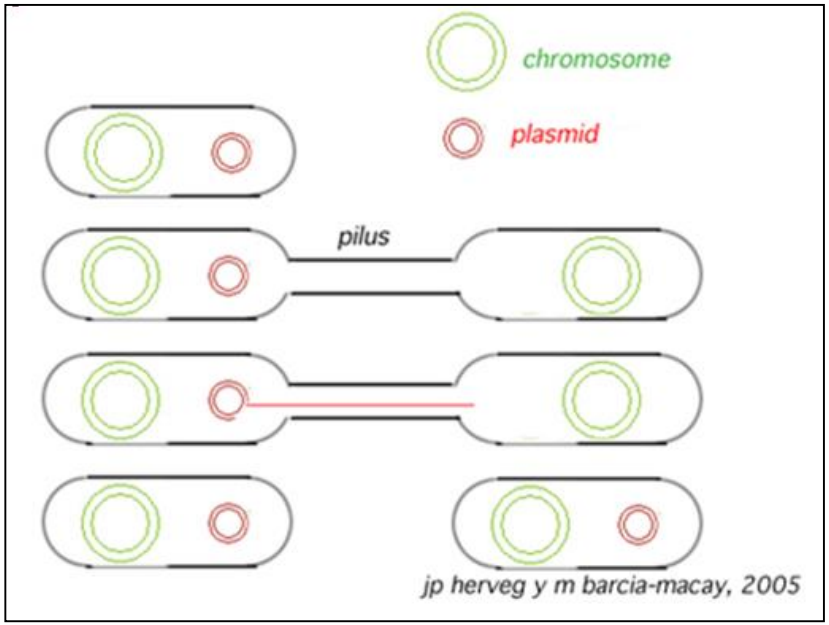
Davidson,2009

Bacterial plasmids

Transferable plasmids

Conjugative plasmids

- They possess the genes required for transfer.
- The sexual transfer of plasmids to another bacterium through a pilus.
- Those plasmids, plasmid F, possess the 25 genes required for transfer.



Smillie *et al.*,2010;..

Bacterial plasmids
Transferable plasmids

- It is not impossible to have horizontal genes transfer between plasmid and chromosomes, and plasmids (alongside different replicons, including bacteriophage, and mobile genetic elements including transposons etc.) are important sources of new genetic material.
- Although, it is not clear whether plasmids evolved from:

1. ancient bacterial chromosome, or
2. viral replication mechanisms.

Plasmid stability
Antibiotic resistance

- In nature, plasmids often carry genes that may benefit the survival of the organism, for example antibiotic resistance.
- Are plasmid genes always expressed?
- If so, then isn't a bacterium wasting it's resources in expressing genes (like antibiotic resistance) which are not required in "normal" conditions?

Biology Stack Exchange,2015

Plasmid stability
Antibiotic resistance

- One of the characteristic features of plasmids is their instability.
- Plasmid-borne features are often lost from a population at a higher frequency than would be expected for the normal processes of mutation.
- However, plasmids may contain genes that confer an important advantage to the population (like in presence of an antibiotic), and if so the plasmid spread. If at some point the ecological pressure cease, the plasmid would turn unstable again.
- The extent of this instability varies enormously from one plasmid to another.

Host range of plasmids

Most plasmids from Gram-negative bacteria have a narrow host range

1. The host range of a particular plasmid is usually limited to closely related genera.
2. Some plasmids, however, are much more promiscuous and have a much broader host range.

- The RK2 plasmids(vectors) have a broad-host-range.
- It is notable for its ability to replicate in a wide variety of single-celled organisms, which makes it suitable as a genetic engineering tool.
- It is capable of transfer, replication, and maintenance in most genera of Gram-negative bacteria.

Types of plasmids
Functions of plasmids

- Plasmids are key vectors of horizontal gene transfer and essential genetic engineering tools.
- They code for genes involved in many aspects of microbial biology, including:

1. Detoxification or detoxication (detox for short),
2. virulence,
3. ecological interactions, and
4. antibiotic resistance: plasmids carrying antibiotic resistance genes produce proteins (enzymes) to confer its antibiotic resistance.

Smillie *et al.*,2010;..

Types of plasmids
Functions of plasmids

- **I. ALL Plasmids:**
  1. Self-replication
- **II. By Some Plasmids:**
  1. Self-transfer
  2. Resistance to antimicrobial agents
     - Antibiotics
     - Chemotherapeutic agents (sulfa drugs)
     - Heavy metals
  3. Pigment production
  4. Toxin production (pathogenicity)
  5. Phage sensitivity
  6. Antibiotic production
  7. Toxin production against closely-related bacterial strains (bacteriocin production)
  8. Induction of plant tumors
  9. Hydrogen sulfide production
  10. Catabolic functions (petroleum fraction degradation)

Types of plasmids

Based on their function plasmids can be classified into many types as:

- Resistant plasmids: They encode enzymes for resistant against antibiotics, heavy metals, radiation, etc. These are large conjugative plasmids that carry one or more antibiotic resistance genes. Resistance plasmid can be conjugative or mobilizable.

- Conjugative plasmids: Plasmids that can be transferred via conjugation is called conjugative plasmids. Cells possessing it (donor) are termed F^+ and those lacking it (recipient) are termed F^-. e.g. F^+ Plasmid - fertility factor (maleness).

- Metabolic plasmids: involved in the production of metabolic enzymes. e.g. 21.5 kb plasmid-encoded lactose fermentation and proteolytic activity properties in *Lactococcus lactis* ssp.

Types of plasmids

Based on their function plasmids can be classified into many types as:

- Bacterocinogenic plasmids: Plasmids those produce bacterocins those act against other bacteria. e.g.
  1. subtilicins by *Bacillus subtilis* and
  2. colicin plasmids, small plasmids which encode the genes to synthesize colicins (bacteriocines).

- Integrative plasmids: These are plasmids that can occasionally integrate into chromosome (previously called episomes). e.g. In *Bacteroides uniformis*, certain sections of chromosome separate themselves from chromosome and become plasmids, which are capable of conjugation.

If plasmids were generated through the breakage of chromosomes, chromosomes would need to contain multiple replicons that could then support each of the new plasmids - but there is no evidence of multiple-replicon chromosomes seen in today's bacteria.

Types of plasmids

Based on their function plasmids can be classified into many types as:

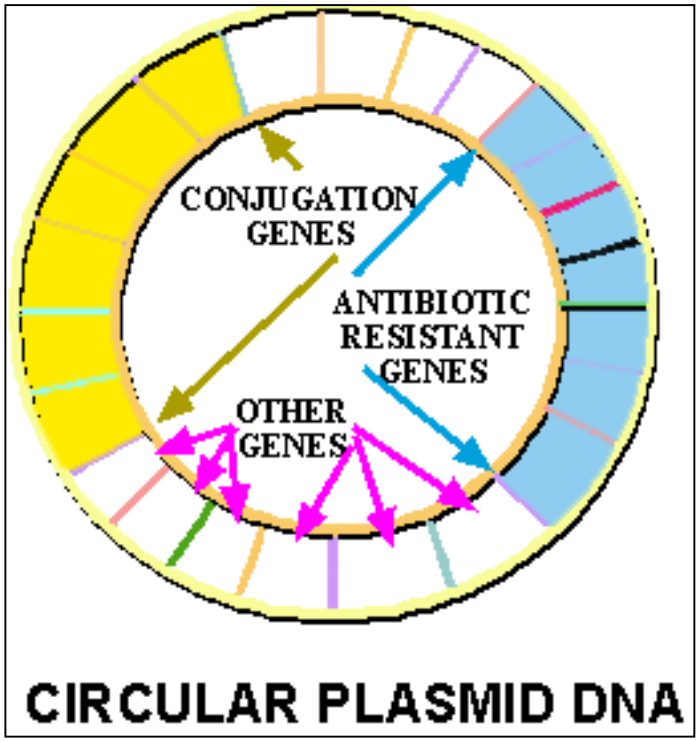
- Virulent plasmids: turn the bacterium into a pathogen.
  Virulence plasmids carry one or several genes that confer virulence properties to pathogens. e.g. by producing toxin or agents that damage host tissues.

- Suicide plasmids: are referred to those plasmids which get transferred to another bacterial cell but does not replicate further. e.g. vectors have also been considered that are suicidal, such that they could not survive transfer to another organism.

Types of plasmids

Fertility plasmid carrying antibiotic resistant genes

- A fertility plasmid carries the genes for conjugation as well as a number of other genes.
- In this figure, the fertility plasmid also carries antibiotic resistant genes.



Hurlbert,1999

Resistant plasmids
How do bacteria become resistant?

- Higher-levels of antibiotic-resistant bacteria are attributed to the overuse and abuse of antibiotics.
- Bacteria may also become resistant in two ways:

1. by a genetic mutation, or
2. by acquiring resistance from another bacterium.

- Some mutations enable the bacteria to produce potent chemicals (enzymes) that inactivate antibiotics, while other mutations eliminate the cell target that the antibiotic attacks.

Resistant plasmids

How does antibiotic resistance spread?

- Genetically, antibiotic resistance spreads through bacteria populations both:

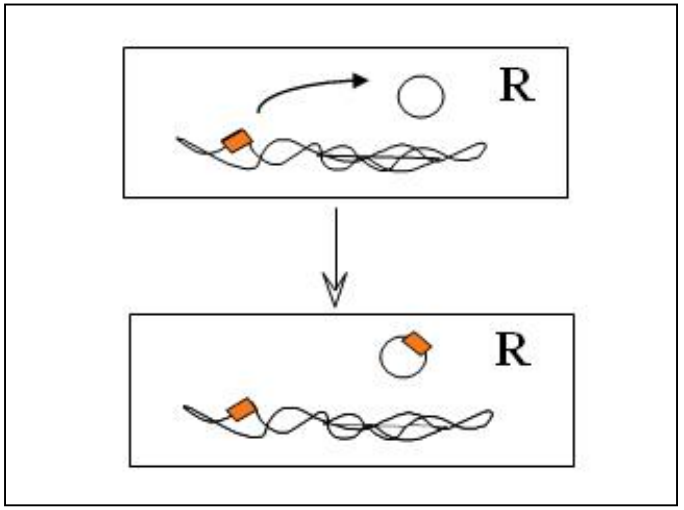
1. vertically, when new generations inherit antibiotic resistance genes, and
2. horizontally, when bacteria share or exchange sections of genetic material with other bacteria.

- Horizontal gene transfer can even occur between different bacterial species.
- Environmentally, antibiotic resistance spreads as bacteria themselves move from place to place; bacteria can travel via airplane, water and wind.

Resistant plasmids
Translocation of resistance gene

- Movement of an antibiotic resistance gene (shown in orange) from the chromosome of a bacterium to a plasmid can facilitate its spread because plasmids can transfer into other bacteria making them resistant as well.
- This process is called translocation.

Antibiotic resistance



CSIRO pedia, University of Waikato,2015

Agrobacterium infection process

The integration of T-DNA region of the Ti plasmid (T-DNA) in plant chromosomal DNA

- The transfer and integration of bacterial T-DNA into plant cells result in tumorogenesis, leading to crown gall disease in plants.

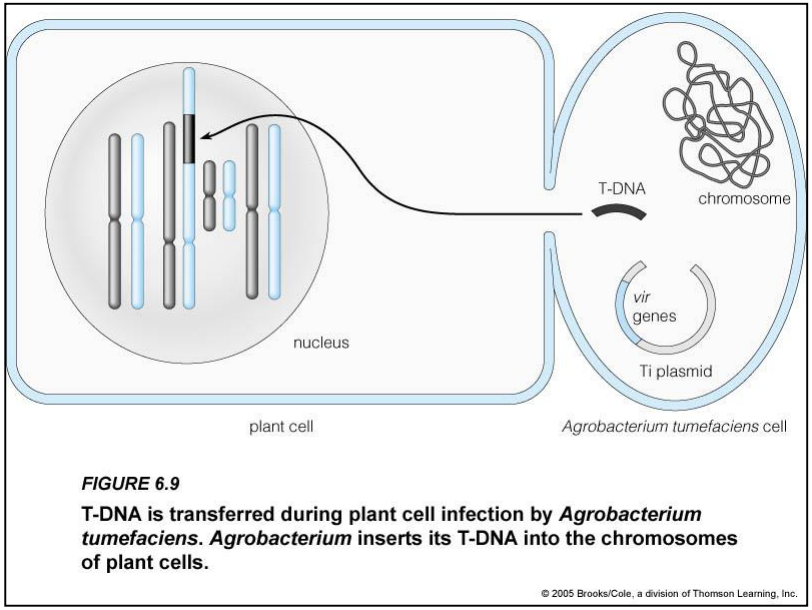
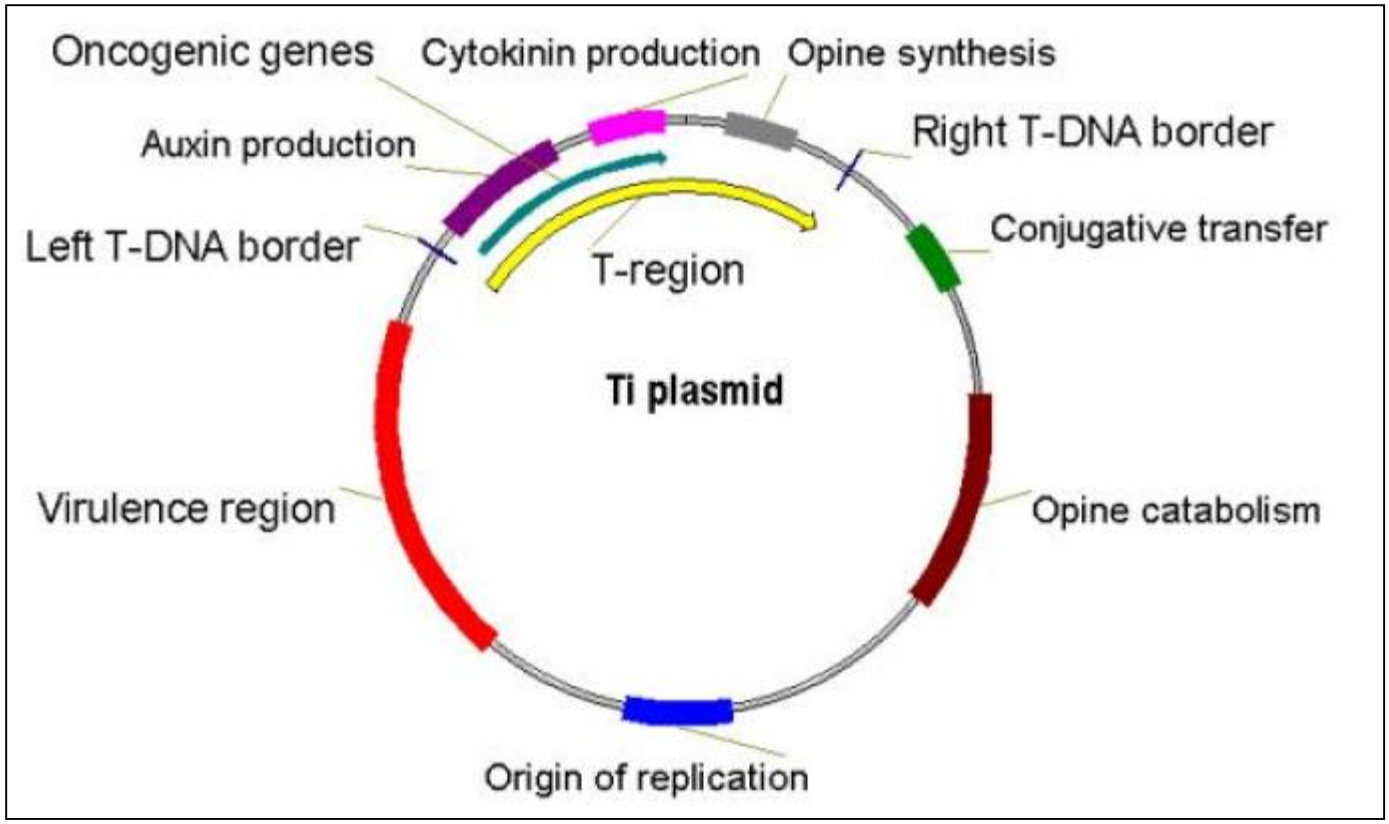


FIGURE 6.9

T-DNA is transferred during plant cell infection by *Agrobacterium tumefaciens*. *Agrobacterium* inserts its T-DNA into the chromosomes of plant cells.

© 2005 Brooks/Cole, a division of Thomson Learning, Inc.

Structure of tumor-inducing plasmid
Ti plamid



An example of Selfish Genes.
A. tumefaciens **genetically engineers plants to make specialized food for it.**

Types of plasmids

Phytopathogenic strains of *P. syringae* containing plasmids

- All strains of *Pseudomonas syringe* pathovars contain the same series of replicons, from the largest chromosome, and smaller plasmids.

Pathovar

- *P. syringae* pv. *angulata*
- *P. syringae* pv. *atrpurpurea*
- *P. syringae* pv. *coronafaciens*
- *P. syringae* pv. *glycinea*
- *P. syringae* pv. *lachrymans*
- *P. syringae* pv. *papulans*
- *P. syringae* pv. *phaseolicola*
- *P. syringae* pv. *savastanoi*
- *P. syringae* pv. *striafaciens*
- *P. syringae* pv. *syringae*
- *P. syringae* pv. *tabaci*
- *P. syringae* pv. *tomato*

OECD,1997;..

Symptoms associated with phytopathogenic bacterial groups

Plasmid-borne genes to the host-pathogen relationship

Symptom	Bacterium	Host plant	Plasmid traits
Blights, cankers, knots	*Ps. syringae* pathovars	Pathovar-specific ranges	*avr*, *vir*, coronatine, Sm^r, Tc^r, Cu^r, UV^r, hormones
Blights	*X. campestris* pathovars	Pepper, cotton and other hosts	*avr* genes, Cu^r, Sm^r, Tc^r
Galls	*Pantoea gypsophilae*	*Gypsophila* (flowering plants in the carnation family)	*hrp* and pathogenicity genes, hormones
Necrosis	*Er. amylovora*	Apple, pear and related ornamentals	Thiamin (Vitamin B1) and EPS biosynthesis
Tissue soft rot	*B. cepacia*	Onion	Endopolygalacturonase production
Tissue soft rot of storage organs	*P. carotovorum*	Wide range, including potato	Cryptic
Wilt	*R. solanacearum*	Potato, tomato, banana, aubergine	Virulence
Wilt	*Pa. stewartii*	Corn	Cryptic

Vivian *et al.*,2001

Types of plasmids

Plasmid-borne coronatine genes in *Ps. syringae* and *Ps. savastanoi*

Pathovar	Plasmid
glycinea	p4180A
atropurpurea	pCOR1
maculicola	pMAC1
morsprunorum	ND
tomato	pPT23A

nd, Not designated.

Vivian *et al.*,2001

Types of plasmids

Phytopathogenic bacteria

Plasmid-borne genes for plant hormone production

Gene	Pathogen	Plasmid	Hormone
efe	*Ps. savastanoi* pv. *glycinea* *Ps. savastanoi* pv. *phaseolicola* *Ps. syringae* pv. *cannabina*	pETH2 pPSP1 pETH1	Ethylene Ethylene Ethylene
etz	*Pantoea gypsophilae*	pPATH	Cytokinin
iaaH	*Pantoea gypsophilae* *Ps. savastanoi* pv. *savastanoi*	pPATH pIAA2	IAA IAA
iaaL IaaM	*Ps. savastanoi* pv. *savastanoi* *Pantoea gypsophilae* *Ps. savastanoi* pv. *savastanoi*	pIAA1 pPath pIAA2	IAA IAA IAA
ipt	*Ps. savastanoi* pv. *savastanoi*	ND	Cytokinin
ptz (*ipt*)	*Ps. savastanoi* pv. *savastanoi*	pCK2*	Cytokinin

nd, Not designated (detected on a plasmid of 44 kb in oleander, and on plasmids ranging from 35 to 97 kb in olive strains).
* Described as pCK1 by Powell & Morris (1986), but as pCK2 by MacDonald *et al.* (1986).

Vivian *et al.*,2001

Types of plasmids

Different types to coexist in a single cell

Compatible and incompatible plasmids

1. Different types plasmids can coexist in a single cell.
2. Several different plasmids have been found in *E. coli*.

- The number of plasmids in a cell is governed by elements encoded within the origin of replication (*ori*).
- Competition for replication factors leads to competition between plasmids.

1. Plasmids are compatible if they can coexist and replicate within the same bacterial cell.
2. Plasmids are incompatible, if they can not coexist and replicate within the same bacterial cell.

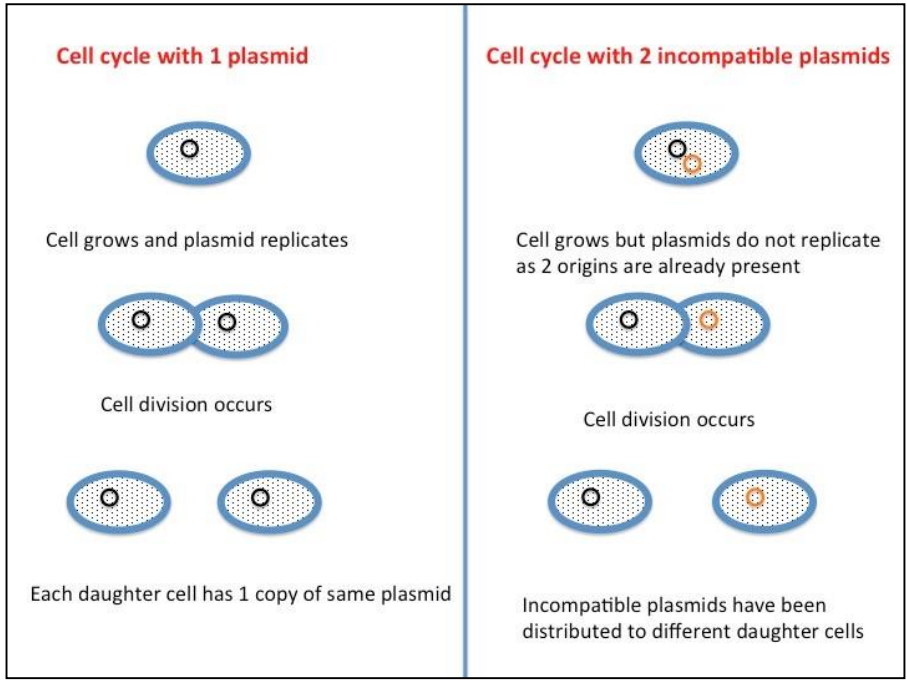
Boundless,2015,..

Types of plasmids

Different types to coexist in a single cell

Compatible and incompatible plasmids

- **Cell with single plasmid:**
- Cell grows and plasmid replicates in two (without any competition for Ori), each daughter cell has 1 copy of same plasmid.
- **Cell with two plasmids** with the same type of Ori (incompatible plasmids)**:**
- Plasmids will compete with each other for replication(Ori).
- Only one of them survives in daughter cell, and the other eliminated.
- Therefor, each cell has only one copy of different plasmids.



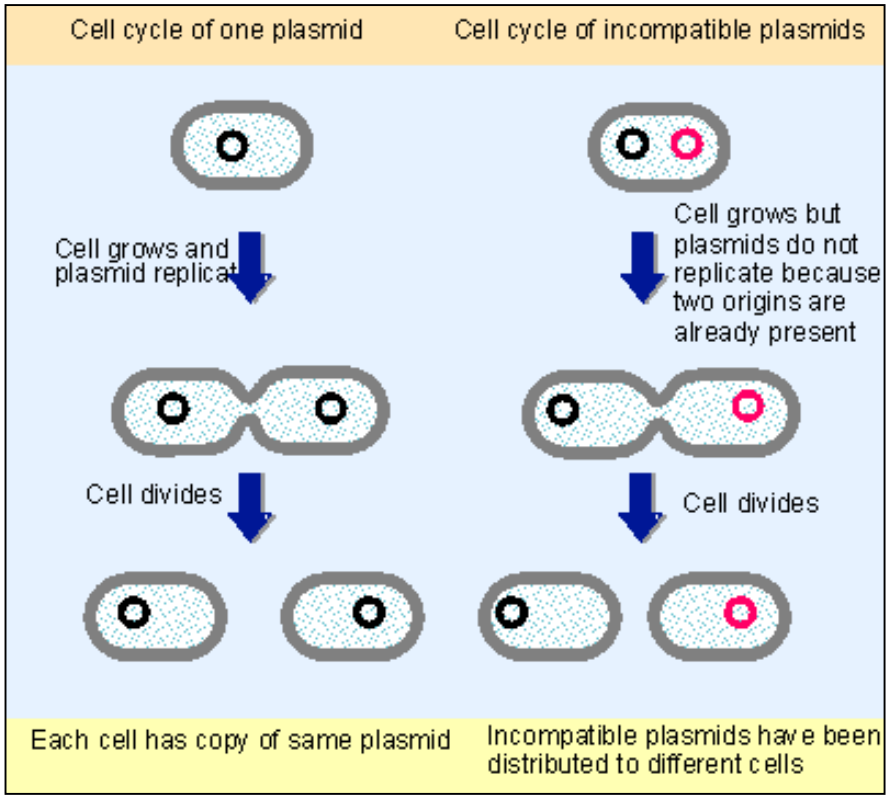
Bitesize Bio;..

Types of plasmids

Different types to coexist in a single cell

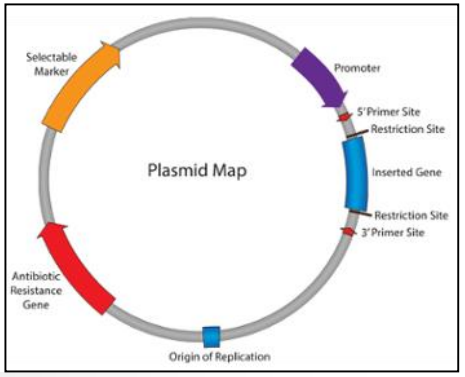
Compatible and incompatible plasmids

- Two plasmids are incompatible if their origins (origin of replication,ori) cannot be distinguished at the stage of initiation.
- The reason for their incompatibility is that they cannot be distinguished from one another at some stage that is essential for plasmid maintenance.



Conjugative plasmids

Conjugational transfer of DNA



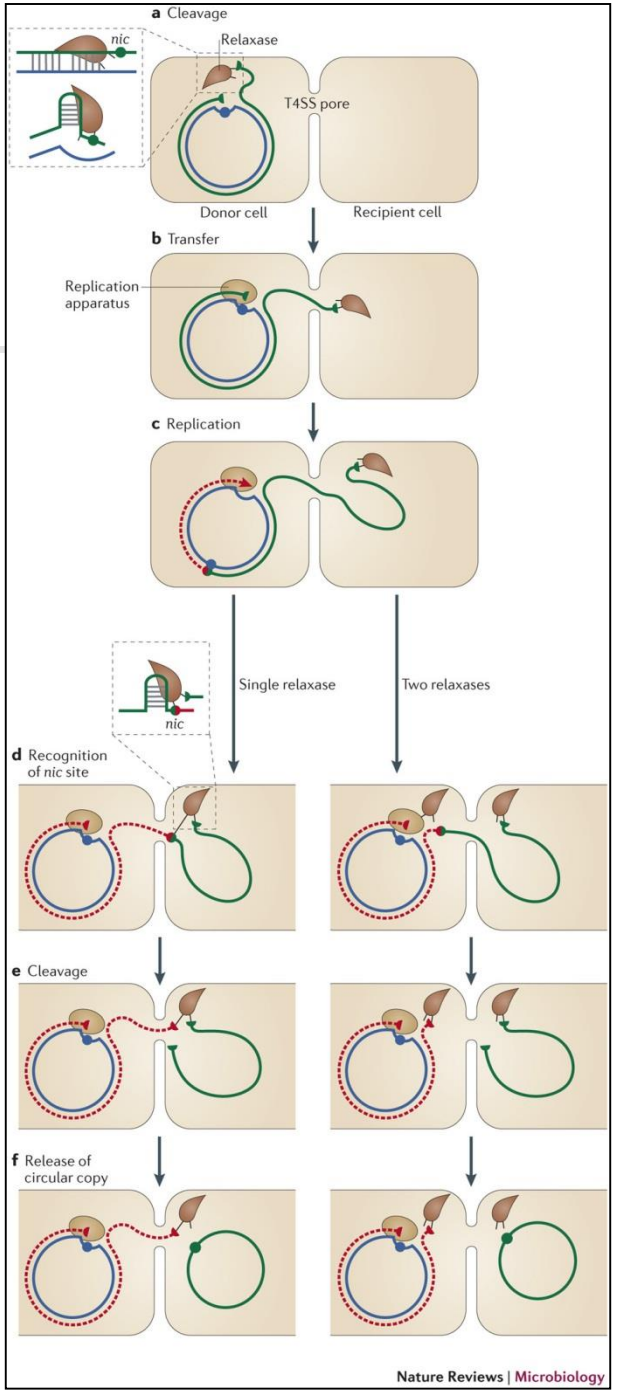
- Thus plasmid replication process can be divided into three stages:

1. initiation,
2. elongation, and
3. termination.

Miljkovic-Selimovic *et al.*,2007

Conjugative plasmids
Conjugational transfer of DNA

1. Relaxases recognize the origin of transfer (*oriT*) of plasmids.
2. The relaxase nicks the double stranded plasmid DNA and linearizes it.
3. The *oriT*/relaxase DNA component gets recruited to the mating pore by the T4CP.
4. Transfer of DNA proceeds through the transport channel synthesized by the T4SS.



Nature Reviews Microbiology,2013;..

3. Bacteriophages

The viruses that infect bacteria

Temperate/virulent

Structure of bacteriophages
T-phages

- The T-phages, T1 through T7, are referred to as lytic phages because they always bring about the lysis and death of their host cell, the bacterium *E. coli*.
- T-phages contain:

1. Double-stranded DNA as their genetic material.
2. Protein coat or capsid (also referred to as the "head"),
3. A tail, and
4. Some related structures.

Todar,2012

Bacterial viruses or bacteriophages

A diagram and electron micrograph of a bacteriophage

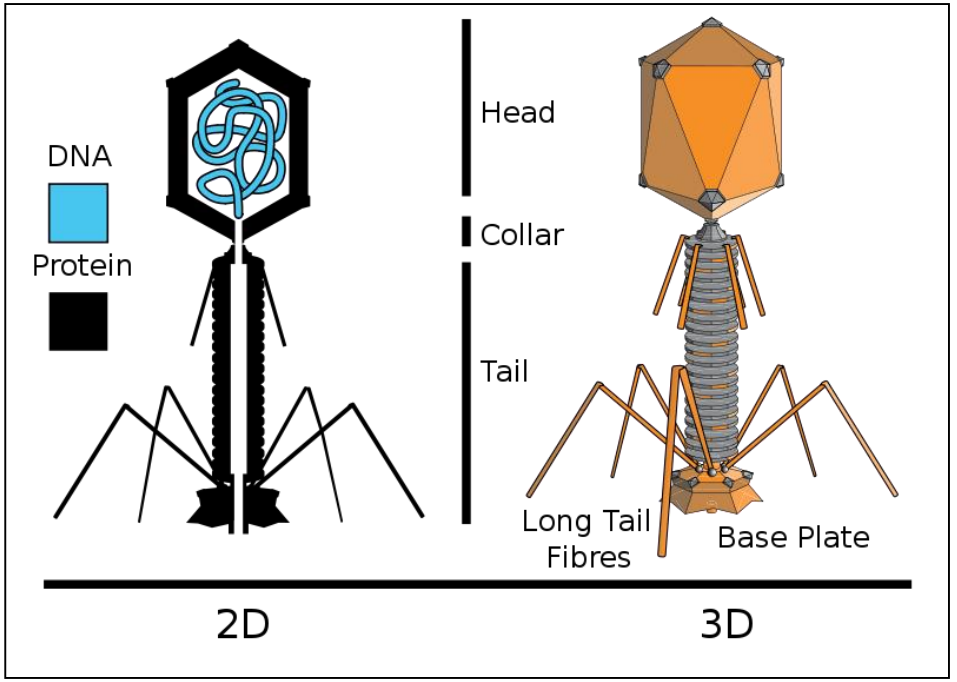
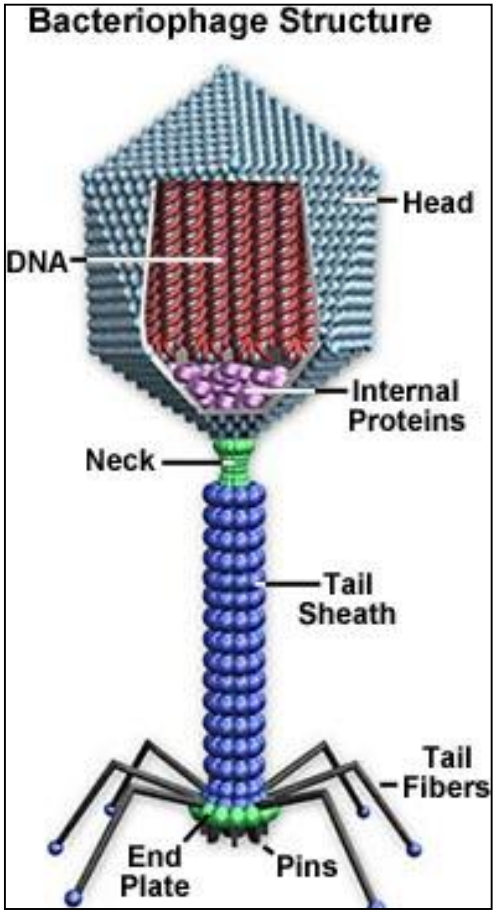


Diagram of a typical tailed bacteriophage structure.

Structure of T4 bacteriophage

- Bacteriophages (bacteriophages) are in many different sizes and shapes.

1. Size: The phage T4 is among the largest phages; it is approximately 200 nm long and 80-100 nm wide.
2. Head or Capsid: All phages contain a head structure inside the head is found the nucleic acid.
3. Tail: The tail is a hollow tube through which the nucleic acid passes during infection.



Mayer,2010;..

Bacterial viruses

Different types/sizes of bacteriophages

- There are at least 12 distinct groups of bacteriophages, which are very diverse structurally and genetically.

1. Simple phages may have only 3-5 genes. For example, ~3.5 kb ssRNA genome in phage MS2.
2. Highly complex phages may have over 100 genes. For example,~500 kb dsDNA genome in *Bacillus* phage G.

Over 96% of phages are tailed and contain dsDNA.

Rao,2006;..

Bacterial viruses or bacteriophages
Distinct biological entities of viruses

1. Prokaryotes as host;
2. Subcellular structure without metabolic machinery;
3. Double stranded DNA, single stranded DNA, RNA;
4. Virulent phage vs. temperate phage.

Viruses lack all metabolic machinery and do not produce ATP because they do not perform energy-requiring processes. Bacteriophages (or phages) can not independently reproduce. They must rely on host cell for reproductive machinery and components.

BC Yang;Todar,2012;..

Bacterial viruses or bacteriophages

Variations in two main type of infections

Lyso-lysis phenomenon in lytic cells of lambda bacteriophage

- There is a tremendous variety of bacteriophages with variations in properties for each type of infection i.e. lytic and lysogenic infections.
- Under certain conditions, some species were able to change the mode of infection, especially if the number of host cells was falling down.
- The lysis-lysogeny phenomenon of bacteriophage lambda is a paradigm for developmental genetic networks.
- Therefore, temperate phages such as lambda phage are not suitable for the phagetherapy.

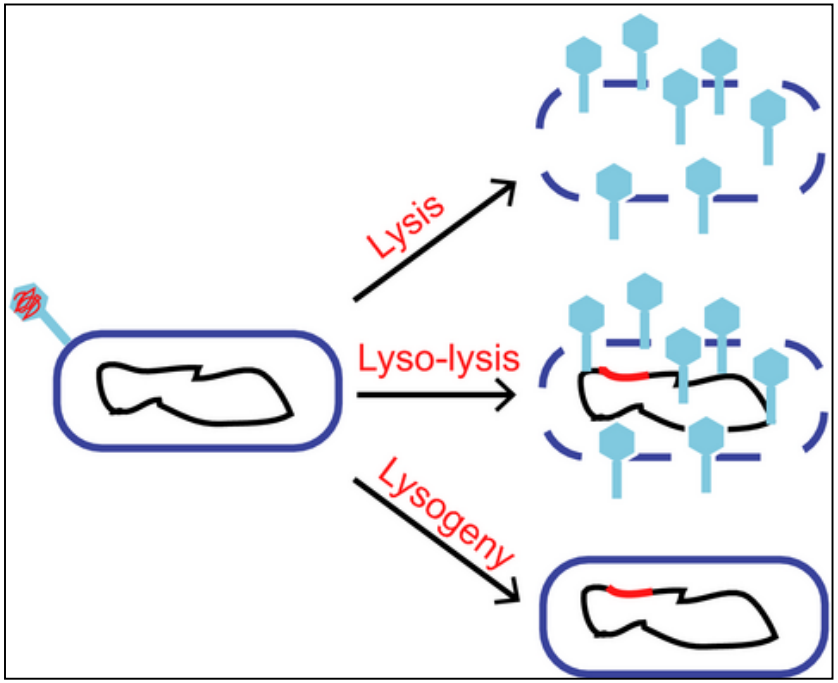
Orlova,2012;..

Phage Therapy

Two types of infections and one intermediate

Lysis-lysogeny coexistence: prophage integration during lytic development in phage lambda

- Historically, it is understood that infection by phage lambda culminates either in:
1. Cell lysis and progeny release, or
2. Cell lysogenic and prophage integration and cell growth.
- Surprisingly, we frequently observed an interesting "lyso-lysis" phenomenon in lytic cells, where phage integrates its DNA into the host, a characteristic event of the lysogenic pathway, followed by cell lysis.



Shao *et al.*,2016

Bacterial viruses or bacteriophages

Two main types of infections

Lytic and lysogenic infections

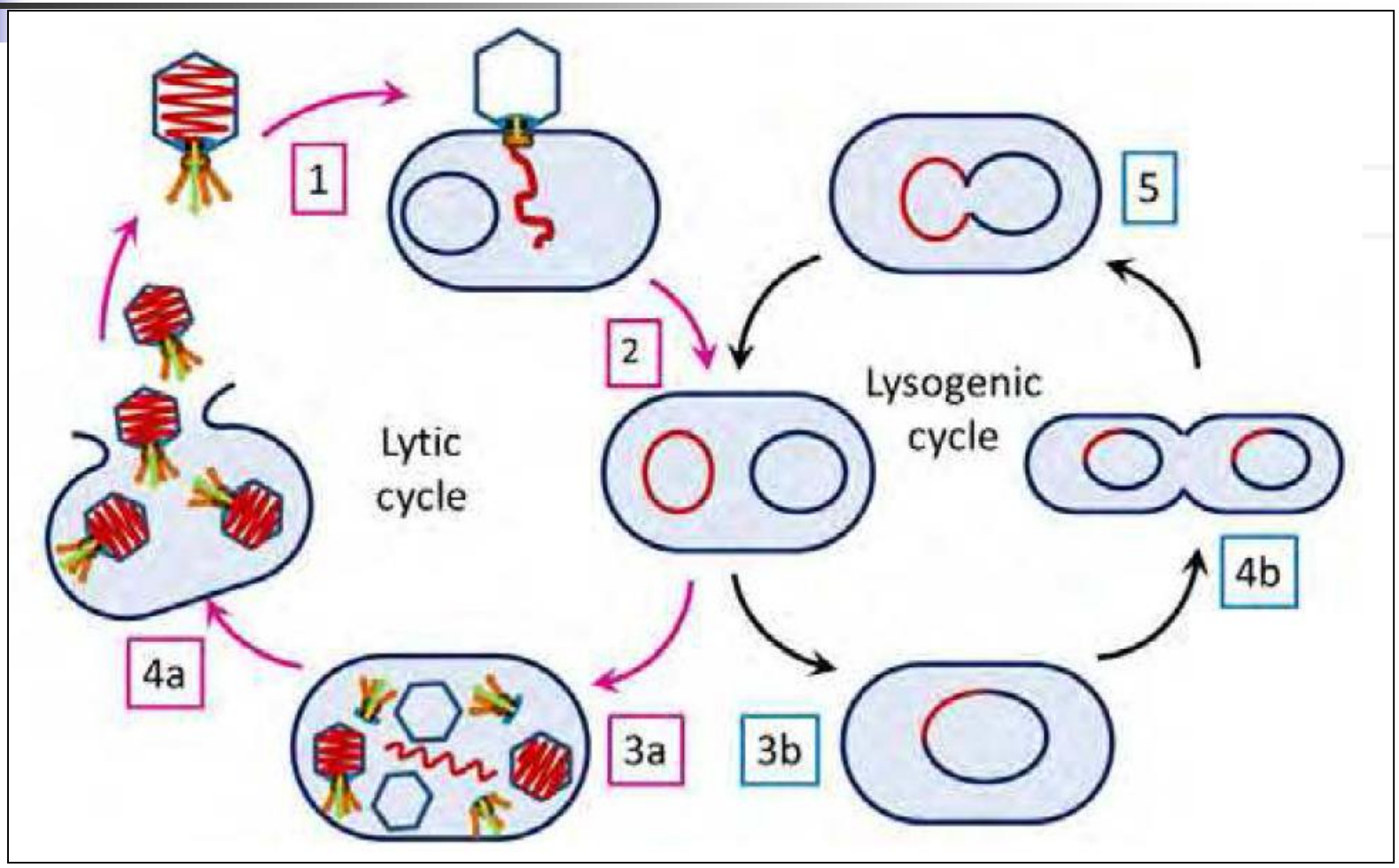
- Bacterial cells can undergo one of two types of infections by viruses:

1. lytic (virulent) infections. e.g. T-even phages such as T2, T4 and T6 that kill the host cells.
2. lysogenic (temperate) infections. e.g. phage lambda (phage λ). Temperate phages establish a persistent infection of the cell without killing the host cells.

BC Yang;Todar,2012

Bacterial viruses or bacteriophages

Differences between lytic and lysogenic cycles

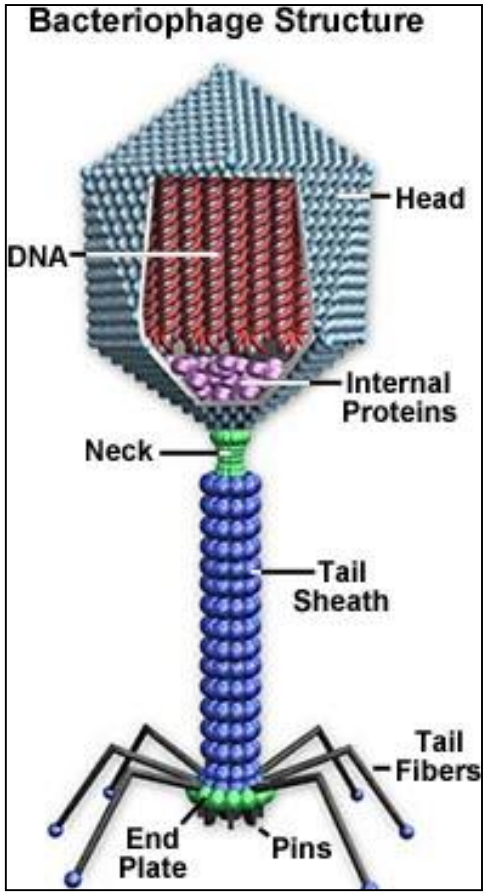


Orlova,2012

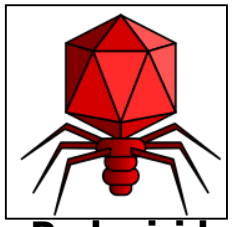
Bacterial viruses

Receptor sites for bacteriophage infection

- Receptor sites for bacterial viruses are:
1. Flagella
2. Fimbriae,
3. Cell wall, and
4. Cell membrane
- The fibers and the tail carry the host cell recognition sensors and are required for attachment of the phage to the cell surface.
- The tail also serves as a channel for delivery of the phage DNA from the head into the host cell cytoplasm.
- Tail has got some tail pins which help the virus particle to sit on the wall.

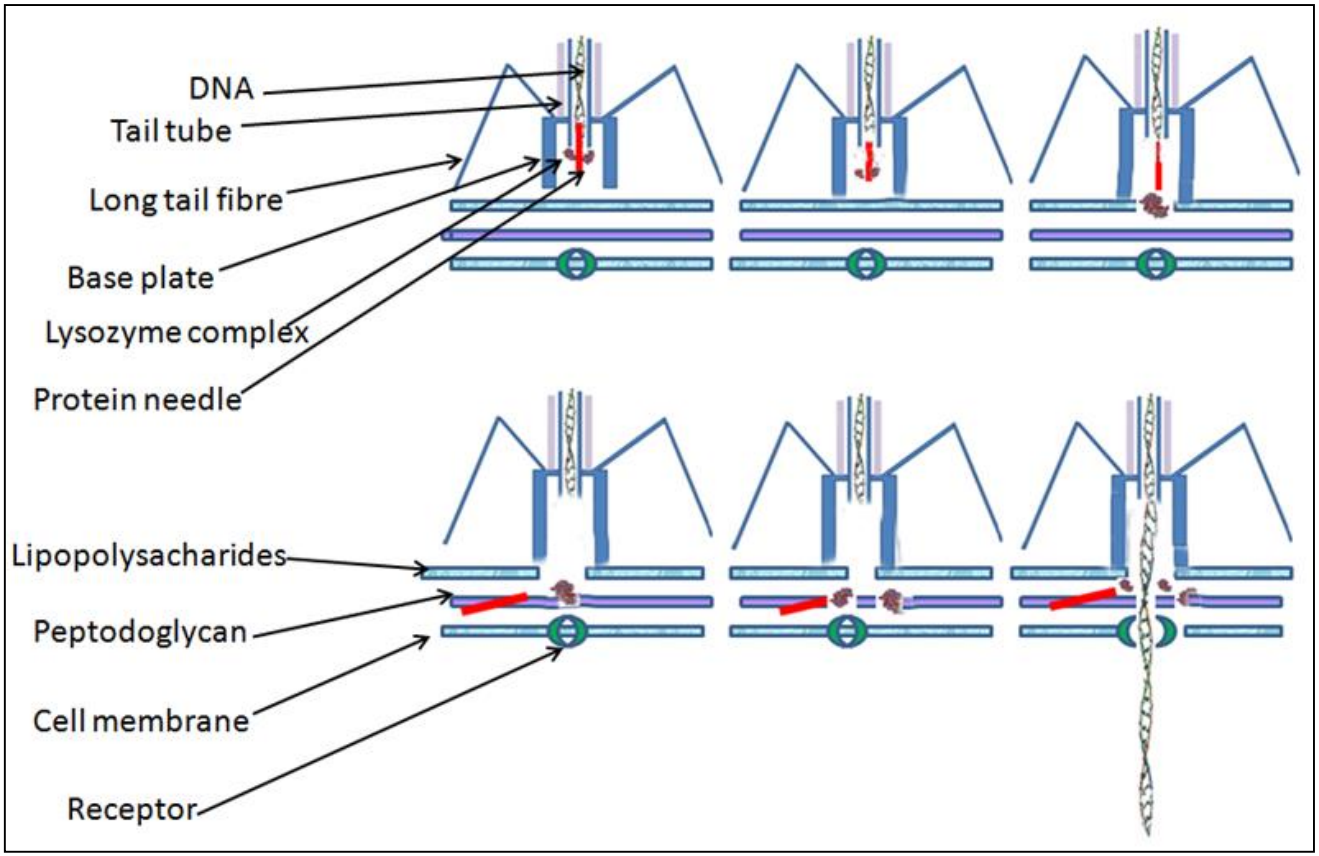


Wikipedia;..



Bacterial viruses

Diagram of the DNA injection process

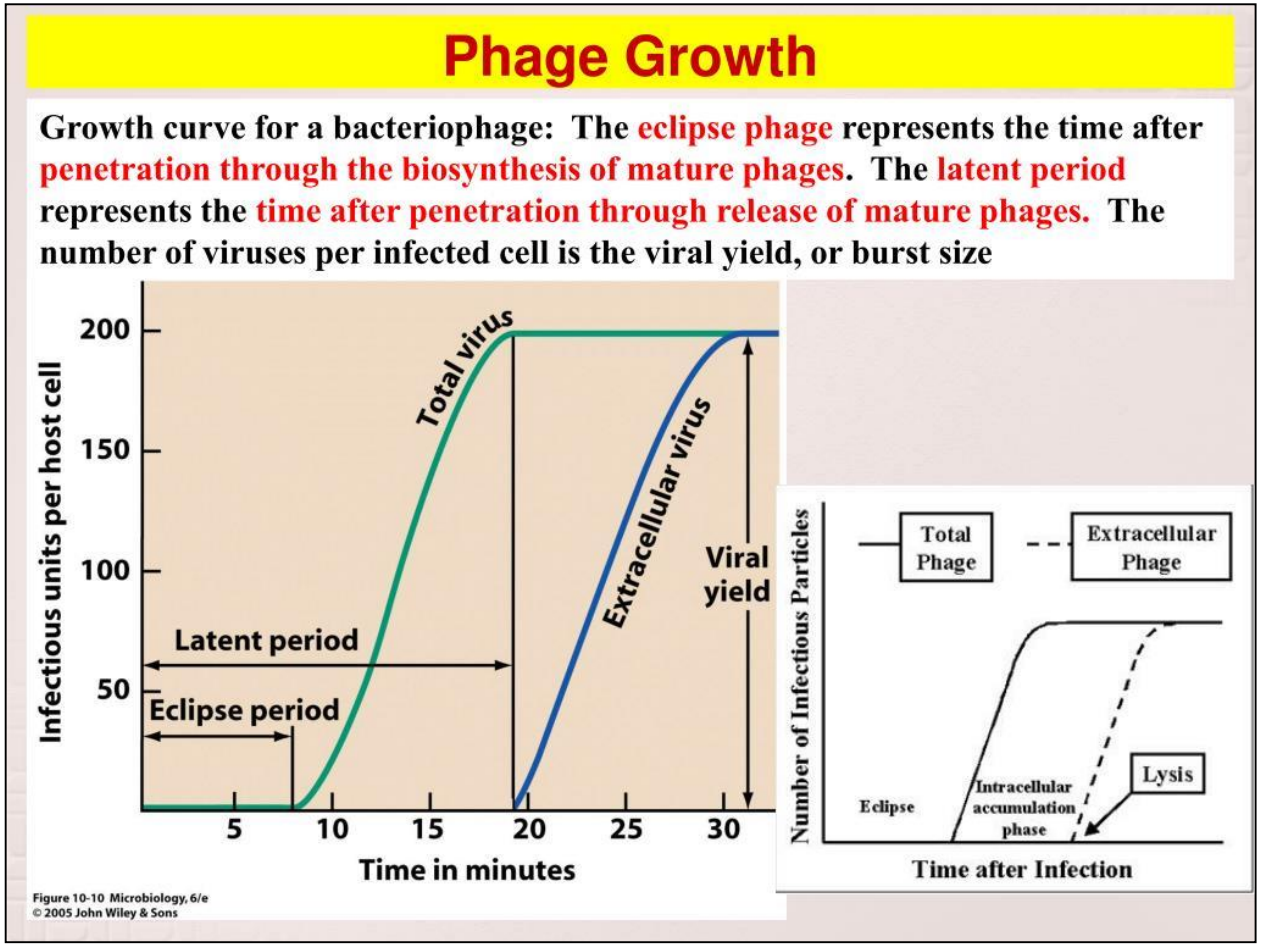


Podoviruses lack an elongated tail sheath similar to that of a myovirus (Any virus of the family Myoviridae, e.g. T4 phage), so they instead use their small, tooth-like tail fibers to enzymatically degrade a portion of the cell membrane before inserting their genetic material.

Bacterial viruses

Multiplication rate

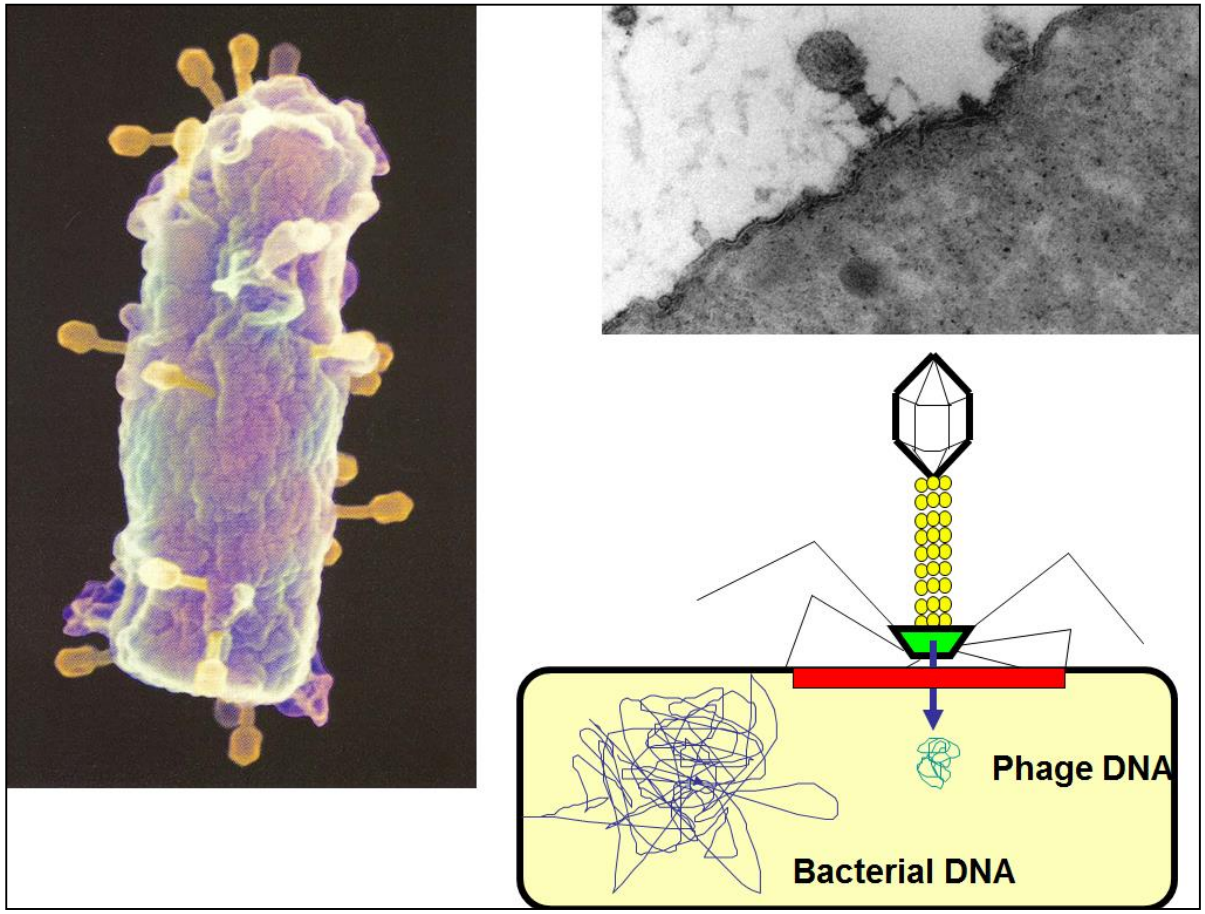
Host cell as factory



Lareina Garrett

Lytic (virulent) phage infection

Kills (lysis) the infected cell



Kelly Doran

Bacteriophage

Escherichia coli phage T4: Coliphage T4

The Lytic Life Cycle of Coliphage T4.

Fig. 2: Adsorption
Tail fibers of Coliphage T4 adsorb to receptors on the cell wall of *E. coli* B.

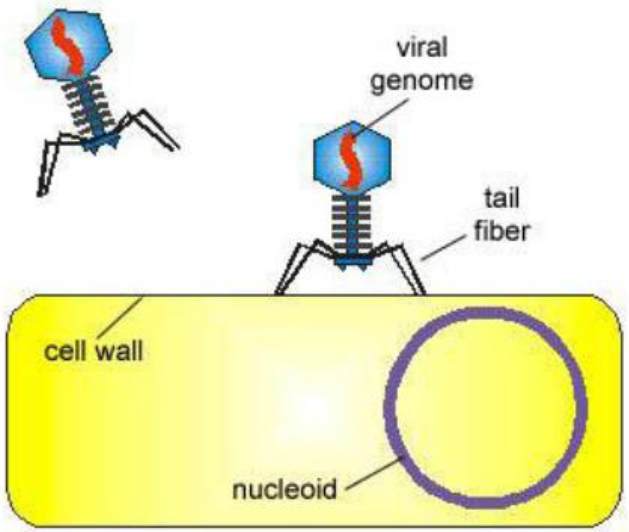
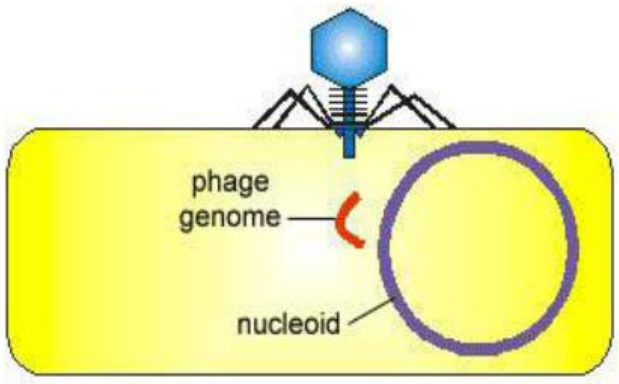


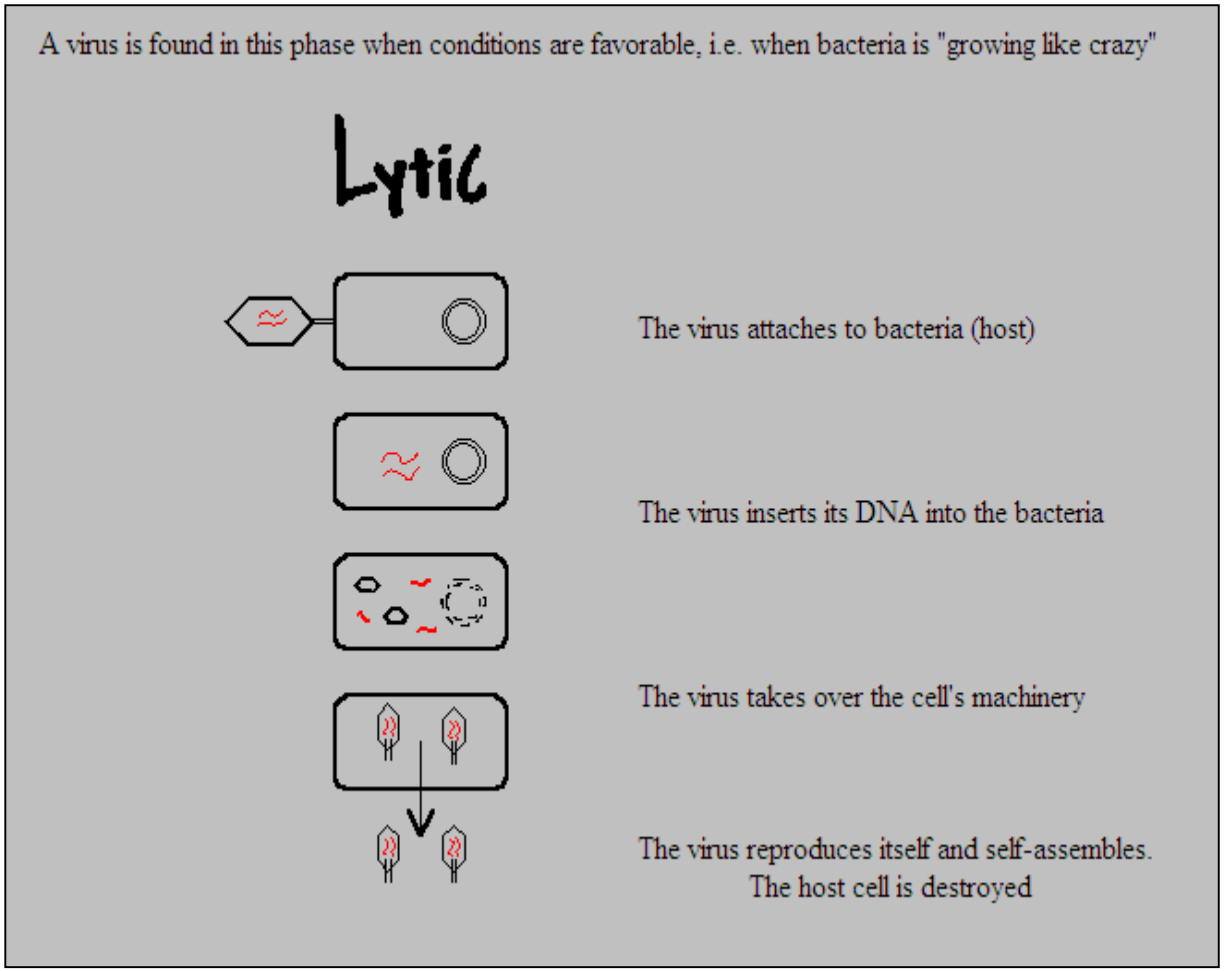
Fig. 3: Penetration
The bacteriophage injects its DNA genome into the cytoplasm of the bacterium.



Kaiser,2009

Lytic (virulent) phage infection

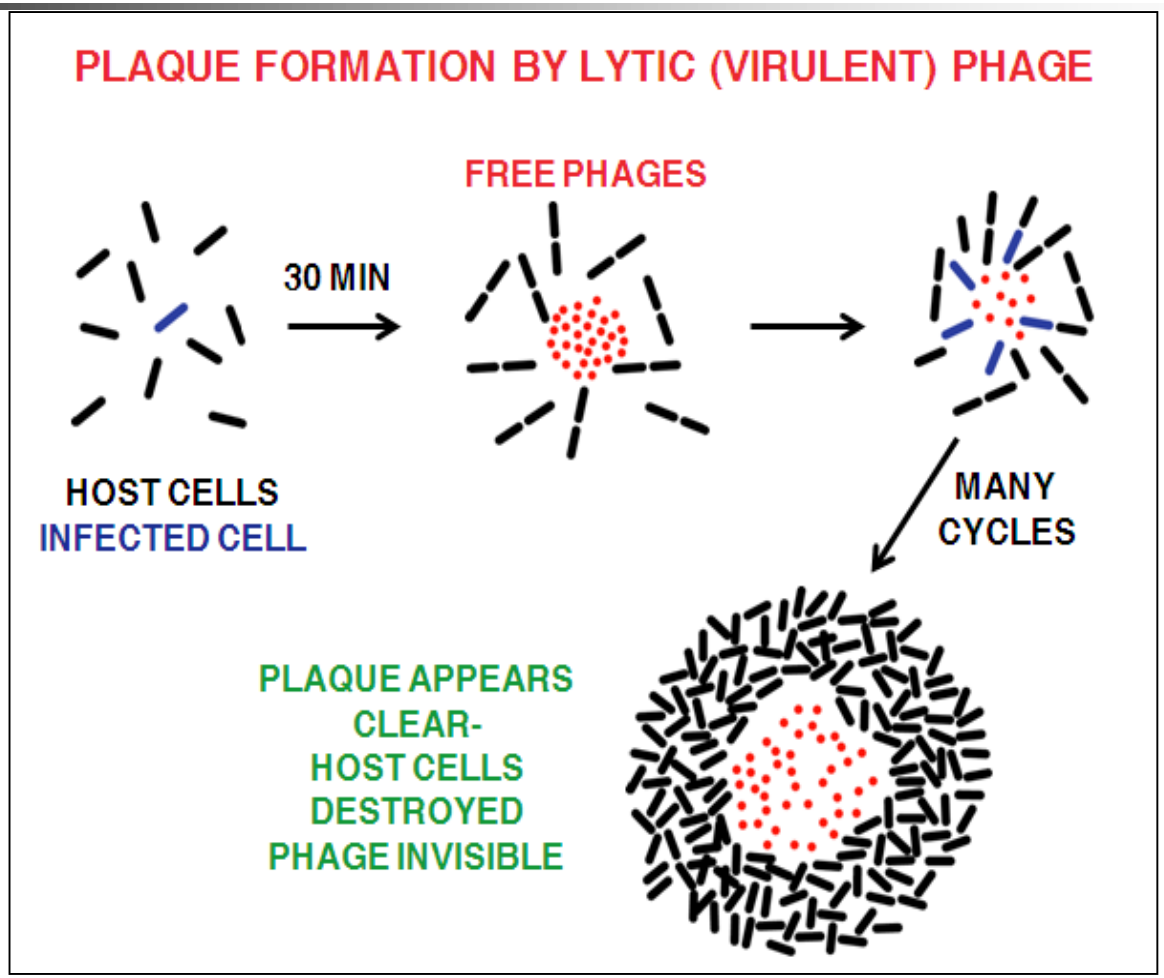
Lysis and plaque formation by lytic phage



2002-2003 Verizon Academic All-District IV Team

Lytic (virulent) phage infection

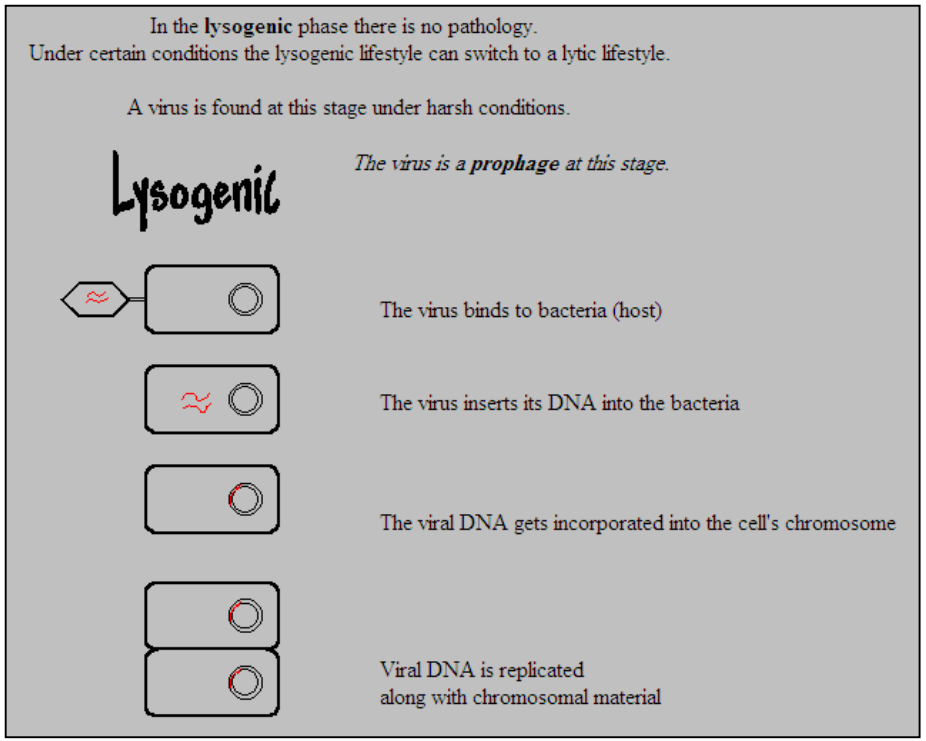
Lysis and plaque formation by lytic phage



web.biosci.utexas.edu

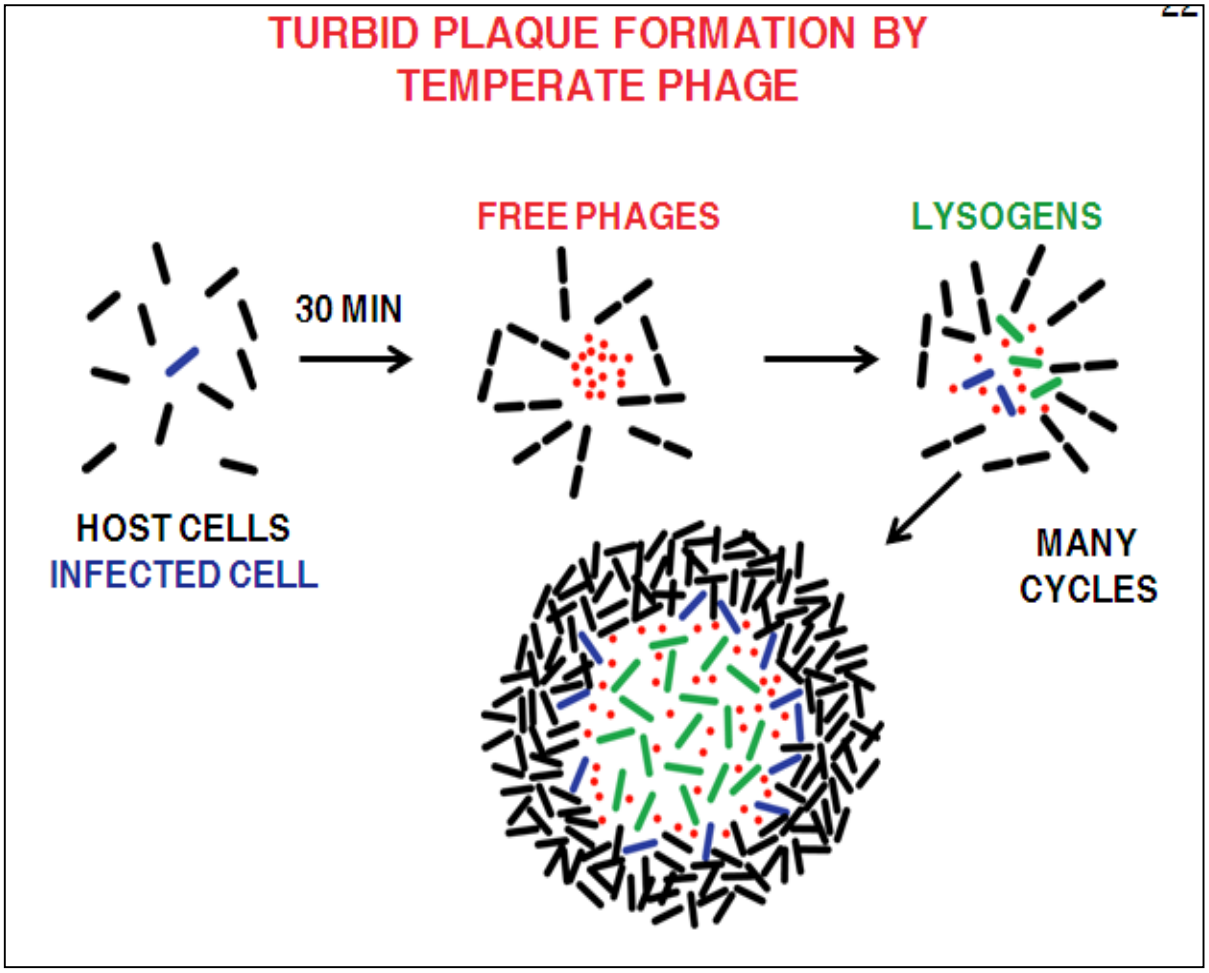
Lysogenic (avirulent) phage infection Without killing(lysis) of the infected cell Known as prophage

- Prophage:
- The latent form of a bacteriophage in which the viral genes are incorporated into the bacterial chromosomes without causing disruption (lysis) of the bacterial cell.
- The host is termed a lysogen when a prophage is present.



Turbid plaque formation by temperate (avirulent) phage

Without bacterial cell lysis



web.biosci.utexas.edu

Bacterial viruses or bacteriophages

Differences between lytic and lysogenic cycles

- **In the Lytic Cycle:**

1. Viral DNA destroys cell DNA, takes over cell functions and destroys the cell.
2. The Virus replicates and produces progeny phages.
3. There are symptoms of viral infection.
4. Virulent viral infection takes place.

- **In the Lysogenic Cycle:**

1. Viral DNA merges with cell DNA and does not destroy the cell.
2. The Virus does not produce progeny.
3. There are no symptoms of viral infection.
4. Temperate viral replication takes place.

Phage progeny release from phage-infected bacterial cells.

Bacterial viruses or bacteriophages
Distinct biological entities of viruses

1. A virus is an infectious, obligate intracellular parasite.
2. The genetic material of a virus is either DNA or RNA.
3. The genetic material of a virus enters a host cell.
4. Within minutes, bacterial ribosomes start translating viral mRNA into protein.
5. Proteins modify the bacterial RNA polymerase so it preferentially transcribes viral mRNA.
6. New virus particles (called virions) were produced.
7. The new virions produced in a host cell then transport the viral genetic material to another host cell or organism to carry out another round of infection.

Virology blog,2010;..

Bacterial viruses or bacteriophages

Virulent (lytic) phages

T-even phages vs. T-odd phages

- Every bacterial species is parasitized by various specific bacteriophages.
- Those which attack *Escherichia coli* are called Coliphages and are designated T type.
- These were numbered T1, T2, T3.....T7 by Max Delbruck (1938).

1. T-even phages are T2, T4, and T6, and
2. T-odd phages are T3, T5,...

The T4 phage of the *Myoviridae* family infects *E. coli* bacteria and is one of the largest phages.

Bacteriophages

Morphology of the T series of phages

Name	Plaque size	Head (nm)	Tail (nm)	Latent period (min)	Burst size
T1	medium	50	150 x 15	13	180
T2	small	65 x 80	120 x 20	21	120
T3	large	45	invisible	13	300
T4	small	65 x 80	120 x 20	23.5	300
T5	small	100	tiny	40	300
T6	small	65 x 80	120 x 20	25.5	200-300
T7	large	45	invisible	13	300

Bacterial viruses or bacteriophages

Temperate phages, integrated phages or prophages

- A temperate phage such as lambda phage can integrate its genome into its host bacterium's chromosome, becoming a lysogen known as a prophage.
- The genetic material of the bacteriophage, called a prophage, can be transmitted to daughter cells at each subsequent cell division.
- Temperate bacteriophages (or phages) have the ability to enter a prophage dormant state upon infection, in which they stably replicate with the bacterial genome.

Paepe *et al.*,2014;..

Temperate phages, integrated phages or prophages Using both the lytic and the lysogenic cycle Horizontal gene transfer

- Temperate phages (such as lambda phage) can reproduce using both the lytic and the lysogenic cycle.
- After infection of the host bacterium, a decision between lytic or lysogenic development is made that is dependent upon:

1. Environmental signals, and
2. The number of infecting phages per cell.

- Genomic comparisons suggest a remarkable level of horizontal gene transfer among temperate phages, favoring a high evolution rate.

Oppenheim *et al.*,2005; Paepe *et al.*,2014

Defective prophage(probacteriophage)

Using both the lytic and the lysogenic cycle

Transduction

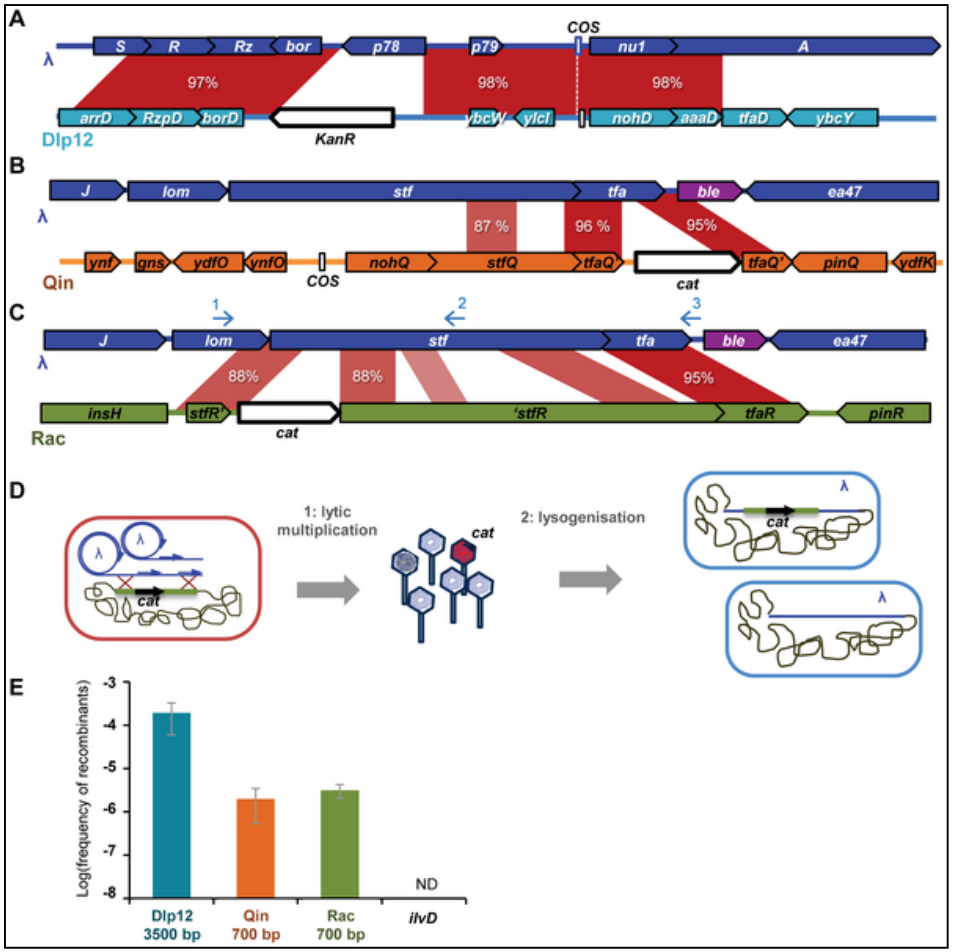
- A temperate bacteriophage mutant that cannot fully infect but that can replicate in the bacterial genome as a defective probacteriophage.
- A temperate bacteriophage mutant the genome of which does not contain all of the normal components and cannot become a fully infectious virus, but it can replicate indefinitely in the bacterial genome as a defective probacteriophage; many defective bacteriophages are mediators of transduction.

Paepe *et al.*,2014;..

Bacterial viruses or bacteriophages
Temperate phages or prophages

Recombinants between λ and defective prophages are formed during lytic cycle.

The mechanisms of genomic exchanges between an invasive infectious phage and defective prophages residing in the host (bacterial) chromosome.



Paepe *et al.*,2014

Bacterial viruses or bacteriophages

Temperate phages, integrated phages or prophages

Their contribution in increasing bacterial virulence

- Integrated phages (prophages) are major contributors to the diversity of bacterial gene repertoires.
- A majority of bacterial genomes contain multiple active or defective prophages, and numerous bacterial phenotypes are modified by these prophages, such as increased virulence.
- These mobile genetic elements are subject to high levels of genetic exchanges, through which new genes are constantly imported into bacterial genomes.

Paepe *et al.*,2014

Classification of Bacteriophages
Phage families and genera

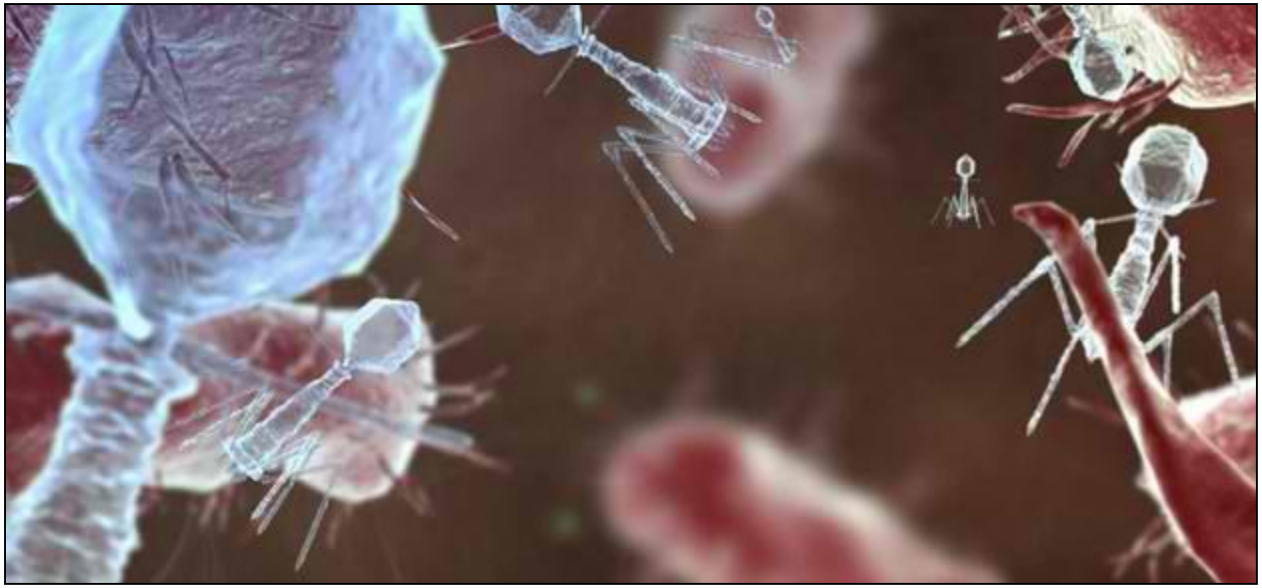
- At least 5,360 tailed and 179 cubic, filamentous, and pleomorphic bacterial viruses have been examined in the electron microscope since the introduction of negative staining in 1959.
- Since at least 100 novel bacterial viruses are described every year, the approximate number of viruses under consideration is over 6,000.
- Numerically, this makes bacteriophages the largest virus group known.

Ackermann,2011

Classification of Bacteriophages

Phage families and genera

- At present, over 5500 bacteriophages have been studied by electron microscopy and can be divided into 14 virus families.



nano.aau.dk/nano3/microbiology,2009; Everything about Bacteriophage,2013

Classification of Bacteriophages
Three order and 19 families

- Phages are presently classified in a hierarchical and holistic system(a combination of analysis, looking at the 'big picture') with Three order and 19 families.
- The cubic, filamentous and pleomorphic phages are small and well defined.
- They contain ds or ssDNA or RNA.
- The Podoviridae and Myoviridae families of tailed phages.

Holistic system is any set (group) of interdependent or temporally interacting parts.

Ackermann,2011;..

ICTV taxonomic classification of bacteriophage infecting bacteria and archaea

Order	Family	Morphology	Nucleic acid	Examples	Subfamilies	Genera
Caudovirales	Ackermannviridae		dsDNA		2	4
	Myoviridae	Nonenveloped, contractile tail	Linear dsDNA	T4 phage, Mu, PBSX, P1Puna-like, P2, I3, Bcep 1, Bcep 43, Bcep 78	6	41
	Siphoviridae	Nonenveloped, noncontractile tail (long)	Linear dsDNA	λ phage, T5 phage, phi, C2, L5, HK97, N15	11	100
	Podoviridae	Nonenveloped, noncontractile tail (short)	Linear dsDNA	T7 phage, T3 phage, Φ29, P22, P37	3	23
Ligamenvirales	Lipothrixviridae	Enveloped, rod-shaped	Linear dsDNA	Acidianus filamentous virus 1		3
	Rudiviridae	Nonenveloped, rod-shaped	Linear dsDNA	Sulfolobus islandicus rod-shaped virus 1		1
Unassigned	Ampullaviridae	Enveloped, bottle-shaped	Linear dsDNA			1
	Bicaudaviridae	Nonenveloped, lemon-shaped	Circular dsDNA			1
	Clavaviridae	Nonenveloped, rod-shaped	Circular dsDNA			1
	Corticoviridae	Nonenveloped, isometric	Circular dsDNA			1
	Cystoviridae	Enveloped, spherical	Segmented dsRNA			1
	Fuselloviridae	Nonenveloped, lemon-shaped	Circular dsDNA			2
	Globuloviridae	Enveloped, isometric	Linear dsDNA			1
	Guttaviridae	Nonenveloped, ovoid	Circular dsDNA			2
	Inoviridae	Nonenveloped, filamentous	Circular ssDNA	M13		7
	Leviviridae	Nonenveloped, isometric	Linear ssRNA	MS2, Qβ		2
	Microviridae	Nonenveloped, isometric	Circular ssDNA	ΦX174	2	6
	Plasmaviridae	Enveloped, pleomorphic	Circular dsDNA			1
	Tectiviridae	Nonenveloped, isometric	Linear dsDNA			2

Phage are classified by the International Committee on Taxonomy of Viruses (ICTV), as of their 2017 update, there are 19 families of phage that infect bacteria and archaea but as more samples from more remote areas are sequenced this is only likely to grow in the future.

Steward,2021

Classification of Bacteriophages

Reclassification of *Podoviridae* phages

Subfamily	Genus	Example	Members	Host
Autographivirinae	T7-like	T7	8	Enterics, *Pseudomonas, Vibrio*
	SP6-like	SP6	4	Enterics
	ϕKMV-like	ϕKMV	3	*Pseudomonas*
	P60-like	P60	3	*Prochlorococcus, Synechococcus*
Nanovirinae	ϕ29	ϕ29	4	*Bacillus*
	44AHJD	44AHJD	7	*Staphylococcus*
(P22-like)	P22-like	P22	7	Enterics
---	BPP-1-like	BPP1	4	*Bordetella, Burkholderia*
---	ε15-like	ε15	2	Enterics
---	N4-like	N4	1	Enterics
---	119-like	119	2	*Pseudomonas*
---	VP2-like	VP2	2	*Vibrio*

Ackermann,2011

Classification of Bacteriophages

Reclassification of *Myoviridae* phages

Subfamily	Genus	Example	Members	Host
Teequatrovirinae	T4-like	T4	15	Enterics, *Acinetobacter, Aeromonas*
	KVP40-like	KVP40	5	*Aeromonas, Vibrio*
	(Cyanophages)	S-PM2	4	*Synechococcus, Prochlorococcus*
Peduovirinae	P2-like	P2	13	Enterics, *Burkholderia, Mannheimia, Pseudomonas, Ralstonia*
	HP1-like	HP1	6	*Aeromonas, Haemophilus, Pasteurella, Vibrio*
Spounavirinae	SP01-like	SP01	1	*Bacillus*
	Twort-like	Twort	7	*Staphylococcus, Listeria*
---	Mu-like	Mu	2	Enterics
---	P1-like	P1	2	Enterics
---	Bcep781-like	Bcep781	5	*Burkholdera, Xanhomonas*
---	BcepMu-like	BcepMu	2	*Burkholderia*
---	Felix 01-like	Felix 01	3	Enterics
---	HAP1-like	HAP1	2	*Halomonas, Vibrio*
---	I3-like *	I3	7	*Mycobacterium*
---	ϕCD119-like	ϕCD119	3	*Clostridium*
---	ϕKZ-like	ϕKZ	2	*Pseudomonas*
---	PB1-like	PB1	7	*Pseudomonas*

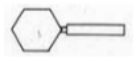
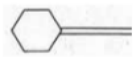
* Renamed I3-like after *Mycobacterium* phage I3.

Ackermann,2011

Classification of Bacteriophages

Overview of phage families

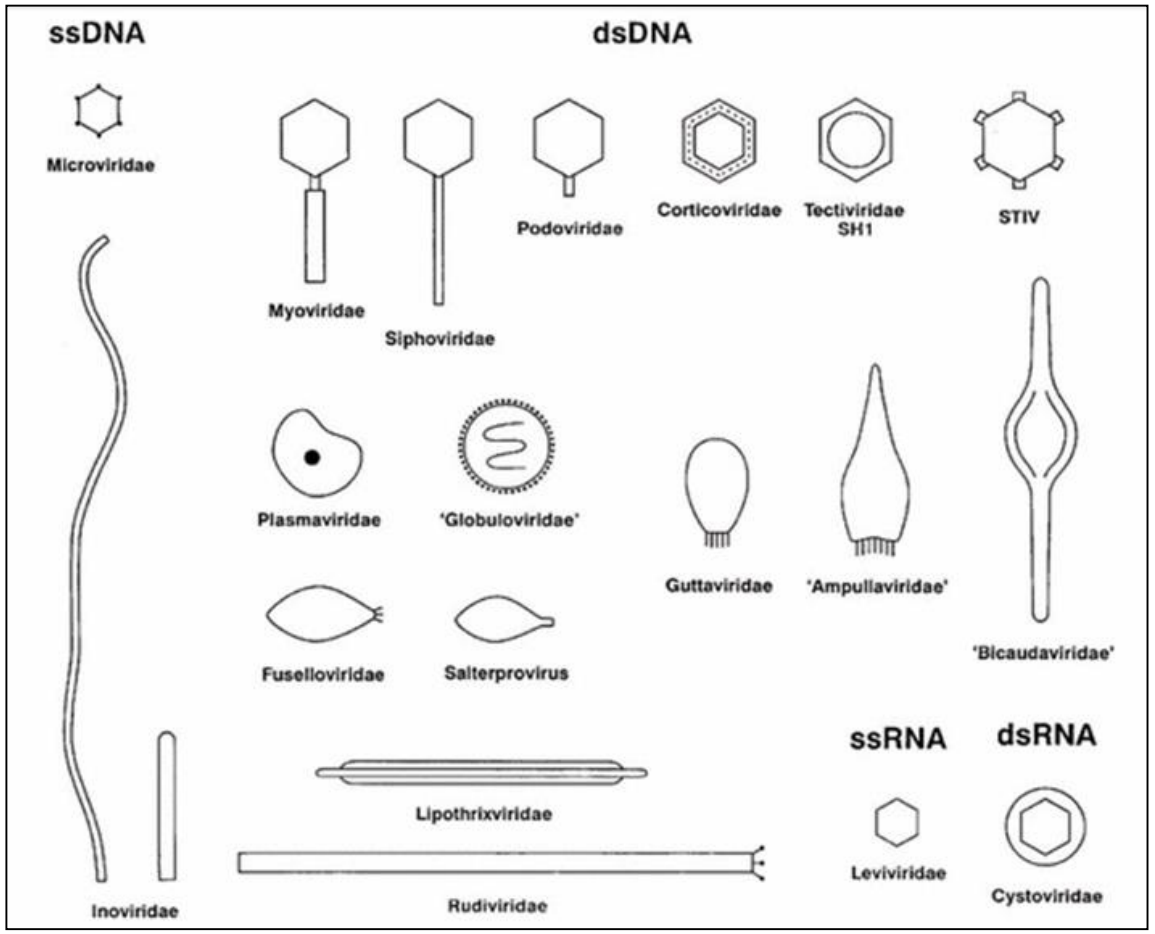
One order and 10 families

Shape	Order or family	Nucleic acid, particulars, size	Member	Number[a]
	Caudovirales	dsDNA (L), no envelope		
	*Myoviridae*	Tail contractile	T4	1312
	*Siphoviridae*	Tail long, noncontractile	λ	3262
	*Podoviridae*	Tail short	T7	771
	*Microviridae*	ssDNA (C), 27 nm, 12 knoblike capsomers	ϕX174	38
	*Corticoviridae*	dsDNA (C), complex capsid, lipids, 63 nm	PM2	3?
	*Tectiviridae*	dsDNA (L), inner lipid vesicle, pseudo-tail, 60 nm	PRD1	19
	*Leviviridae*	ssRNA (L), 23 nm, like poliovirus	MS2	38
	*Cystoviridae*	dsRNA (L), segmented, lipidic envelope, 70–80 nm	ϕ6	3
	*Inoviridae*	ssDNA (C), filaments or rods, 85–1950 x 7 nm	fd	66
	*Plasmaviridae*	dsDNA (C), lipidic envelope, no capsid, 80 nm	MVL2	5

[a] From reference 1. C, circular; L, linear.

Ackermann,2011

Images of bacteriophages



Kelly Doran

Classification of Bacteriophages
Nomenclature

- Nomenclature is inseparable from classification.

1. The ICTV uses latinised terms for order, family, subfamily and genus names.
2. Families are characterized by the suffix, -*viridae*.
3. Species epithets are not latinised; for example, phage T4 is and will remain T4.

Ackermann,2011

Bacteriophages
Phage families and genera

1. Tailed phages:

- Comprise 96% of phages and are the largest virus group known. e.g. T-even lytic or virulent phages such as T2, T4 and T6 that infect *E.coli*.

2. Cubic, filamentous, and pleomorphic phages:

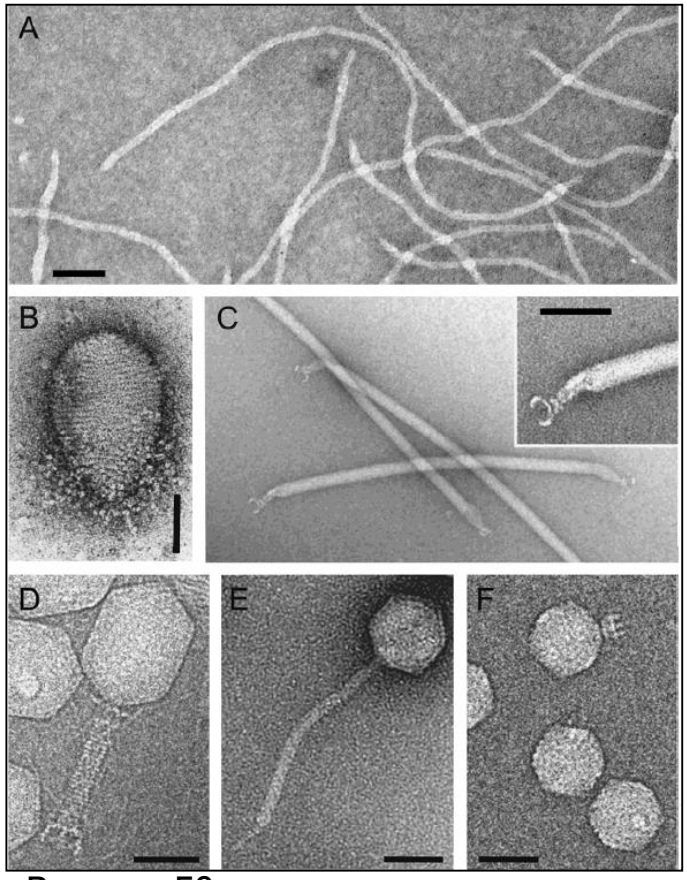
- This group includes 10 small phage families that correspond to approximately 4% of phages.
- These phages differ greatly in:

1. nucleic acid nature;
2. particle structure;
3. sometimes have a single member;
4. Host ranges are mostly narrow.

Ackermann,2004

Images of bacteriophages

A. Filamentous phage B5;
B. Droplet shaped virus;
C. Filamentous virus 1 with tail structures in their native conformation;
D. Bacteriophage T4;
E. Bacteriophage SPP1virulent *Bacillus subtilis* dsDNA phage;
F. Bacteriophage P22.



Bars are 50 nm.

Bacteriophages

Main properties and frequency of phage families[a]

Talked, cubic, filamentous and pleomorphic shaped

Shape	Nucleic acid	Family	Genera	Particulars	Example	No. of Members[b]
Tailed	DNA, ds, L	Myoviridae	6, see text	Tail contractile	T4	1143
		Siphoviridae	6, see text	Tail long, noncontractile	λ	3011
		Podoviridae	3, see text	Tail short	T7	698
Cubic	DNA, ss, C	Microviridae	*Microvirus* *Bdellomicrovirus* *Chlamydiomicrovirus* *Spiromicrovirus*	Conspicuous capsomers	ϕX174	40
	DNA, ds, C, S	Corticoviridae	*Corticovirus*	Complex capsid, lipids	PM2	3?
	DNA, ds, L	Tectiviridae	*Tectivirus*	Double capsid, pseudo-tail, lipids	PRD1	18
	RNA, ss, L	Leviviridae	*Levivirus* *Allolevirirus*		MS2	39
	RNA, ds, L, M	Cystoviridae	*Cystovirus*	Envelope, lipids	ϕ6	1
Filamentous	DNA, ss, C	Inoviridae	*Inovirus*	Long filaments	fd	57
			Plectrovirus	Short rods	L51	
	DNA, ds, L	Lipothrixviridae	*Lipothrixvirus*	Envelope, lipids	TTV1	6?
	DNA, ds, L	Rudiviridae	*Rudivirus*	Stiff rods, no envelope, no lipids	SIRV1	2
Pleomorphic	DNA, ds, C, S	Plasmaviridae	*Plasmavirus*	Envelope, no capsid, lipids	MVL2	6?
	DNA, ds, C, S	Fuselloviridae	*Fusellovirus*	Lemon-shaped, envelope, lipids	SSV1	8?

[a]Modified from Ackermann (1987) with permission of Blackwell Scientific Publications Ltd. C, circular; L, linear; M, multipartite; S, supercoiled; ss, single-stranded; ds, double-stranded.
[b]Exluding phage-like bacteriocins and known defective phages. Computed October 31, 2000.

Phi (Φ) = most often pronounced fi is simply an irrational number like pi (p= 3.14).

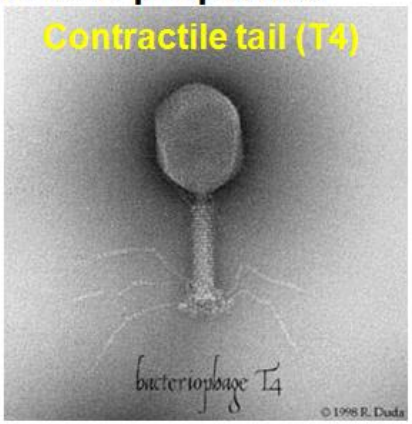
Ackermann,2004

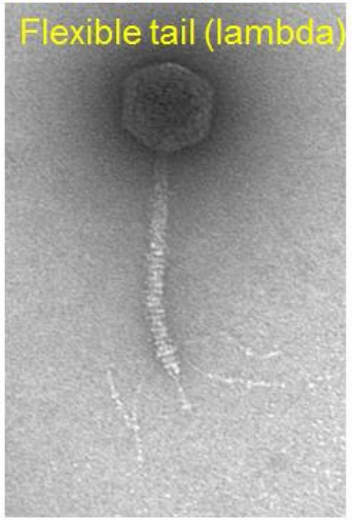
Classification of Bacteriophages

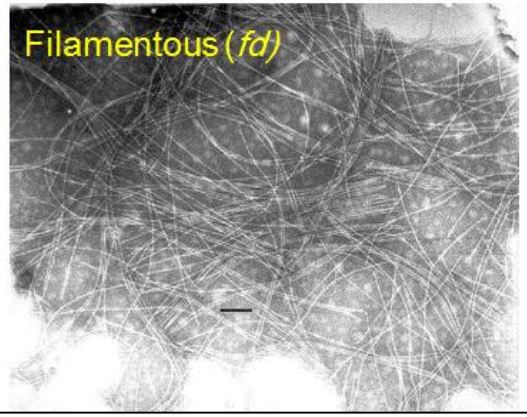
Phage families and genera

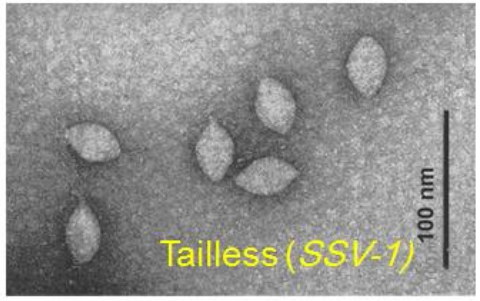
The most important criteria used for classification are phage morphology and nucleic acid properties

- Morphology
  - Head shape
  - Contractile tails
  - Noncontractile tails
  - Tailless
  - Filamentous
- dsDNA
- ssDNA
- ssRNA
- dsRNA



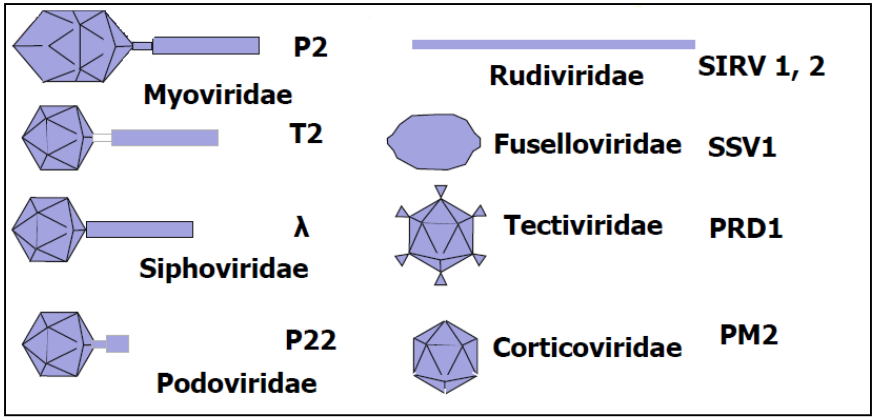
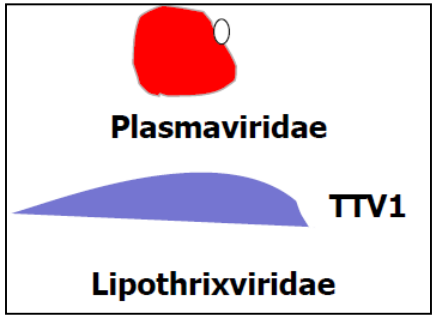


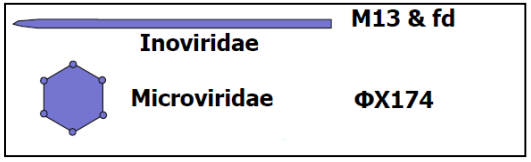




Kelly Doran

Bacteriophage families

Double stranded DNA, Non-enveloped	Double stranded DNA, Enveloped
	

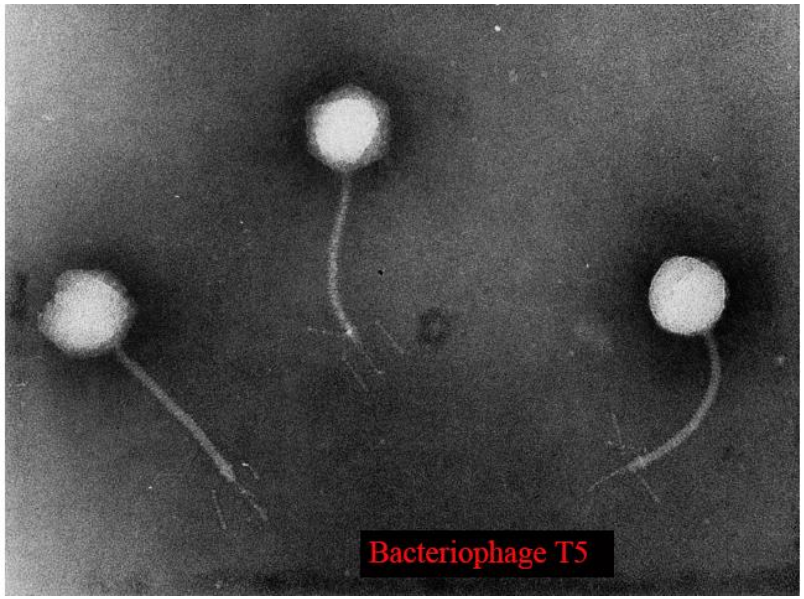
Single-stranded DNA	Single stranded RNA	Double stranded RNA
		

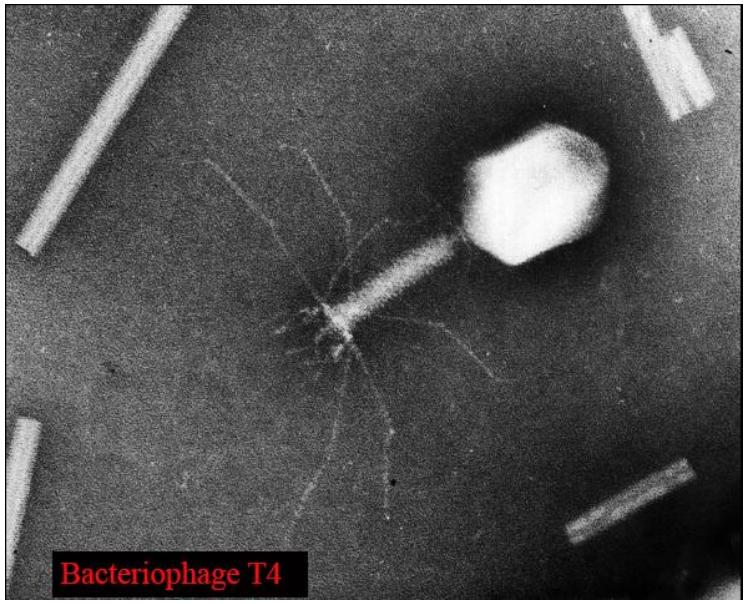
nano.aau.dk/nano3/microbiology,2009

Bacteriophages

Morphology of the T series of phages

T4 and T5

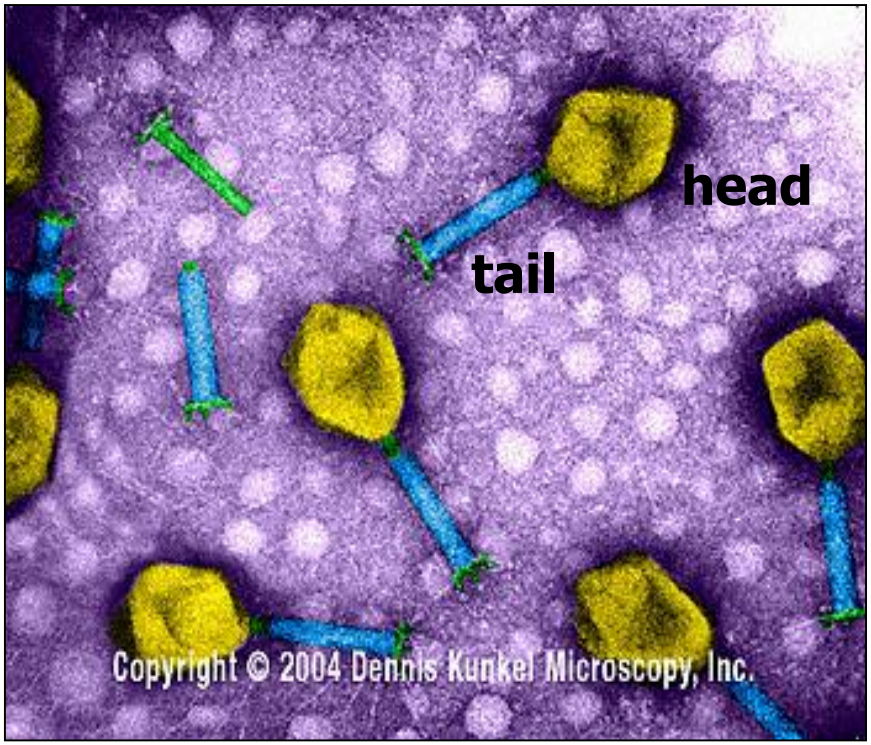


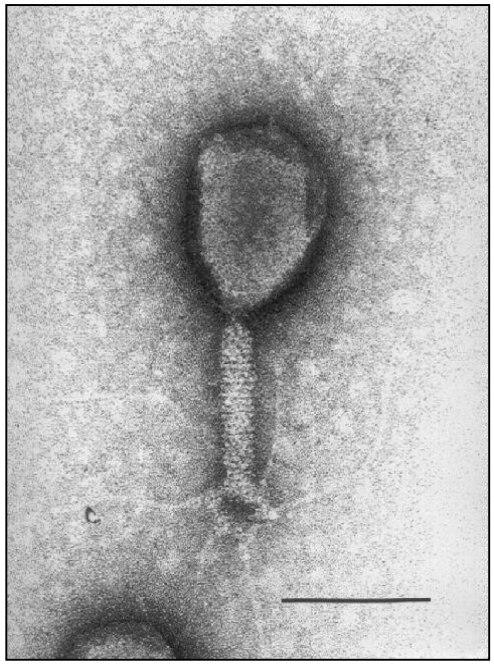


nano.aau.dk/nano3/microbiology,2009

Electron micrographs of bacteriophage T4

TEM (x390,000)





Negative stain electron micrograph

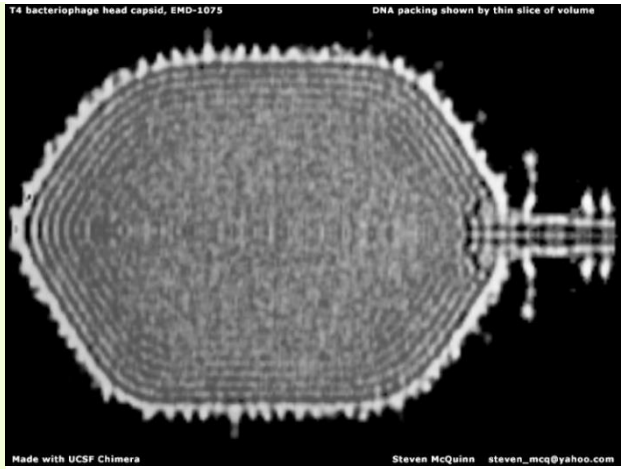
Mayer,2010

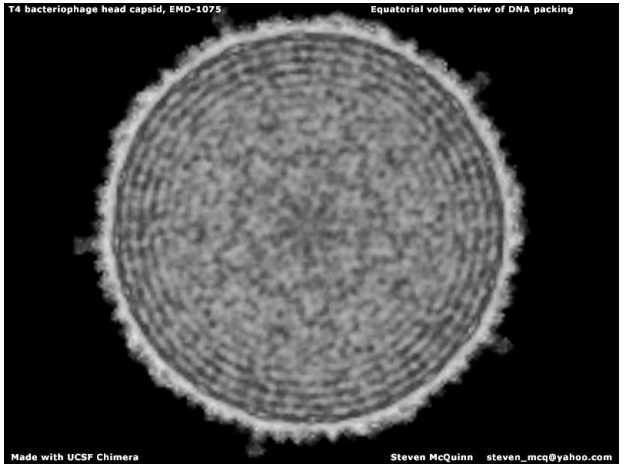
T4 bacteriophage head capsid

Surrounding this nucleic-acid genome is a protein-based capsid

DNA packing shown by thin slice of volume

- The head or capsid is composed of many copies of one or more different proteins.
- Inside the head is found the nucleic acid.
- The phage DNA actually integrates into the host chromosome and is replicated along with the host chromosome and passed on to the daughter cells.





Mayer,2010

Bacteriophages
Applications

- Temperate phages have various biological and molecular applications.
- They can be used to genetically manipulate eukaryotic cells, especially species that have large genomes like plants and mammals.
- Gene therapy, manipulation of cell lines, and construction of transgenic organisms can also employ phage enzymes.
- The temperate phage Mu-1 has a DNA-modifying function, which is of great importance especially in virology.

4. Transposable elements(TEs)

Segments of DNA that move from one genomic location to another and create phenotypically significant mutations and alter the cell's genome size

1. **Simple or smaller transposons (Insertion sequence, IS)**
2. **Complex or larger transposons (Transposons,Tn)**
3. **Bacteriophage Mu**

Conjugative transposons such as Tn916 are transposon-like elements, different from well-studied transposons such as Tn*5* and Tn*10*. Resistance genes need not be carried on the conjugative transposon to be transferred.

Transposition

Transposable genetic elements in Bacteria

Common feature of mobile elements

- Transposable elements cause genetics changes and make important contributions to the evolution of genomes.
- In the process, they may cause:

1. Mutations;
2. Increase (or decrease) the amount of DNA in the genome;
3. Promote genome rearrangements;
4. Regulate gene expression;
5. Induce chromosome breakage and rearrangement.

Mobile genetic elements were discovered in bacteria in the late 1960s.

Transposition

Transposable genetic elements in Bacteria

Common feature of mobile elements

1. Create more genetic diversity.
2. Range from several hundred to tens of thousands of base pairs.
3. Typical lab *E. coli* contains 10-20 different transposons, with many having multiple copies.

- Transposons move within a DNA molecule by using a special recombination enzymes - transposases encoded by the transposon.
- Transposition initiates when transposase recognizes inverted terminal repeats (ITR) or simply inverted repeats(IR) located on both ends of the transposon.

Transposable elements

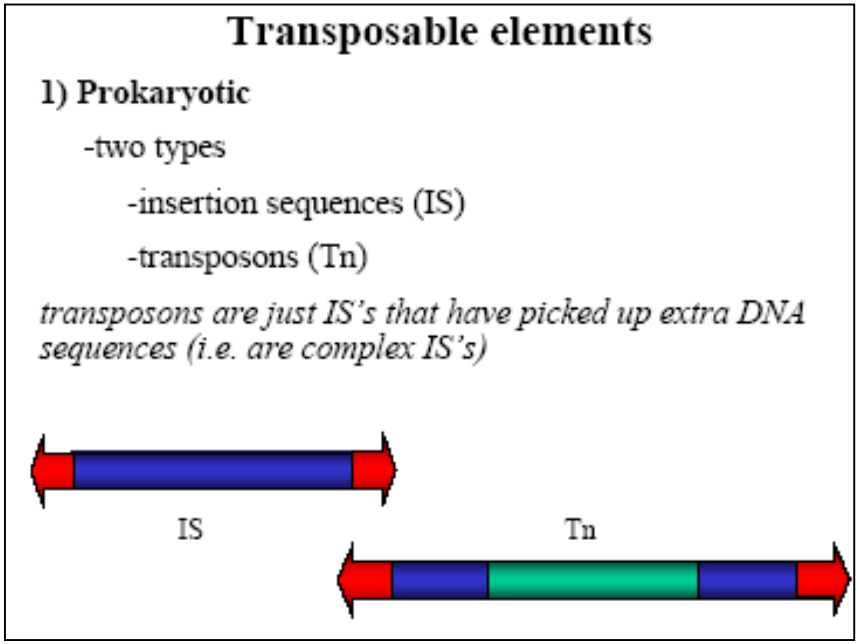
The two classes of bacterial elements are insertion sequences or IS elements and Tn elements

1. Simple transposons (Insertion sequences, IS):

- Small bacterial transposon lack additional genes. they are known as insertion sequences (IS family).

2. Complex transposons (Transposons,Tn):

- carry an additional gene for function belong to the Tn family.



Transposable elements

Jumping genes

IS elements and Tn elements

- These mobile segments of DNA are sometimes called "jumping genes".
- In the process, they may cause mutations increase (or decrease) the amount of DNA in the genome of the cell.
- The two main classes of bacterial transposable elements are:

1. Simple or smaller transposons (Insertion sequences, IS);
2. Complex or larger transposons (Transposons, Tn).

Transposable elements
IS elements vs. Tn elements

- **Simple or smaller transposons (Insertion sequences, IS):**
- The simple transposons consist of genes needed for insertion, particularly the gene coding for transposases, which are enzymes that catalyze their insertion.
- **Complex transposons (Transposons,Tn):**
- The complex transposons are those consisting of other genes apart from those needed for insertion.

Transposable elements
IS elements vs. Tn elements

- Simple transposons (Insertion sequences, IS):

1. carry only one gene;
2. short sequences (800 to 1500 bp);
3. do not code for proteins (resistance or other markers).

- Complex transposons (Transposons,Tn):

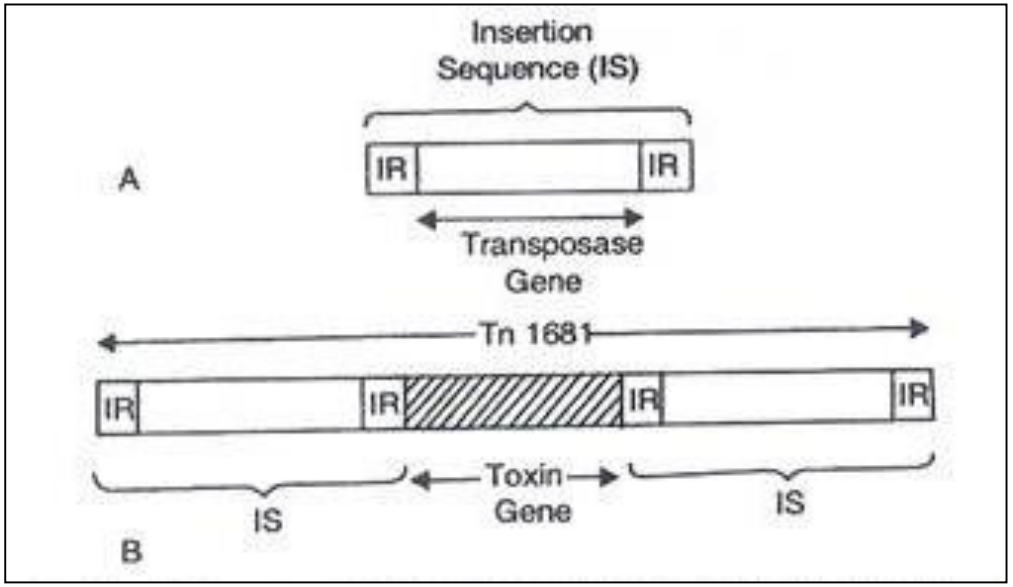
1. carry many genes;
2. long sequences (several thousand base pair long);
3. code for one or more proteins (including resistance factors in bacteria which act against antibiotics).

Transposable elements

IS elements vs. Tn elements

A. Simple transposons (Insertion sequences,IS): do not code for proteins.

1. Complex transposons (transposons,Tn): code for one or more proteins. E.g. toxin.

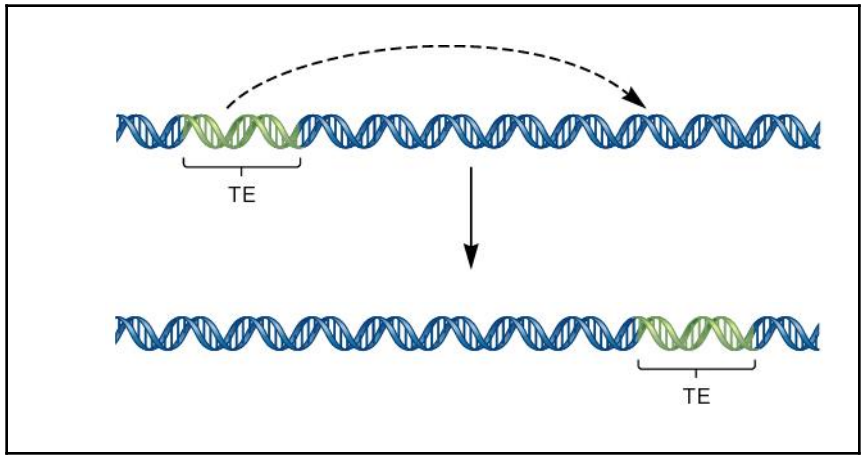


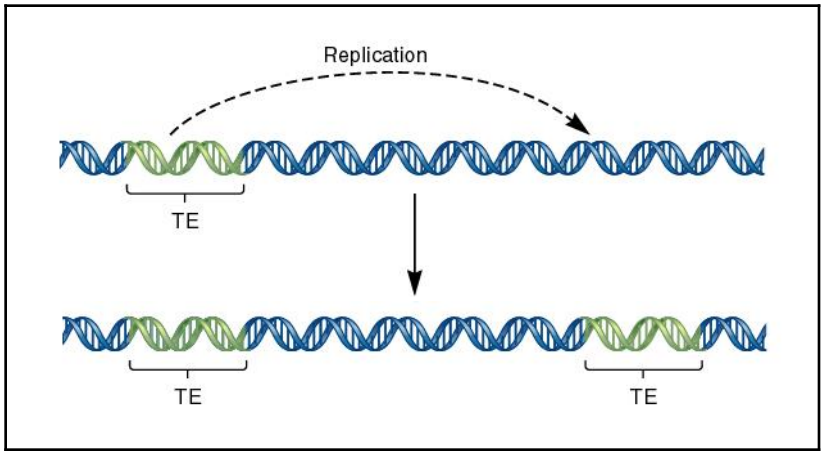
Transposition

Transposable elements (TEs)

IS elements (not utilize replicative transposition mechanism)

Tn elements (utilize a replicative transposition mechanism)





Simple transpositions

Transposable element (TE)

(Insertion sequences, IS):

Involves excision of the Transposable element (TE) and reintegration at a new site without replicative transposition (direct transposition).

Replicative transpositions

Transposable element (TE)

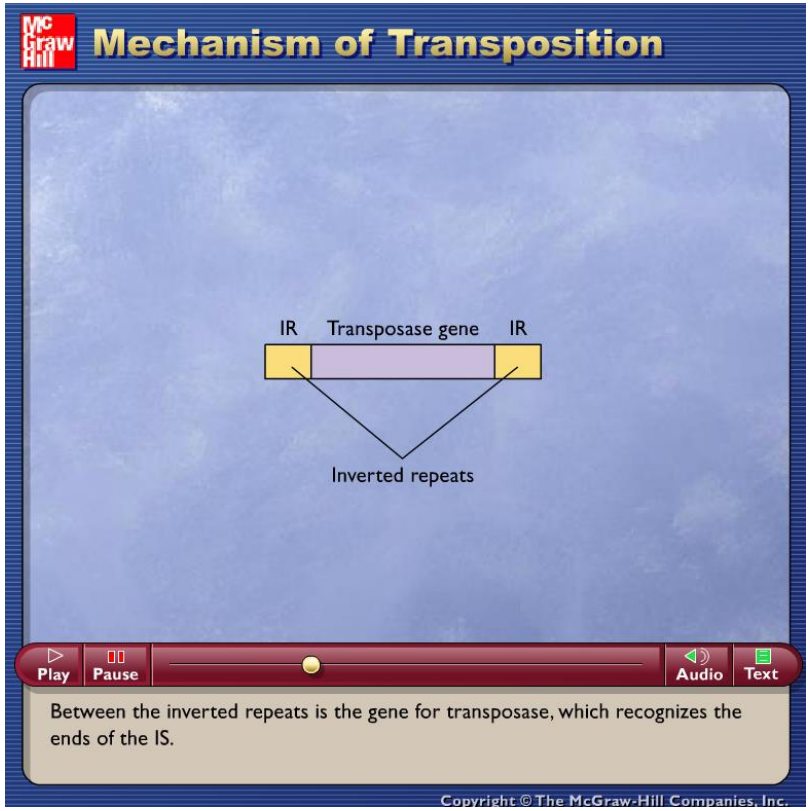
(transposons,Tn):

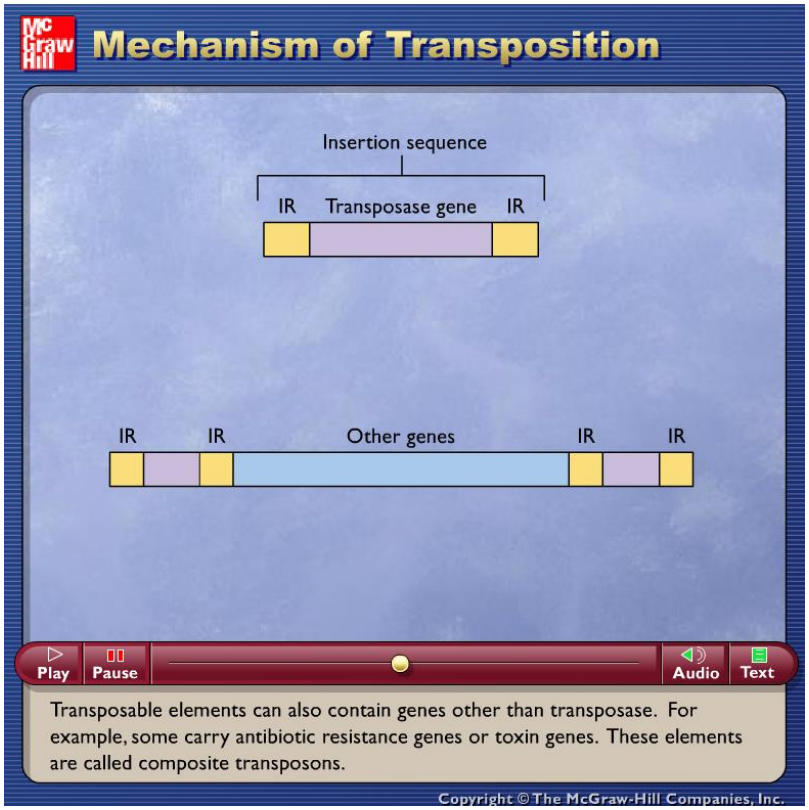
Involves replication of the Transposable element (TE) and insertion of the copy into another chromosomal location.

Only found in bacteria.

Transposable elements

The two classes of bacterial elements are insertion sequences or IS elements and Tn elements





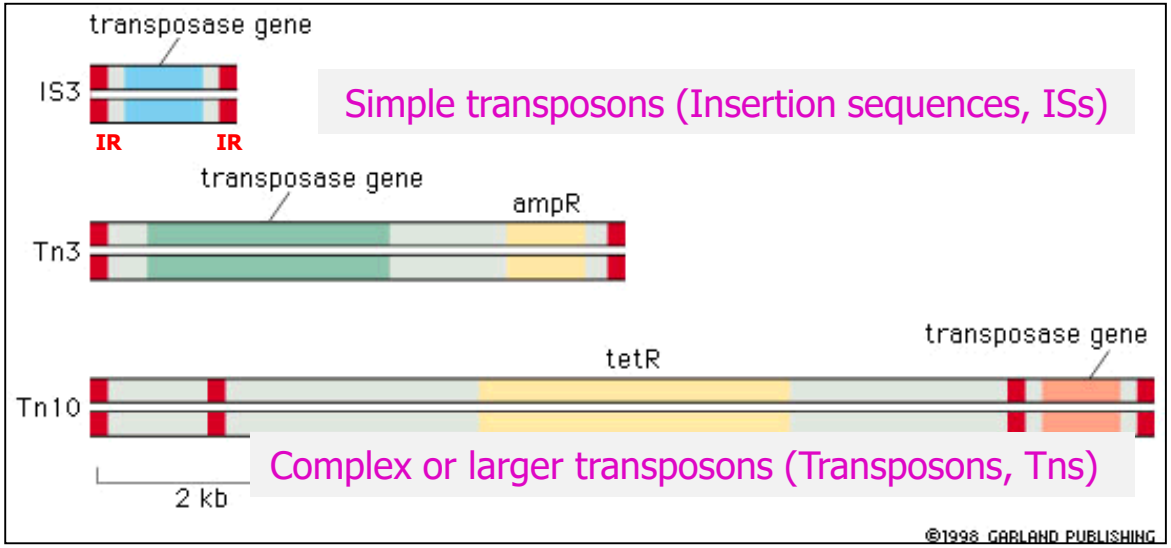
At the ends of some transposons such as IS and Tn contain inverted repeats (IR). Between inverted repeats there is the gene which codes for enzyme transposes. The enzyme transposes recognize the ends of IS i.e. IR.

Transposition

Transposable genetic elements in Bacteria

Inverted repeats

- Inverted repeats are recognized only by the particular transposase encoded.
- Inverted repeats define the boundaries in transposons.
- Inverted repeats are principally found at the origins of replication of cell organism

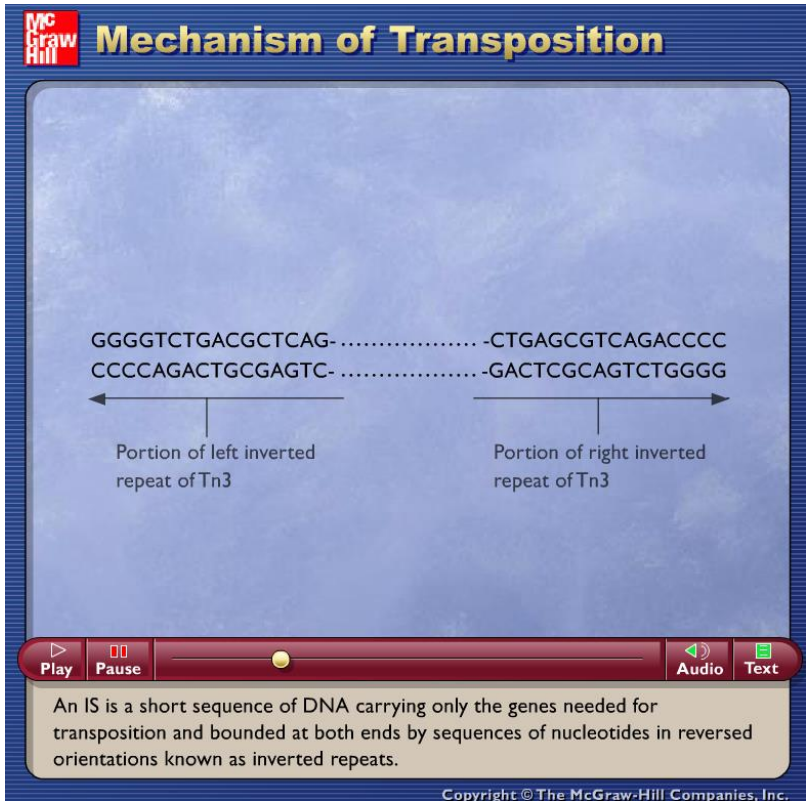


Transposition

Transposable genetic elements in Bacteria

Inverted repeats

- The ends of a simple transposable element contain inverted repeats (IR).
- These sequences (IR) bound both ends of transposons.
- Since these are in reversed orientations these were known as inverted repeats (IR/ITR).
- The inverted repeats are usually ~15 to 25 base pairs in length.



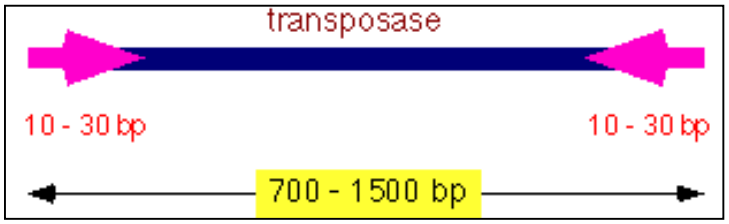
McGraw-Hill Higher Education,2008;..

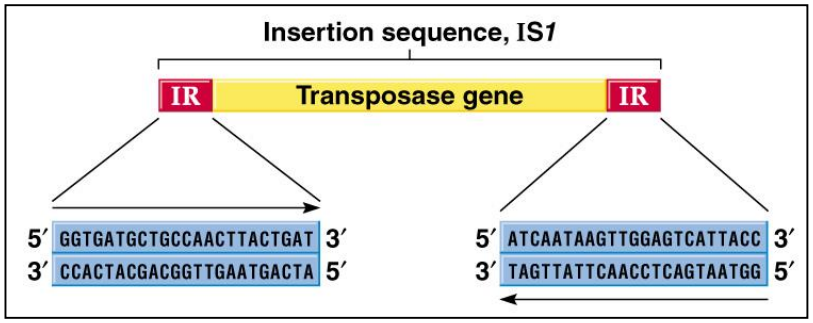
Transposition

Transposable genetic elements in Bacteria

Transposases

- Between the inverted repeats is the gene for transposase.
- Transposase is the enzyme activity involved in insertion of transposon at a new site.
- Both simple & replicative transposons contain a gene encoding a transposase enzyme.
- Transposase binds transposon inverted repeats (IR) and mediates excision.





Transposition

Transposable genetic elements in Bacteria

Transposases

- These enzymes, generically called transposases and site-specific.
- The transposase protein is extremely specific in the sequences that it binds.
- A transposase is responsible:

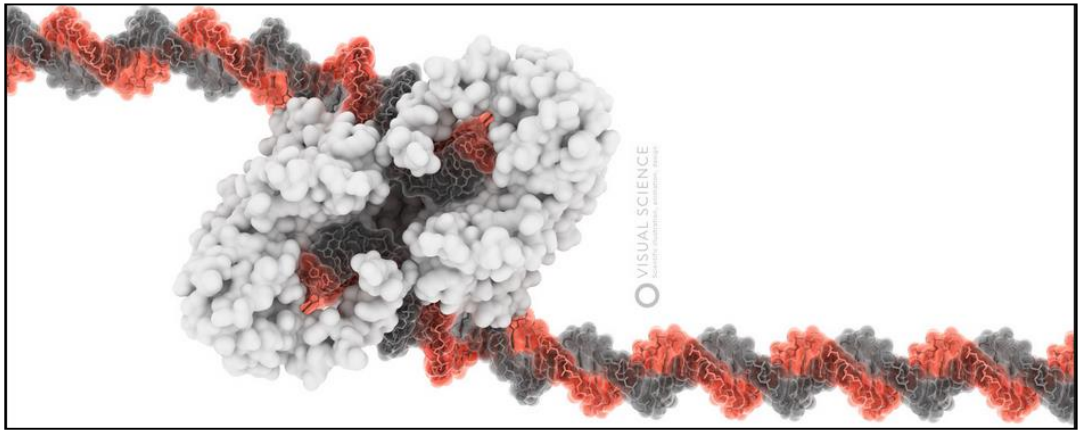
1. recognizes short sequences(IRs) at the ends of a transposon,
2. makes breaks at those locations (excision), and
3. reinserts the element at another location.

Transposition

Transposable genetic elements in Bacteria

Tn5 transposase

- Tn5 transposase is the enzyme that helps bacteria to share antibiotic resistance genes.
- Tn5 is among the first described transposons.
- It was discovered in *Escherichia coli* during studies of bacterial resistance to the antibiotic kanamycin.

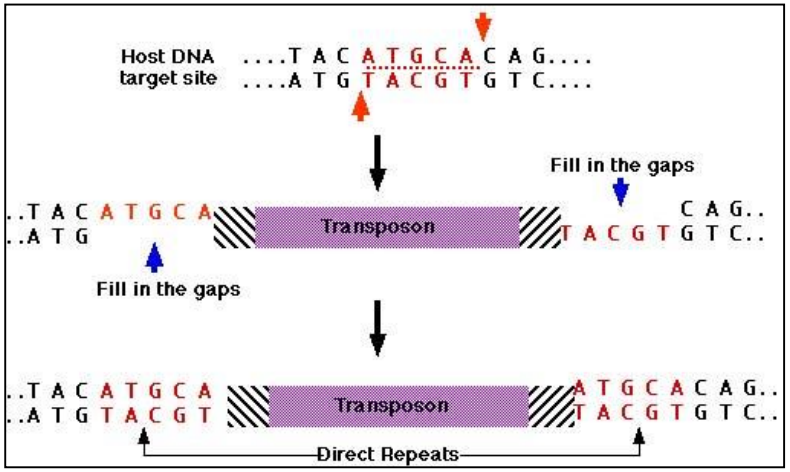


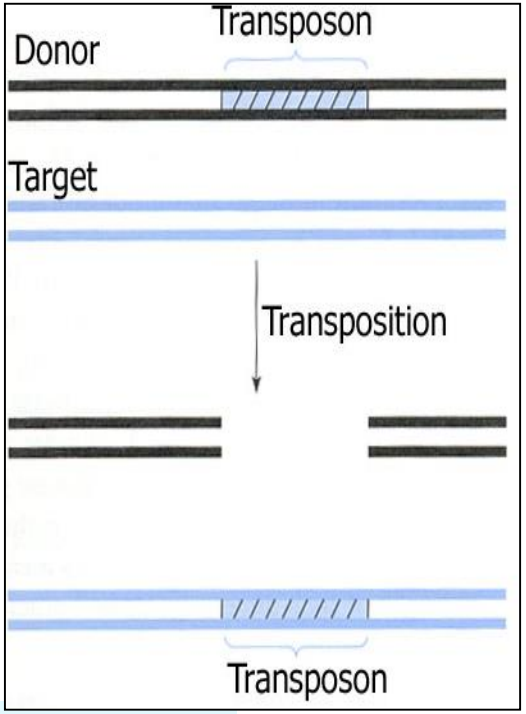
Transposition

Transposable genetic elements in Bacteria

Duplication of target sequence

- Generation of short direct repeats flanking the newly inserted element.
- This results for a staggered cut being made in the DNA strands at the site of insertion.

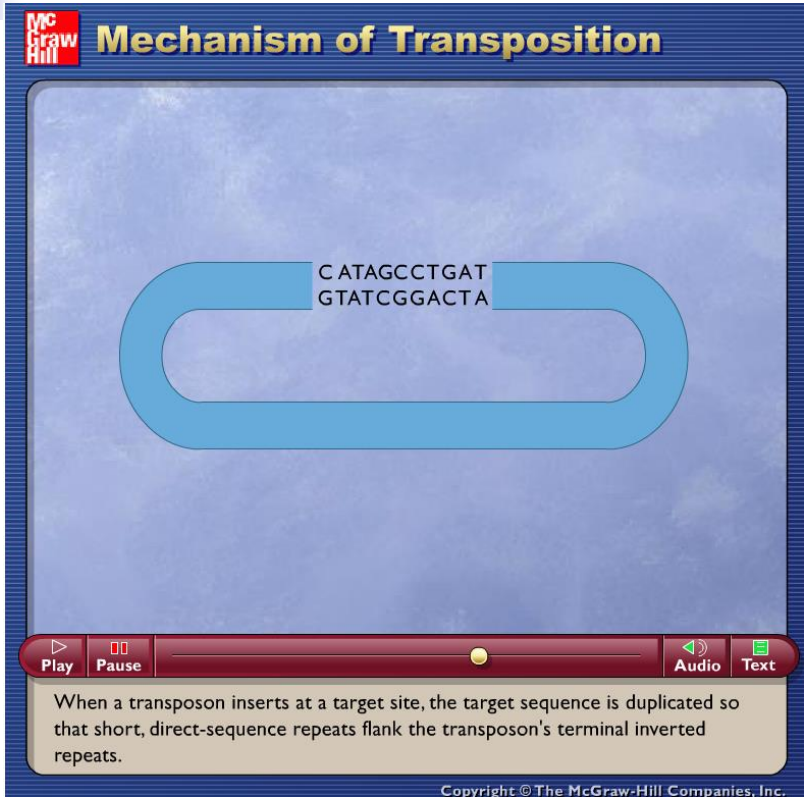


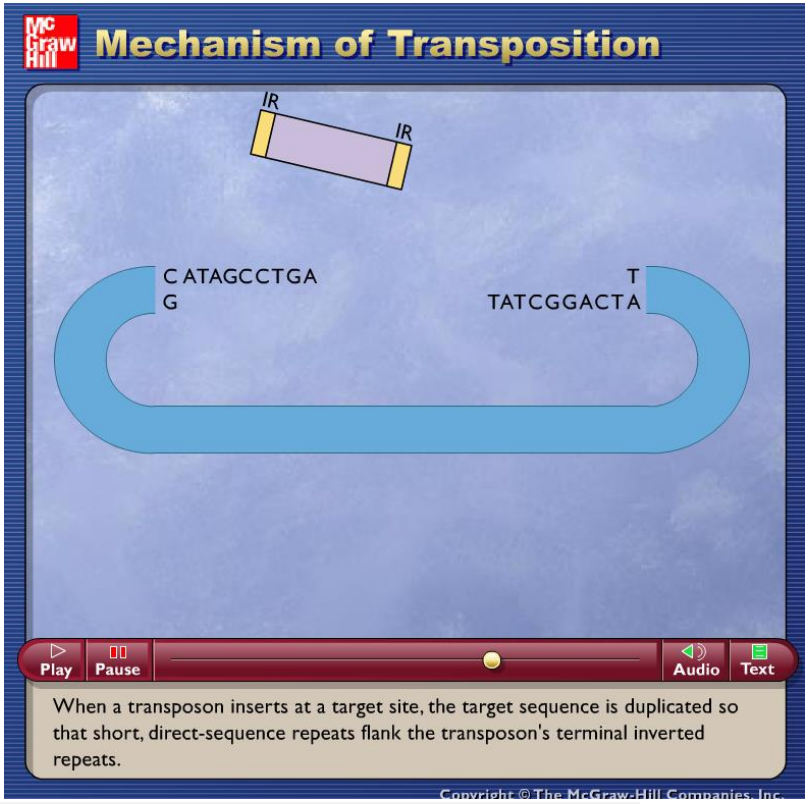


Transposition

Transposable genetic elements in Bacteria

Duplication of target sequence



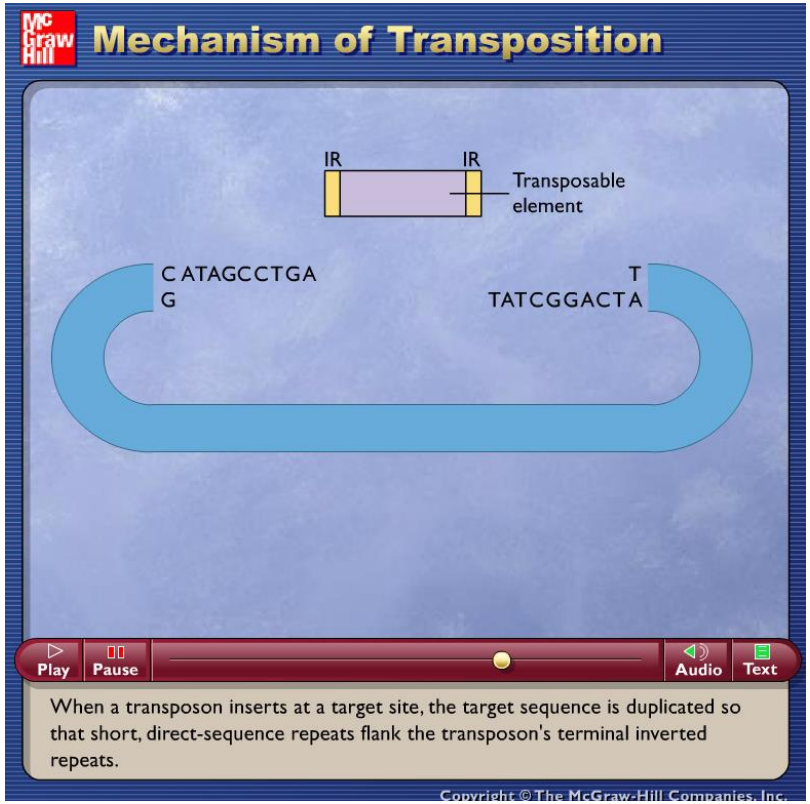


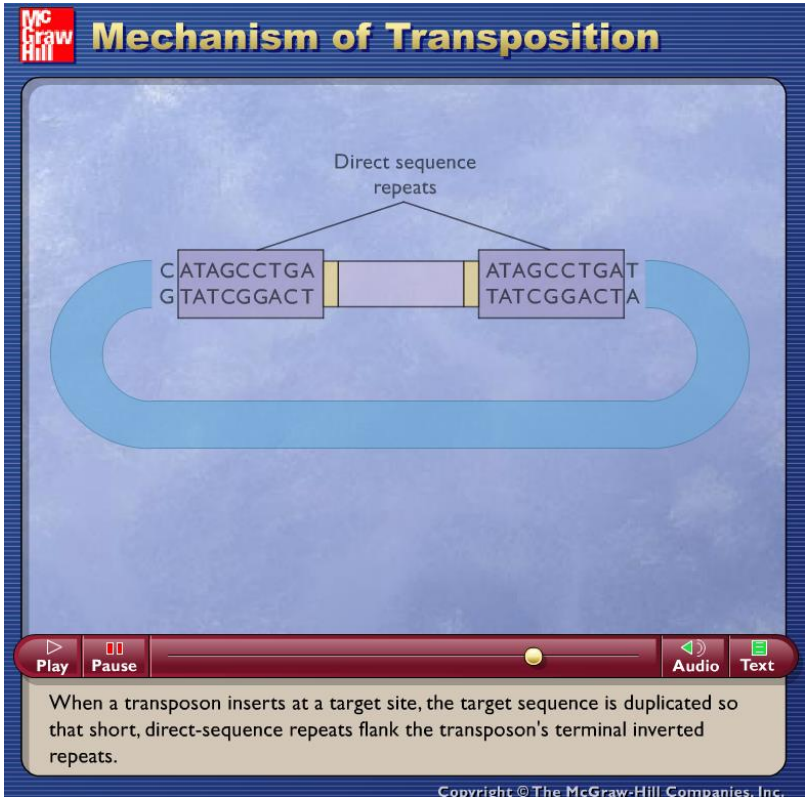
When a transposon inserts as a target site, the target sequence is duplicated so that short, direct- sequence repeats flank the transposon's terminal inverted repeats (IR/ITR). The sequence into which the transposable element insert is called target sequence.

Transposition

Transposable genetic elements in Bacteria

Duplication of target sequence



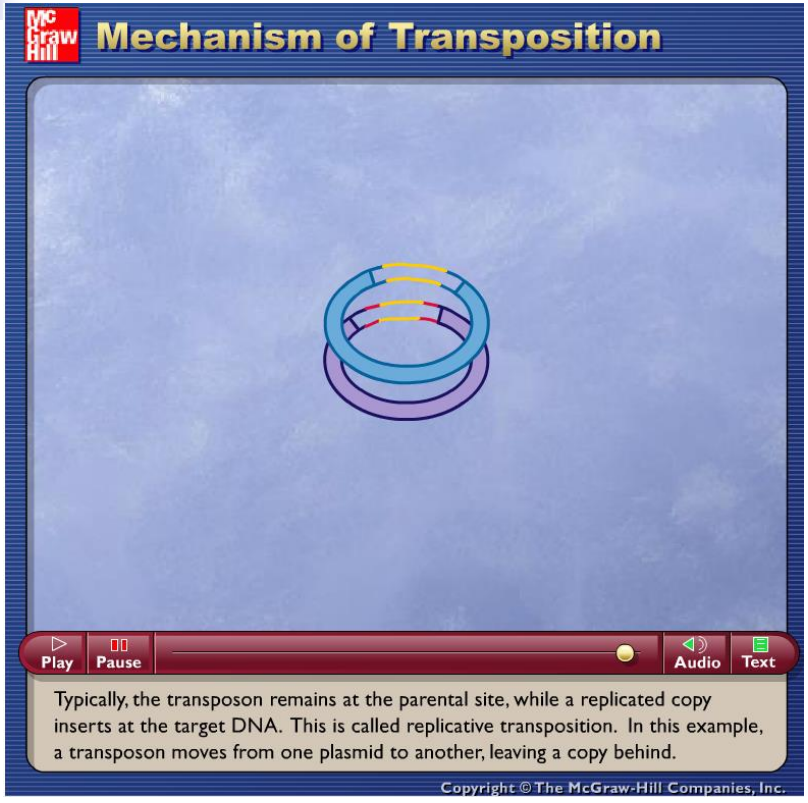


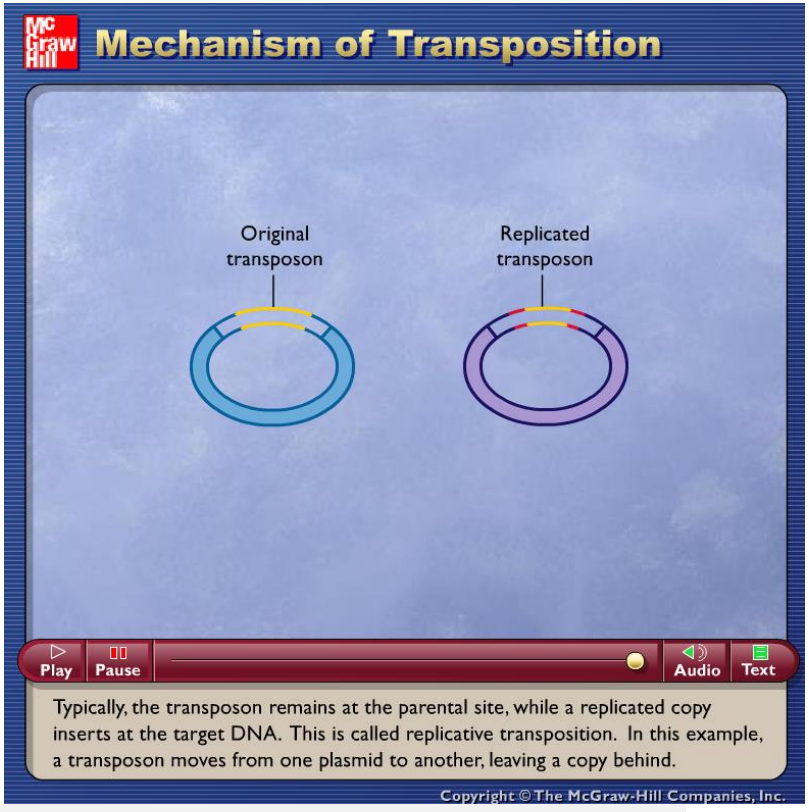
When a transposon inserts as a target site, the target sequence is duplicated so that short, direct- sequence repeats flank the transposon's terminal inverted repeats (IR/ITR).

Transposition

Transposable genetic elements in Bacteria

Duplication of target sequence





Typically, the transposon remains at the parental site, while a replicated copy inserts at the target (receiver) DNA. This is called a replicative transpositon. In this example, a transposon moves from one plasmid to another, leaving a copy behind.

Transposable elements

1. Insertion Sequences(ISs)

In more detail

- Abundance in Prokaryotic Genomes.
- Widespread in archaea as well as eukaryotes, viruses, and plasmids.
- Some bacteria with small genome in size, also lacked IS elements. E.g. *Bacillus subtilis*.
- Encode only genes for mobilization and insertion.
- In bacteria these are located in chromosomes and plasmids.
- ISs are regarded as key determinants of genome plasticity.

Insertion sequences have been found in bacteriophages, in F factor plasmid and many bacteria.

Transposable elements

Insertion Sequences (ISs)

In more detail

- IS elements are transferred between genomes by all the classical mechanisms of horizontal gene transfer (HGT).
- designated by a number identifying its type IS1, IS2, IS3, and so on.
- IS1 first identified in *E. coli*'s galactose operon is 768 bp long and is present with 4-19 copies in the *E. coli* chromosome.

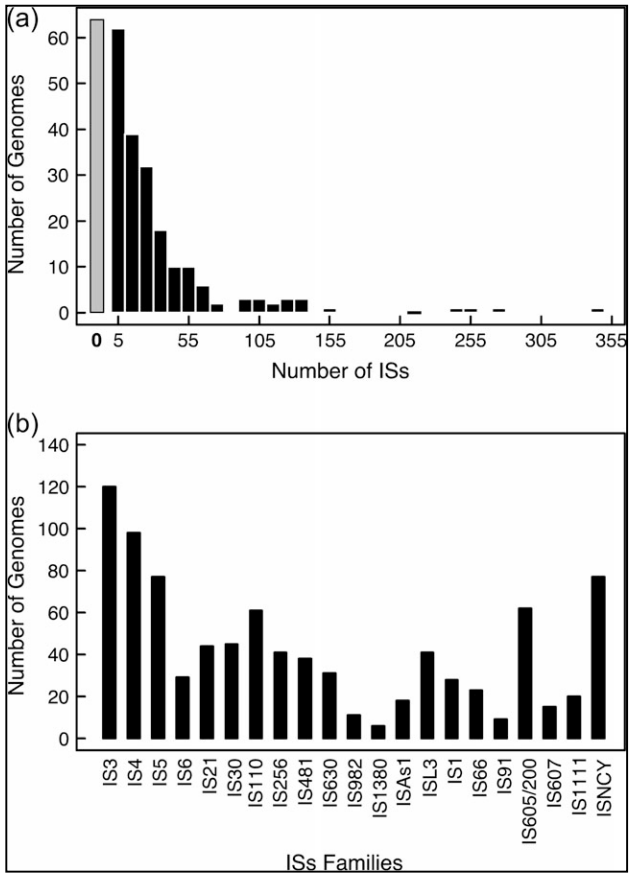
Insertion sequences

Examples of some *E. coli* insertion elements

IS Element	Length (bp)	Inverted Repeats	Target Site size	No. of copies in *E. coli*	
				chromosome	F plasmid
IS1	768	20/23	9	5 - 8	
IS2	1327	32/41	5	5	1
IS3	1258	39/39	3	5	2
IS4	1426	16/18	11, 12 or 14	1 or 2	
IS5	1195	15/16	4	abundant	
IS10 R	1329	17/22	9		

Transposable elements
Insertion Sequences(ISs)
In more detail

- a) Distribution of the number of complete IS elements in the 262 genomes.
- The first gray bar corresponds to the number of genome having no complete IS elements.
- The black bars correspond to the histogram of prokaryotic genomes (archaea and bacteria) harboring at least one IS, with class interval sizes of 10 ISs.
- b) Distribution of the number ofgenomes containing at least one IS, for each IS family.
- We show the data for each of the 20 distinct ISs families, as well as for the group of unclassified elements (ISNCY).



Touchon *et al.*,2007

Transposable elements

Insertion Sequences(ISs)

Distribution of IS elements among the 262 genomes

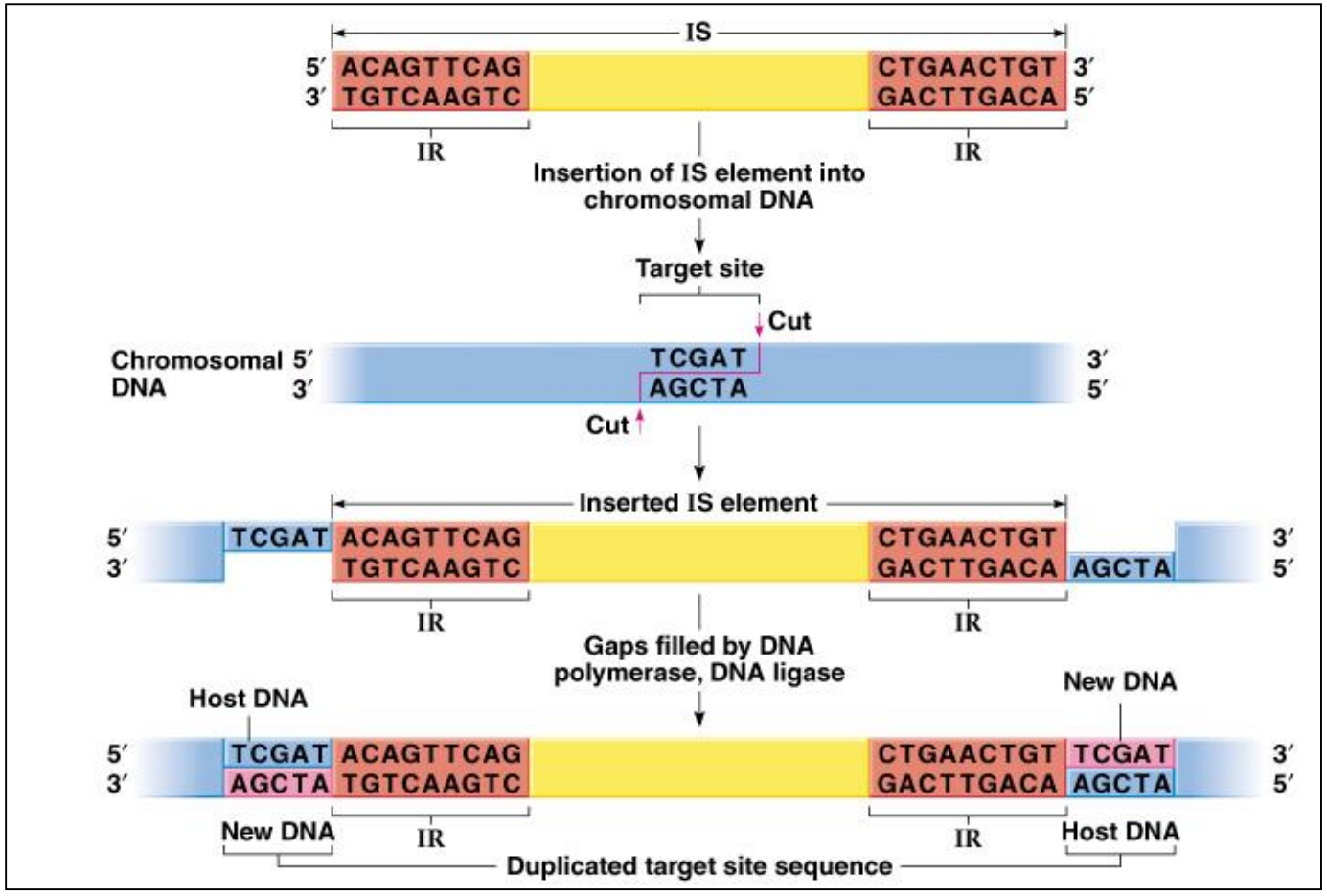
Group	Number genomes	Number genomes IS = 0	% Genomes IS = 0	Median number IS	Min:Maximal number IS	Median genome size (Mb)	Median IS density (Mb)	Min:Maximal IS density (Mb)
Archaea	23	8	34.8	4	0:123	2.01	1.91	0:41.1
Firmicutes	52	3	5.8	12	0:91	2.81	4.29	0:24.2
Actinobacteria	20	5	25	23	0:77	0.30	5.50	0:31.3
Mollicutes	12	5	41.7	2	0:63	0.84	2.43	0:52
Cyanobacteria	11	5	45.4	1	0:53	2.59	0.37	0:20
Spirochaetes	5	3	60	0	0:51	1.14	0	0:10.9
α-Proteobacteria	27	12	44.4	5	0:105	2.06	1.46	0:30.9
β-Proteobacteria	19	0	0	32	2:246	4.30	11.03	0:60.2
δ-Proteobacteria	5	0	0	13	1:33	3.66	3.64	0.3:9.0
ε-Proteobacteria	6	4	66.6	0	0:14	1.72	0	0:6.6
γ-Proteobacteria	59	10	17	24	0:342	4.61	6.68	0:70.9
Chlamydiae	10	9	90	0	0:19	1.20	0	0:7.9
Cytophaga	3	0	0	39	8:42	5.28	6.24	1.5:17.9
Others	10	0	0	7	0:8	2.02	3.85	0:4.3
Total	262	63	24	12	0:342	2.75	3.52	0:70.9

They include: 8 archaea, 9 chlamydia, 5 cyanobacteria, 5 actinobacteria, 3 firmicutes,5 mollicutes, 12 α-proteobacteria, 4 ε-proteobacteria, 10 γ-proteobacteria, and 3 spirochaetes. They also cover a broad spectrum of life styles: 27% are free living and commensal, 11% facultative pathogens, 51% are obligatory pathogens, and 11% are obligatory mutualists.

Touchon *et al.*,2007

Insertion sequences

Integration of IS element in chromosomal DNA



Transposable elements

2. Transposons (Tns)

In more detail

- Transposons are located:

1. in the chromosomes of organisms,
2. in plasmids, and
3. in the genetic material of viruses.

- Tn are several thousands base pair long.
- Contain insertion sequences (IS) or closely related sequences at each end, usually as inverted repeats(IR).
- In yeasts on either side of a transposon is a short direct repeats (DR).

Transposable elements
Transposons(Tns)

- Transposons are small segments of DNA that can move:

1. from one region of a chromosome to another region of the same chromosome, or
2. to a different chromosome or a plasmid.

- Transposons are found in both eukaryotes and prokaryotes and are the most common source of mutation in many organisms.

Difference: Only eukaryotic genomes contain a special type of transposable elements, called retroposons, which use reverse transcriptase to transpose through an RNA intermediate.

Transposable elements

Bacterial transposons(Tns)

Normally, the number associated with the transposon name is italicized

- Tn*3* - beta-lactamase for Ampicillin resistance
- Tn*5* - kanamycin resistance
- Tn*7* - depending on the antibiotic cassette, resistance to trimethoprim, streptothricin, spectinomycin, streptomycin
- Tn*9* - chloramphenicol resistance
- Tn*10* - tetracycline resistance
- Tn*903*
- Tn*1681*
- Tn*gamma delta*

Transposon	Antibiotic or other resistance marker	Length (bp)	Inverted Repeat
Tn5	kanamycin	5700	IS50
Tn9	chloramphenicol	2638	IS1
Tn10	tetracycline	9300	IS10

Bacterial Transposon List

Transposons(Tns)

In Gram-ve and Gram+bacteria

- This list includes many transposons from various Gram negative and Gram positive bacteria and some accessions.

Transposon Name	Family	Organism	Accession (if available or known)	Reference(s)	Notes
Tn3		*E. coli*	V00613		
Tn5		*E. coli*	U00004		
Tn7		*E. coli*			
Tn10		*E. coli*, but accession is from a *Shigella* isolate	AF162223		
Tn21	Tn3	*E. coli*	AF071413		
Tn1721	Tn3	*E. coli*	X61367		
Tn gamma delta (Tn1000)	Tn3	*E. coli*	D16449		
Tn2501	Tn3	*E. coli*			
Tn1545		*Streptococcus pneumoniae*			Gram positive element

Bacterial Transposon List
In Gram-ve and Gram+bacteria
Continued

- This list includes many transposons from various Gram negative and Gram positive bacteria and some accessions.

Tn4555		*Bacteriodes*			
Tn5397		*Clostridium difficile*			
Tn5398		*Clostridium difficile*			
Tn4451		*Clostridium difficile*			
Tn4453a		*Clostridium difficile*			
Tn4453b		*Clostridium difficile*			
Tn544		*Staphylococcus aureus*			Gram positive element
R391		*Proteus*, *Vibrio*, *Pseudomonas*			
R997		*Proteus*, *Vibrio*, *Pseudomonas*			
Tn501	Tn3				
Tn512		*Psuedomonas*	EU306744		
Tn502		*Pseudomonas aeruginosa*	EU306743		

Bacterial Transposon List
In Gram-ve and Gram+bacteria
Continued

- This list includes many transposons from various Gram negative and Gram positive bacteria and some accessions.

Tn5468	Tn7	*Thiobacillus*			
Tn5053	Tn7	*Pseudomonas*			
Tn5075		*E. coli*	AF457211		
Tn552	Tn7	*Staphylococcus aureus*	X52734		
Tn402	Tn7	*Pseudomonas*	GQ857074		
Tn1	Tn3				
Tn2	Tn3				
Tn801	Tn3	*Pseudomonas*	AF080442		
Tn917	Tn3	*Enterococcus faecalis*	M11180		Gram positive element
Tn551	Tn3	*Staphylococcus aureus*	Y13600		Gram positive element
Tn*phoA* (Tn5*phoA*)	Synthetic Tn5 derivative	*E. coli*	U25548		This is a SYNTHETIC transposon & is NOT naturally occurring!!

Bacterial Transposon List
In Gram-ve and Gram+bacteria
Continued

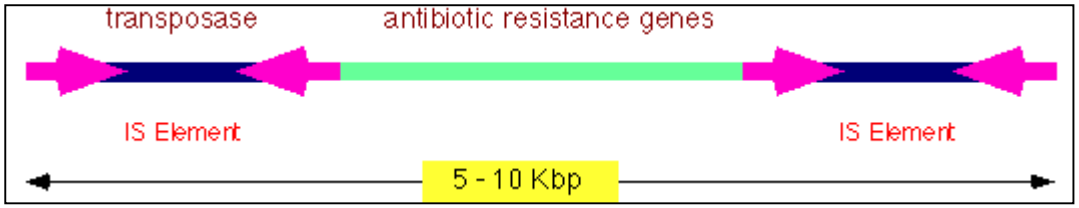
- This list includes many transposons from various Gram negative and Gram positive bacteria and some accessions.
- This table is NOT comprehensive and if you have other transposons to add or would like to correct an entry, please sign in or email the EcoliWiki Team at ecoliwiki@gmail.com for an account.

Tn9		*E. coli*	V00622		
Tn554		*Staphylococcus aureus*	X03216		Gram positive element
Tn2555		*E. coli*	AY485150	PMID:15063507 [1]	
edit table					

Transposable elements
Transposons (Tns)

- These are different from IS elements because have complex structure and carry additional genes.
- Two types of transposons (Tn) found in bacteria:

1. Composite transposons: carry genes such as antibiotic resistance genes and terminate with IS element at each end).



2. Noncomposite transposons (carry genes which do not terminate with IS elements).

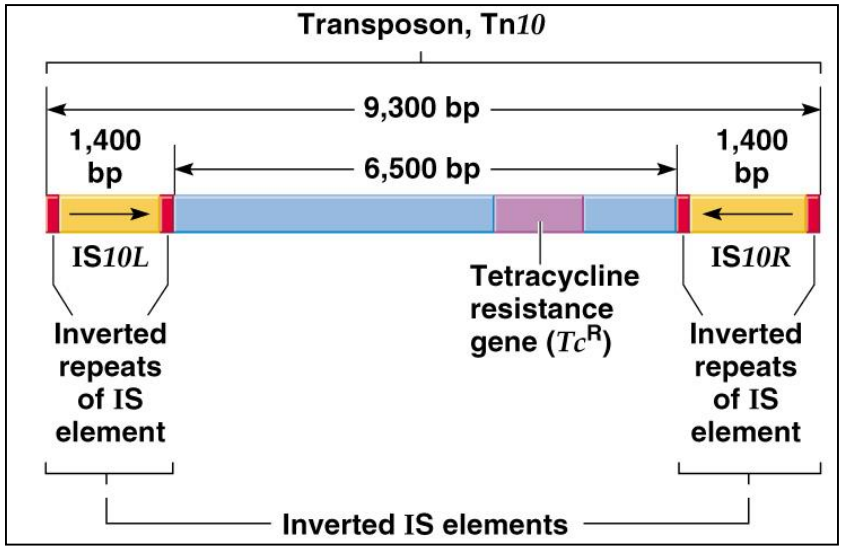
Transposable elements

Composite transposons (Tn)

Transposons, Tn*10*

- Carry genes (example might be a gene for antibiotic resistance) flanked on both sides by IS elements.
- Tn*10* is 9.3 kb and includes 6.5 kb of central DNA (includes a gene for tetracycline resistance) and 1.4 kb inverted IS elements.

1. IS elements supply transposase and
2. ITR recognition signals.



Transposon terminate with IS element at each end.

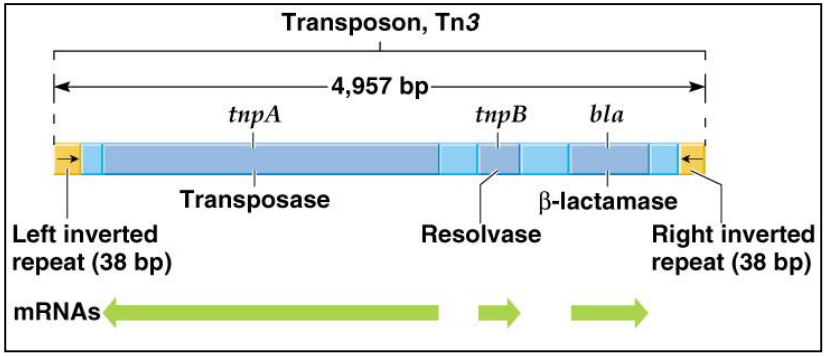
Transposable elements

Noncomposite transposons (Tn)

Transposons, Tn*3*

- Carry genes (example might be a gene for antibiotic resistance) but do not terminate with IS elements.
- Ends are non-IS element repeated sequences.
- Tn3 is 5 kb with 38-bp ITRs and includes 3 genes:

1. *bla* (β-lactamase),
2. *tnpA* (transposase), and
3. *tnpB* (resolvase, which functions in recombination).



Transposon do not terminate with IS element at each end.

Transposable elements

3. Bacteriophage Mu

A phage that reproduces by transposition

- Bacteriophage Mu is essentially a 40kb transposon carrying capsid genes which replicates via transposition and is then packaged into virions.
- Bacteriophage Mu is a temperate bacteriophage that uses DNA-based transposition in its lysogenic cycle.
- It can infect *E. coli* as a normal phage.
- It is a temperate phage so it has both lytic and lysogenic growth cycles.
- The Mu prophage integrates into the host genome.

bacteriophage Mu (Mu=mutator), is a temperate bacteriophage causes mutations in its host bacterium *Escherichia coli* and related phages.

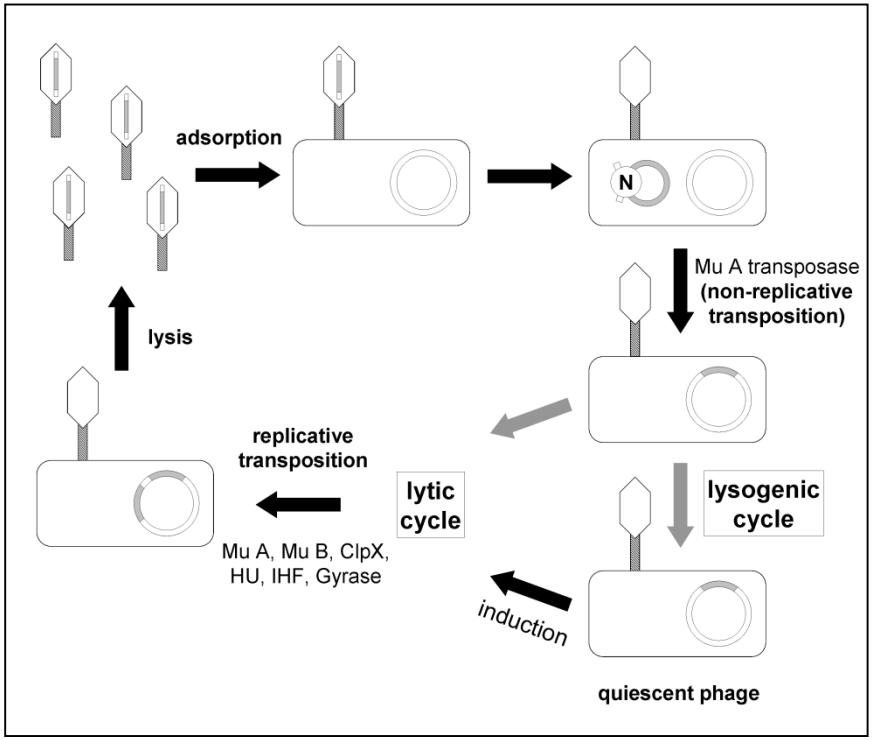
Transposable elements
Bacteriophage Mu

- Bacteriophage Mu, also known as mu phage or mu bacteriophage, is a mulikevirus (the first of its kind to be identified) of the family Myoviridae which has been shown to cause genetic transposition.
- It is of particular importance as its discovery in *E. coli* by Larry Taylor in the late 1950s.
- He first observed the mutagenic properties of Mu.
- This discovery opened up the world to an investigation of transposable elements and their effects on a wide variety of organisms.

Wikipedia

Transposable elements
Bacteriophage Mu

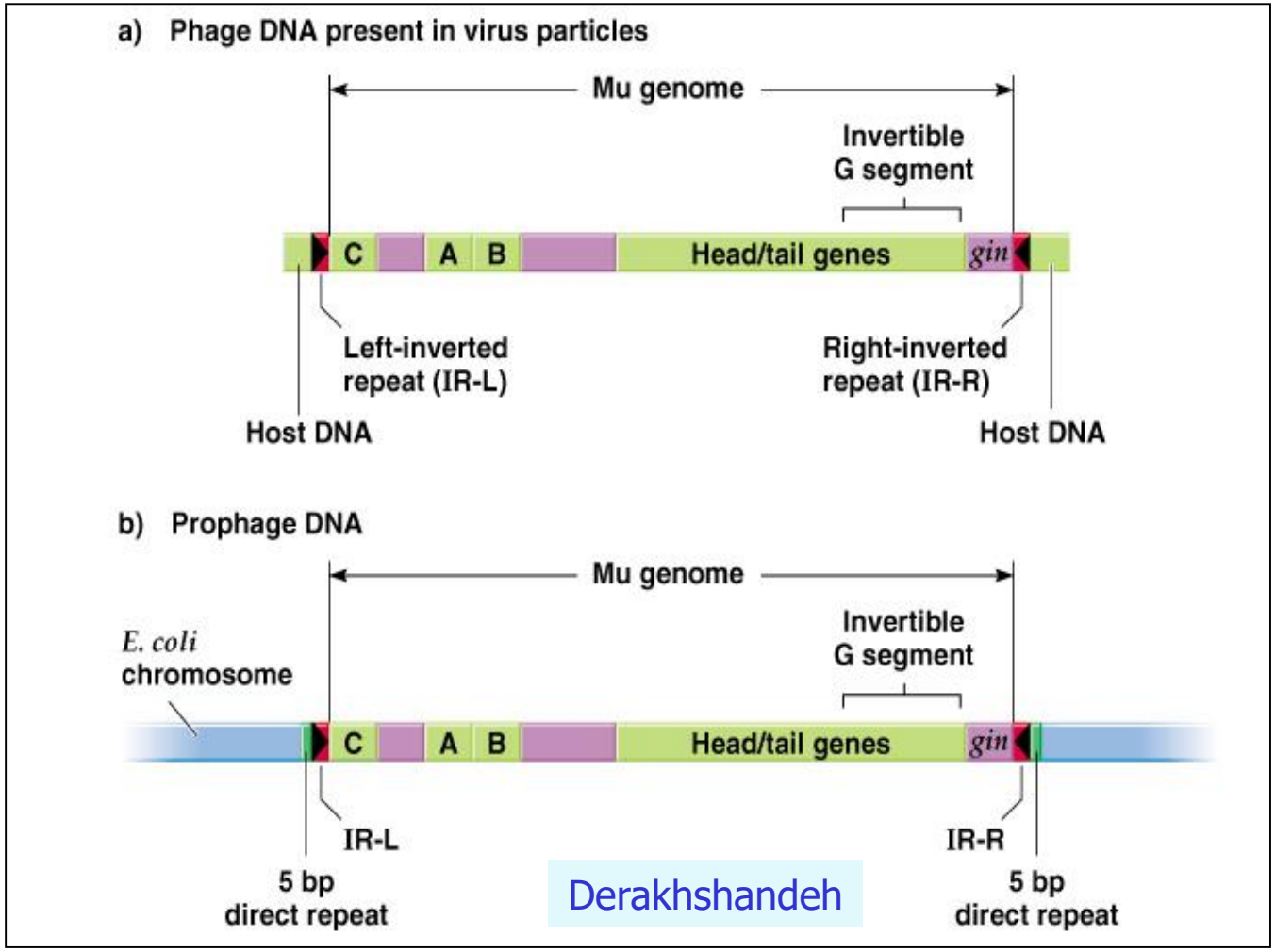
1. 37 kb linear DNA with central phage DNA and unequal lengths of host DNA at each end;
2. Mu integrates by transposition;
3. replicates when *E. coli* replicates;
4. During the lysogenic cycle, Mu remains integrated in *E. coli* chromosome.



Transposable elements

Bacteriophage Mu

When Mu replicates, it does so by replicative transposition events in which each new copy of Mu is inserted somewhere else in the chromosome



Derakhshandeh

Bacteriophage Elements

Bacteriophage Mu

The advantages/disadvantage of Mu

- The advantages of the use of Mu are:

1. it is not normally found in the bacterial genome, therefore there are few problems with homology to existing sequences in the chromosome; in contrast to most other transposons.
2. Mu does not need a separate vector system since it is itself a vector.
3. A wide variety of useful mutants of Mu have been generated.

- The disadvantage of Mu:

1. it is a bacteriophage and therefore can kill the host(bacterial) cell.

Mechanism of transposition

Two distinct mechanisms of transposition

- The mechanism of transposition in bacteria can be either "copy and paste" or "cut and paste".

1. Conservative (cut and paste) transposition:

- The element itself moves from the donor site into the target site, involving excision of the element and reintegration at a new site.

2. Replicative (copy and paste) transposition:

- The transposable element is copied, and one copy remains at the original site while the other inserts itself at a new. The element moves a copy of itself to a new site via a DNA intermediate.

Types and mechanisms of transposition

Conservative (cut and paste) transposition
Replicative (copy and paste) transposition

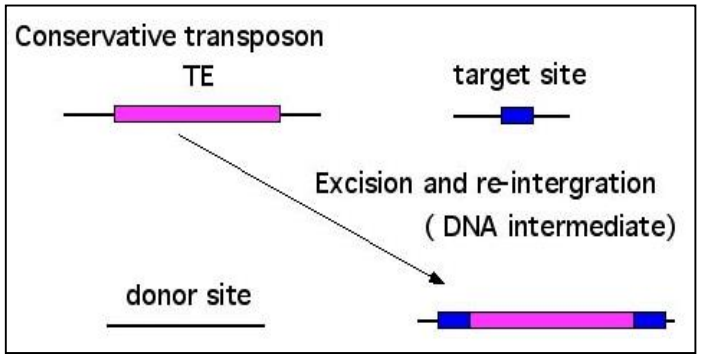
1. **Conservative transposition** (direct or simple or non-replicative transposition): Transposable element (TE) itself moves, "cut & paste". The conservative mechanism is used by elements like Tn*10*.
2. **Replicative transposition**: Transposable element (TE) is copied then moved, "copy & paste". The replicative mechanism used by elements like Tn*3* and bacteriophage Mu.

- Elements can use one or both types of transposition.

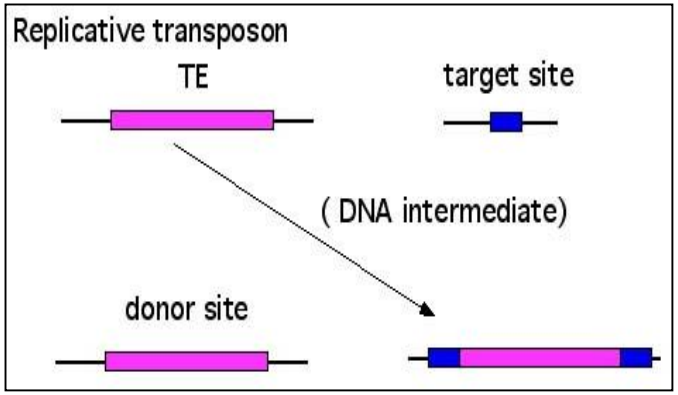
Types and mechanisms of transposition

Conservative (cut and paste) transposition

Replicative (copy and paste) transposition



The conservative mechanism is used by elements like Tn*10*.



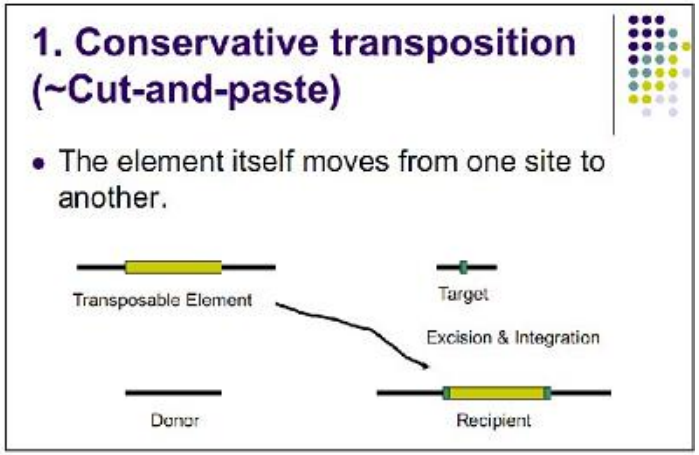
The replicative mechanism used by elements like Tn*3* and bacteriophage Mu.

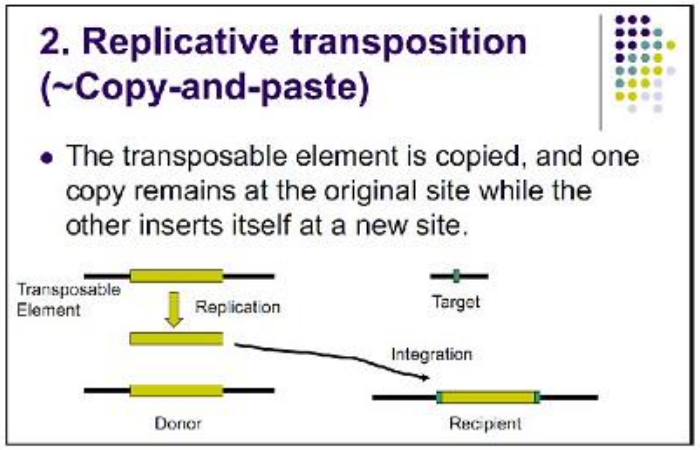
The sequence into which the transposable element insert is called target sequence.

Types and mechanisms of transposition

Conservative (cut and paste) transposition

Replicative (copy and paste) transposition



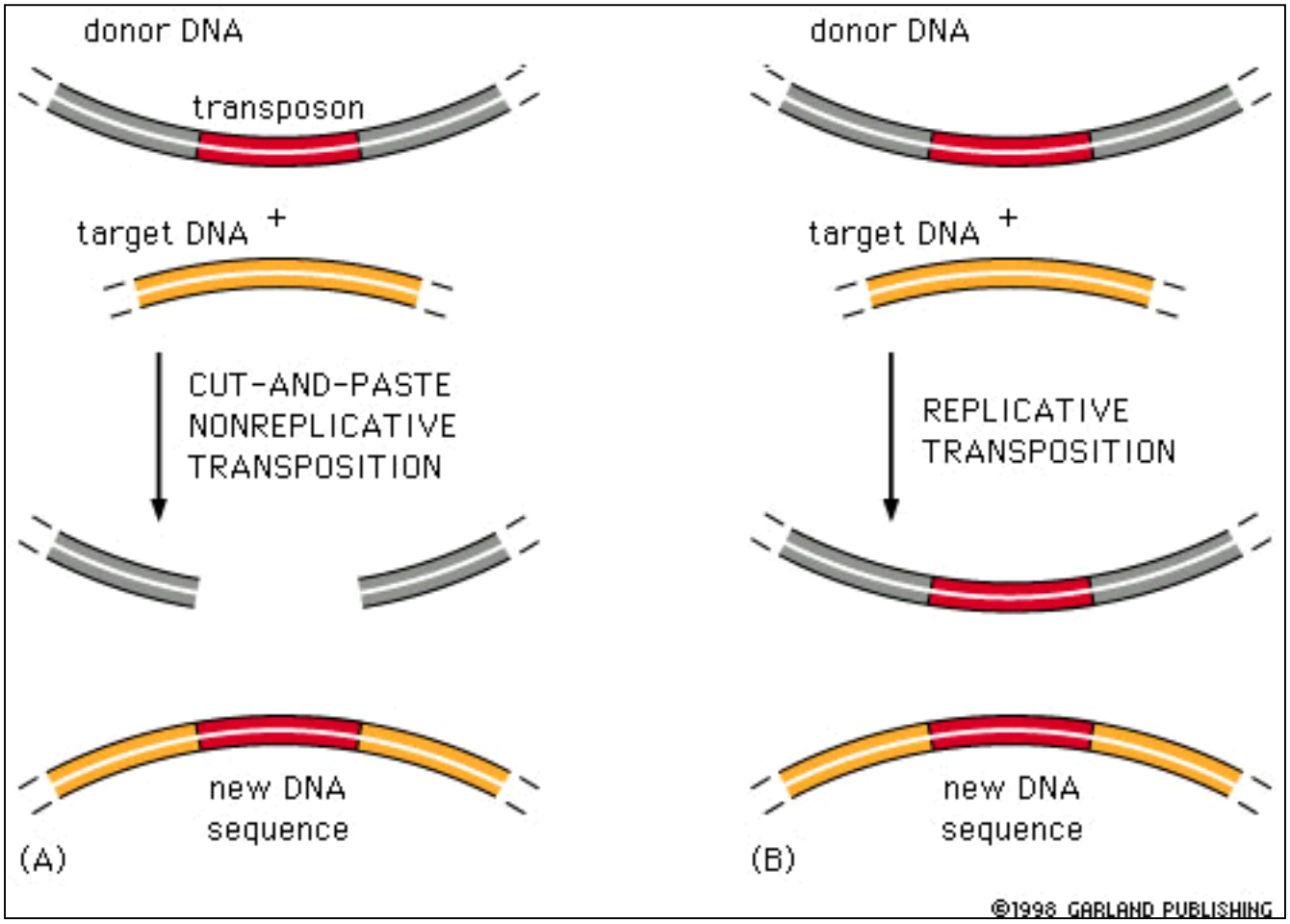


The conservative mechanism is used by elements like Tn*10*.	The replicative mechanism used by elements like Tn*3* and bacteriophage Mu.

Types and mechanisms of transposition

Conservative (cut and paste) transposition

Replicative (copy and paste) transposition



Transposition
Evolution of Transposons
Horizontal gene transfer

- Transposon(IS and Tn) abundance correlates positively with the frequency of:

1. horizontal gene transfer (HGT),
2. genome size,
3. pathogenicity,
4. nonobligatory ecological associations, and
5. human association.

So far, there is no empirical evidence that pathogens have a particularly high frequency of IS elements.

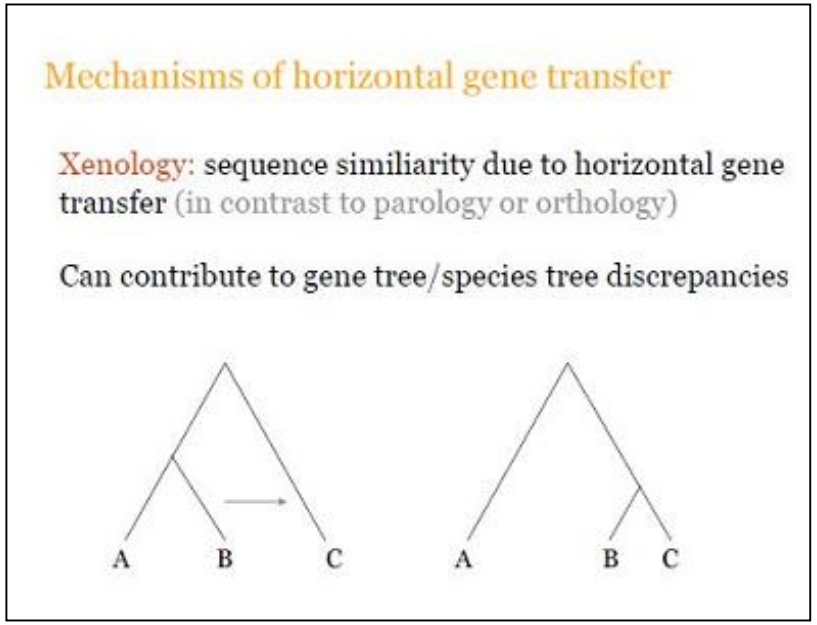
Touchon *et al.*,2007;..

Transposition

Evolution of Transposons

Horizontal gene transfer

- Transposons are found in all major branches of life.
- Duplications and DNA rearrangements contributed greatly to the evolution of new genes.
- Transposons play an important role in the evolution of bacterial genomes.
- Transposible elements also facilitate horizontal gene transfer between unrelated organisms.



Transposition

Transposons causing diseases

- Transposons have been shown to cause mutations in bacteria, fungi, plants and animals.
- Human diseases caused by transposons include:

1. hemophilia A and B
2. severe combined immunodeficiency
3. Porphyria
4. Cancer
5. Duchenne muscular dystrophy

Transposition

Direct and indirect effects of transposons in disease incidence/mutation/overexpression or mis-expression of a gene, disease management

- Plant bacterial diseases in relation with transposons:
- Virulence factors are often associated with transposable elements or genomic (pathogenicity islands) and may acquired by other microbes via HGT.
- Bacterial mutant strains affected in their interactions with plants after transposon mutagenesis.
- Transposon-encoded streptomycin resistance in plant pathogenic bacteria.Transposon Tn*5393* carrying streptomycin resistance genes(*strA* and *strB*) in *Erwinia amylovora* and other gram-negative bacteria.

Transposition
Applications

- The features of transposons have made them an important experimental tool in molecular biology.
- Researchers use transposons as a means of mutagenesis.

1. To identifying the mutant allele;
2. To study the chemical mutagenesis methods;
3. To study gene expression(the introduction of transposons into a cell is a convenient way to abolish the expression of a gene).

Methods for generation of mutant populations

- The most reliable method to ascertain gene function is:

1. to disrupt the gene, and
2. determine the phenotype change in the resulting mutant individual.

- Two most popular methods to generate mutants:

1. Insertional mutagenesis;
2. Deletion mutagenesis.

Technical limitations of transposon mutagenesis

EZ::TN™ Transposome™ system

- Transposon mutagenesis can have the following technical limitations:

i. The transposase gene must be expressed in the target host;

ii. The transposon must be introduced into the host on a suicide vector;

iii. The transposase gene on the transposon can be expressed in subsequent generations, which may result in genetic instability.

- The EZ::TN ™ Transposome ™ system(EZ-Tn5™ Transposase) was developed to overcome the above problems.

How do we use a transposon for mutagenesis?

Use of the EZ::TN™ Transposome™

- **EZ-Tn5™ Transposon Construction Vectors:**
- EZ::TN Transposomes provide an efficient and reliable method for generating a library of random gene knockouts *in vivo*.
- EZ-Tn5™ Transposase is a hyperactive form of Tn5 transposase.
- The highly purified, single-subunit enzyme can be used to randomly insert (transpose or "hop") any EZ-Tn5 Transposon into any target DNA *in vitro* with an efficiency up to >10^6 insertion clones per standard reaction.

How do we use a transposon for mutagenesis?

Use of the EZ::TN™ Transposome™

- The Transposome is so stable that it can be electroporated into living cells.
- Once in the cell, the Transposome is activated by intracellular Mg^{2+} and the EZ-Tn5 Transposon component is randomly inserted into the host's genomic DNA.

How do we use a transposon for mutagenesis?

Use of the EZ::TN™ Transposome™

- A typical EZ-Tn5 transposition reaction requires four components:

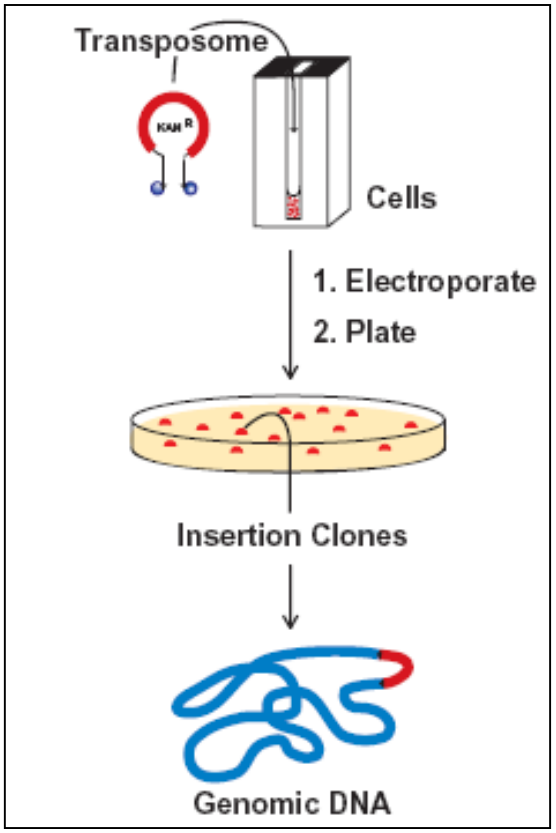
1. the EZ-Tn5 Transposase;
2. an EZ-Tn5 Transposon;
3. a target DNA; and
4. the presence of Mg^{2+}.

- The highly random insertion of an EZ-Tn5 Transposon into the target DNA proceeds by a cut-and-paste mechanism, catalyzed by the EZ-Tn5 Transposase, and results in a 9-bp duplication of target DNA sequence immediately adjacent to both ends of the Transposon.

How do we use a transposon for mutagenesis?

Use of the EZ::TN™ Transposome™

- Because the EZ::TN Transposase is degraded in the cell, no subsequent transposition occurs.
- EZ::TN Transposome insertions have been demonstrated for a number of species, including:
- *Escherichia*
- *X. fastidiosa*
- *Pseudomonas*.



The transposition can be identified when a known DNA sequence or selection markers are inserted within the elements.

2. Genetic elements expression

Gene expressions

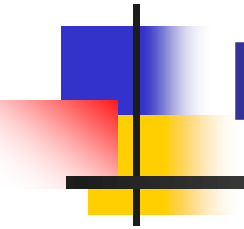
1. Protein synthesis
2. Gene transfer

1. Mechanisms of protein synthesis

- Replication
- Transcription
- Translation

2. Mechanisms of gene transfer

- Transformation
- Transduction
- Conjugation

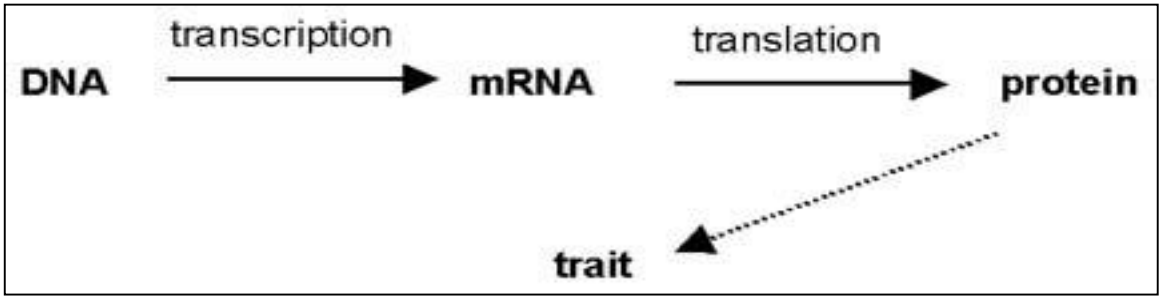


DNA replication

DNA to DNA information transfer

The central dogma of molecular biology
Replication, transcription and translation

DNA → RNA → Protein → Small Molecules



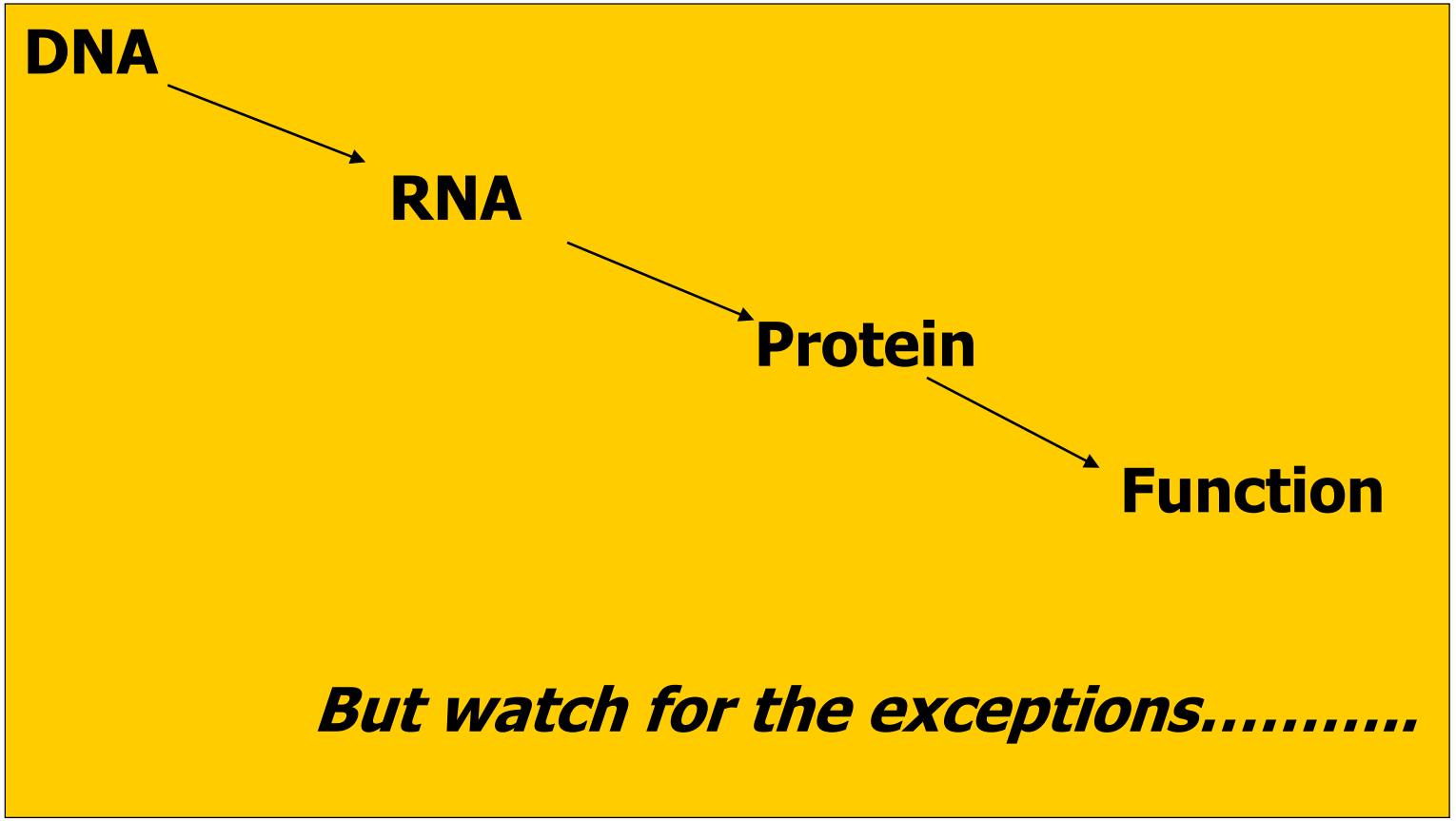
Transcription is very similar to DNA replication but there are some important differences. E.g.

DNA replication is the process of producing two identical replicas from one original DNA molecule. The enzyme that catalyzes DNA replication is DNA Polymerase.

RNA polymerase (RNAP or RNA pol), is an enzyme that catalyzes the reaction and produces RNA.

The Central Dogma

Replication, transcription and translation



Campbell and co-workers

DNA vs. RNA

- DNA and RNA share many similarities. However, a few key differences stand out.

1. DNA serves as the primary storage of genetic information for all cells, whereas RNA molecules assist to execute the instructions encoded by that DNA.
2. DNA always stays in the nucleus, whereas RNA travels from the nucleus to the cytoplasm.

1. RNA contains the base uracil (U) instead of the thymine (T) in DNA.
2. It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.
3. Inosine (a precursor to adenosine) is a purine nucleoside containing the base hypoxanthine and the sugar ribose, which occurs in tRNAs.

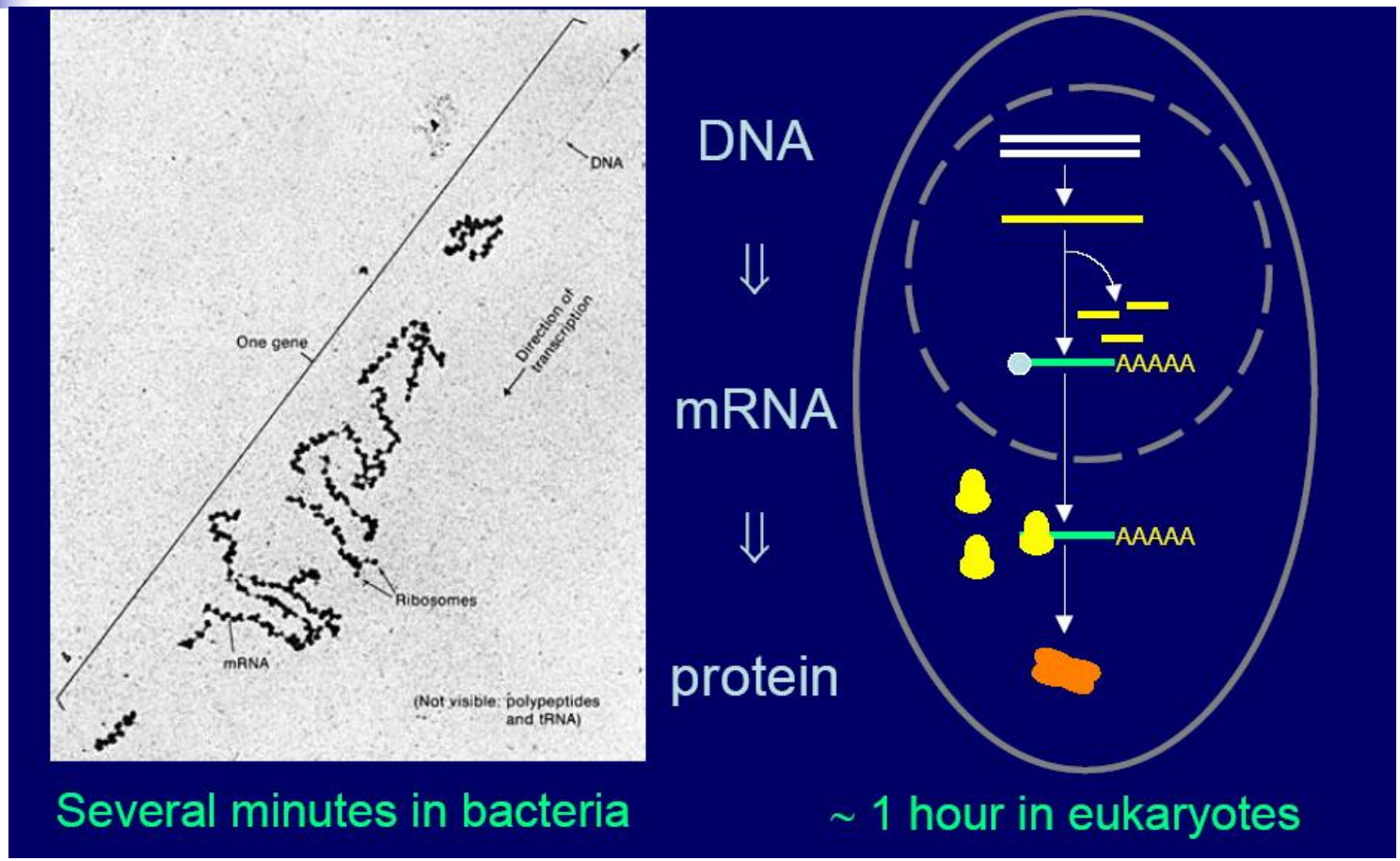
Comparison of DNA, RNA and proteins
The three essential macromolecules of life

	DNA	RNA	Proteins
Encodes genetic information	Yes	Yes	No
Catalyzes biological reactions	No	Yes	Yes
Building blocks (type)	Nucleotides	Nucleotides	Amino acids
Building blocks (number)	4	4	20
Strandedness (having a strand or strands)	Double	Single	Single
Structure	Double helix	Highly complex	Highly complex
Stability to degradation	Extremely high	Variable	Variable
Repair systems	Yes	No	No

These building blocks are made of three parts: 1) a phosphate group, 2) a sugar group and 3) one of four types of nitrogen bases. RNA less stable than DNA because it is more susceptible to hydrolysis.

The Central Dogma

Gene expression

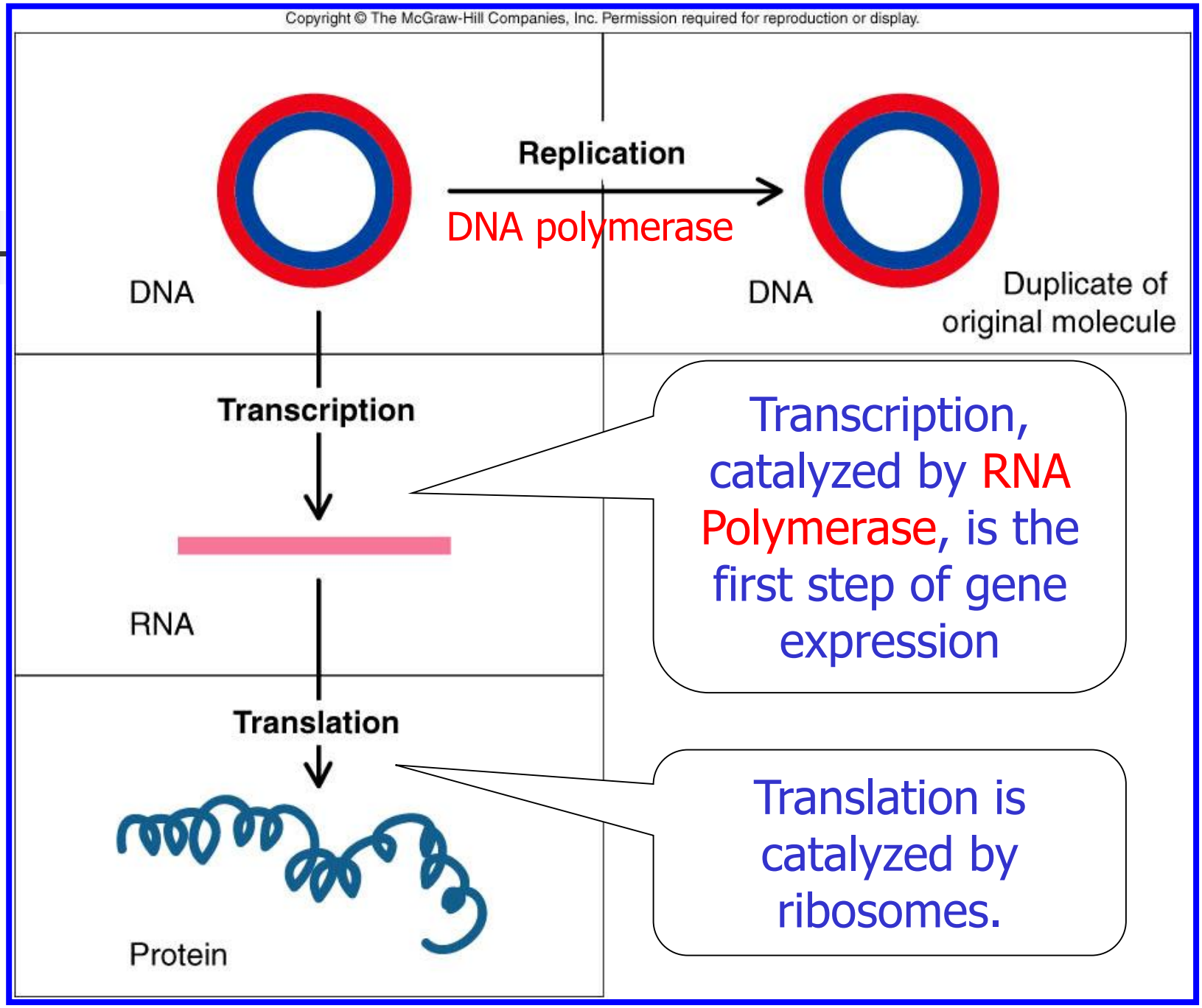


The Central Dogma

Gene expression

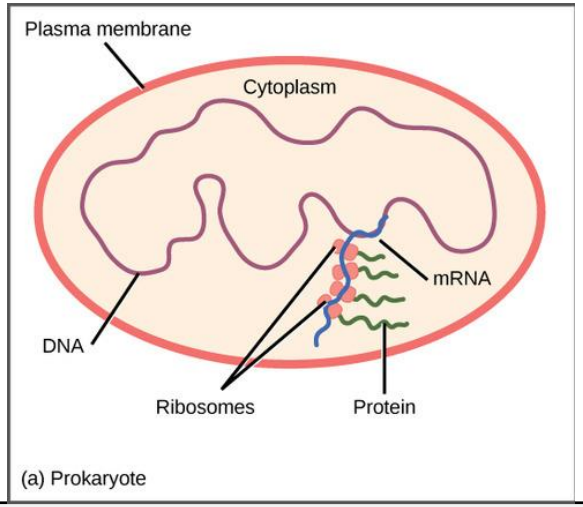
- The first step for expression of any gene function is transcription of RNA.
- Thus dsDNA is transcribed into ss mRNA.
- The mRNA then is translated into protein by ribosomes.
- To make protein, lots of kinds of RNA are needed.
- All kinds of RNA (tRNA, mRNA & rRNA) are made in the same way from a DNA template.
- Non-protein coding genes (e.g. rRNA genes, tRNA genes) are not translated into protein.
- RNA is not used as template to make more RNA.
- Synthesis of DNA is catalyzed by DNA polymerase (and associated enzymes).
- Synthesis of RNA is catalyzed by RNA polymerase.

Central Dogma

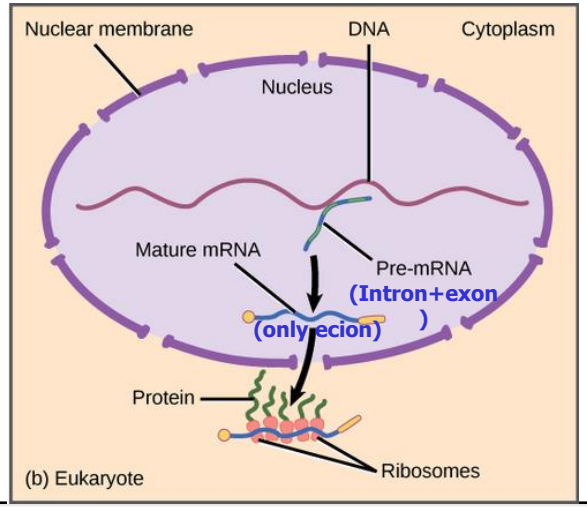


DNA replication, transcription and Translation

Prokaryotic vs. Eukaryotic transcription



In a prokaryotic cell, transcription and translation are coupled; that is, translation begins while the mRNA is still being synthesized. This is because there is no nucleus to separate the processes of transcription and translation. When bacterial genes are transcribed, their transcripts can immediately be translated.



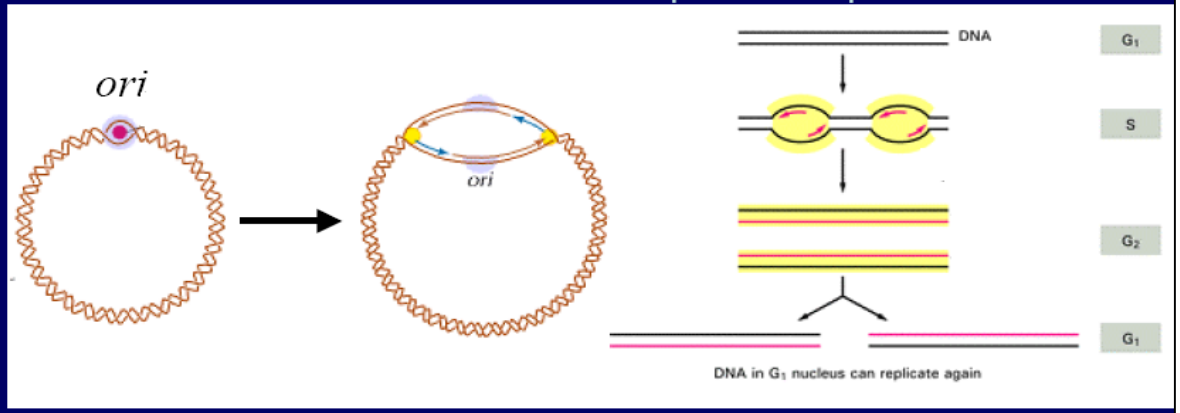
In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs in the cytoplasm. This is because transcription occurs in the nucleus to produce a pre-mRNA molecule. The pre-mRNA (immediate product of the transcription of one strand) is typically processed to produce the mature mRNA (only exons). The mRNA transcript is subsequently exported through nuclear pores (pores in the nuclear envelope) to the cytoplasm for translation.

Pearson Education;..

DNA replication
Bacteria vs. Eukaryotes

Replication -

bacteria	eukaryotes
Single ori Replication biparallel	Multiple oris, 10-330kb apart Oocytes 4-7kb between oris Replication 'in parallel'

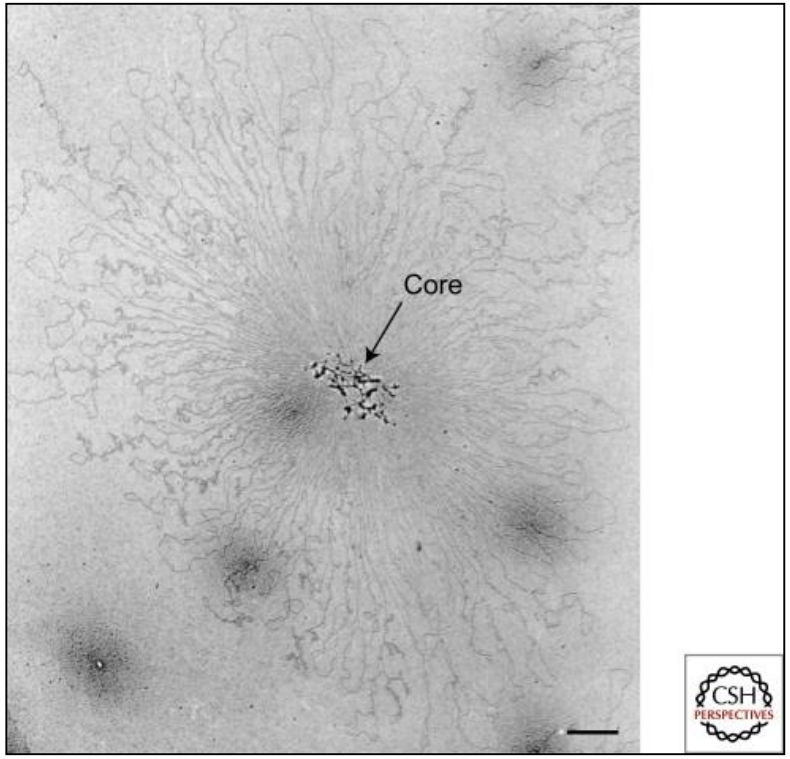


E. coli bacterium:	*Fruit fly embryonic nuclei:*
Replication fork ~ 1000nt s^{-1}	Replication fork ~ 50-100nt s^{-1}
Okazaki fragments ~ 1000-2000nt	Okazaki fragments ~ 200nt
Replication time ~ 1hr	Replication time ~ 3 minutes
Division during exp. Growth ~ 20 min	Nuclear division time ~ 9 min

The bacterial chromosome

Bacterial chromosomes are generally ~1000 times longer than the cells in which they reside

- Electron micrograph of isolated membrane-free chromosomes from *E. coli*.
- The central core, from which several tens of loops(domains of supercoiling) radiate out from this core, is sensitive to RNAse.
- Bar=1 μm.



Reprinted from Kavenoff and Bowen,1976

Toro and Shapiro,2010

Chromosome replication

Sites of DNA replication

Prokaryotes vs. Eukaryotes

1. Most bacterial chromosomes are circular with one replication origin.
2. Eukaryotic chromosomes each contain one linear DNA molecule and multiple origins of replication.

- Replication of a bacterial chromosome normally starts at a fixed point (the origin of replication, *oriC*) and proceeds in both directions to a termination point (*ter*) that is approximately opposite to the origin, producing two double-stranded molecules.

Dale and Park,2004

Chromosome replication

DNA replication

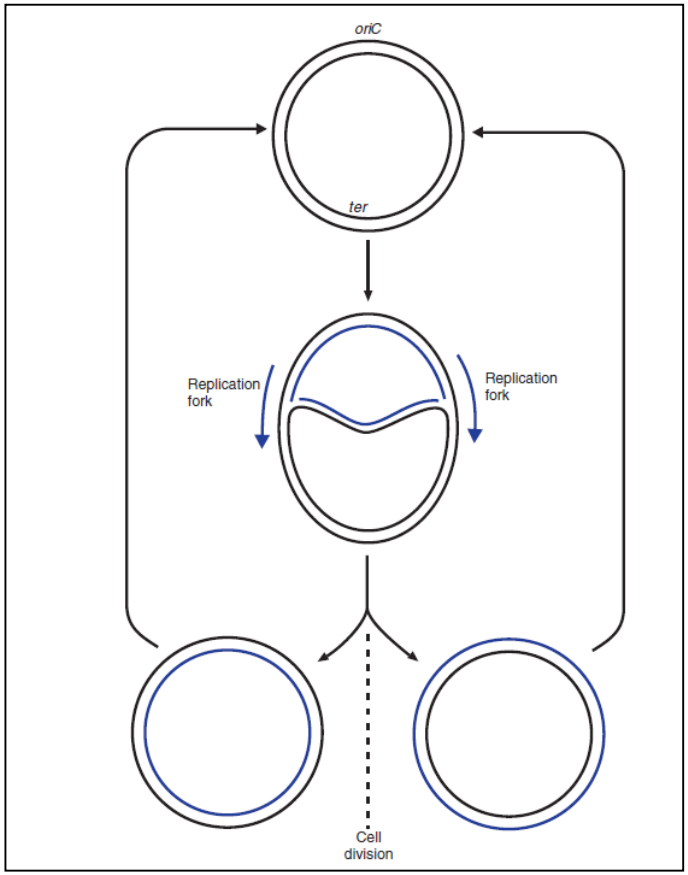
Bidirectional replication

- Cellular DNA replicates before cell division.
- Each daughter cell receives identical chromosome to parent cell.
- Replicating DNA follows a process called semiconservative, which means that each "copy" is half new and half original.
- DNA (Chromosome) replicated is bidirectional from the 3′ end.
- Many enzymes involve in DNA replication.

Note: Some plasmids have a bidirectional mode of replication and others replicate unidirectionally.

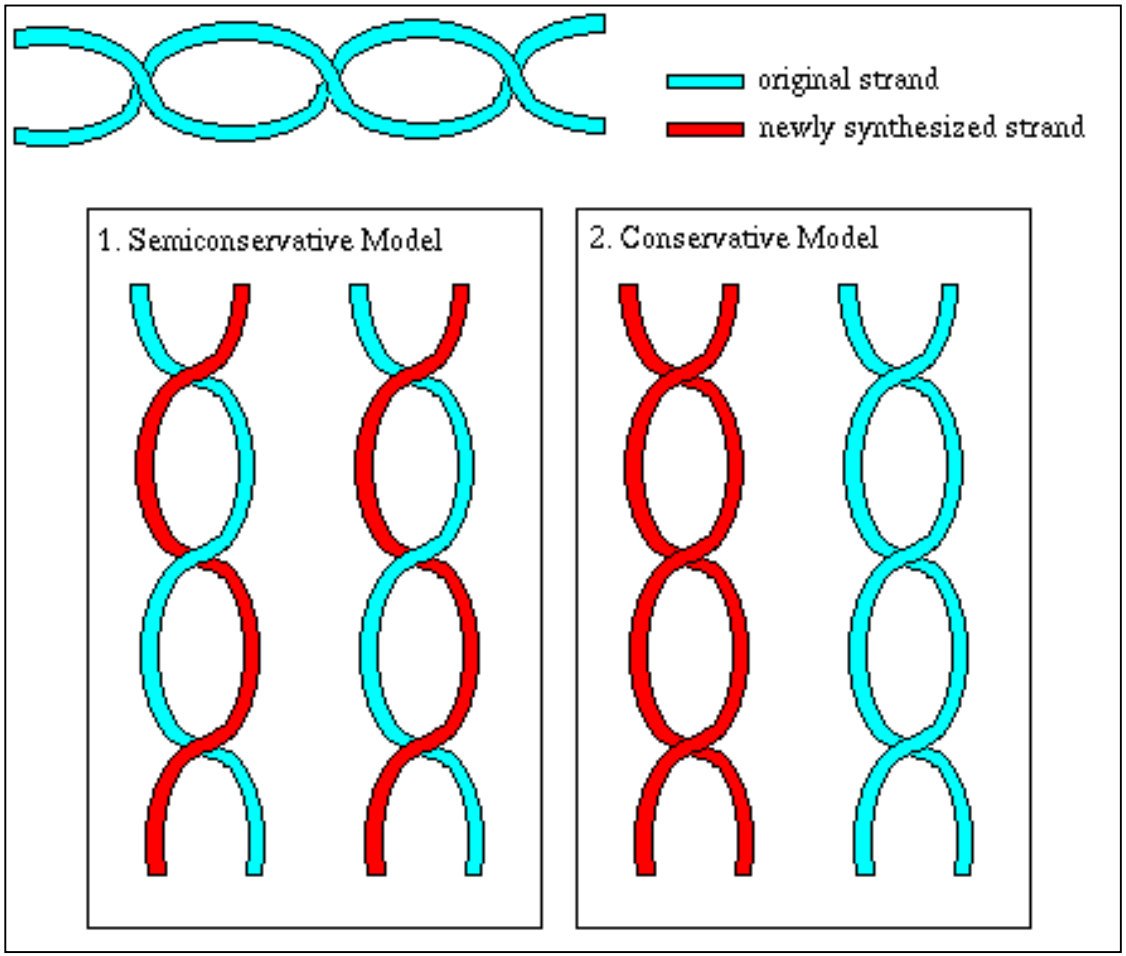
Chromosome replication
Bidirectional replication

- Bidirectional replication starts at *oriC* and continues to the termination site *ter*, producing two double-stranded molecules.



Dale and Park,2010

Conservative/Semi conservative replication

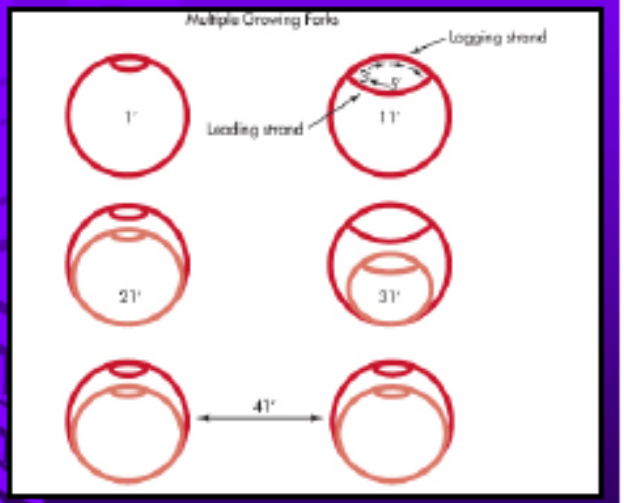


DNA replication

Semiconservative model

Bacterial DNA Replication

- Begins at a certain point (OriC)
- Semi-conservative replication
- Bi-directional
- Enzymes involved
  - Helicase
  - Primase
  - DNA polymerase
  - Ligase
  - Topoisomerase/gyrase

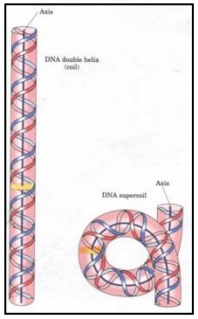


DNA replication
Relaxed DNA

- The double stranded DNA when is unzipped it is "supercoiled" (full of tension).
- This tension must be relived prior any gene expressions and that is done by another enzyme called DNA gyrase which cuts the individual strands, allowing them to unwind to a "relaxed state" (without the tension!).

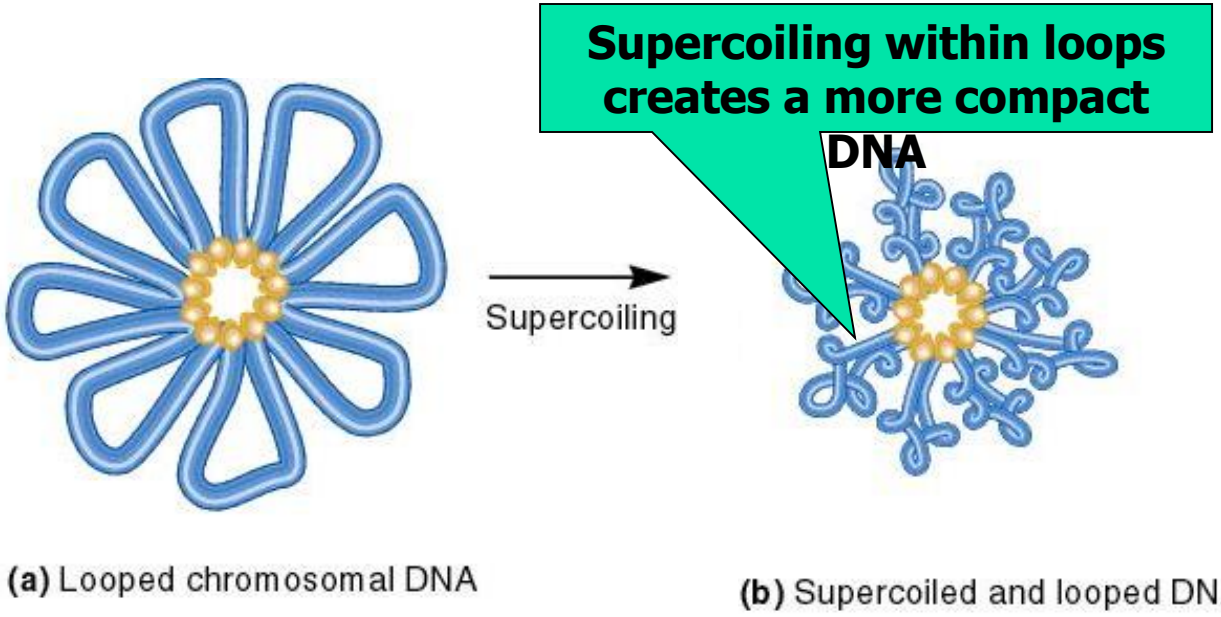
Bacterial chromosome

DNA supercoiling



Supercoiling is the twisting of the DNA axis upon itself.

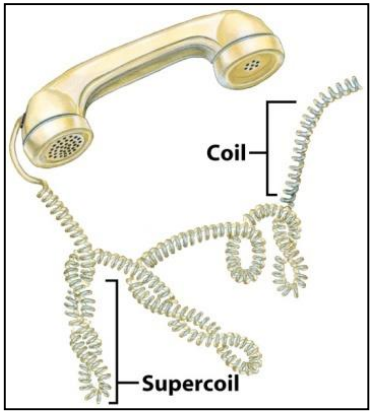
- DNA supercoiling is a second important way to compact the bacterial chromosome. These domains supercoil independently.
- Supercoiling and other interactions causes further compaction, such that it fills an area of about 1 μm.



DNA supercoil

What is supercoiling?

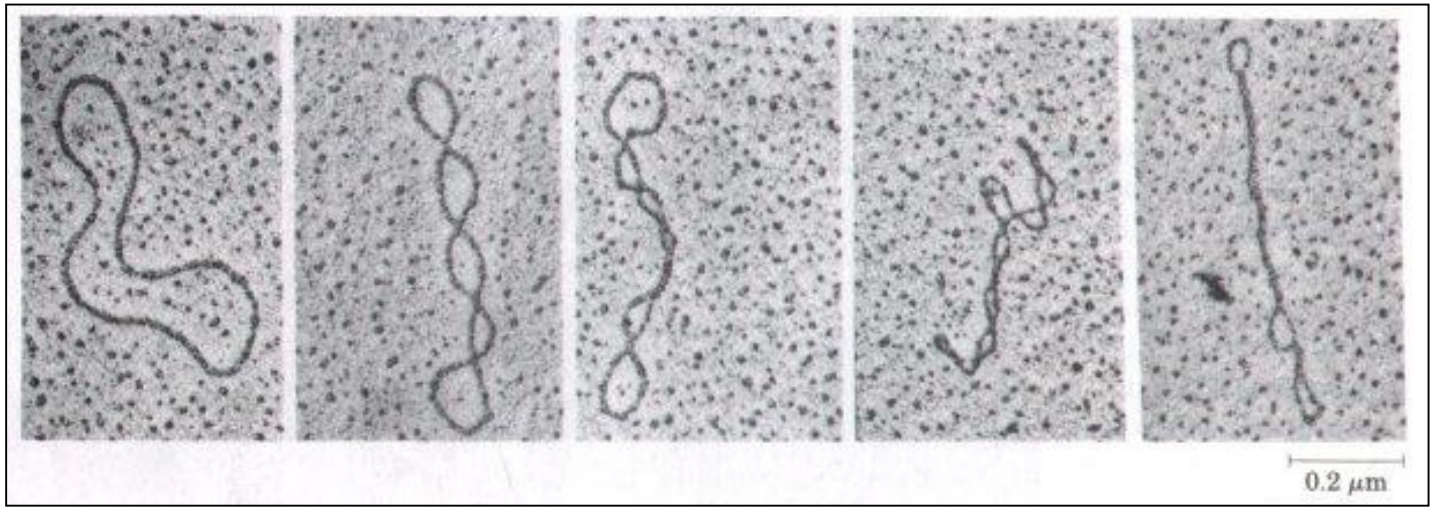
- Supercoiling induced by separating the strands of a helical structure.
- Twist two linear strands of rubber band into a right-handed double helix as shown. Fix the left end by having a friend hold onto it.
- If the two strands are pulled apart at the right end, the resulting strain will produce supercoiling as shown.



DNA supercoil

What is supercoiling?

- Electron micrographs of relaxed and supercoiled plasmid DNAs.
- The molecule at the left is relaxed, and the degree of supercoiling increases from left to right.



DNA supercoil

What is supercoiling?

- **The helix of normal DNA is right-handed.**
- A right-handed coil is assigned a negative number (negative supercoiling). and
- A left-handed coil is assigned a positive number (positive supercoiling).

1. An excess of duplex turns would give rise to positive supercoiling(overwinding).
2. The loosely coiling is referred to as negative supercoiling(underwinding or unwinding).

Negative supercoiling
DNA replication and transcription

- Negative supercoiling makes it easier to denature base pairs.
- Negative supercoils favour:
- Local unwinding (underwound) of the DNA, allowing processes such as:

1. DNA replication;
2. Transcription;
3. Recombination.

DNA supercoils
Classes of topoisomerases

- Most topoisomerases simply relax (or remove) negative or positive supercoils.
- DNA topoisomerases (topos)are divided into two classes:

1. Type I enzymes (topoisomerases I, III and V) break single-strand DNA. i.e. removes supercoils from one strand and allowing the double helix to rotate.
2. Type II enzymes (topoisomerases II, IV and VI) break both strands of the DNA. i.e. removes supercoils two at a time.

Bacterial topoisomerases

Classes of topoisomerases

Enzyme	Type	Structure	Genes	Mass (kD)	Viable Mutants?*	DNA binding	$MgCl_2$?	ATP?
Topo I	Ia	monomer	topA	97	yes	5'	yes	no
Gyrase	II	heteromdimer (AB)(AB)	gyrA gyrB	97 90	yes	5'	yes	yes
Topo III	Ia	monomer	topB	74	yes	5'	yes	no
Topo IV	II	heteromdimer (AB)(AB)	parC parE	75 70	no	5'	yes	yes

DNA supercoil

Functions of topoisomerase I and II

- The control of supercoiling in bacteria is accomplished by two main enzymes:

1. **DNA topoisomerase I**
    - Relaxes negative supercoils
2. **DNA gyrase (also termed DNA topoisomerase II)**
    - Introduces negative supercoils using energy from ATP.
    - It can also relax positive supercoils when they occur.

- The competing action of these two enzymes governs the overall supercoiling of bacterial DNA.

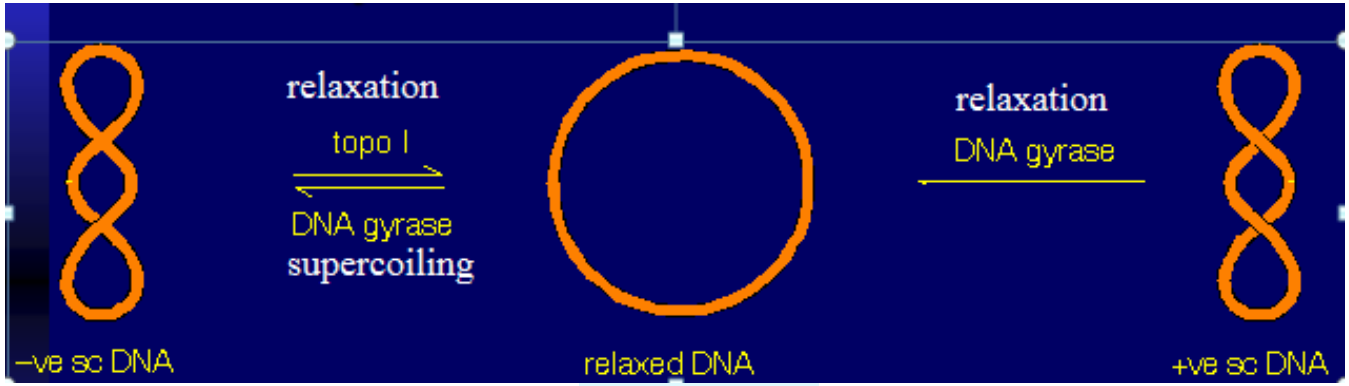
DNA supercoil
Functions of topoisomerase I and II

- Topoisomerase I:

1. Remove only negative supercoils(relax only negative supercoils).

- Topoisomerase II (two functional enzyme):

1. Introduces negative supercoils, and
2. Remove(relax) the positive supercoils during DNA replication.



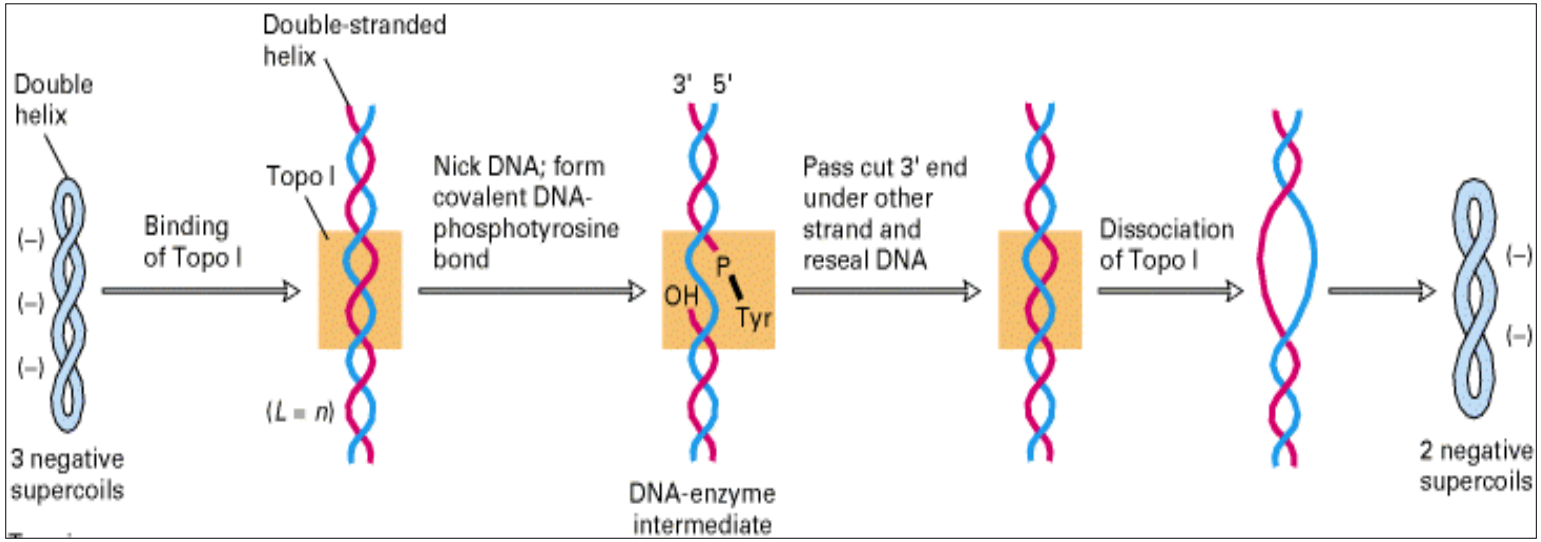
A left-handed coil

A right-handed coil

DNA supercoils

Mechanism of topoisomerase I

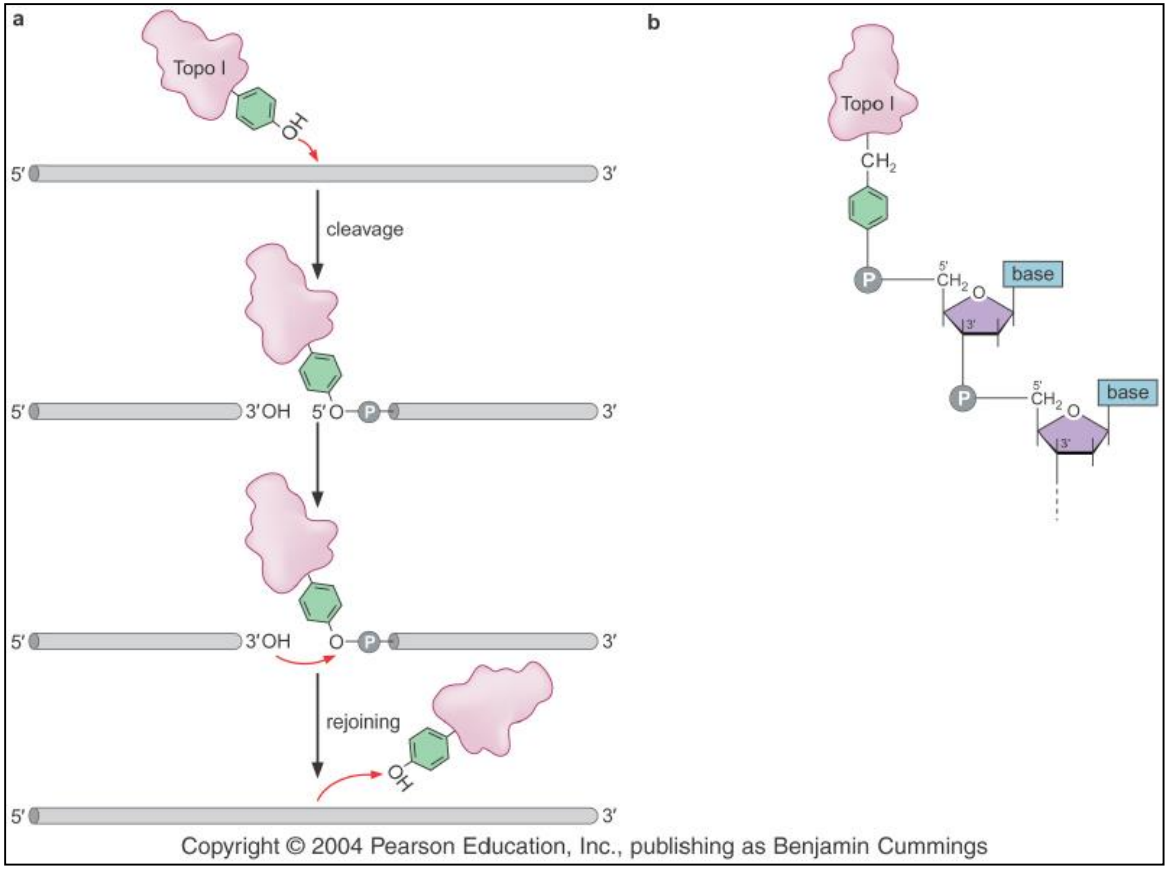
- They cut one strand of the double helix, pass the other strand through, then rejoin the cut ends.



DNA supercoils

Mechanism of topoisomerase I

Break single-strand DNA



Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings

All topoisomerases cleave DNA using a covalent Tyrosine-DNA intermediate. Because the relaxation (removal) of DNA supercoils by Topo I is energetically favorable, the reaction proceeds without an energy requirement.

DNA replication and transcription
DNA topoisomerases and helicases

- The supercoiled DNA must be uncoiled and relaxed in order for:

1. DNA polymerase to bind for DNA replication, and
2. RNA polymerase to bind for transcription of the DNA.

- During transcription, RNA polymerase makes a copy of a gene from the DNA to mRNA as needed.
- This process is similar in eukaryotes and prokaryotes.

DNA replication

DNA topoisomerases and helicases

- Antiparallel DNA strands have 2 different modes of replication:

1. Continuous replication (leading strand);
2. Discontinuous replication in fragments (lagging strand).

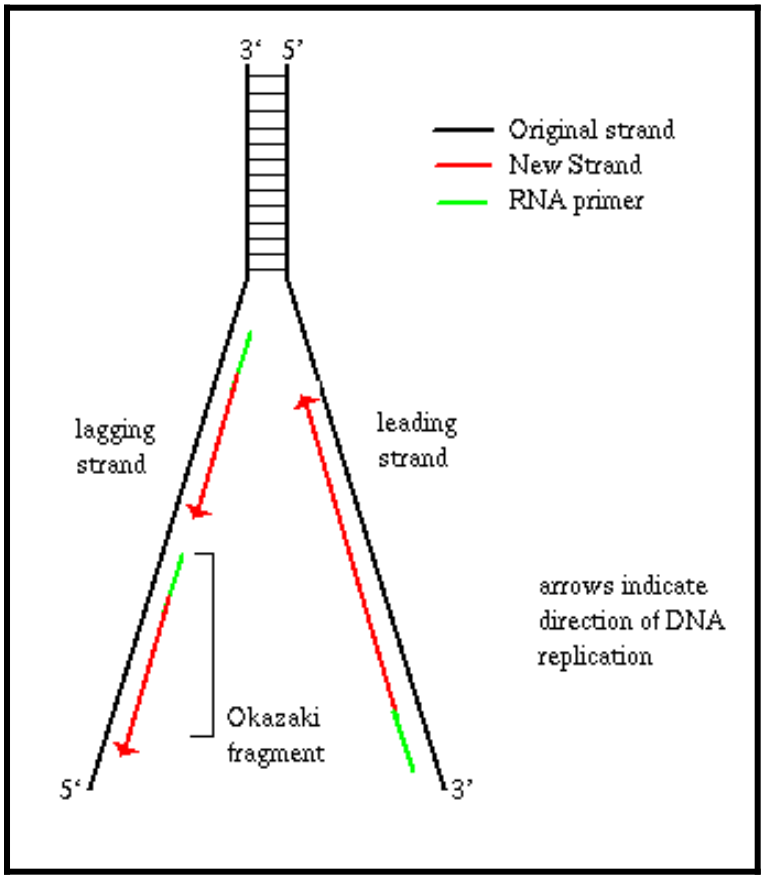
- Replication fork (unzipping area)

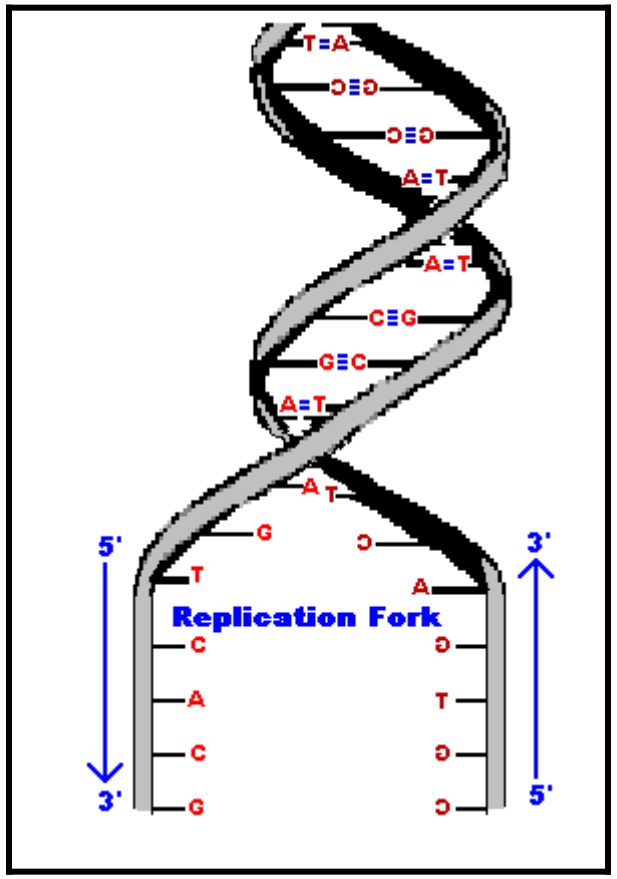
1. Leading strand (template 3' --> 5') replication continuously;
2. Lagging strand (template 5' --> 3') replication in fragments.

- DNA polymerase use a strand of DNA as a template upon which to build the new (polymerized) strand.

DNA replication

DNA topoisomerases and helicases





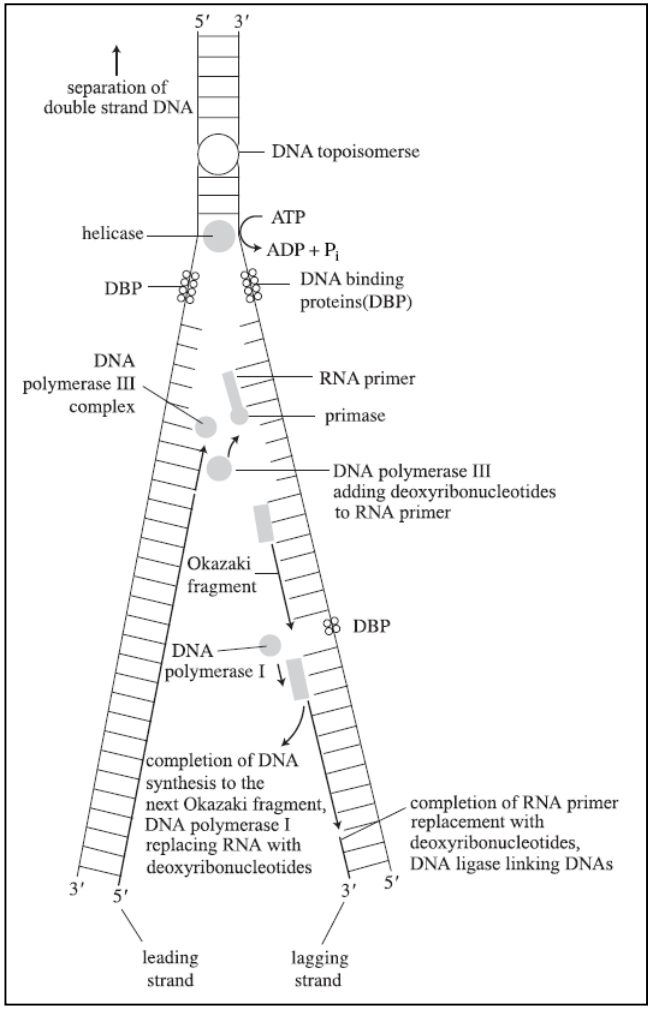
As the two DNA strands separate (unzip) and the bases are exposed, the enzyme DNA polymerase moves into position at the point where synthesis will begin.

DNA replication

DNA topoisomerases and helicases

1. Topoisomerases relaxes the supercoiled DNA.
2. Helicases separates the double strand to single strand.

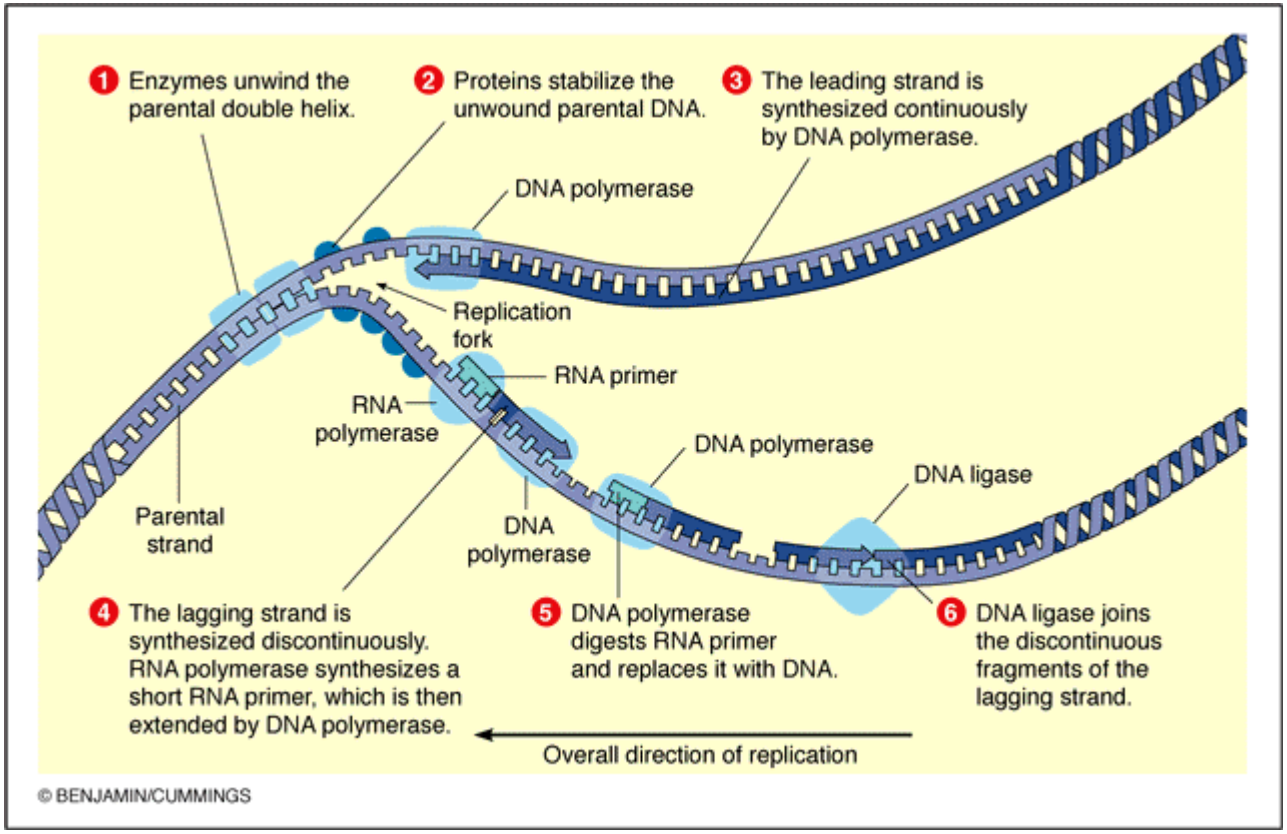
- If the supercoiling is not relieved, it will physically prevent the movement of helicase.
- DNA topoisomerase converts supercoiled DNA to a relaxed form before the helicase separates the double strand to single strands, thereby breaking the hydrogen bonds to start replication.



DNA replication

DNA topoisomerases and helicases

DNA and RNA polymerases



RNA polymerase synthesizes a short RNA primer (a short strand of RNA) that initiates DNA replication. DNA ligases are essential for the joining of Okazaki fragments during replication. Reiji Okazaki, the Japanese molecular biologist who discovered the role of Okazaki fragments.

DNA replication enzymes

- DNA replication enzymes(proteins) include:
- Topoisomease: This enzyme unzip (unwinds) most of the heavy unravelling of DNA from its coiled form.
- Helicase: A protein that unwinds (unzip) double helix DNA between base pairs, at replication forks, causes the strands to separate.
- Primase (RNA polymerase): This enzyme lays down the primers that are necessary for DNA polymerase activity. The primers are composed of RNA. Starts replication with short RNA primer.
- DNA polymerases: Extends polymer from 3' end.
- Ligase: Joins discontinuous fragments with sugar phosphate bonds.
- Telomerase: Makes sure the chromosome ends are complete. Bacteria do not need for telomerase because its DNA is circular.

DNA replication enzymes

Telomeres are specialized structures comprised of DNA and protein which cap the ends of eukaryotic chromosomes

- Telomerase is the enzyme that regenerates telomeres.
- Telomeres form the end pieces of linear DNA strands in chromosomes.
- Telomeres as repetitive nucleotide sequences are located at the end of linear chromosomes of most eukaryotic organisms.
- The sequence of nucleotides in telomere of vertebrates is TTAGGG.
- Telomeres protect the end of the chromosome:

1. from deterioration, or
2. from fusion with neighboring chromosomes.

Wikipedia,2015

DNA replication enzymes
Problems with linear chromosomes

- Two problems arise with the presence and use of linear chromosomes compared to circular chromosomes in prokaryotes.

1. Since free double-stranded DNA ends are very sensitive to degradation by intracellular nucleases, there must be a mechanism to protect the ends.
2. Secondly, the telomeres, which are the ends of the linear DNA molecules, will require a different process for DNA replication.

Wikipedia,2015

DNA replication enzymes

Telomeres are specialized structures comprised of DNA and protein which cap the ends of eukaryotic chromosomes

1. Circular nature of bacterial chromosome provides a place for machinery to bind ALWAYS => no shortening=> no need for telomere.
2. But linear chromosomes have an end problem.

- After DNA replication, each newly synthesized DNA strand is shorter at its 5' end than at the parental DNA strand's 5' end.
- Growing shorter chromosomes during replications cause loss of genetic information.

Capping: The cap structure is added to the 5' of the newly transcribed mRNA precursor in the nucleus prior to processing and subsequent transport of the mRNA molecule to the cytoplasm.

DNA replication enzymes

Telomeres in few bacteria

- Not all bacteria have a single circular chromosome.
- Some bacteria have linear chromosomes:
- *Streptomyces*,
- *Agrobacterium*, and
- *Borrelia*
- These bacteria possess telomeres, which are very different from those of the eukaryotic chromosomes in structure and functions.

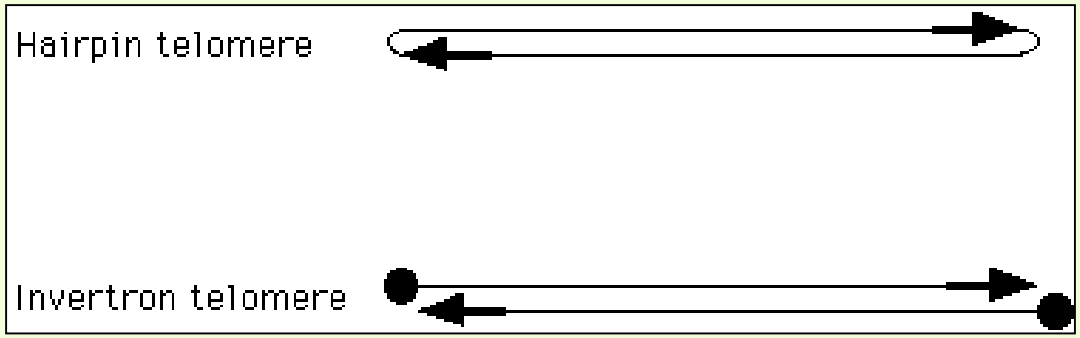
Wikipedia,2015

DNA replication enzymes

Telomeres are repetitive, noncoding sequences that cap the ends of linear chromosomes

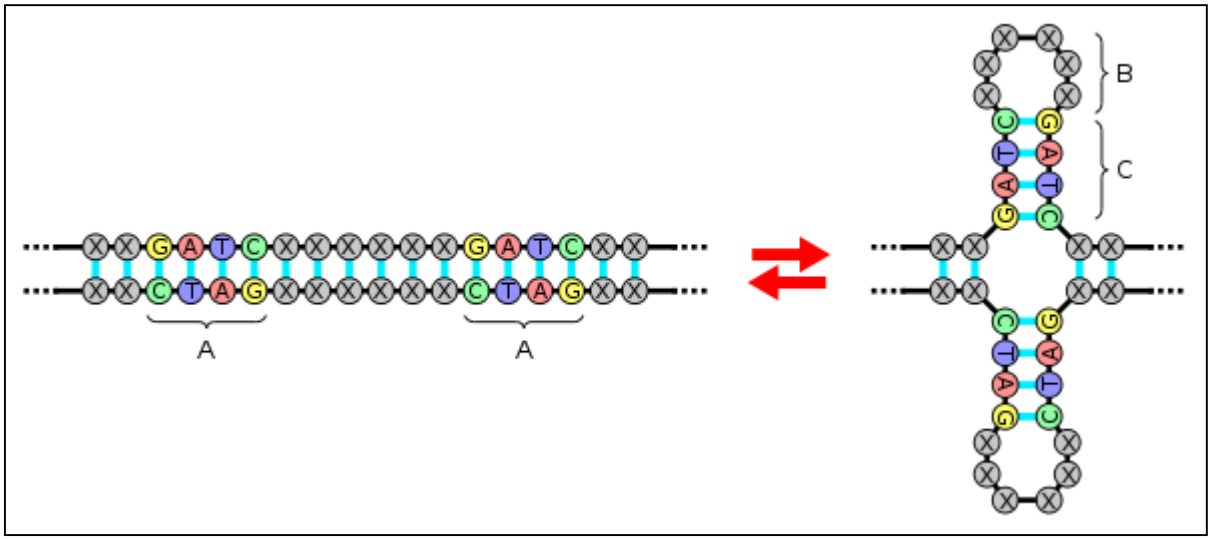
- These problems are solved by features of the telomeres.
- Two different types of telomeres have been observed in bacteria:

1. The first one is called the palindromic hairpin loops, in which there are no free double-stranded ends available.
2. The other type is invertron telomeres where the chromosomes contain a protein that binds to the 5'-ends.



Microbial Genetics supplement,2002 ;MicrobeWiki,2012

Image of a palindrome in a DNA structure



A = Palindrome, B = Loop, C = Stem

The DNA sequence GAGTC is palindromic because its nucleotide-by-nucleotide complement is CTAG. Reading 5' to 3' forward on one strand matches the sequence reading backward 5' to 3' on the complementary strand.

MicrobeWiki,2012

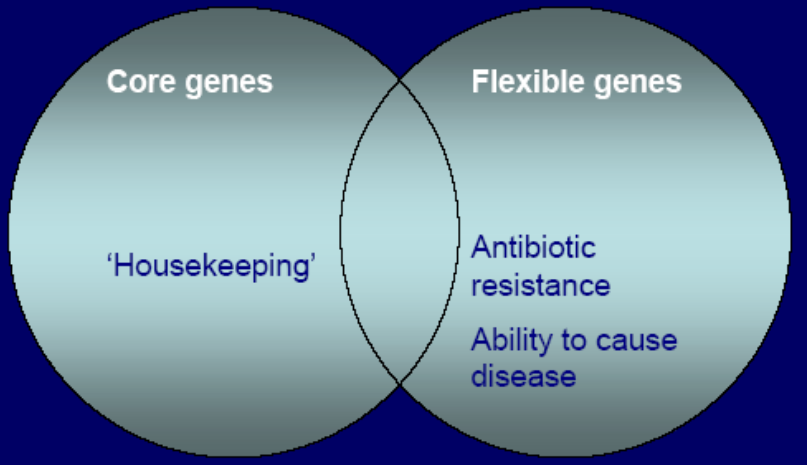
Nutrient fluctuation and mode of replication

When nutrient availability varies drastically, high growth rates selected

Nutrient fluctuation & circular mode of replication have had two main effects on bacteria (& archaea)

1. 'In series' replication selects for loss of genes not in immediate use
2. Continual pressure for fast growth when resources are abundant has resulted in fast gene expression

'Use it or lose it' may have driven evolution of HGT



DNA is dynamic and has high energy
But not stiff or static as first thought

- DNA appears a perfect spring that can be stretched and then spring back to its original conformation.
- DNA is also accompanied by a cloud of counterions (charged particles that neutralize the genetic material's very negative charge) and, of course, the protein macromolecules that affect DNA activity.
- It is dynamic with high energy existing naturally in a slightly underwound state and its status changes in waves generated by normal cell functions such as:
- DNA replication, transcription, repair and recombination.

Baylor College of Medicine,2009

DNA is dynamic and has high energy
But not stiff or static as first thought

- DNA appears a perfect spring that can be stretched and then spring back to its original conformation.
- It is dynamic with high energy existing naturally in a slightly underwound (supercoiled) state and its status changes in waves generated by normal cell functions such as:

1. DNA replication,
2. Transcription,
3. Repair and recombination.

Baylor College of Medicine,2009

DNA is dynamic and has high energy
But not stiff or static as first thought

- An important property of DNA is that it can replicate, or make copies of itself.
- Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases.
- This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell.

Genetics Home Reference,2015

DNA is dynamic and has high energy
But not stiff or static as first thought
Replication

- An important property of DNA is that it can replicate, or make copies of itself.
- Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases.
- This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell.

Genetics Home Reference,2015;..

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

1. Each cell expresses, or turns on, only a fraction of its genes.
2. The rest of the genes are repressed, or turned off.

- Gene regulation(transcription + translation) is an important part of normal development.
- Genes are turned on and off in different patterns during development to make a brain cell act different from a liver cell or a muscle cell.

All these cells contain the same DNA, but their genes are expressed differently (turned "on" or "off"), which creates the different cell types.

Genetics Home Reference,2015;..

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- Humans are complex organisms made up of trillions of cells, each with their own structure and function.
- Scientists have come a long way in estimating the number of cells in the average human body.
- Most recent estimates put the number of cells at around 30 trillion. Written out, that's 30,000,000,000,000!
- Some belief, we all are made up of trillions of cells (37 trillion cells) with nearly identical 20000 genes.

Wydeven,2011; jacusers.johnabbott.qc.ca,2014; Cranage,2020

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- There are over 200 different cell types in the human body.
- Each type of cells is specialized to carry out a particular function.
- The cells in our heart, for instance, work very differently from the cells in our eyes.
- The same is true for our skin and our liver cells.

Human cells have little colour and so resolution and contrast would be low without the use of stains. Methylene blue stains the cell and its organelles differentially so that the cell is more clearly seen and some organelles stand out more than others, e.g. the nucleus.

Wydeven,2011; jacusers.johnabbott.qc.ca,2014; Cranage,2020

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- These cells are different because they use the same set of genes differently.
- So even though each of our cells has the same 20,000 or so genes, each cell can select:

1. which ones it wants to "turn on" and
2. which ones it wants to keep "turned off".

- Genes tell each cell in our body what to do and when to do it.
- For example, genes have the instructions for making liver cells, carrying oxygen in our blood, or helping us break down sugar.
- How to they do all of this? By making proteins.

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- But not every protein can do every job.
- Proteins are specialized to do certain things like carry oxygen in the blood or recognize bacteria in our bodies.
- And because of this, some proteins are only needed during certain times or in certain areas of the body.
- For example, proteins that build bone don't need to be in our hair!
- So how do the proteins know which places in the body to be?
- The answer is that genes are able to be turned on and off in different places and at different times.

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- Just like a chef reads a recipe to make a dish, cells read a gene to make a protein.
- Each gene makes a specific protein that does a specific job.

1. Cells need to make thousands of different proteins in the right amounts,
2. at the right time, and
3. in the right place in order to work properly.

- This is where promoters come in. Promoters are what turns genes on and off.

Wydeven,2011; Rodriguez,2013

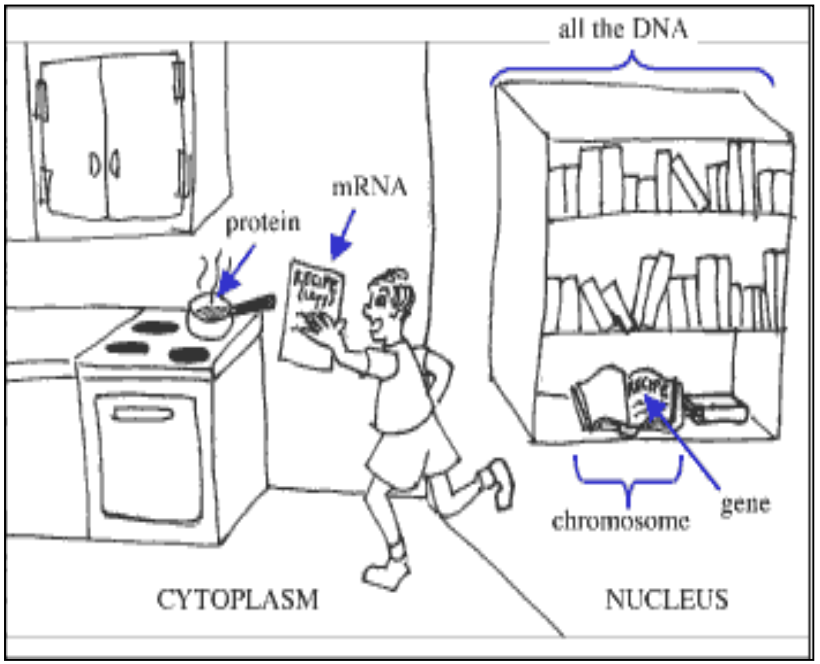
DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- Gene regulation can occur at any point during gene expression, but most commonly occurs at the level of transcription (when the information in a gene's DNA is transferred to mRNA).
- Signals from the environment or from other cells activate proteins called transcription factors.
- These proteins bind to regulatory regions of a gene and increase or decrease the level of transcription.
- By controlling the level of transcription, this process can determine the amount of protein product that is made by a gene at any given time.

Genetics Home Reference,2015;..

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- It might help to think about our set of genes as a cookbook and each of our cell types as a different cook.
- So the breakfast cook focuses only on breakfast recipes and the lunch chef only on lunch ones.



Wydeven,2011

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

1. Most genes contain the information needed to make functional molecules called proteins.
2. A few genes produce other molecules that help the cell assemble proteins. e.g. rRNA (in synthesis of ribosomes an proteins).

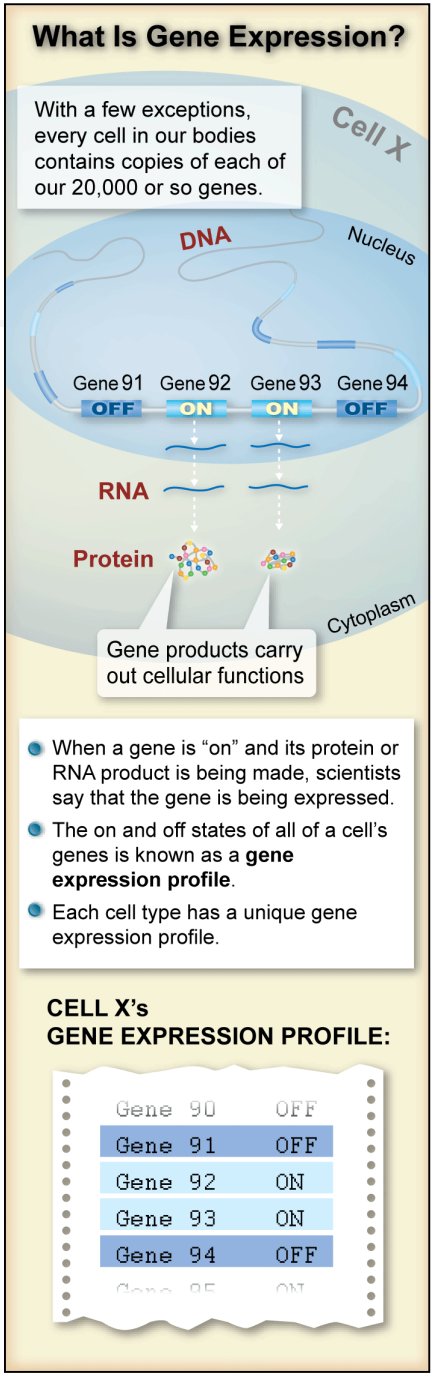
- The journey from gene to protein is complex and tightly controlled within each cell.
- It consists of two major steps:

1. Transcription, and
2. translation.

- Together, transcription and translation are known as gene expression.

Genetics Home Reference,2015;..

DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation



DNA is dynamic and has high energy
But not stiff or static as first thought
Gene regulation

- So when everything is working OK, each cell type only reads the genes it needs. That way it only makes the proteins it needs too.
- For example, early on our cells need to divide and divide so we can grow from one to trillions of cells.
- All of our cells have the growth genes churning our proteins to make the cells keep dividing.
- But when we are done growing, we don't want our cells to keep dividing. These genes stop making proteins and cells stop dividing so much.
- If the cells didn't stop growing, this growth would cause cancer.

Wydeven,2011;..

DNA replication and transcription

Gene regulation

What is the epigenome

- Epigenetics is the study of mechanisms that switch genes on or off.
- It refers to external modifications to DNA that turn genes "on" or "off."
- "epi-" means above in Greek. Epigenetics literally means "above" or "on top of" genetics.
- Epigenetic changes alter the physical structure of DNA.
- These modifications do not change the DNA sequence, but instead, they affect how cells "read" genes.
- Chemical modifications to the DNA or proteins, can alter gene expression.

Genetics Home Reference; Live Science;.

DNA replication and transcription

Gene regulation

Epigenetics

- Not all of the genes are turned on at the same time in every cell, and so they must be regulated.
- Reading and controlling the genome to express genes can come from many different sources such as:

1. DNA methylation- a process by which methyl groups are added to DNA. Methylation modifies the function of the DNA, typically acting to suppress gene transcription.
2. Histone modifications(in eukaryotes), and
3. Transcription factors.

DNA replication and transcription

Epigenetic changes

DNA methylation in bacteria

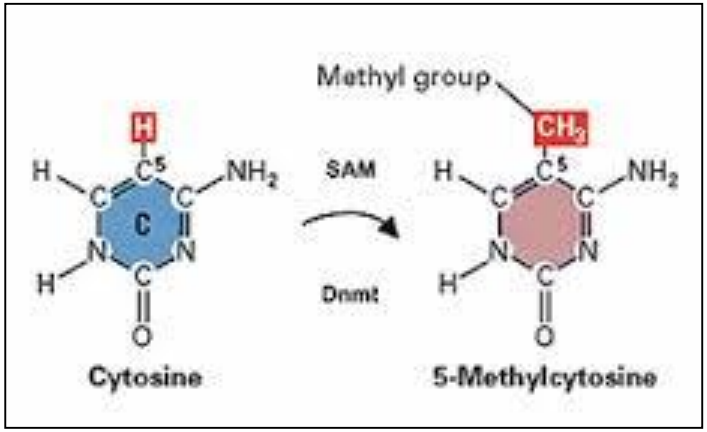
- Little is known concerning how widespread epigenetic control is in the bacterial world and the roles that epigenetic regulatory systems play in bacterial biology, including pathogenesis.
- Epigenetic mechanisms evolved as a form of cellular defense, targeting incoming viral and other foreign DNA sequences for degradation.
- Bacteria also use epigenetic mechanisms to control phase variation.

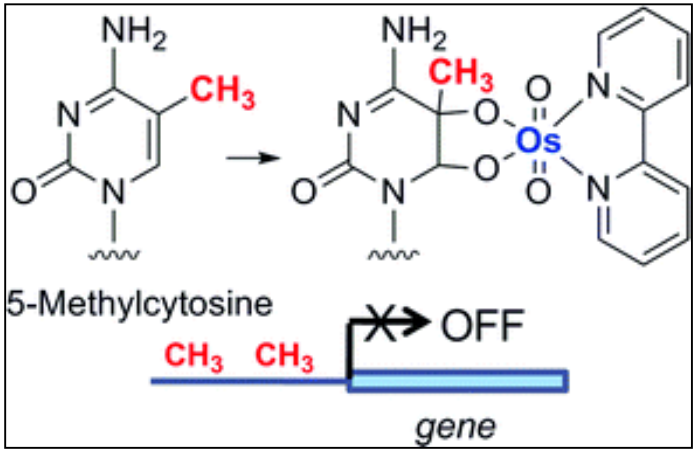
Casadesu´s and Low,2006

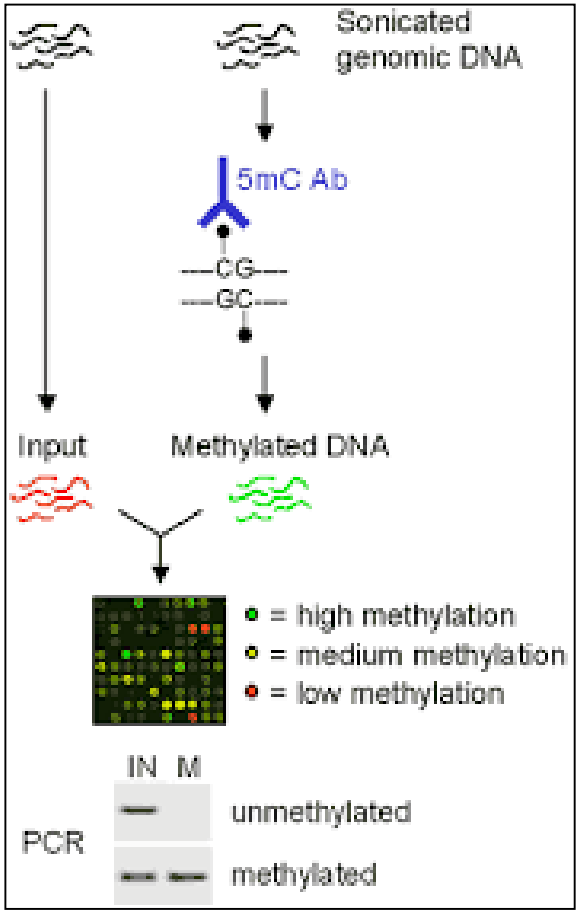
DNA replication and transcription

Gene regulation

DNA methylation in bacteria





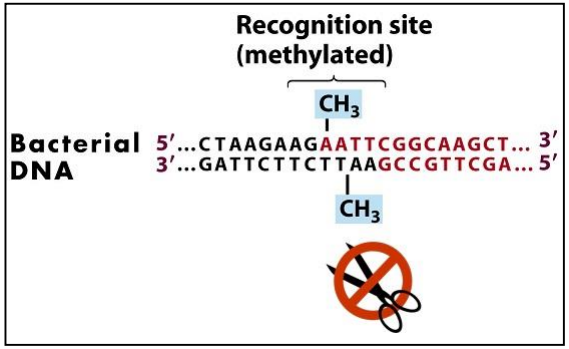


DNA replication and transcription

Epigenetic changes

DNA methylation in bacteria

- A common type of epigenomic modification is called methylation.
- DNA methylation is the addition of a methyl group, or a "chemical cap," to part of the DNA molecule, which prevents certain genes from being expressed.
- Methylation involves attaching small molecules called methyl groups, each consisting of one carbon atom and three hydrogen atoms, to segments of DNA.



DNA replication and transcription

Epigenetic changes

DNA methylation in bacteria

- When methyl groups are added to a particular gene, that gene is turned off or silenced, and no protein is produced from that gene.
- Methylation modifies the function of the DNA, typically acting to suppress gene transcription.
- In the bacterial kingdom, the most prevalent base modifications are in the form of DNA methylations, specifically to adenine and cytosine.
- The methylation of native DNA acts as a sort of primitive immune system, allowing the bacteria to protect themselves from infection by bacteriophage.

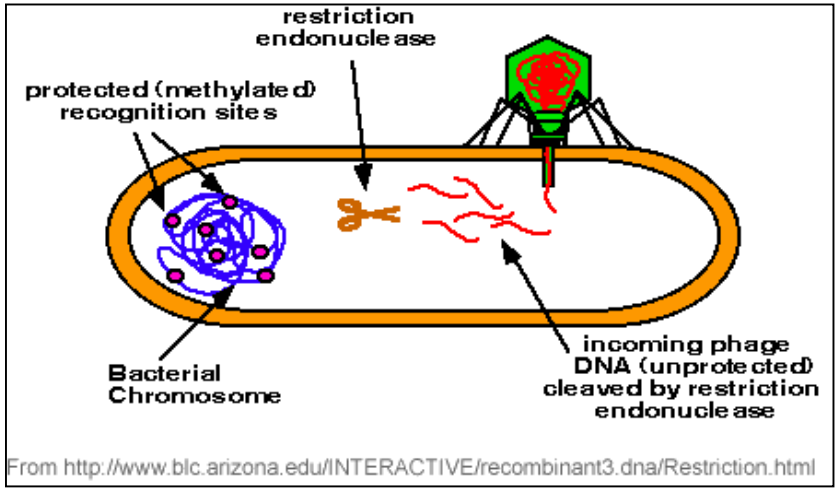
Genetics Home Reference;..

DNA replication and transcription

Epigenetic changes

DNA methylation in bacteria

- When foreign phage DNA enters a bacterium, it is recognized as foreign and these cellular enzymes cut it up, restricting the ability of the phage to infect the bacterium.
- The phage genome is cut at specific sites by the restriction enzyme. e.g. EcoR1.



Genetics Home Reference,...

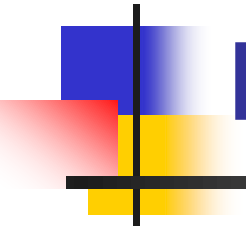
DNA replication and transcription

Epigenetic changes

Demerits

- Because errors in the epigenetic process, such as modifying the wrong gene or failing to add a compound to a gene, can lead to abnormal gene activity or inactivity, they can cause genetic disorders. Conditions including cancers, metabolic disorders, and degenerative disorders have all been found to be related to epigenetic errors.
- Note: Environmental influences, such as a person's diet and exposure to pollutants, can also impact the epigenome (negative epigenetic modifications).

Genetics Home Reference;..

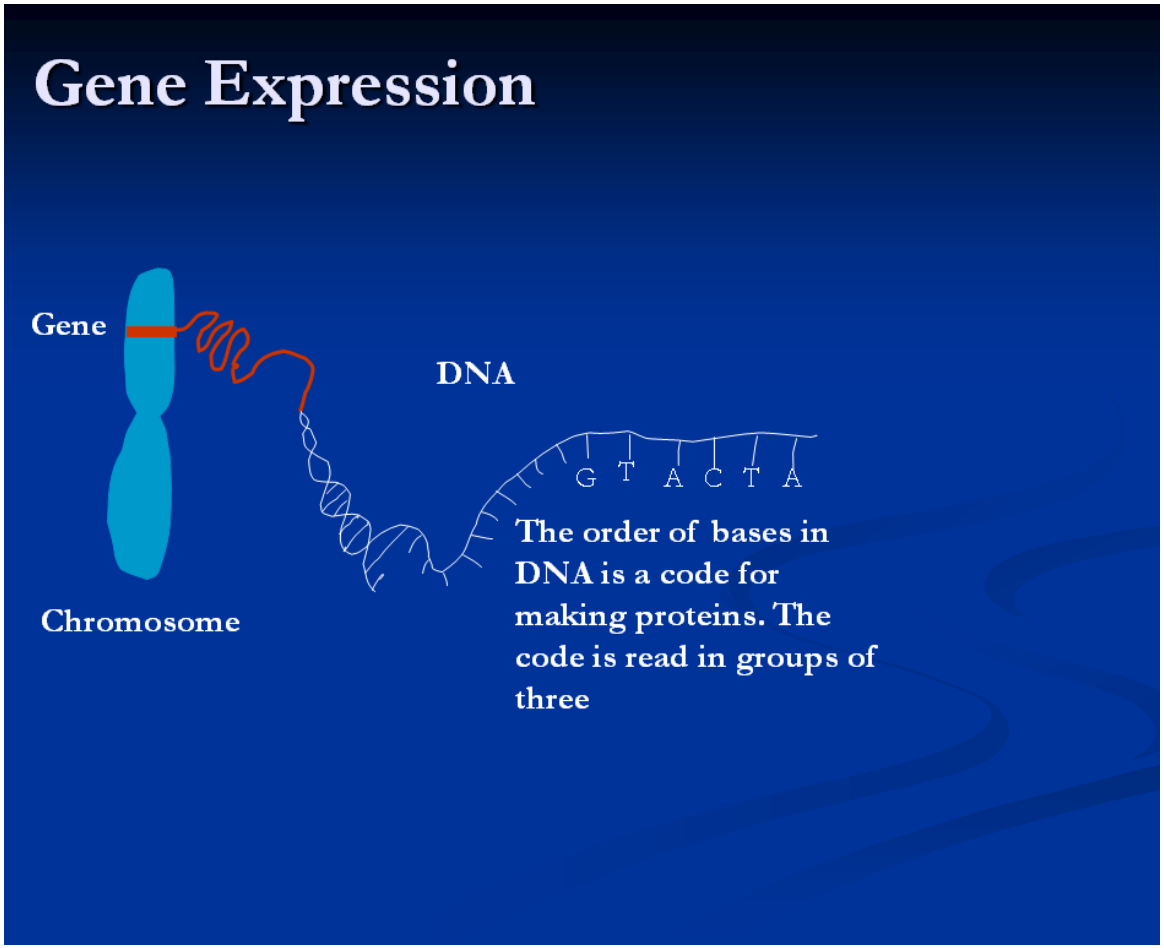


DNA Transcription

RNA synthesis from DNA

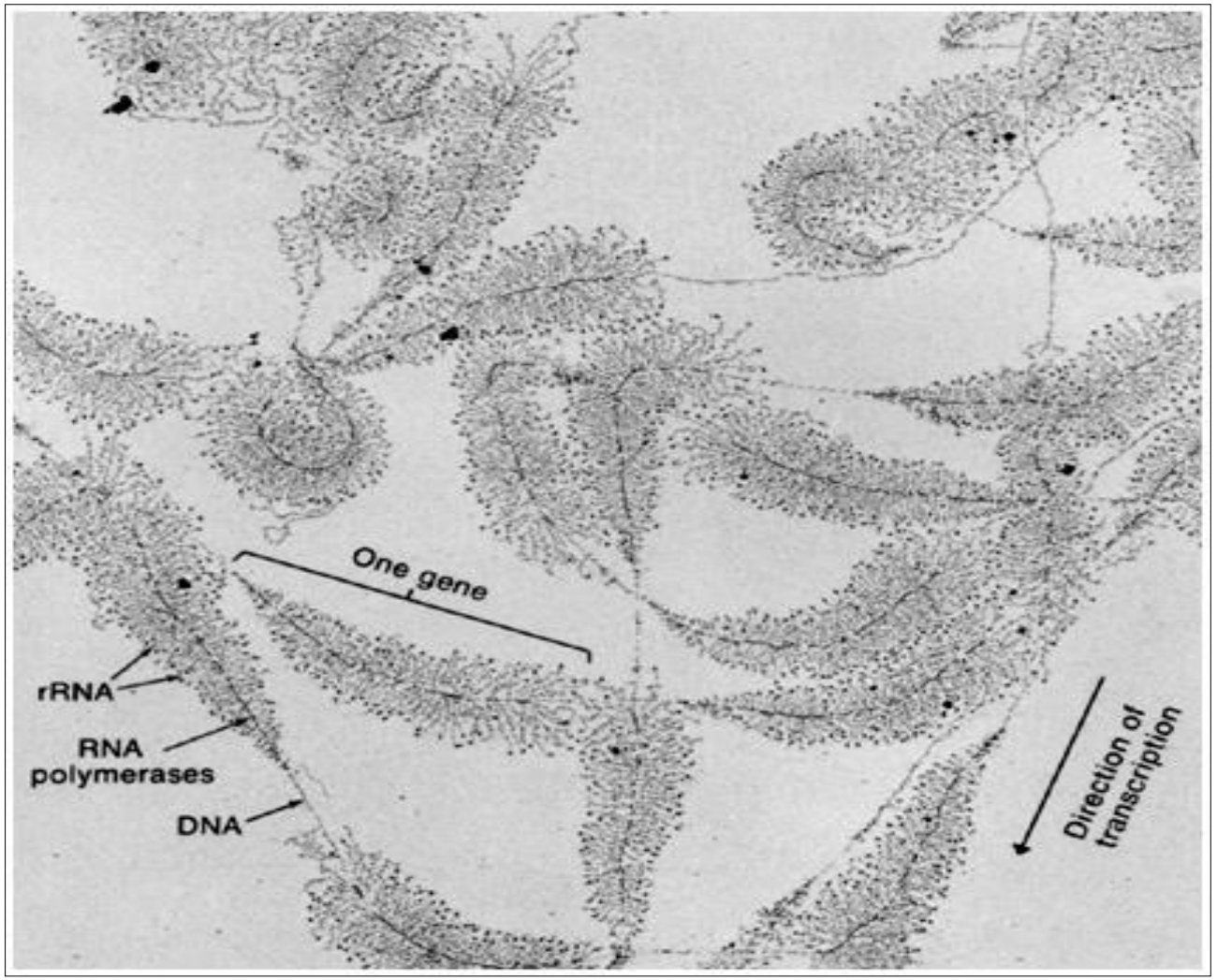
(DNA to RNA information transfer)

Gene expression
For protein synthesis

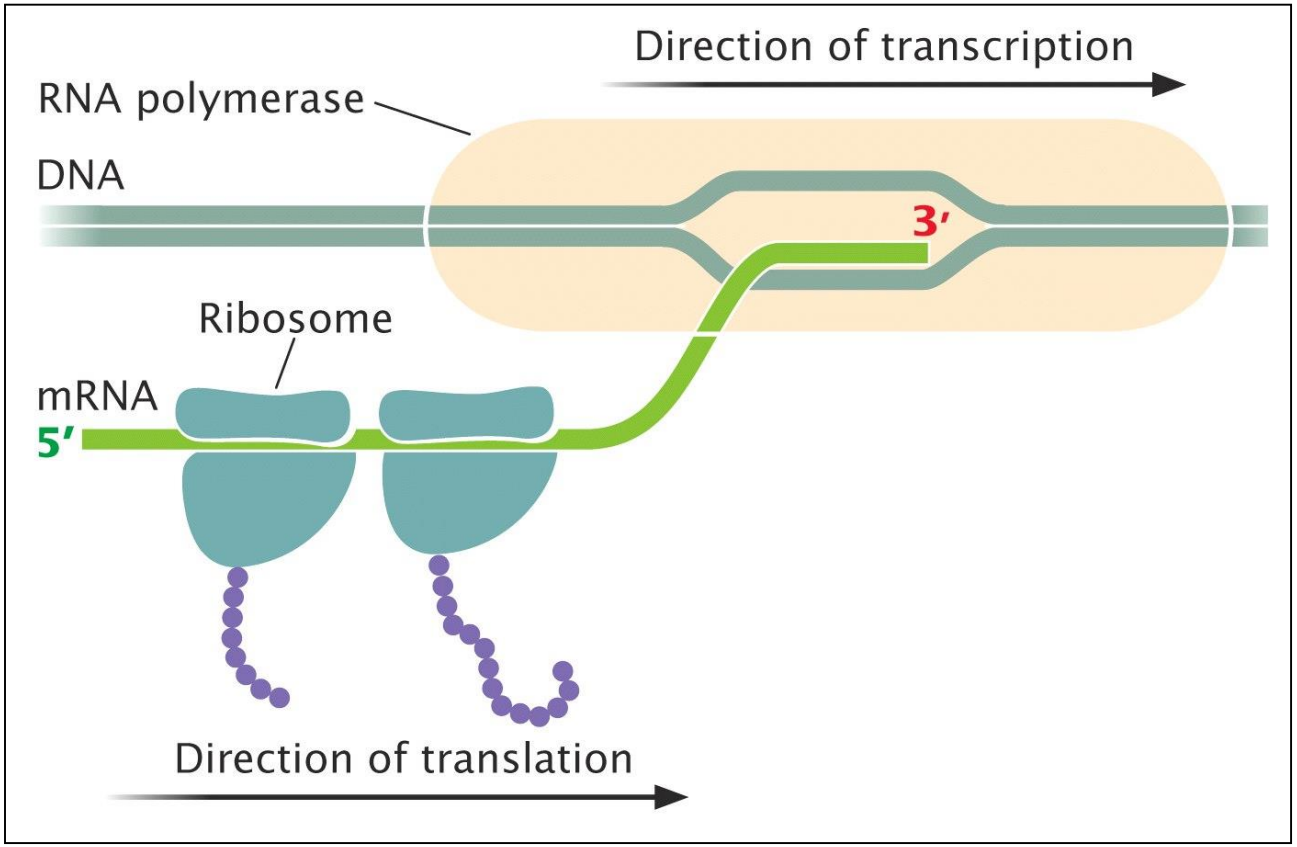


Transcription of ribosomal RNA

rRNA



Simultaneous transcription and translation in prokaryotic cells



Transcription by RNA polymerase proceeds in a series of steps

RNA polymerase and the transcription cycle

1. Initiation
2. Elongation
3. Termination

Prokaryotic vs. Eukaryotic RNA Polymerase

RNA polymerases come in different forms, but share many features

- RNA polymerases performs essentially the same reaction in all cells.
- Bacteria have only a single RNA polymerases.
- This one RNA polymerase synthesizes all classes of RNA i.e. mRNA, rRNA, tRNA.
- While in eukaryotic cells there are three RNA polymerases(I-III).

Prokaryotic		Eukaryotic		
Bacterial	**Archaeal**	**RNAP I**	**RNAP II**	**RNAP III**
Core	**Core**	**(Pol I)**	**(Pol II)**	**(Pol III)**
β'	A′/A″	RPA1	RPB1	RPC1
β	B	RPA2	RPB2	RPC2
α^{I}	D	RPC5	RPB3	RPC5
α^{II}	L	RPC9	RPB11	RPC9
ω	K	RPB6	RPB6	RPB6
	[+6 others]	[+9 others]	[+7 others]	[+11 others]

Note: The subunits in each column are listed in order of decreasing molecular weight.

Source: Data adapted from Ebright R.H. 2000 *J. Mol. Biol.* **304:** 687–698, Fig. 1, p. 688. © 2000 Academic Press.

Copyright © 2004 Pearson Education, Inc., publishing as Benjamin Cummings

The role of ω(omega) has been unclear.

Eukaryotic RNA Polymerase

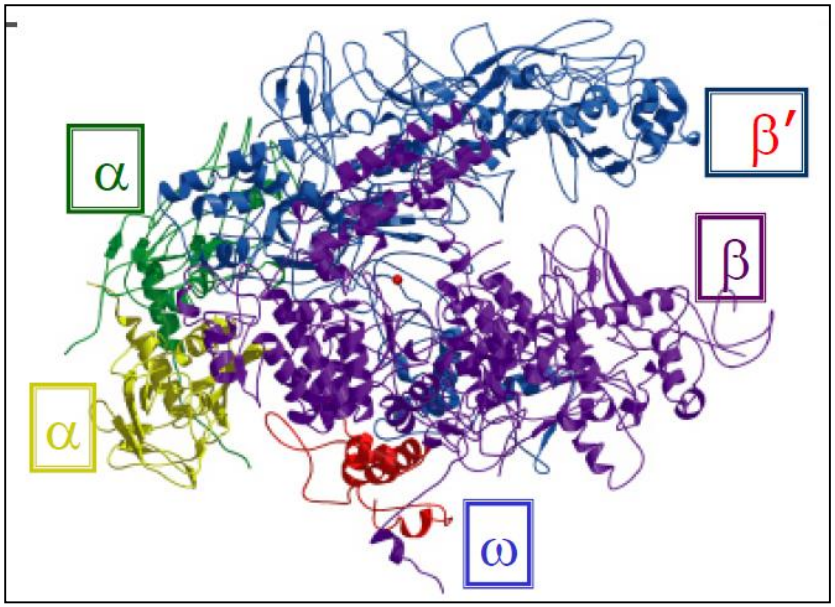
- Three types exist;
- Protein is greater than 500 kd in size;
- Two large subunits; <10 small subunits; largest subunit is homologous to beta', second largest subunit is homologous to the beta subunit;
- Many non-polymerase factors required for binding of the enzyme to DNA.

Type of Polymerase	Product	Location
RNA Polymerase I	rRNA	nucleolus
RNA Polymerase II	hnRNA (heterogeneous nuclear RNA)	nucleoplasm (a fluid found in the nucleus of eukaryotic cells)
RNA Polymerase III	tRNA	nucleoplasm

Prokaryotic RNA polymerase

The enzyme without the sigma subunit is called the 'core' enzyme

- One type exists;
- The core enzyme has five subunits:
- alpha, beta, beta', sigma and omega.
- Only the first three subunits (alpha, beta, beta') are required for polymerase activity and RNA synthesis.
- The sigma factor recognizing the core promoter region.
- The sigma factor is only needed for transcriptional initiation.



RNA polymerase core enzyme is a multimeric protein α, β, β′, ω.
The β′ subunit is involved in DNA binding. The β subunit contains the polymerase active site.
The α subunit acts as scaffold on which the other subunits assemble.
ω may facilitates assembly of RNAP and stabilizes assembled RNAP.

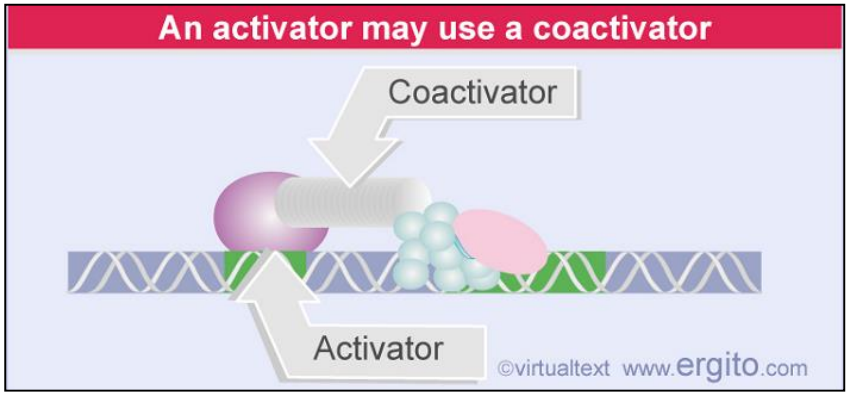
Transcription by RNA polymerase proceeds in a series of steps

Additional proteins playing crucial roles in gene regulation

- Additional proteins also playing crucial roles in gene regulation. e.g.

1. Coactivators: A protein that works with transcription factors to increase the rate of gene transcription.
2. Corepressors: A protein that works with transcription factors to decrease the rate of gene transcription.

- Since they lack DNA-binding domains, these are not classified as transcription factors.



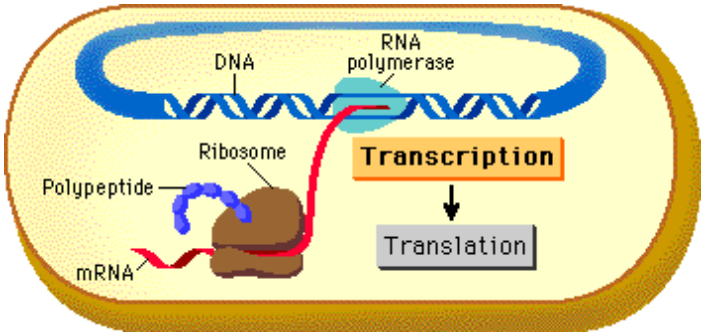
DNA Transcription

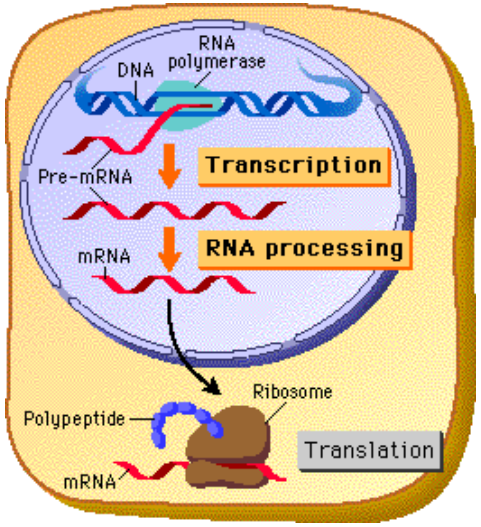
- Transcription is a fundamental cellular process.
- RNA polymerases "transcribe" the genetic information on DNA into RNA strands.
- All cells have RNA polymerases (RNAP).

Prokaryotic vs. Eukaryotic transcription

- **Prokaryotic transcription:**
- Occurs in the cytoplasm alongside translation.
- **Eukaryotic transcription:**
- Is primarily localized to the nucleus, where it is separated from the cytoplasm by the nuclear membrane.
- In prokaryotes, mRNA is not modified.
- Eukaryotic mRNA is modified.

Prokaryotic vs. Eukaryotic transcription





In a prokaryotic cell, transcription and translation are coupled; that is, translation begins while the mRNA is still being synthesized. This is because there is no nucleus to separate the processes of transcription and translation, when bacterial genes are transcribed, their transcripts can immediately be translated.

In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs in the cytoplasm. This is because transcription occurs in the nucleus to produce a pre-mRNA molecule.
The pre-mRNA is typically processed to produce the mature mRNA, which exits the nucleus and is translated in the cytoplasm.

The process of transcription
RNA(mRNA) synthesis

- A transcription unit is the distance between sites of initiation and termination by RNA polymerase; may include more than one gene.
- Transcription of any gene usually involves three distinct stages:
- There are three steps in transcription:

1. Initiation,
2. Elongation, and
3. Termination.

The process of transcription

RNA synthesis

- However, prior to the first steps, there will have to be localized unwinding of the chromosome to allow the RNA polymerase to access the DNA.
- Either strand of the DNA may be the template strand for RNA synthesis for a given gene.
- For any given gene,

1. the template strand is also referred to as the antisense (or non-coding) strand and
2. the non-template strand as the sense (or coding) strand.

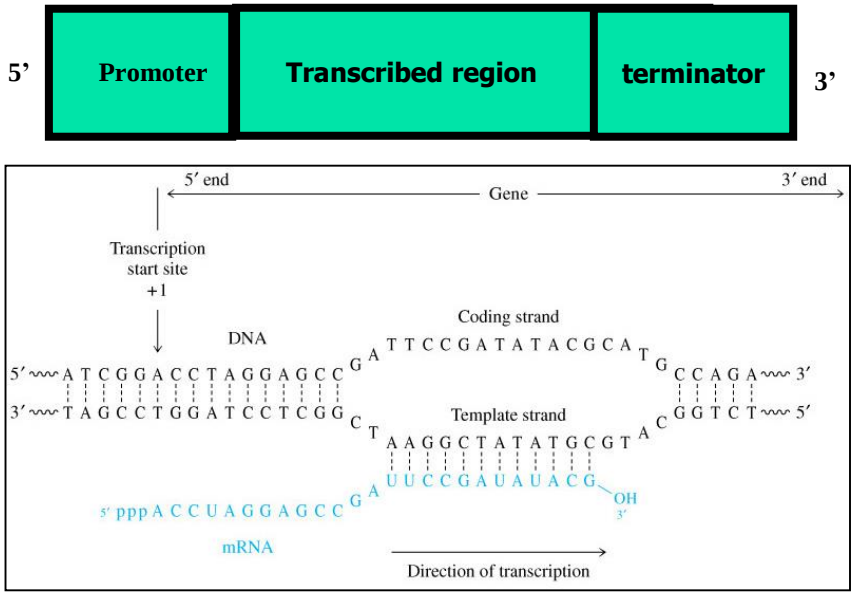
The three steps in transcription

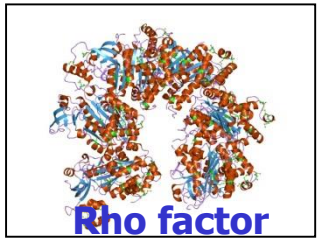
- **Initiation:** Transcription is initiated at the promoter.
- The promoter is a key feature for control of gene expression.
- It is a RNA polymerase binding site, and promoters vary in "strength".
- Later, in the section on plant transformation, we'll see the key role that promoters play in terms of magnitude and timing of gene expression.
- **Elongation:** The mRNA is synthesized in a 5'-3' direction.
- **Termination:** RNA polymerase recognizes specific signals for chain termination and release of the mRNA transcript from the DNA template.

Transcription and Gene Expression

General gene structure

- Promoter: sequences recognized by RNA polymerase as start site for transcription.
- Transcribed region: template from which mRNA is synthesized.
- Terminator: sequences signaling the release of the RNA polymerase from the gene.





Rho factor

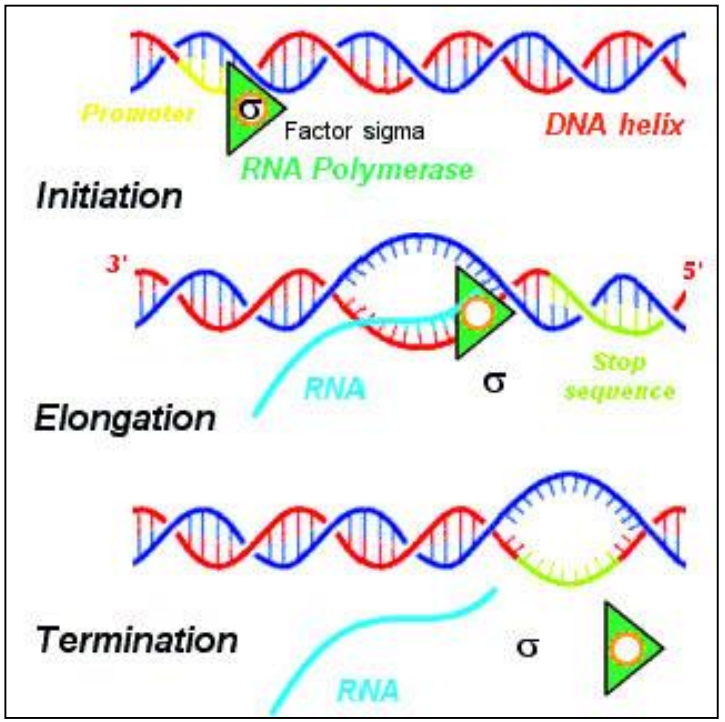
A ρ factor (Rho factor) is a prokaryotic protein involved in the termination of transcription. It is a ~274.6 kD hexamer of identical subunits. Rho's key function is its helicase activity. It terminates transcription in *Escherichia coli*.

Prokaryotic DNA transcription

mRNA transcription is recognized by the sigma protein

Transcription proceeds in one direction

- In bacteria, all transcription is performed by a single type of RNA polymerase.
- The sigma subunit dissociates from the RNA polymerase core enzyme shortly after transcription begins.
- Nucleotides are added onto the 3' end of the growing RNA chain.



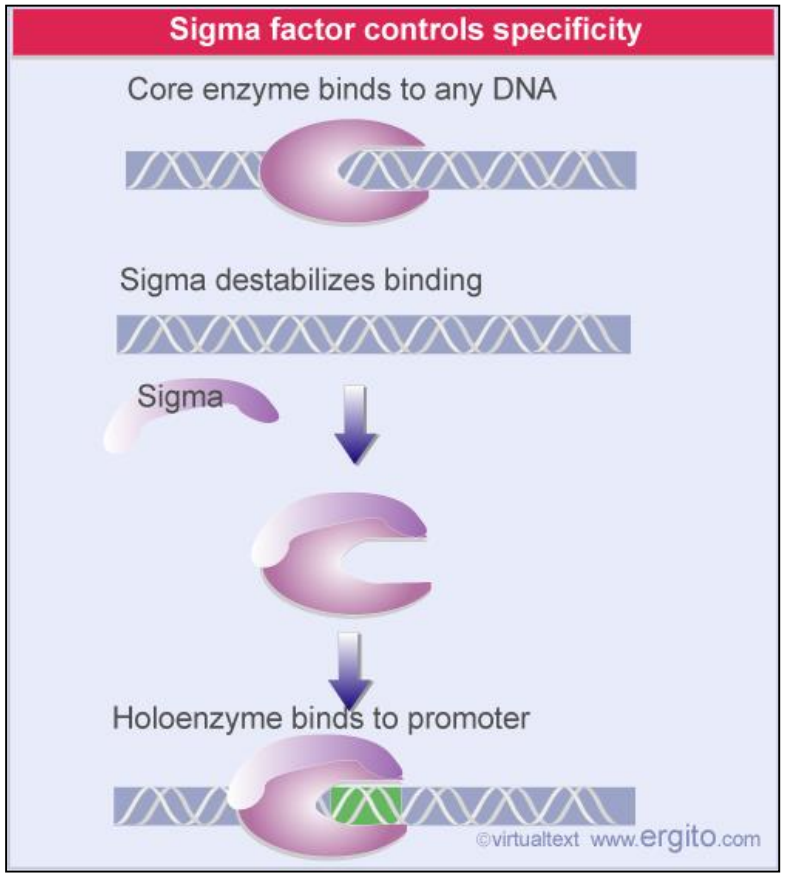
However, a recent study has shown that σ^{70} can remain attached in complex with the core RNA polymerase, at least during early elongation. It was indicates that sigma plays roles during early elongation.

Prokaryotic DNA transcription

Initiation

Sigma factor are subunits of RNA polymerase

- Sigma factors are specialized bacterial proteins that bind to RNA polyerases for transcription initiation(RNA synthesis).
- Several distinct sigma factors have been identified.
- Sigma factors are discriminatory, as each binds a distinct set of promoter sequences.



Holoenzyme is the core enzyme saturated with sigma factor 70.

Rasul Chaudhry;..

Transcription and Gene Expression

Transcription in bacteria

RNA polymerase is an enzyme that synthesizes RNA from a DNA template

A sigma factor (σ factor) is a prokaryotic initiation enzyme factor that must be part of RNA polymerase for attachment to promoter sites on DNA. Shortly thereafter, RNA polymerization begins (the first nucleotide bears a triphosphate on it 5' OH) and the sigma subunit is released.

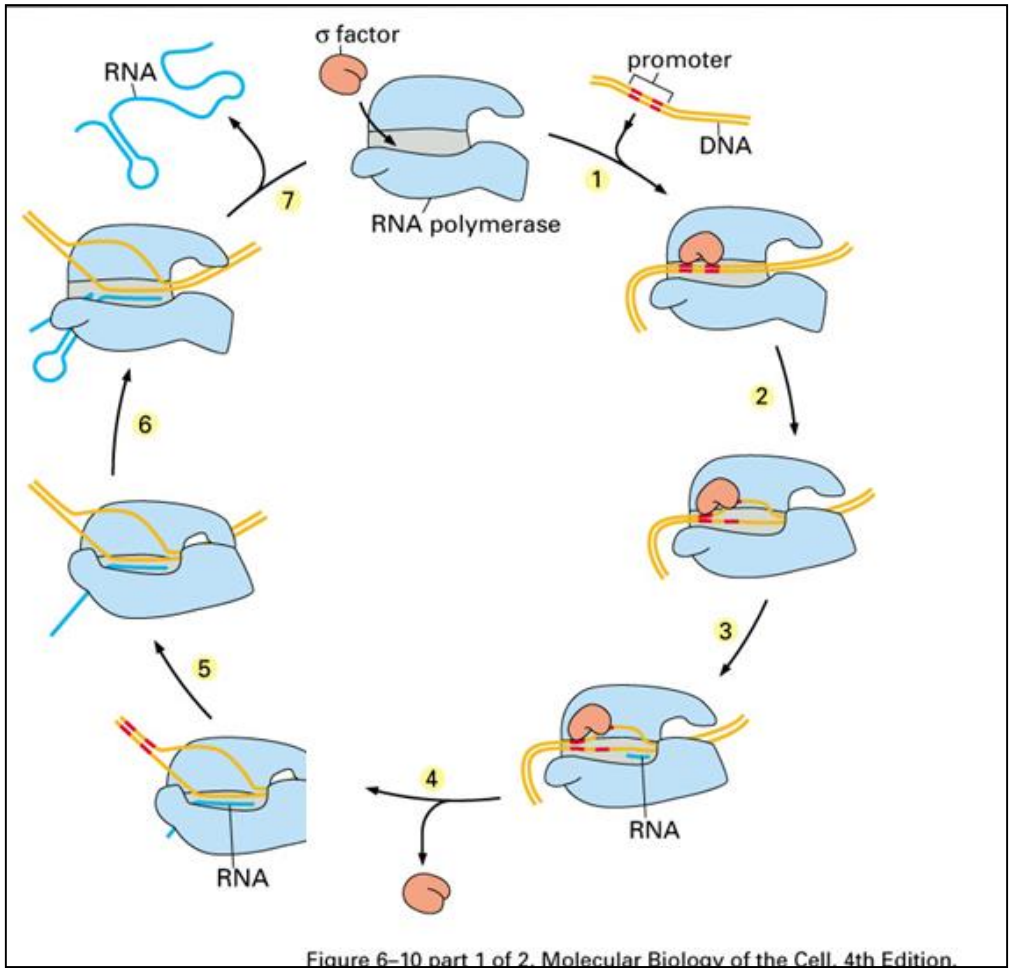
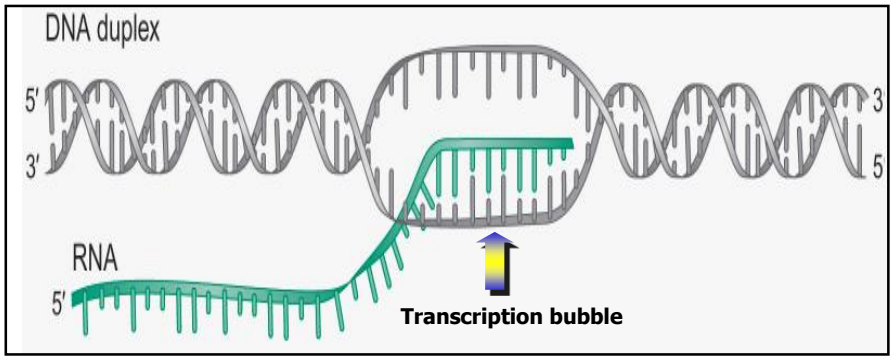


Figure 6–10 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

Transcription by RNA polymerase proceeds in a series of steps

Initiation

- Promoter: the DNA sequence that initially binds the RNA polymerase.
- The structure of promoter-polymerase complex undergoes structural changes to proceed transcription.
- DNA at the transcription site unwinds and a "bubble" forms.
- Direction of RNA synthesis occurs in a 5'-3' direction (3'-end growing).



Prokaryotic DNA transcription

mRNA transcription is recognized by the sigma protein

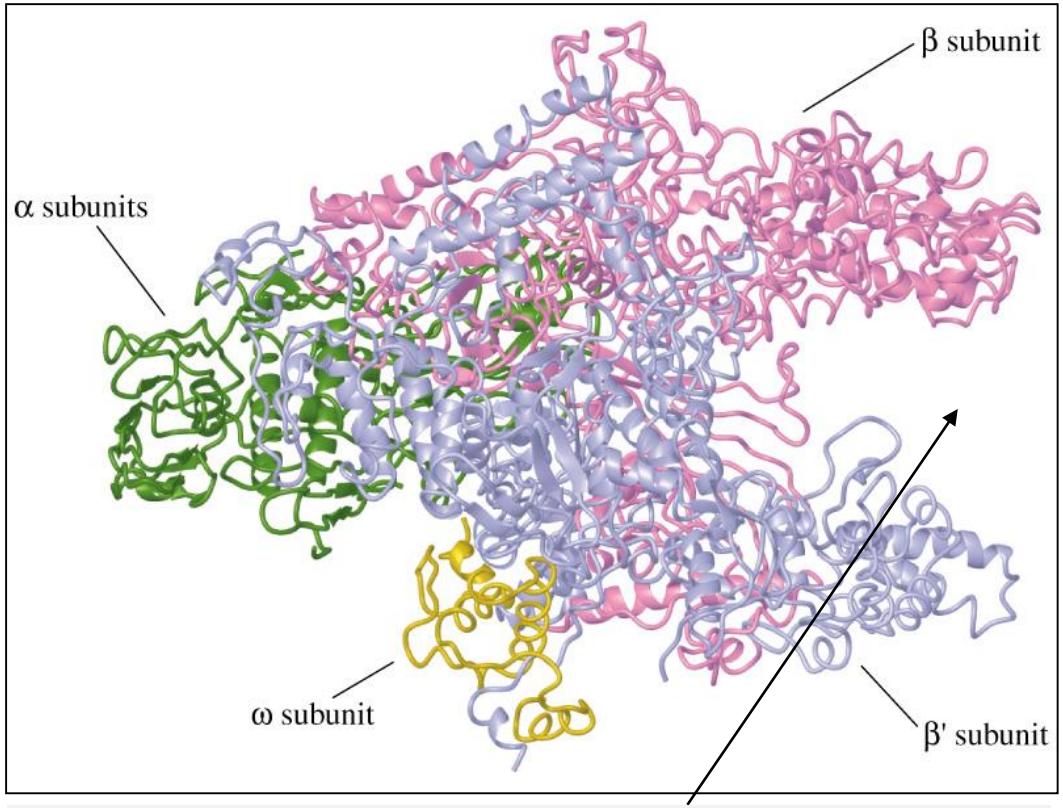
RNA polymerase copy DNA to RNA(mRNA)

- In initiation process, RNA polymerase holo-enzyme (RNA polymerase+sigma factor) binds to DNA and scans for promoter sequences (start sequences).
- The sigma factor is the subunit of the RNA polymerase complex that recognizes the specific promoter sequence of DNA that the RNA polymerase complex should bind to.
- Scanning occurs in only one dimension, 100 times faster than diffusion limit.
- During scanning enzyme is bound non-specifically to DNA.
- Can quickly scan 2000 base pairs.

Holoenzyme is the core enzyme saturated with sigma factor 70. A biochemically active compound formed by the combination of an enzyme (core enzyme) with a coenzyme or cofactor(sigma factor).

Prokaryotic RNA polymerase

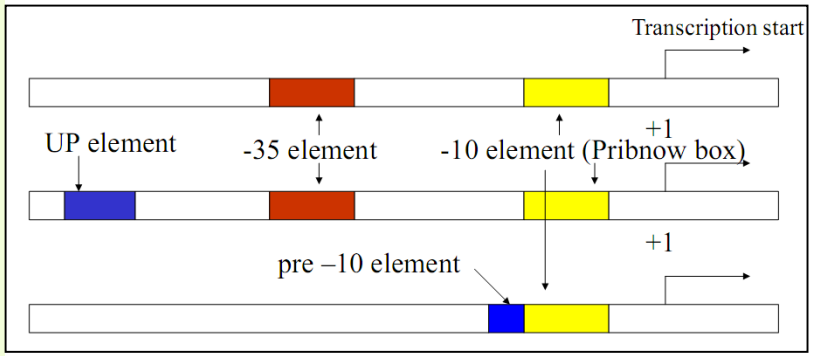
The enzyme without the sigma subunit is called the 'core' enzyme



Site of DNA binding and RNA polymerization.

Transcription by RNA polymerase proceeds in a series of steps
Bacterial promoters

- Most bacterial promoters have –35 and –10 elements (regions).
- Some have UP element (upstream promoter element).
- Some lack –35 element, but have extended –10 region.
- Note also that molecular biologists use a numbering system which has no zero!
- The first nucleotide of the RNA transcript is numbered +1;
- Position +1 is the transcription start site.
- The nucleotide immediately upstream from that is numbered -1.



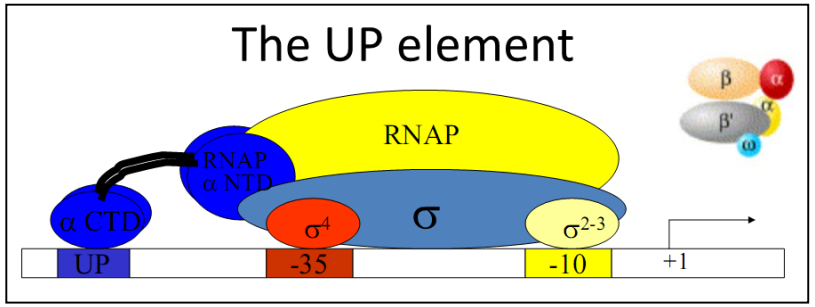
A conserved sequence centered 10 bp upstream of the start point of transcription was identified. The sequence was initially called the Pribnow Box. However, it is more usual now to refer to it simply as the -10 region.

In another word, the Pribnow box is the sequence of six nucleotides thymine-adenine-thymine-etc.(TATAAT) that is an essential part of a promoter site on DNA for transcription to occur in bacteria.

Transcription by RNA polymerase proceeds in a series of steps

Bacterial promoters- UP element

- Some bacterial promoters have UP element (upstream promoter element).
- UP element is an AT rich motif present in some strong (e.g. rRNA) promoters.
- UP element interacts directly with C-terminal domain of RNA polymerase a subunits.



- RNAP(RNA polymerase);
- σ (sigma factor);
- The alpha subunit consists of a N-terminal domain (NTD) and a C-terminal domain (CTD), connected by a short linker.

Prokaryotic DNA transcription
Elongation

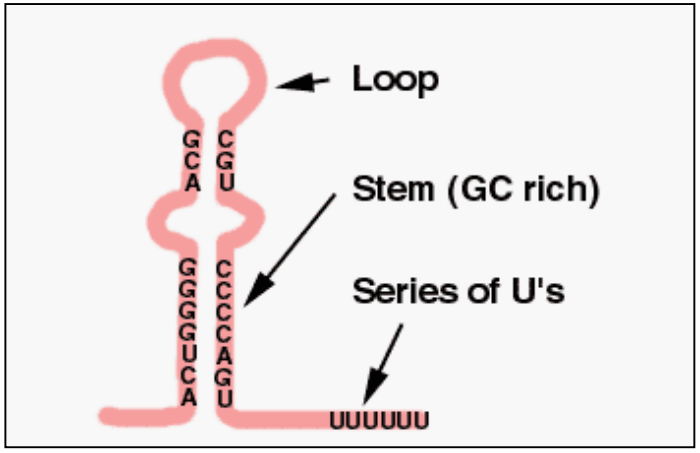
- During the elongation phase, the enzyme moves along the DNA.
- A bubble representing the separation of the two strands of DNA in the vicinity of the enzyme accompanies this movement.
- This means that the DNA is being unwound ahead of the bubble and rewound behind it.
- Remember, RNA synthesis(transcription) is 5' to 3‘.
- Synthesis continues until the enzyme comes to a terminator sequence on the DNA, at which point RNA polymerase ceases RNA synthesis, releases the RNA, and falls off the DNA.

Sofer,1999

Prokaryotic DNA transcription
Termination

- Termination occurs in two ways in bacteria.
- When RNA polymerase encounters a characteristic two part sequence:

1. One part of the sequence causes the newly synthesized RNA to form a stem-loop structure.
2. The second part of the sequence produces a series of uridines (ribose and uracil) usually 6 or more.



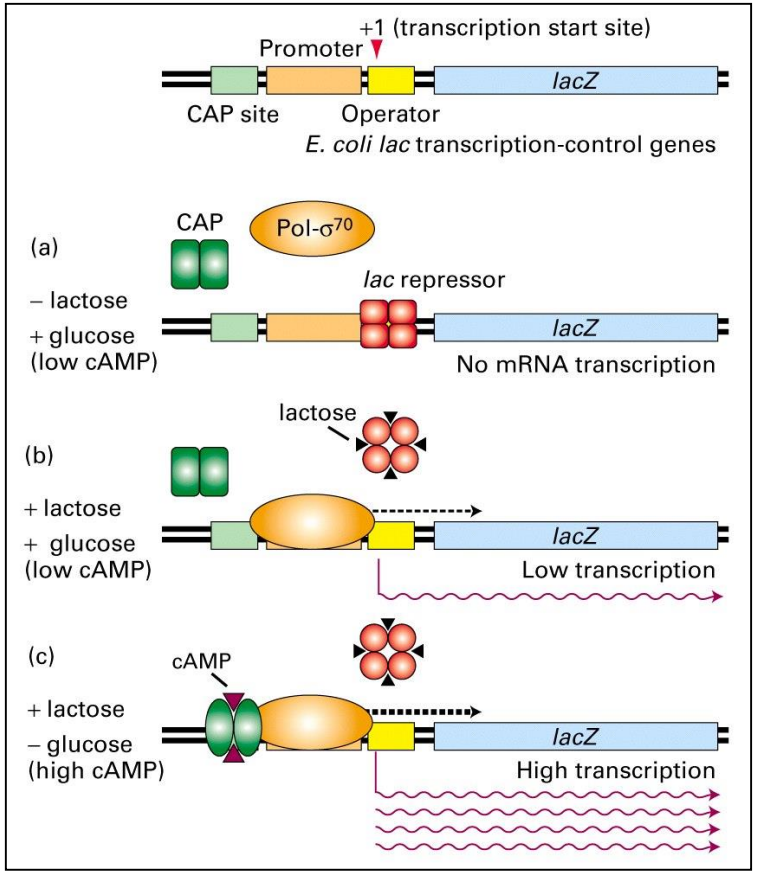
The formation of the stem-loop structure in the RNA appears to cause the RNA polymerase to pause, and during this pause, the DNA/RNA hybrid dissociates and so does the enzyme, thereby stopping RNA synthesis.

Sofer,1999

Transcription by RNA polymerase proceeds in a series of steps

Bacterial promoters, elongation and termination

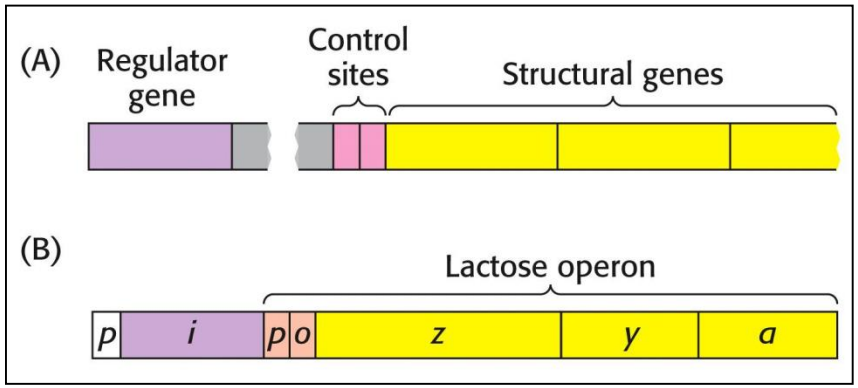
- A promoter is a region of DNA that initiates transcription of a particular gene.
- Promoters are located near the transcription start sites of genes.
- Promoters can be about 100-1000 base pairs long.
- All CAP–dependent promoters have weak –35 sequence, so that RNAP-σ is unable to bind the promoter without CAP assistance (Catabolite Activator Protein).



Transcription of the lac operon

- Lac - Lactose (dissacharide): Glucose and Galactose
- Operon - series of structural genes all under the control of a Regulatory Gene

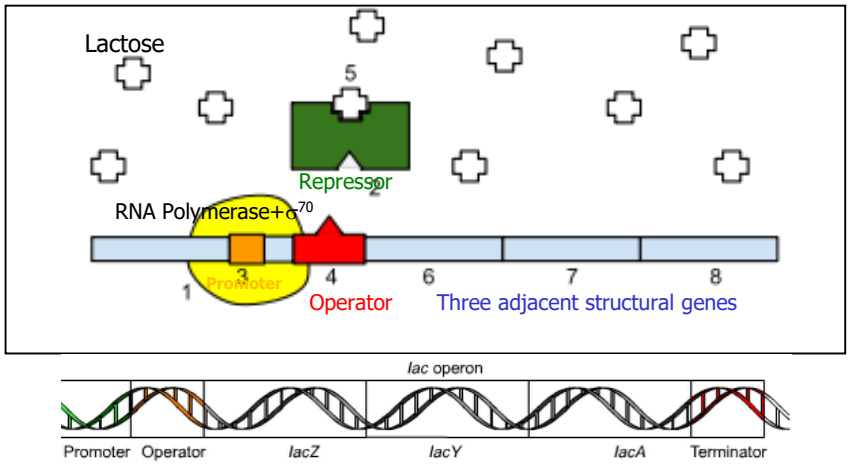
1. lac operon is normally turned off.
2. lac operon is an Inducible operon.



Transcription of the lac operon

The whole mechanism of transcription of the lac operon depends highly on the cooperation of transcription factors

- The *lac* operon (lactose operon) is an operon required for the metabolism of lactose.
- It has three adjacent structural genes, *lacZ*, *lacY*, and *lacA*.
- RNA polymerase to bind with the promoter, and express the genes, which synthesize lactase.
- The repressor will then bind to the operator, stopping the manufacture of lactase.



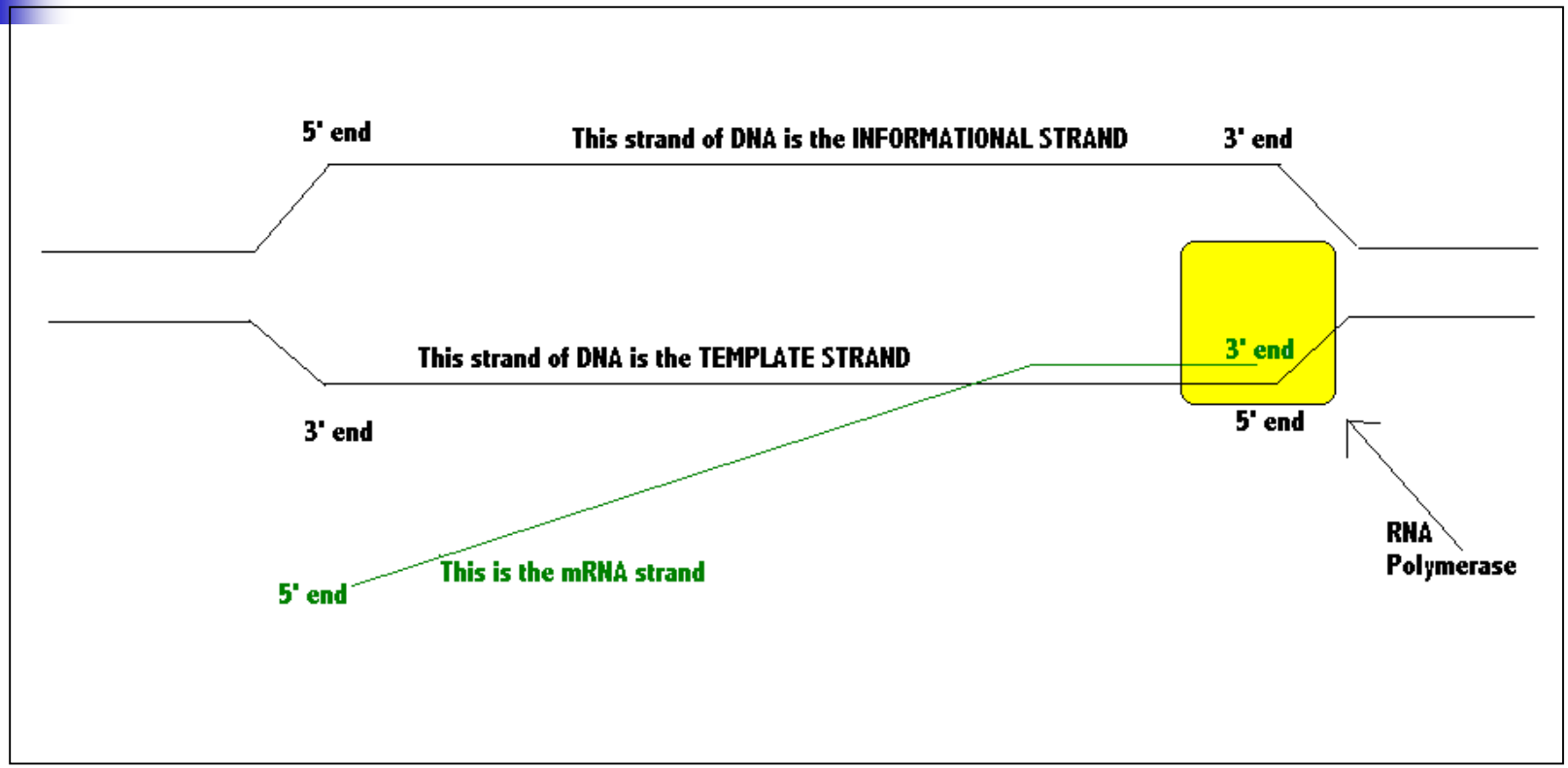
1: RNA Polymerase, 2: Repressor, 3: Promoter, 4: Operator, 5: Lactose, 6: lacZ, 7: lacY, 8: lacA.

The general structure of an operon:

1. Promoter: A nucleotide sequence that enables a gene to be transcribed.
2. Operator: A segment of DNA that a regulator binds to.
3. Structural genes: The genes that are co-regulated by the operon.

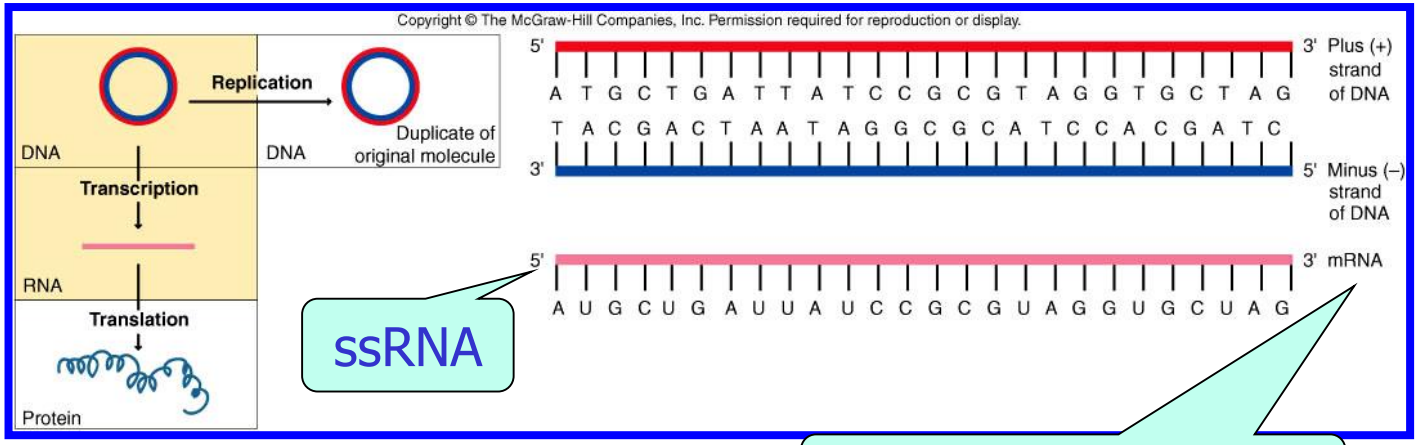
The process of transcription in cartoon form

RNA transcription



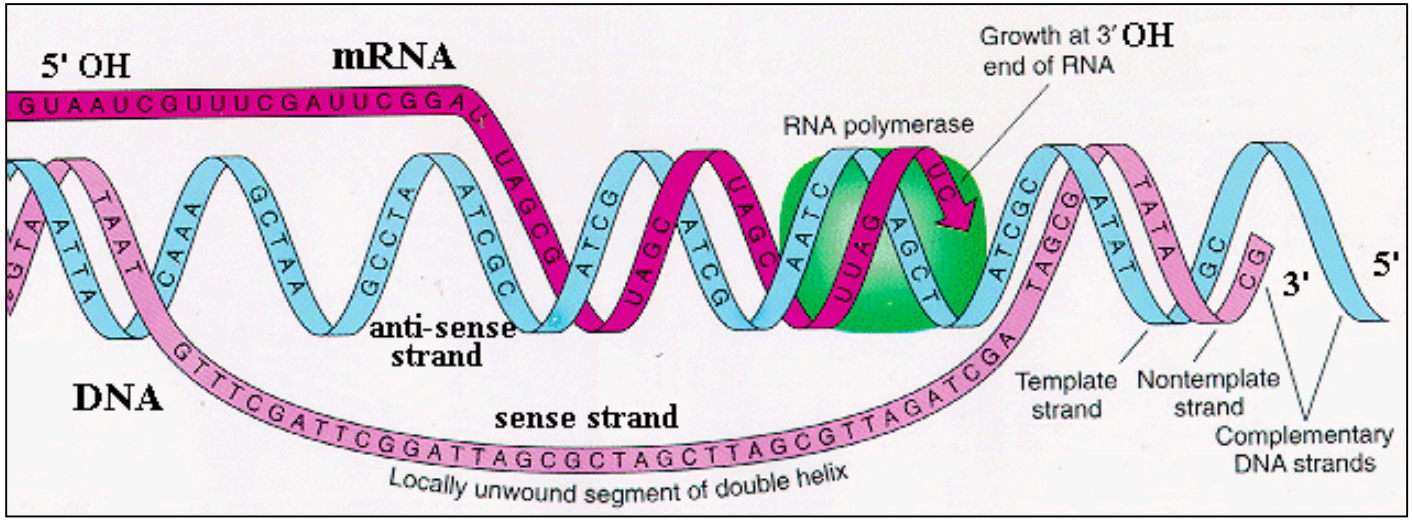
RNA transcription

RNA synthesis

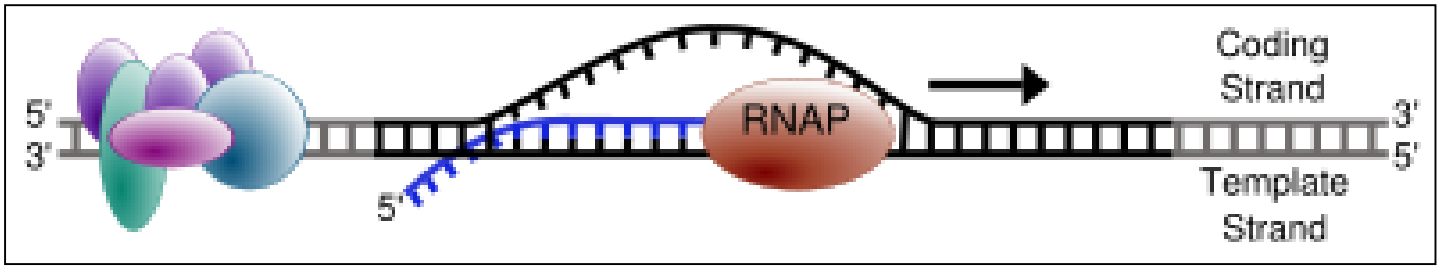


RNA transcription

RNA synthesis



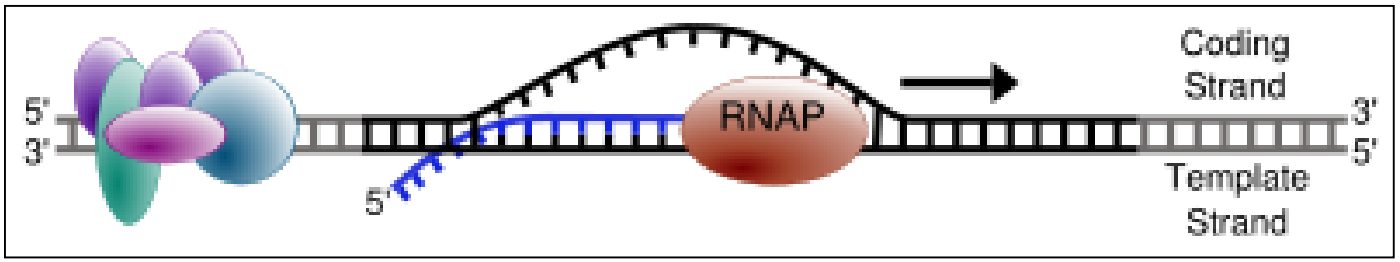
Simple diagram of elongation



Wikimedia,2008

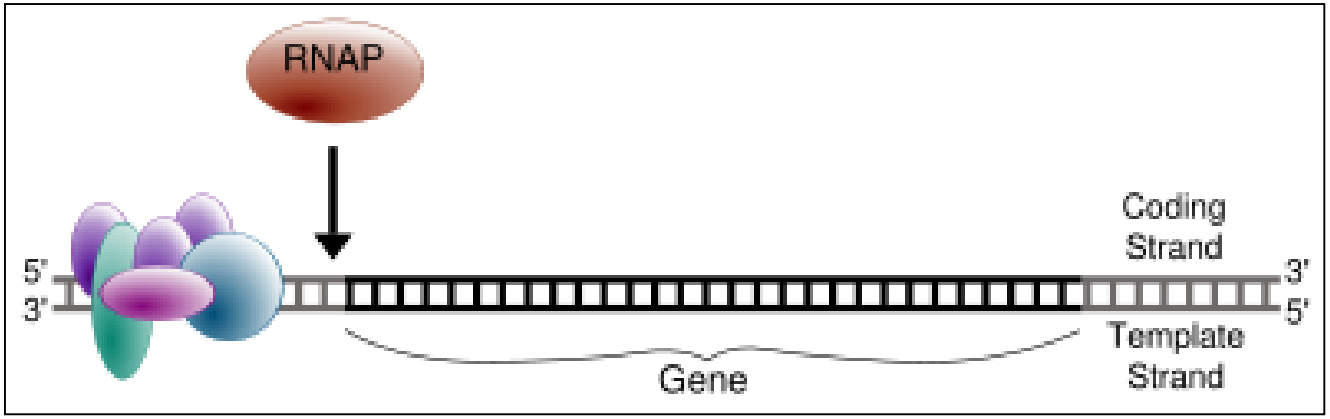
Simple diagram of elongation

- One strand of DNA, the template strand (or non-coding strand), is used as a template for RNA synthesis.
- As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarily with the DNA template to create an RNA copy.



Wikimedia,2008

Simple diagram of initiation

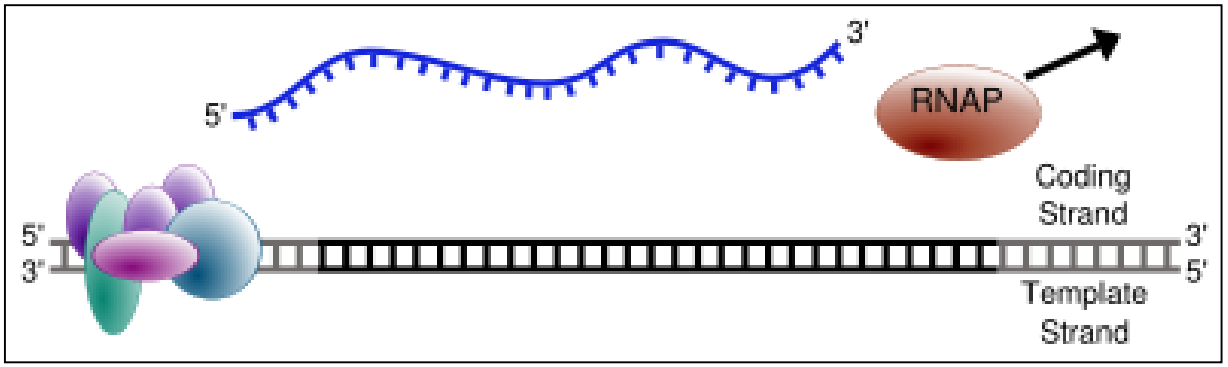


In bacteria, transcription begins with the binding of RNA polymerase to the promoter in DNA.

RNAP = RNA polymerase

Simple diagram of termination

- The final stage of the transcription of a gene is "termination", after the stop codon of the gene.
- RNA polymerase recognizes specific signals for chain termination and release of the mRNA transcript from the DNA template.



Wikimedia,2008

Transcription

Components of transcription in prokaryotes

- + strand
- + strand
- Promoter
- RNA polymerase
- Sigma(σ) factor
- Terminator

Sense(+) strand complement of antisense strands (-)

- The coding strand (sense strand) of DNA has the same sequence as the mRNA and is related by the genetic code to the protein sequence that it represents.
- The mRNA is defined as a "sense strand".
- The antisense strand (template strand) of DNA is complementary to the sense strand, and is the one that acts as the template for synthesis of mRNA.

Genes along a particular length of double-stranded DNA might have this "sense/antisense" arrangement:

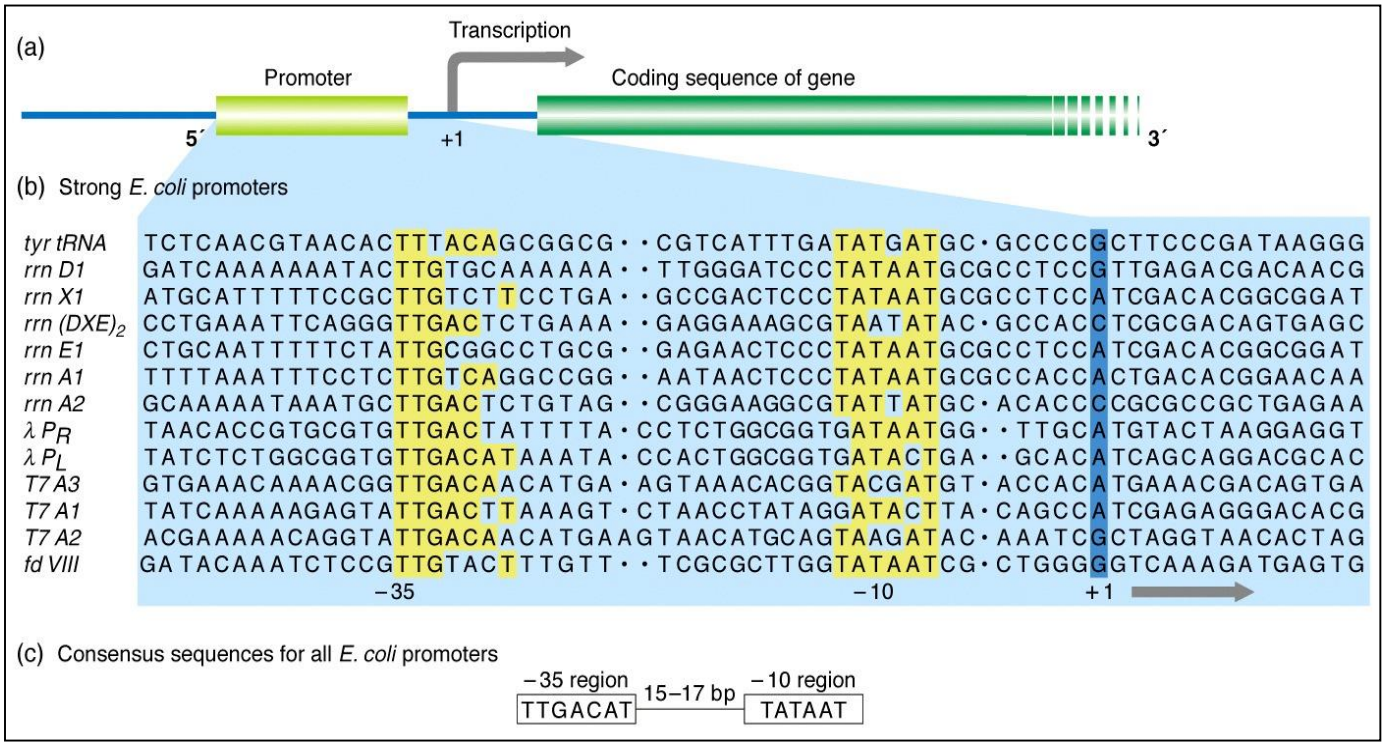
A Strand = 5' +-+++-+---3'

Opposite Strand= 3'-+---+-+++5'

Promoter

A region of DNA where RNA polymerase binds to initiate transcription

- Promoter: Sequences in the DNA strand which accept RNA polymerase and initiate transcription.
- Transcription proceeds in the 3' to 5' direction along the DNA template coding strand (RNA strand grows in the 5' to 3' direction).



Promoter

A region of DNA where RNA polymerase binds to initiate transcription

- In genetics, a promoters is a region of DNA that initiates transcription of a particular gene.
- Promoters are located near the transcription start sites of genes, on the same strand and upstream on the DNA (towards the 5' region of the sense strand).
- Promoters can be about 100–1000 base pairs long.
- In bacteria, the promoter is recognized by RNA polymerase and an associated sigma factor, which in turn are often brought to the promoter DNA by an activator protein's binding to its own DNA binding site nearby.

Promoter

A region of DNA where RNA polymerase binds to initiate transcription

Bacterial vs. eukaryotic/archaeal promoters

- Promoter region of DNA, contains the transcription start site(s).

1. In bacteria, the sigma factor recognizes and binds to the promoter sequence.
2. In eukaryotes, the transcription factors (equivalent of sigma) perform this role and recognize and bind to the promoter sequence.

- The general transcription factors (GTFs) are a set of factors in eukaryotes that are required for all transcription events.

Promoter

A region of DNA where RNA polymerase binds to initiate transcription

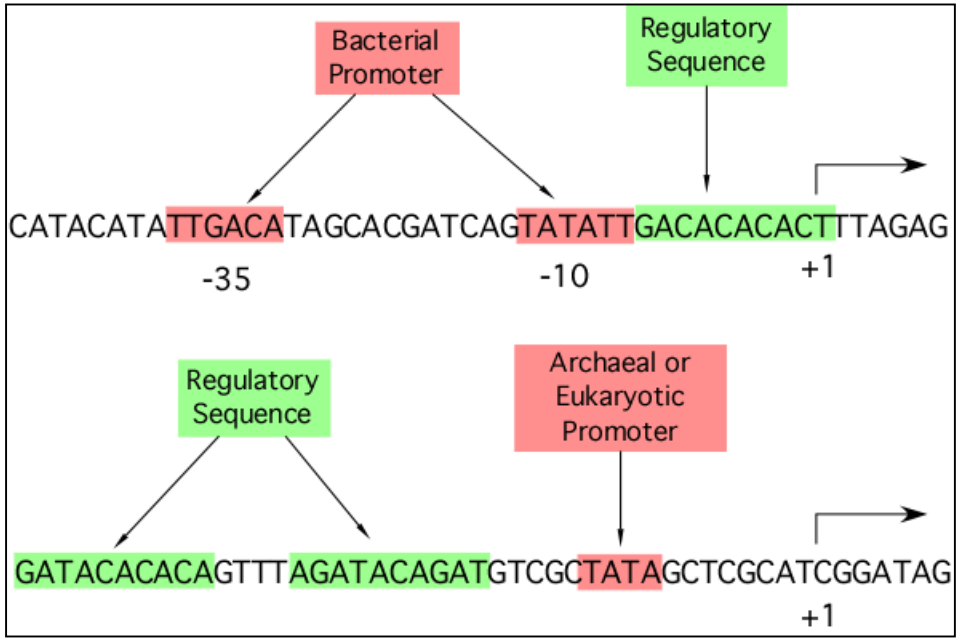
Bacterial promoters

1. In bacteria, a promoter for mRNA transcription is recognized by the sigma protein and has two recognition zones about 10 and 35 bases before the transcription start site.
2. A sigma factor (σ factor) is a prokaryotic transcription initiation factor that enables specific binding of RNA polymerase to gene promoters. Different sigma factors are activated in response to different environmental conditions.
3. Several distinct sigma factors have been identified.
4. σ factors are always associated with RNA polymerase.
5. Every molecule of RNA polymerase contains exactly one sigma factor subunit.

Promoter

A region of DNA where RNA polymerase binds to initiate transcription

- In bacteria, a promoter for mRNA transcription is recognized by the sigma protein and has two recognition zones about 10 and 35 bases before the transcription start site.



Transcription by RNA polymerase proceeds in a series of steps

The σ factors

- In bacteria the promoter is recognized by RNA polymerase and an associated sigma factors.
- Sigma factors are required for:

1. Promoter recognition, and
2. Transcription initiation in prokaryotes.

- A variety of σ factors exist in *E. coli*.
- σ^{70} is essential for cell growth in all conditions.
- Other sigmas are required for special events, like:
- nitrogen regulation (σ^{54}),
- response to heat shock (σ^{32}), sporulation, etc.

Sigma factor (rotcaf σ)

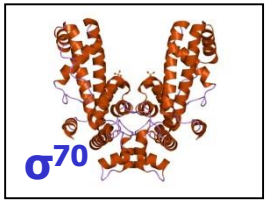
Bacterial transcription initiation factor

Found in bacteria

- It is the σ-factor that allows the RNA polymerase to recognize specifically the two conserved nucleotide motifs in the promoter region.
- Thus the sigma factor determines the specificity of the enzyme.
- All sigma factors are distinguished by their characteristic molecular weights.
- For example,
- σ^{70} refers to the sigma factor with a molecular weight of 70 kDa.

Sigma factor (rotcaf σ)

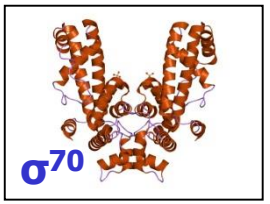
Sigma factors or sigma factor regulons



- Sigma factor (also known as the regulon) is a protein needed only for initiation of RNA synthesis.
- Sigma factors are a major regulator of prokaryotic gene expression.
- It is well known that bacteria use different sigma factors to control the initiation specificity at different promoters, including those promoters whose genes encode virulence factors.

Paget *et al.*,2003; Wilson *et al.*,2015

Sigma factor (rotcaf σ)
Types and numbers



- Including both:

1. Major or abundant primary (housekeeping) sigmas, and
2. Minor alternative (secondary) sigmas from diverse organisms.

- Every molecule of RNA polymerase holoenzyme contains exactly one housekeeping" sigma factor (primary sigma factor), transcribes most genes in growing cells.
- Primary sigma factor is present in all growth conditions.

1. The number of sigma factors varies between bacterial species.

- Thus there are members of the σ^{70} family of sigma factors.

Paget *et al.*,2003;..

Sigma factors

Bacterial transcription initiation factor

Different numbers and specialized sigma factors

- Sigma factor (σ factor) is a protein needed only for initiation of RNA synthesis.
- It is a bacterial transcription initiation factor that enables specific binding of RNA polymerase to gene promoters.
- The number of sigma factors varies between bacterial species.
- Different sigma factors are utilized under different environmental conditions.
- These specialized sigma factors bind the promoters of genes appropriate to the environmental conditions, increasing the transcription of those genes.

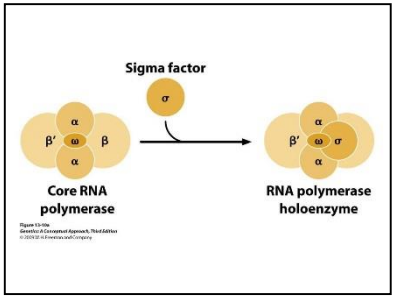
Wikipedia

Sigma factors
Types and numbers

- Every molecule of RNA polymerase holoenzyme contains exactly one sigma factor subunit.
- Sigma factors allow sequence-specific binding of RNA polymerase to bacterial promoters.
- The number of sigma factors varies between bacterial species.
- *E. coli* has seven sigma factors.
- The 10 sigma factors thus far identified in *B. subtilis*.
- *Streptomyces avermitilis* and *Streptomyces coelicolor* with 13 and 14 σ^{70} genes, respectively.
- Sigma factors are distinguished by their characteristic molecular weights.
- For example, σ^{70} refers to the sigma factor with a molecular weight of 70KDa.

Sigma factors

Major and minor sigma factors



- A sigma factor (σ factor) is a prokaryotic initiation enzyme factor that must be part of RNA polymerase for attachment to promoter sites on DNA.
- Different sigma factors can be expressed when a cell is exposed to different conditions.
- ***E. coli* sigma factors:**
- σ70 (RpoD) - The "housekeeping" sigma factor, transcribes most genes in growing cells.
- σ38 (RpoS) - The starvation/stationary phase sigma factor
- σ28 (RpoF) - The flagellar sigma factor
- σ32 (RpoH) - The heat shock sigma factor
- σ24 (RpoE) - The extracytoplasmic stress sigma factor
- σ54 (RpoN) - The nitrogen-limitation sigma factor
- σ19 (FecI) - The ferric citrate sigma factor

The sigma factor RpoH (σ^{32}) is the key regulator of the heat shock response in *Escherichia coli*.

Sigma factors

Major and minor sigma factors

Name	Function
σ^{70}(RpoD) (major)	The sigma factor 70 (the rpoD gene product) is responsible for transcription of most genes expressed during the exponential cell growth
σ^{19} (FecI)	the ferric citrate sigma factor, regulates the fec gene for iron transport
σ^{24}(RpoD)	the extracytoplasmic/extreme heat stress sigma factor
σ^{28}(RpoD)	the flagellar sigma factor
σ^{32} (RpoH)	the heat shock sigma factor, it is turned on when the bacteria are exposed to heat
σ^{38} (RpoS)	the starvation/stationary phase sigma factor
σ^{54} (RpoN)	the nitrogen-limitation sigma factor

Note: There are also anti-sigma factors that inhibit the function of sigma factors and anti-anti-sigma factors that restore sigma factor function.

Sigma factors

Minor sigma factor

Sigma factor RpoH (σ^{32})

- The sigma factor RpoH (σ^{32}) is the key regulator of the heat shock response in *Escherichia coli*.

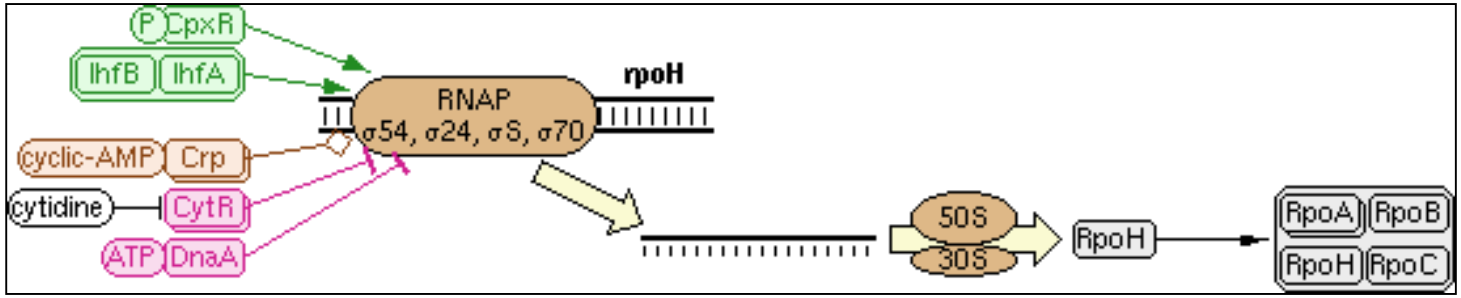
1. During heat shock stress condition, intracellular concentration of free RpoH protein increases in response to heat shock, due to slightly increased transcription, increased synthesis and stabilization of the protein.
2. RNA polymerase (RNAP) also involve in initiation of transcription of heat shock genes/heat shock proteins.
3. In the absence of heat shock, or after heat shock, sigma factor RpoH (σ^{32}) activity is inhibited by transient association with DnaK and DnaJ, which reduces the amounts of free active RpoH (makes it unstable since RpoH is quickly degraded by the FtsH protease), leading to a decrease in the rate of synthesis of heat shock proteins and shut-off of the heat shock response.

Sigma factors

Sigma 32 (sigma H) factor

Heat shock control

- *rpoH* encodes σ^{32}, the primary sigma factor controlling the heat shock response during log-phase growth.
- It is subject to tight control via a multivalent regulatory system that reponds to temperature and the abundance of misfolded proteins within the cell.
- σ^{32} directly controlled by the DnaK-DnaJ-GrpE chaperone system.

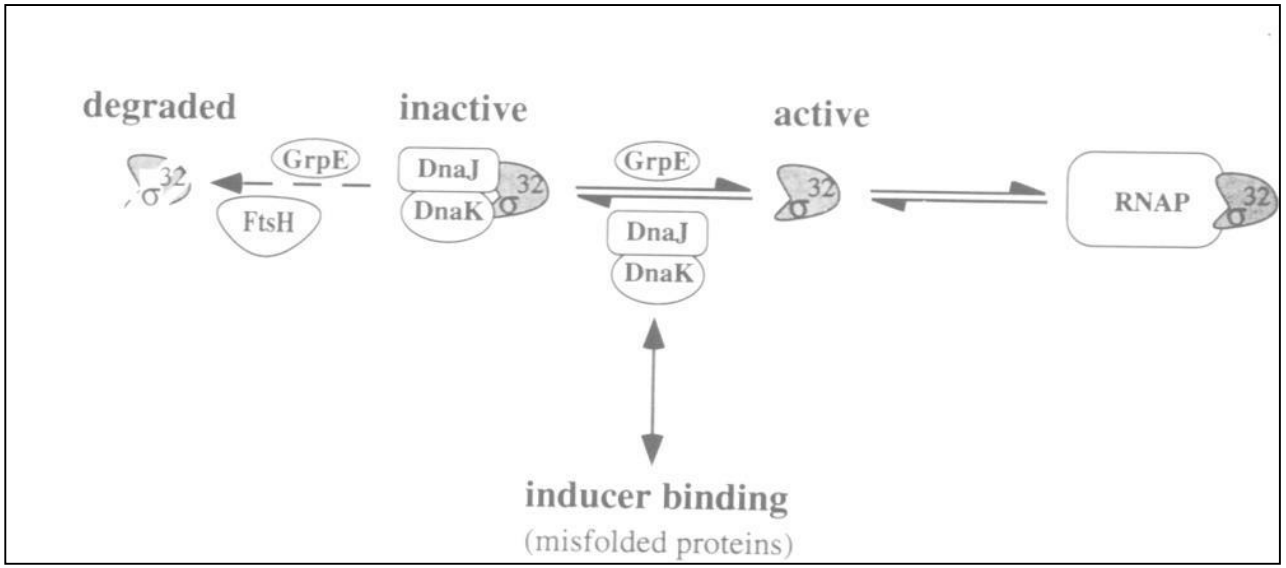


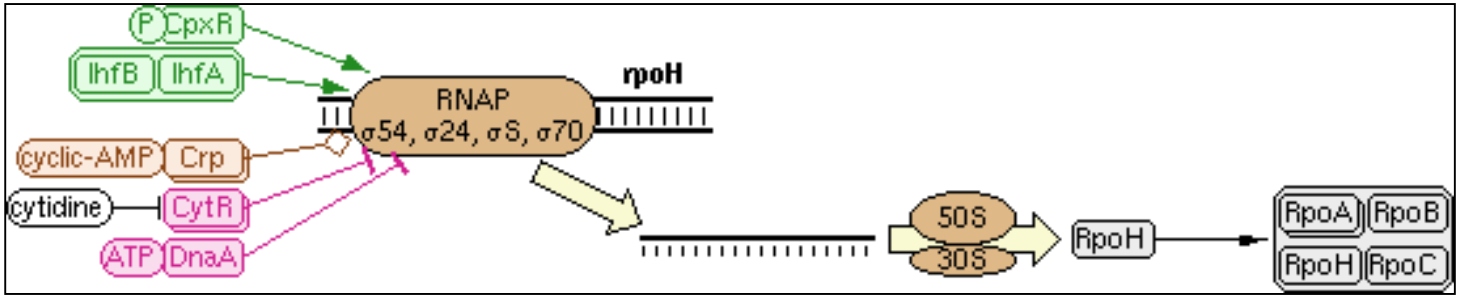
Biocyc.org

Protein Folding

Heat-shock proteins (HSPs)

The DnaK chaperone system



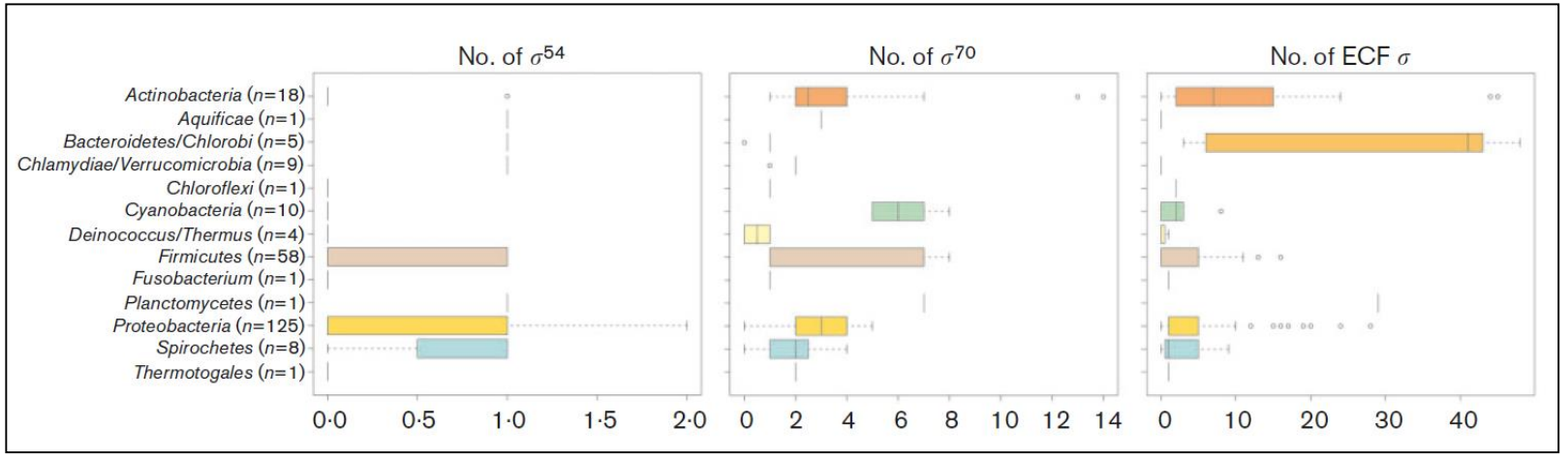


Gamer *et al.*,1996;Biocyc.org

Prokaryotic DNA transcription

Box and whisker plot of the number of sigma factor proteins in 13 different bacterial phyla

- *P. fluorescens* Pf5 is the largest genome and it also has by far the largest number of sigma factors – a total of 33 (1 σ^{54}, 4 σ^{70} and 28 ECF σ^{E}).
- The σ^{70} group is shown in the middle.
- The extracytoplasmic function (ECF) Group IV sigma factors are conserved across both Gram-positive and Gram negative species. σ^{E}, an ECF sigma factor, important for bacterial responses at the cell surface
- nitrogen-responsive σ factor.



Kill *et al.*,2015

Prokaryotic DNA transcription

Alternative sigma factors

Minor alternative (secondary) sigmas

- Bacteria and especially those capable of persisting in diverse environments, such as *Escherichia coli* provide particularly valuable models for exploring how single-celled organisms respond to environmental stresses.
- Bacteria have developed sets of specific response genes that are regulated by a subset of the σ^{70}-like sigma factors in order to respond to a changing environment. E.g.

1. When the cell is put under stress by high temperature.
2. Sporulation sigma factors in *Bacillus subtilis*.

- Because it is the sigma subunit that is responsible for promoter recognition, different sigma subunits may allow different promoters to be recognized.

Sofer,1999;..

Sigma factors

Alternative sigma factors

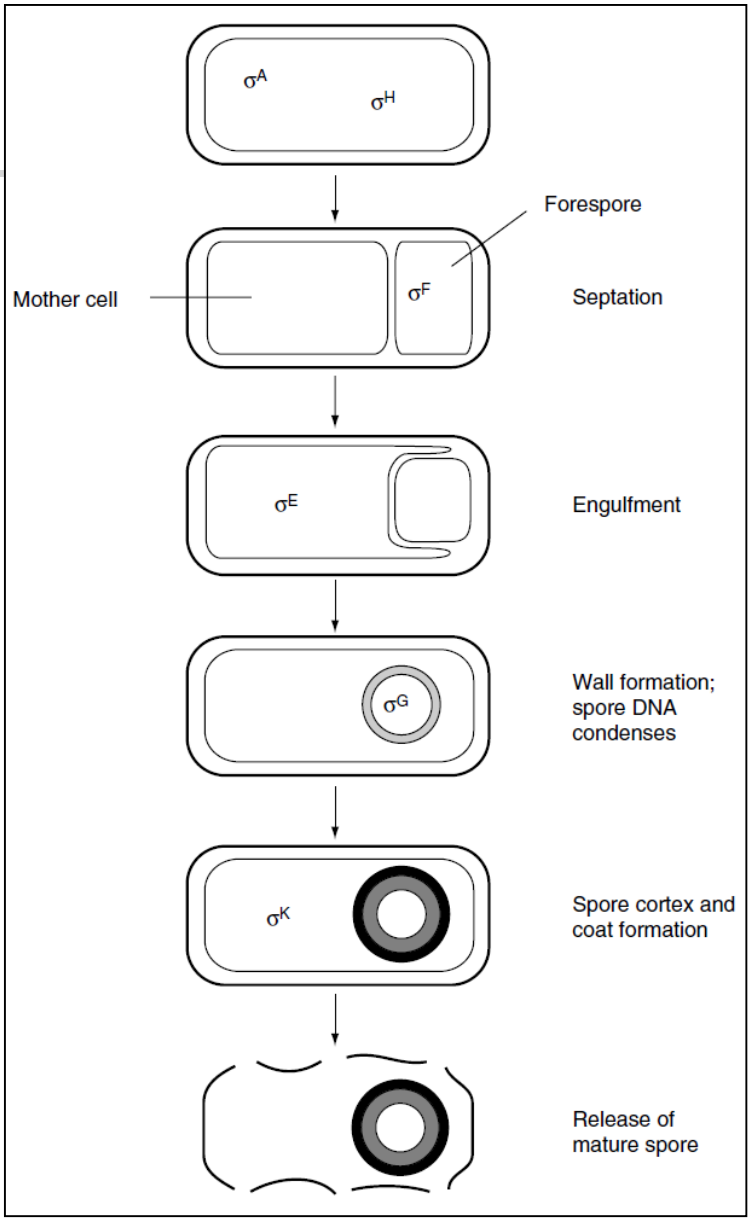
- Alternative sigma factors and promoter recognition sequences.
- *In *B. subtilis*.
- All others are in *E. coli*.

-35 region	Spacer	-10 region	Sigma factor	Regulated genes
TTGACA	N_{16-18}	TATAAT	σ^{70}	General housekeeping genes
CTTGAAA	N_{13-15}	CCCATNT	σ^{32}	Heat shock genes
CTGGNA	N_6	TTGCA	σ^{54}	Nitrogen assimilation genes
		CTACACT	σ^{38}	Stationary phase/General stress genes
GTTTAA	N_{12-14}	GGGTAT	σ^{B}	General stress genes*
TAAA	N_{15}	GCCGATAA	σ^{28}	Flagella genes

Dale and Park,2004

Sigma factors
Sporulation in *Bacillus*

- The diagram shows the stages at which each sigma factor first becomes active.
- Sporulation is governed by a complex transcriptional regulatory programme that controls the expression of more than 100 genes and involves the sequential activation of six different σ-factor.



Dale and Park,2004

Prokaryotic DNA transcription

RNA polymerase holoenzymes containing other sigma factors recognize different core promoter sequences

Factor	Gene	Use	-35	Separation	-10
σ^{70}	*rpoD*	General	TTGACA	16 - 19 bp	TATAAT
σ^{32}	*rpoH*	Heat Shock	CCCTTGAA	13 - 15 bp	CCCGATNT
σ^{28}	*fliA*	Flagella	CTAAA	15 bp	GCCGATAA
σ^{54}	*rpoN*	Nitrogen starvation	CTGGNA	6 bp	TTGCA

Alternate sigma factors are used to express specific sets of genes.

Prokaryotic DNA transcription

Alternative sigma factors

Minor alternative (secondary) sigmas

- σ^S(RpoS) and σ^B (SigB) have been identified as general stress responsive alternative sigma factors in Gram-negative and in Gram-positive bacteria, respectively.
- σ^B, a Group III sigma factor encoded by *sigB*, was initially identified and characterized in *B. subtilis*, but has also been identified in other Gram-positive bacteria.
- The *B. subtilis* σ^B-dependent general stress regulon is large: over 200 genes are expressed following bacterial exposure to heat, acid, ethanol, salt stress, entry into stationary phase, or starvation for glucose, oxygen, or phosphate.
- *sigB* mutants are sensitive to oxidative stress.

Boor,2006

Alternative sigma factors and their roles in bacterial virulence

The stationary phase sigma factor RpoS

- Heat-shock proteins are also essential for stationary phase.
- The alternative sigma factors RpoS (σ^{38}) has been shown to regulate the expression of genes in response to stationary phase, nutrient deprivation, and oxidative and osmotic stress.
- These are environments which are physiologically relevant to those encountered by many microbial pathogens during the natural course of infection.
- The RpoS sigma factor has been shown to be important for virulence in a number of plant and animal bacterial pathogens including *P. aeruginosa*.

Kazmierczak *et al.*,2008

Alternative sigma factors and their roles in bacterial virulence

The stationary phase sigma factor RpoS

- Among the pathogenic bacteria, *Pseudomonas aeruginosa* is perhaps the best understood in terms of the virulence factors regulated and the role the Quorum sensing plays in pathogenicity.
- Regulation of Quorum sensing by RpoS in *Pseudomonas aeruginosa*.
- RsmA, RpoS, QocR all negatively regulate the Rhl or Las Quorum sensing systems, thus preventing early activation of these systems.

Whiteley *et al.*,2000; Deep *et al.*,2000

Alternative sigma factors and their roles in bacterial virulence

PvdS of *Pseudomonas aeruginosa*

- Some bacterial pathogens are known to express virulence genes in low iron environments.
- In *P. aeruginosa*, the siderophore pyoverdine is a virulence factor.
- PvdS is required for virulence and appears to regulate only virulence-related genes.
- The genes involved in pyoverdine synthesis are located in three clusters on the *P. aeruginosa* chromosome and regulated by ECF sigma factor PvdS.

Extracytoplasmic function (ECF) sigma factors, σ^E

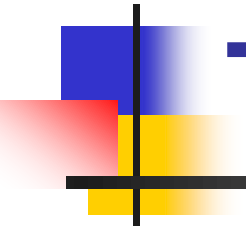
Kazmierczak *et al.*,2008; Brooks and Buchanan,2008

Alternative sigma factors involved in virulence

- As alternative sigma factors have been shown to regulate expression of both virulence and virulence-associated genes, these sigma factors can contribute both directly and indirectly to bacterial virulence.

Family, class, and sigma factor[a]	Bacterial species[b]
σ^{70} family	
Stress response	
σ^{B}	*B. anthracis*, *L. monocytogenes*, *M. tuberculosis*, *S. aureus*, *S. epidermidis*
σ^{S}	*E. coli*, *P. aeruginosa*, *S. enterica* serovar Typhimurium, *S. enterica* serovar Typhi
σ^{F}	*M. tuberculosis*
ECF	
RpoE	*H. influenzae*, *S. enterica* serovar Typhimurium, *V. cholerae*
AlgU	*P. aeruginosa*
PvdS, FpvI	*P. aeruginosa*
σ^{C}	*M. tuberculosis*
σ^{D}	*M. tuberculosis*
σ^{E}	*M. tuberculosis*
σ^{H}	*M. tuberculosis*
HrpL	*Erwinia* spp., *P. syringae*
σ^{28}	
FliA	*C. jejuni*, *H. pylori*, *S. enterica* serovar Typhimurium, *V. cholerae*, *Y. enterocolitica*
σ^{54} family	
σ^{N}	*C. jejuni*, *H. pylori*, *L. monocytogenes*, *P. aeruginosa*, *P. syringae*, *V. cholerae*, *V. parahaemolyticus*

Kazmierczak *et al.*,2008



Translation

RNA to protein information transfer

Bacterial ribosome

Ribosomes (nucleoprotein particles)

Sites or workplaces of protein synthesis

- Prokaryotic bacteria contain 70S ribosomes, nearly 20,000 per cell. This accounts for about 25% of the total cell mass.
- Ribosomes are the most dynamic cellular machine with 2.5 MD mass. About two-thirds of ribosome's mass made up of RNA.
- Ribosomes are classified as eukaryotic or prokaryotic in type, on the basis of:

1. Sensitivity to various antibiotics;
2. Functional interchangeability of soluble factors and ribosomes from different sources; and
3. Structure and sedimentation characteristics, e.g.

- 70s (prokaryotic), or
- 80s (eukaryotic).

Eukaryotes have 80s type of ribosomes, 10-20 million per cell, are found in cytoplasm. Viruses use host ribosomes to make proteins for them.

Three stages to translation

- **Initiation:** Initiates translation.
- rRNA polymerase of ribosome binds to mRNA strand.
- 1st tRNA is bonded to mRNA.
- **Elongation:** Ribosome reads mRNA chain in three nucleotide groups (codon) & inserts 2nd tRNA.
  - tRNA anti-codon (with amino acid) binds to mRNA codon.
  - Translocation: 5' to 3' sequence continues building amino acid polymers (protein), one codon at a time.
- **Termination:** tRNA recognizes release factors of nonsense codon.
- Newly completed polypeptide is released from ribosome.

Players with the process of translation

- The ribosome
- tRNA (transfer RNA)
- The A site
- The P site
- Codons
- Anticodons
- Amino Acids

Translation

- Free-floating pieces of DNA, called tRNA, bind to the mRNA at the ribosome.
- All tRNA have amino acids attached to them.
- When the tRNA binds to the mRNA, the amino acids are linked into a protein.
- Each tRNA has a "codon" and each type of codon always carries a particular amino acid.
- A "codon" is a small piece of DNA with 3 nucleotide bases.
- Specified by the DNA.

Translation

- In DNA, there are 4 types of nucleotide bases.
- An "A" (Adenine) only bonds with a "T" (Thymine) in double bond(A=T) and triple bond(G≡C) when "C" (Cytosine) matches with a "G" (Guanine).
- In RNA, uracil (U) bond with Adenine(A=U).
- Thus, the codon on the tRNA can only match specific codons on the mRNA.
- This forms the basis of the language in the DNA, allowing the amino acids to be strung together in the sequence.

Codon/anticodon

- A set of three nucleotide bases on an mRNA molecule is called a codon.
- A set of three nucleotide bases on a tRNA molecule is called an anticodon.
- Anticodon in tRNA complementary to a codon on mRNA.
- An anticodon in a tRNA molecule aligns with a particular codon in mRNA under the influence of the ribosome, so that the amino acid carried by the tRNA is added to a growing protein chain
- Even though there are only 20 amino acids that exist, there are actually 64 possible tRNA molecules:

  $4 \times 4 \times 4 = 64$ possible combinations

- The 20 amino acids can be linked in different combinations and in different numbers. For example:

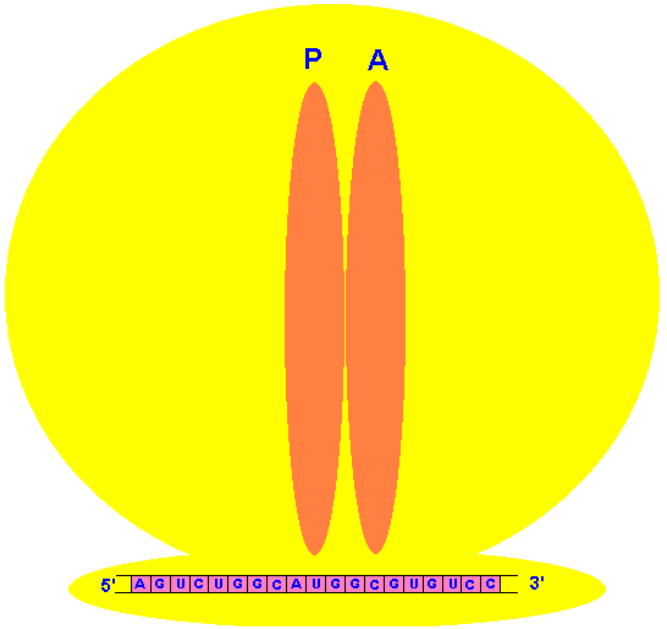
  alanine-valine-tryptophan.....serine

Translation

- in bacteria start codon (AUG) codes for formylmethione where AUG codes for methionine in other regions of the protein.
- tRNA assists in the transport of required amino acids during protein synthesis. In fact, tRNAs serve as adaptor molecules that align amino acids along the mRNA template.
- In prokaryotes, translation of mRNA into protein can begin before transcription is complete since mRNA produced in the cytoplasm.
- **A & P sites:** The two binding sites for loaded tRNA are different:
  - A site binds **a**mino acyl tRNA
  - P site binds **p**eptidyl tRNA
- **Translocation:** Movement of mRNA (& tRNA's) relative to the Ribosome.

A & P sites

- The A site is where the "charged" tRNAs enter the ribosome.
- The P site, holds onto the growing peptide (protein) chain.



Genetic code

DNA codon table(64 codons)

Standard genetic code

1st base	2nd base T		2nd base C		2nd base A		2nd base G		3rd base
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine	TCA		TAA	Stop (*Ochre*)	TGA	Stop (*Opal*)	A
	TTG		TCG		TAG	Stop (*Amber*)	TGG	(Trp/W) Tryptophan	G
C	CTT		CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

Wikepedia,2015

Genetic code

RNA codon table(64 codons)

Standard genetic code

1st base	2nd base U		2nd base C		2nd base A		2nd base G		3rd base
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GUG		GCG		GAG		GGG		G

Wikepedia,2015

Genetic code

RNA codon table (64 codons)

- A codon is a set of 3 bases(nucleotides) which code for an amino acid.
- There are in total 64 different codons which produce 20 different amino acids. These include:

1. 61 possible codons;
2. 1 Start codon;
3. 3 stop codons.

1st letter	Second Letter: U	Second Letter: C	Second Letter: A	Second Letter: G	3rd letter
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Genetic code

- DNA: triplet code
- mRNA: codon (complimentary to triplet code of DNA)
- tRNA: anticodon (complimentary to codon)

Johnson,2003

Genetic code

Standard and nonstandard genetic code

- The standard genetic code was established in the 1960s.
- It consists of:
- 61 triplet nucleotides (codons) coding for the common 20 amino acids, and
- 3 codons coding chain termination (anticodons).
- The genetic code is degenerate as some amino acids are coded by up to 6 "synonymous" codons.
- These deviated genetic code from the standard genetic code are known as nonstandard genetic codes.

Chan *et al.*,2006

Genetic code

Standard and nonstandard genetic code

- In bacteria, nonstandard genetic codes have been found in *Micrococcus*, *Mycoplasma*, and *Spiroplasma*.
- *e.g.* the stop codon UGA in genome of *Mycoplasma* encode tryptophan in *Spiroplasma*.
- Therefore, different genomes have their own characteristic pattern of synonymous codon usage(codon usage diversity among microbes).
- This property has been used to identify pathogenicity islands and HGT clusters in bacterial genomes.

Chan *et al.*,2006

Genetic code

Start codon

- Start codon (initiation codon or initiator, promotor codon):
- Codon on mRNA identifies the first amino acid in the polypeptide sequence (methionine in eukaryotes and N-formylmethionine in prokaryotes).

1. In eukaryotes, the start codon AUG codes for methionine and signals translation to begin.
2. In prokaryotes, the start codons AUG or GUG can act as initiators, and codes for formylmethionine and signals translation to begin.

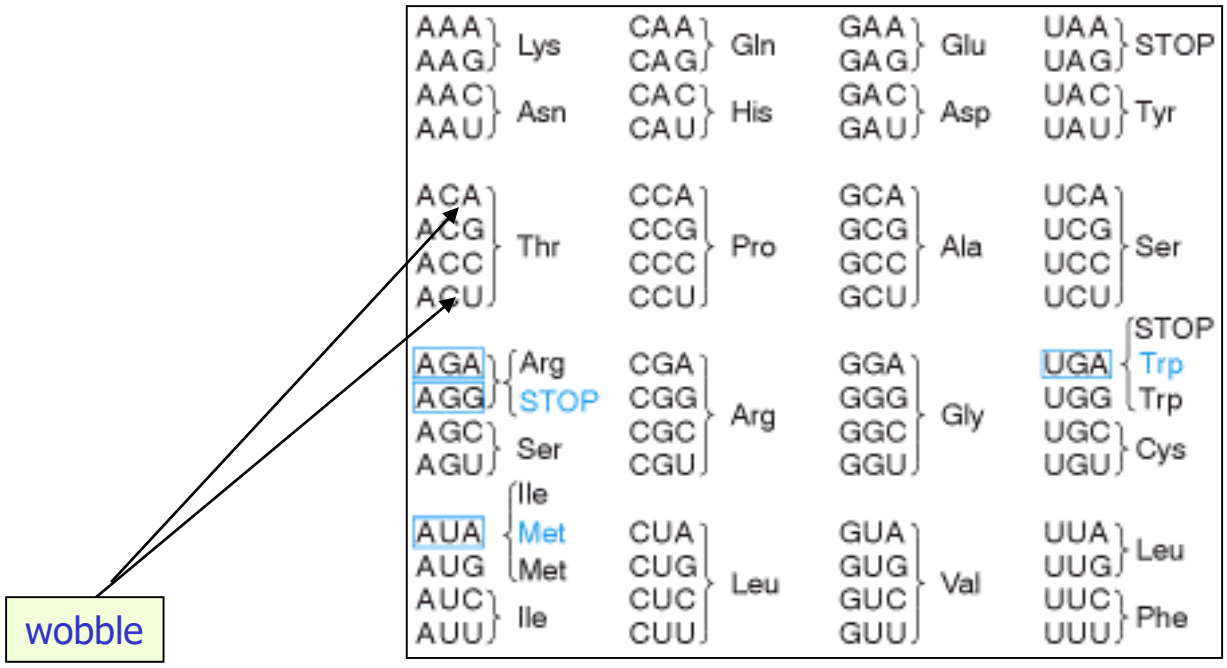
Genetic code

Stop codon

- Stop codon (Termination codon or nonsense codons):
- Any codon that signals the termination of genetic translation.
- Three stop codons are found:
- UAA (ochre),
- UAG (amber), and
- UGA (opal).

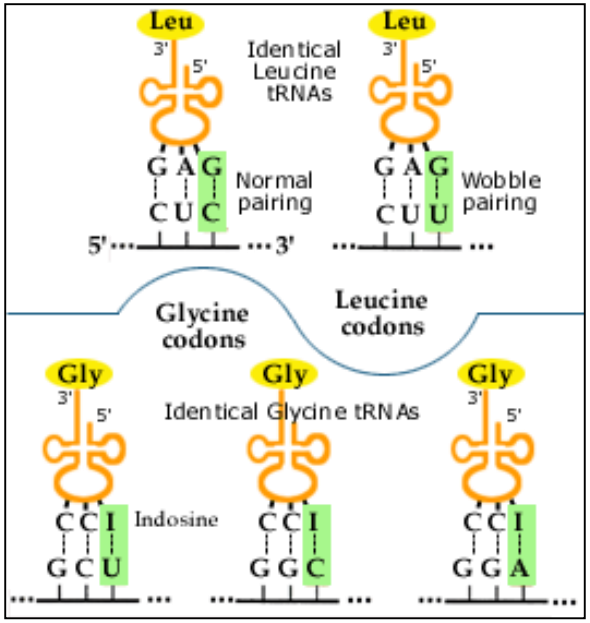
Genetic code

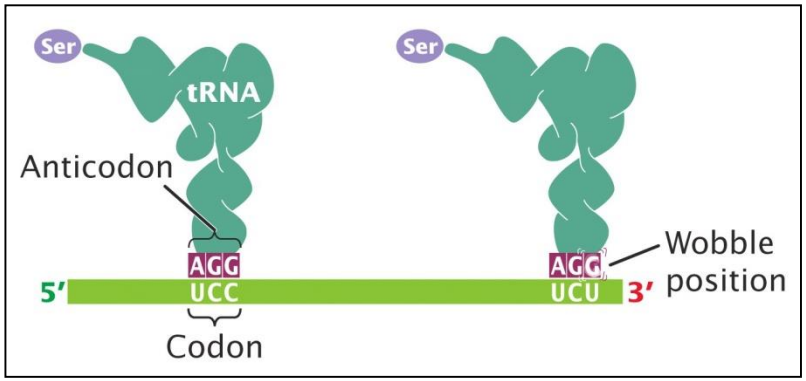
- Each coding triplet is called a codon.
- 61 possible codons:
- 1 Start codon AUG. 3 stop codons, 20 amino acids.



Genetic code
Wobble position

- However, due to the degeneracy of the genetic code, the third base is less discriminatory for the amino acid than the other two bases.
- This third position in the codon is referred to as the wobble position.

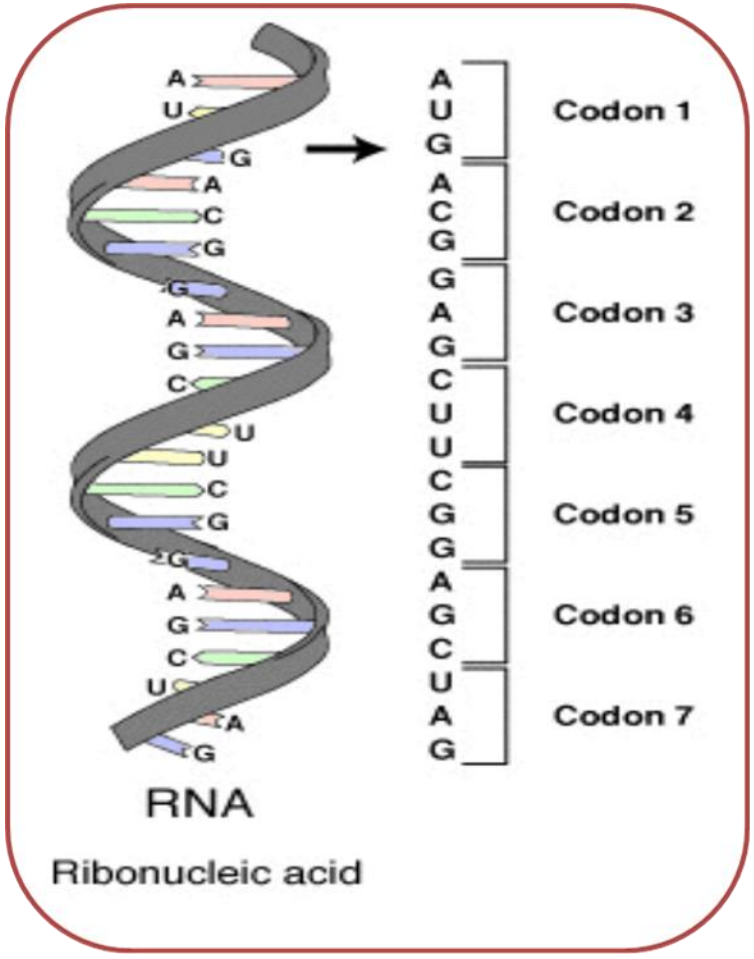




The table shows which codons code for which amino acids

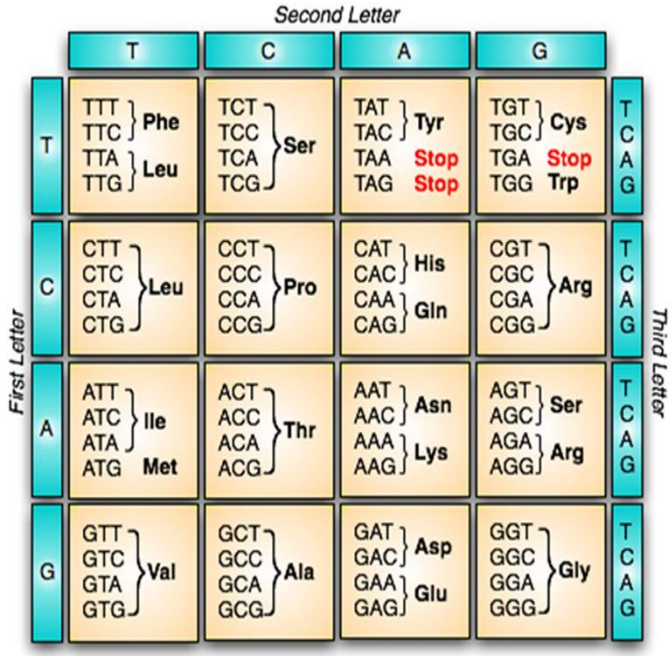
AMINO ACID	RNA CODON
ALANINE	GCC, GCA, GCG, GCU
ARGININE	AGA, AGG, CGU, CGA, CGC, CGG
ASPARAGINE	AAC, AAU
ASPARTIC ACID	GAC, GAU
CYSTEINE	UGC, UGU
GLUTAMIC ACID	GAA, GAG
GLUTAMINE	CAA, CAG
GLYCINE	GGA, GGC, GGG, GGU
HISTIDINE	CAC, CAU
ISOLEUCINE	AUA, AUC, AUU
LEUCINE	UUA, UUG, CUA, CUC, CUG, CUU
LYCINE	AAA, AAG
METHIONINE (INITIATION)	AUG
PHENYLALANINE	UUC, UUU
PROLINE	CCA, CCC, CCG, CCU
SERINE	UCA, UCC, UCG, UCU, AGC, AGU
THREONINE	ACA, ACC, ACG, ACU
TRYPTOPHAN	UGG
TYROSINE	UAC, UAU
VALINE	GUA, GUC, GUG, GUU
STOP	UAA, UAG, UGA

Genetic code
Codons

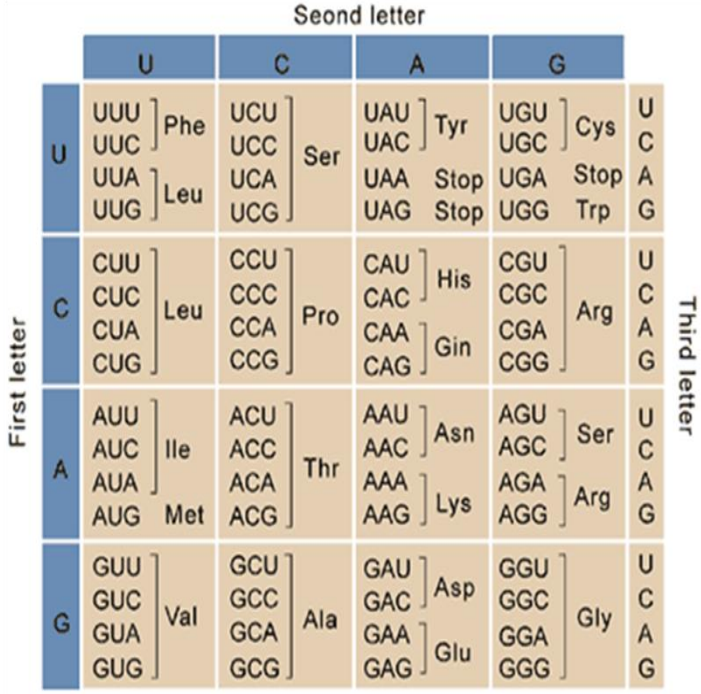


Genetic code

DNA Codon and DNA Codon



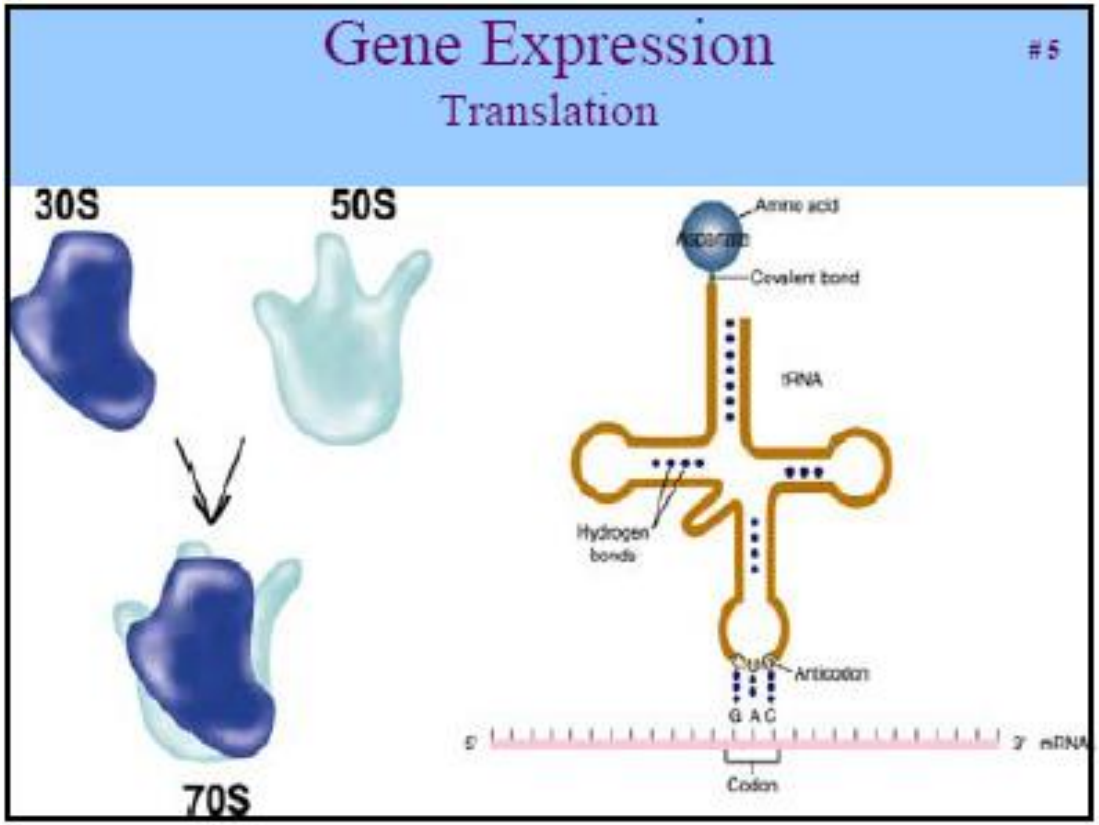
DNA Codon

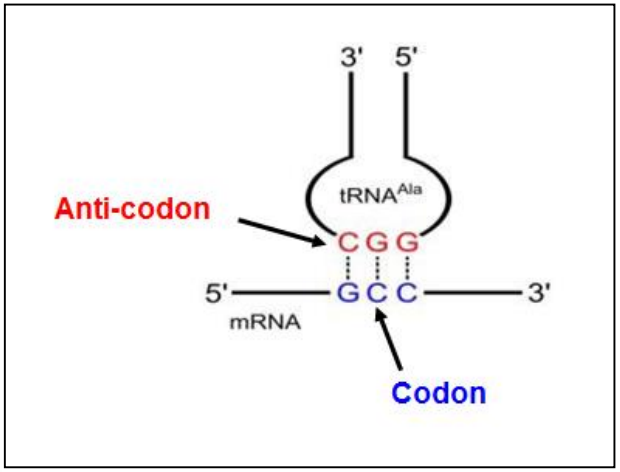


RNA Codon

Translation

Takes place on a ribosome





Protein synthesis

Strands and directions of synthesis

- All strands are synthesized from the 5' ends > > > to the 3' ends for both DNA and RNA.
  Protein chains are synthesized from the amino ends > > > to the carboxy ends.
  Color mnemonic: the old end is the cold end (blue); the new end is the hot end (where new residues are added) (red).

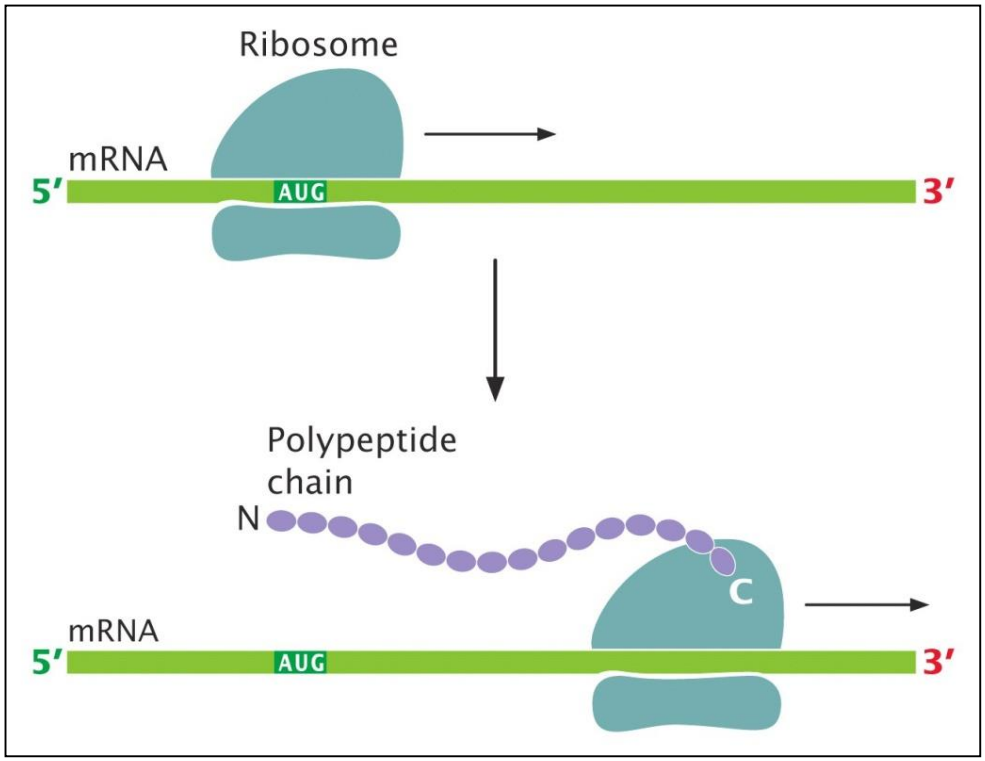
DNA	Coding Strand (Codons)	**5'** > > > - - - - - - T T C - - - - - - > > > **3'**
	Template Strand (Anti-codons)	**3'** < < < - - - - - - A A G - - - - - - < < < **5'**
mRNA	Message (Codons)	**5'** > > > - - - - - - U U C - - - - - - > > > **3'**
tRNA	Transfer (Anti-codons)	**3'** < < < A A G < < < **5'**
Protein	Amino Acid	**Amino** > > > Phenylalanine > > > **Carboxy**

Sandler and Nüsslein

Translation

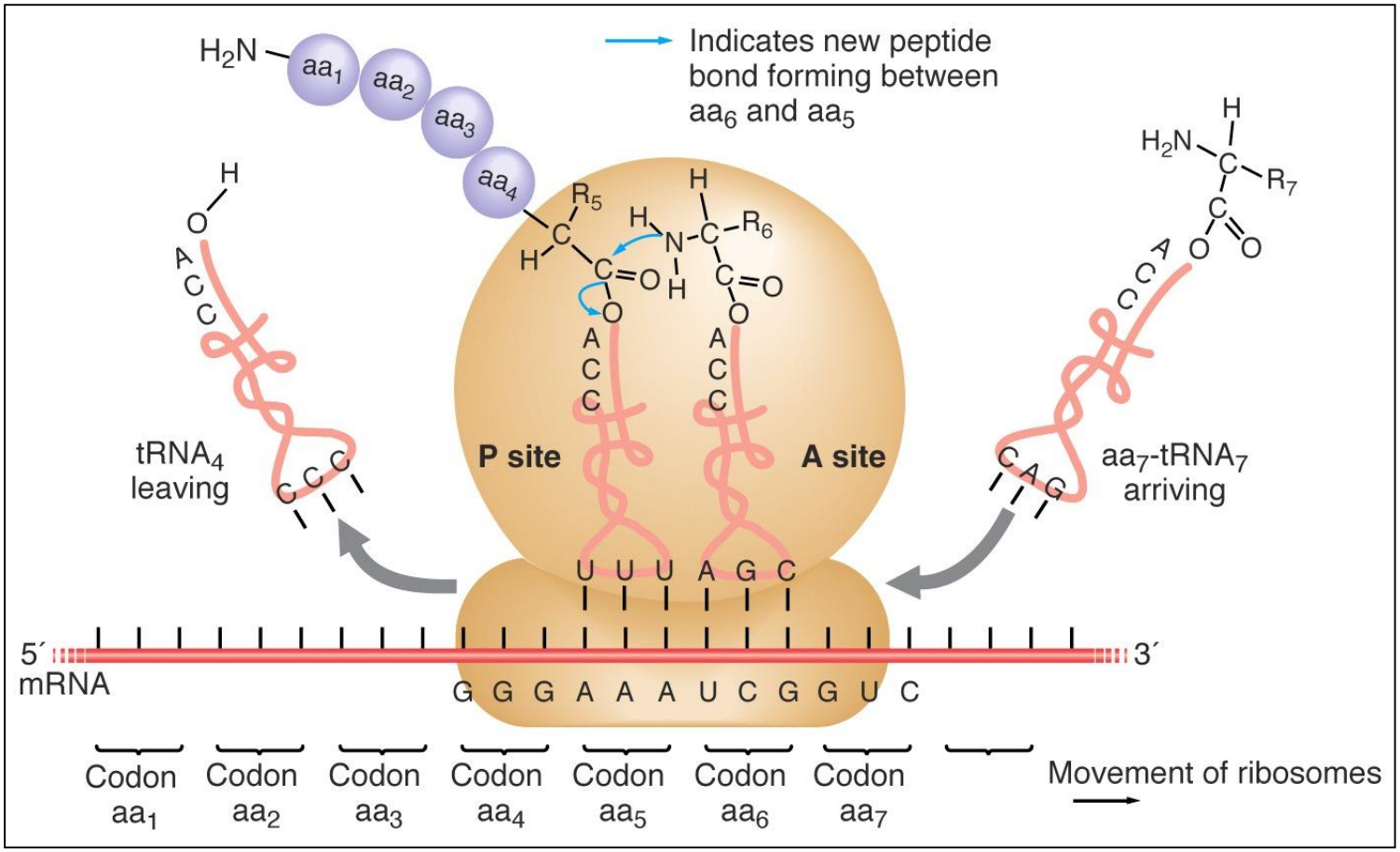
Takes place on a ribosome

- The mRNA and the polypeptide chain are colinear.
- The AUG near the 5′ end of the mRNA encodes the N-terminus of the protein.



Translation

Takes place on a ribosome



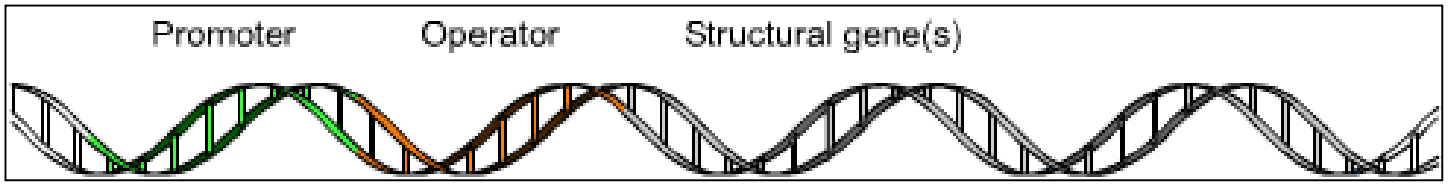
Operon

A group of genes expressed together

A unit of transcription

- The operon is simply defined as a cluster of genes transcribed into a single mRNA molecule.
- Operon is common in bacteria but rare in eukaryotes.
- The general structure of an operon:

1. Structural genes: The genes that are co-regulated by the operon.
2. Promoter: A nucleotide sequence that enables a gene to be transcribed.
3. Operator: A segment of DNA that a repressor binds to.



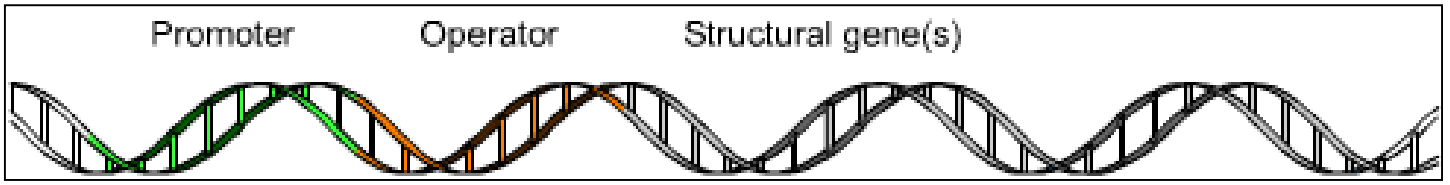
The promoter is recognized by RNA polymerase, which then initiates transcription.

Operon

A unit of transcription

- Close to the promoter lies a section of DNA called an operator.

1. Promoter: is a region of DNA where transcription of a gene initiates.
2. Operator: is a segment of DNA adjacent to a gene; a repressor is a protein that binds to the operator and controls the expression of the gene.
3. Enzymes in many biosynthetic pathways of bacteria and viruses are encoded by operons.

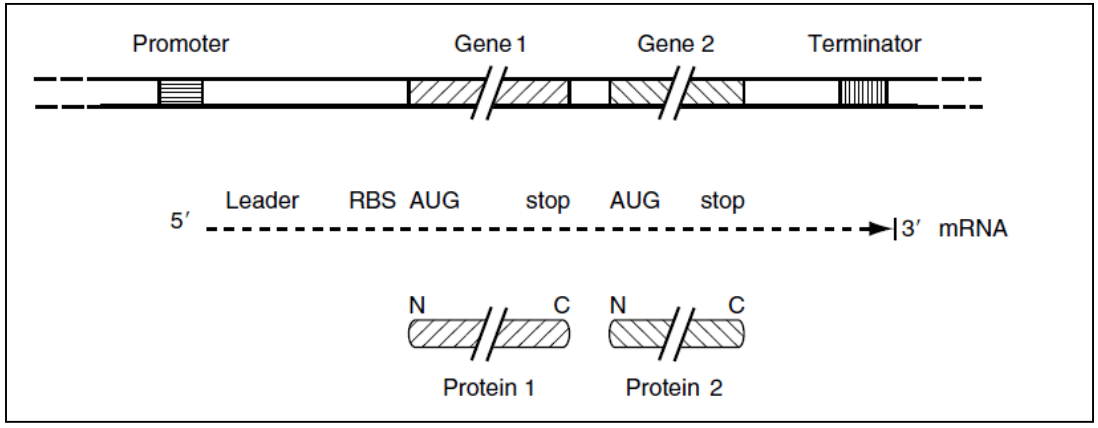


Dale and Park,2004;..

Operon

The general structure of an operon

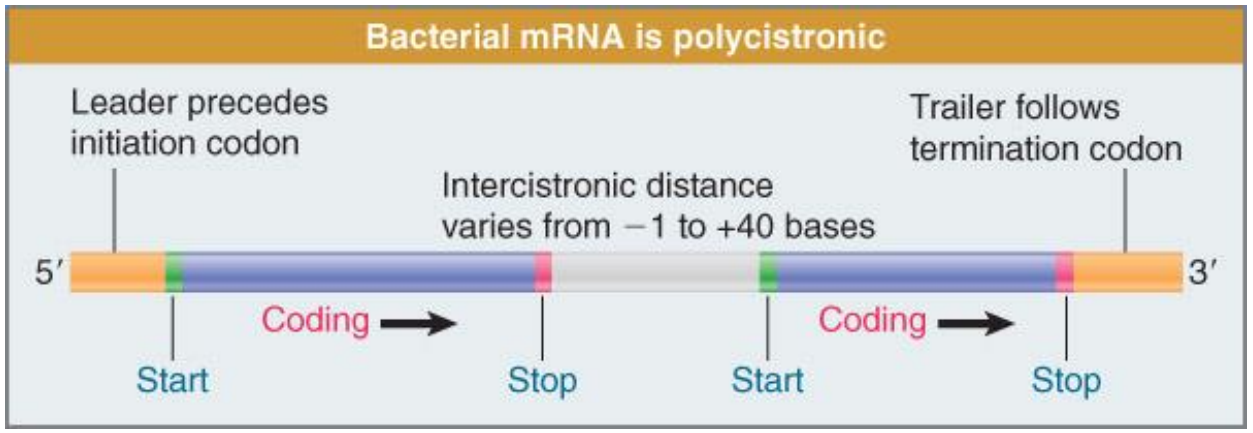
- In genetics, an operon is a functioning unit of genomic material containing a cluster of genes under the control of a single regulatory signal or promoter.
- The genes are transcribed together into a polycistronic mRNA (several strands of mRNA that each encode a single gene product) from which several independent polypeptides can be translated.



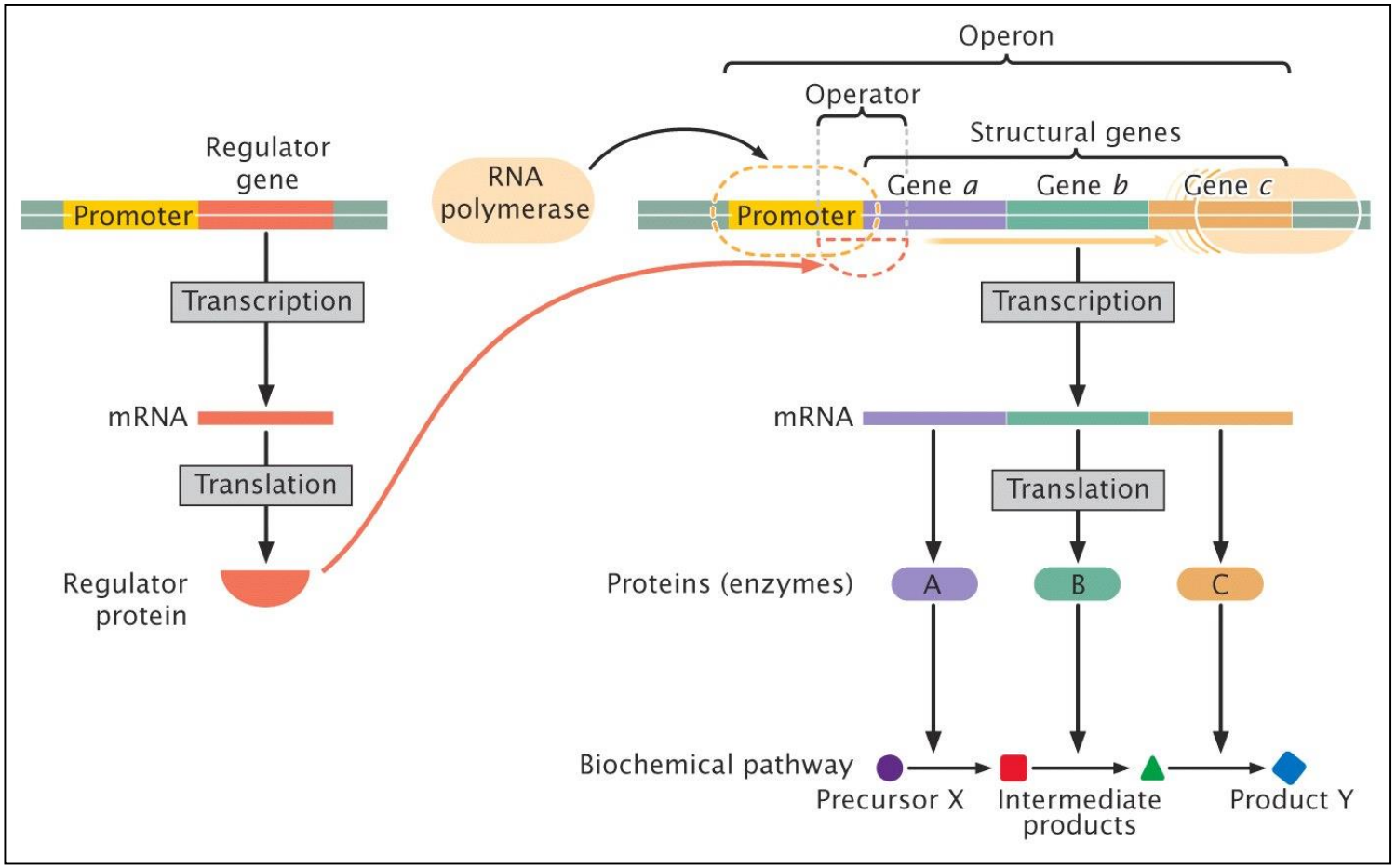
Wikipedia,2010; Dale and Park,2004

Operon
Polycistronic mRNA

- A bacterial mRNA may be polycistronic in having several coding regions that represent different genes.
- A mRNA found in prokaryotes that encodes more than one protein.
- Each eukaryotic mRNA contains information coding for only one protein, hence monocistronic, whereas prokaryotic mRNAs may encode more than one protein and are said to be polycistronic.



Operon
A unit of transcription



Operon

lac operon- a typical operon

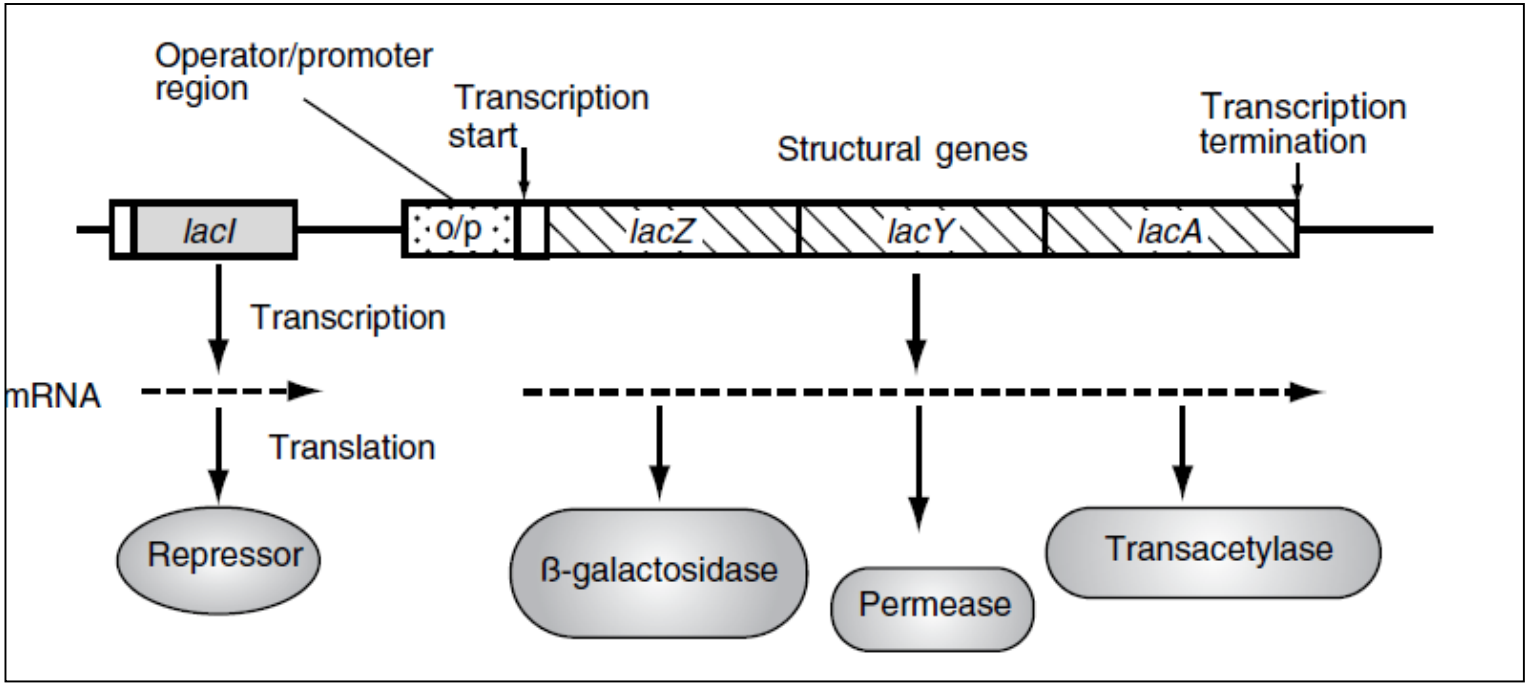
- An operon e.g. *lac* operon consists of:

1. Few regulated genes(*lac* Z, *lac* Y and *lac* A),
2. A promoter,
3. An operator, and
4. The regulator gene encoding a repressor.

- The transcription of each operon is initiated at a promoter region and controlled by a neighbouring regulatory gene.
- This regulatory gene specifies a regulatory protein (a repressor or apoinducer) that binds to the operator sequence in the operon to either prevent or allow its transcription.

Operon

Structure of the *lac* operon



Dale and Park,2004

Operon
Repression
Control of gene expression

- Control of an operon is a type of gene regulation that enables organisms to regulate the expression of various genes depending on environmental conditions.
- Operon regulation can be either negative or positive by induction or repression.

1. With positive control, an activator protein stimulates transcription by binding to DNA (usually at a site other than the operator).
2. Negative control involves the binding of a repressor to the operator to prevent transcription.

Wikipedia,2010

Gene regulation
Positive vs. negative regulation

- All genes are not always being expressed.
- Genes turned on all the time – Constitutive.
- Other genes can be regulated:

1. Turned On
2. Turned Off

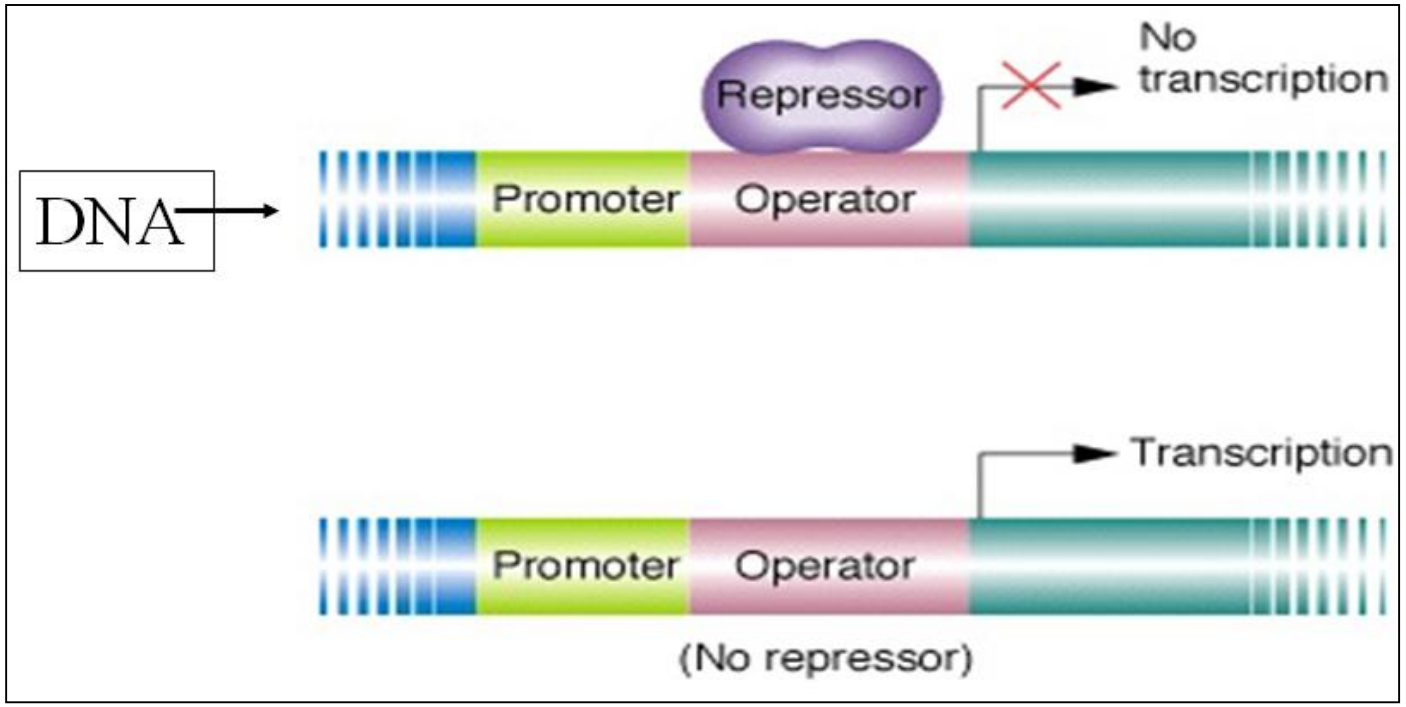
- Gene is turned/expressed, or turned off/silenced.

1. It can be a positive regulation in which transcription factors bind to the promoter and enhancer ... and therefore turn on its expression, or
2. It could also could be a negative regulation in which the gene is repressed/silenced.

Operon

Repression

Control of gene expression



Repressor (repressor protein) binds to the operator therefore transcription is inhibited.

Proteins
Ribosomal proteins

- Proteins:
- Composed of units of about 20 different amino acids, which, in turn, are composed of carbon, hydrogen, oxygen, nitrogen, and sometimes sulfur.
- In a protein molecule these acids form peptide bonds-bonds between amino and carboxyl (COOH) groups-in long strands (polypeptide chains).
- Ribosomal proteins:
- Proteins found in ribosomes.
- They are believed to have a catalytic function in reconstituting biologically active ribosomal subunits.

Proteins
Non-ribosomal proteins

- Nonribosomal peptide synthetases (NRPSs) are multimodular enzymes, found in fungi and bacteria, which biosynthesize peptides without the aid of ribosomes.
- Some of these metabolite products are:

1. Peptide antibiotics such as the gramicidins , polymyxins, bacitracins, glycopeptides, etc;
2. Siderophores,
3. Toxins of syringomycin and syringopeptin.

Proteins
Non-ribosomal proteins

- Nonribosomal peptides are a very diverse family of natural products with an extremely broad range of biological activities (e.g. virulence factors) and pharmacological properties.
- They are often:

1. Toxins,
2. Siderophores, or
3. Pigments.

- Nonribosomal peptide antibiotics are in commercial use.

Wikipedia,2012;..

Proteins
Non-ribosomal proteins

- Nonribosomal peptides (NRP) are a class of peptide secondary metabolites synthesized by one or more specialized nonribosomal peptide-synthetase (NRPS) enzymes.
- The NRPS genes for a certain peptide are usually organized in one operon in bacteria and in gene clusters in eukaryotes.

Bushley and Turgeon,2010;Wikipedia,2012

Proteins
Non-ribosomal proteins

- Nonribosomal peptides:

1. Are not synthesized by ribosome,
2. Unlike the ribosomes, are independent of messenger RNA(mRNA).

- Each nonribosomal peptide synthetase can synthesize only one type of peptide.

Bushley and Turgeon,2010;Wikipedia,2012

Proteins
Non-ribosomal proteins

1. Nonribosomal peptides often have a cyclic and/or branched structures,
2. Can contain non-proteinogenic amino acids including D-amino acids.

- Thus:

1. Nonribosomal peptides can include both D and L forms of the 20 amino acids used in ribosomal protein synthesis, as well as
2. Non-proteinogenic amino acids such as ornithine, imino acids, and hydroxy acids such as α-aminoadipic and α-butyric acids.

Bushley and Turgeon,2010;Wikipedia,2012

Proteins
Non-ribosomal proteins

- Nonribosomal peptide enzymes (NRPSs) use a set of core domains, known as a module, to accomplish peptide synthesis.
- A minimal module consists of three core domains:

1. Initiation or starting module;
2. Elongation or extending modules, and
3. Termination or releasing module.

Bushley and Turgeon,2010;Wikipedia,2012

Proteins

Non-ribosomal proteins

A-T-C module

- The three core domains of a module are:

1. An adenylation (A) domain which recognizes and activates the substrate via adenylation with ATP;
2. A thiolation (T) or peptidyl carrier protein (PCP) domain which binds the activated substrate to a 4'-phosphopantetheine (PP) cofactor via a thioester bond and transfers the substrate to a condensation domain(C);
3. Condensation (C) domain which catalyzes peptide bond formation (amide bond) between adjacent substrates on the megasynthase complex.

Bushley and Turgeon,2010

Proteins

Non-ribosomal proteins

Single or repeated A-T-C modules

- NRPSs may be:

1. Monomodular, consisting of a single A-T-C module, or
2. Multimodular, consisting of repeated A-T-C modules.

Bushley and Turgeon,2010

Gene expression

What is meant by gene expression?

Gene expression efficiency

- Gene expression refers to genes being 'turned on' and producing a product.
- The product could be an enzyme, a structural protein, or a control molecule.
- Studies of gene expression typically measure the production of mRNA.
- Most mechanisms that control gene expression do so by controlling transcription, the synthesis of mRNA.

Gene expression in prokaryotes
Gene expression efficiency

- Prokaryotes only transcribe genes that their end-proteins are needed at the time.
- They do this in order to save up energy and increase efficiency.
- The regulation of gene expression is depended mainly on their immediate environment, for example on the presence and absence of nutrients.
- Gene expression in prokaryotes occurs primarily at the level of transcription.

Gene expression in prokaryotes
Gene expression efficiency

- Genes can be expressed with different efficiencies.
- Gene A is transcribed and translated much more efficiently than gene B.
- This allows the amount of protein A in the cell to be much greater than that of protein B.

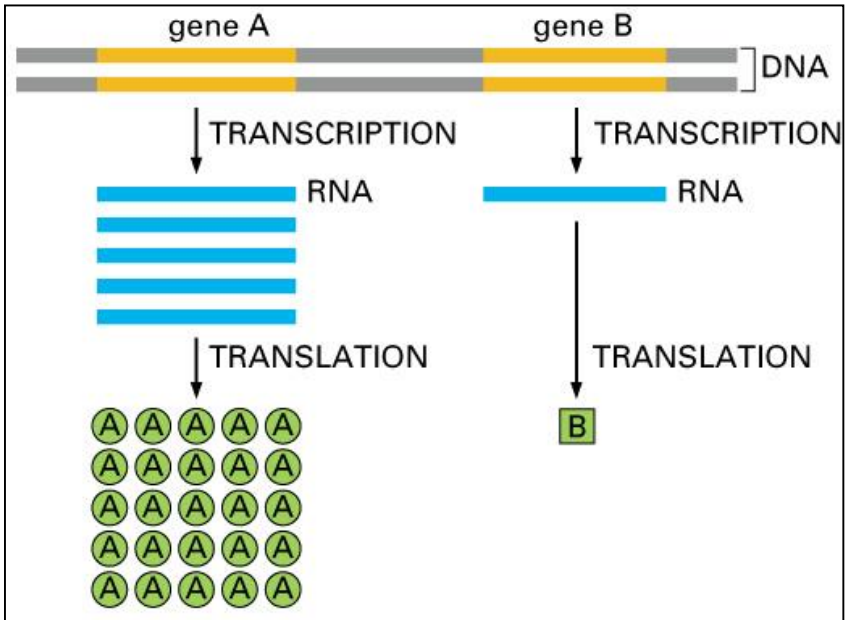


Figure 6–3. Molecular Biology of the Cell, 4th Edition.

2. Mechanisms of gene transfer

- The three forms of bacterial DNA exchange (genetic transfer) lead genetic recombination:

1. Transformation (competence);
2. Conjugation (cell-cell contact);
3. Transduction (transducting particles/lysates).

Plasmid replication and copy number was also considered as a tool for exchange of genetic material (See also Part II- Plasmid vectors section).

Bacterial Exchange of DNA

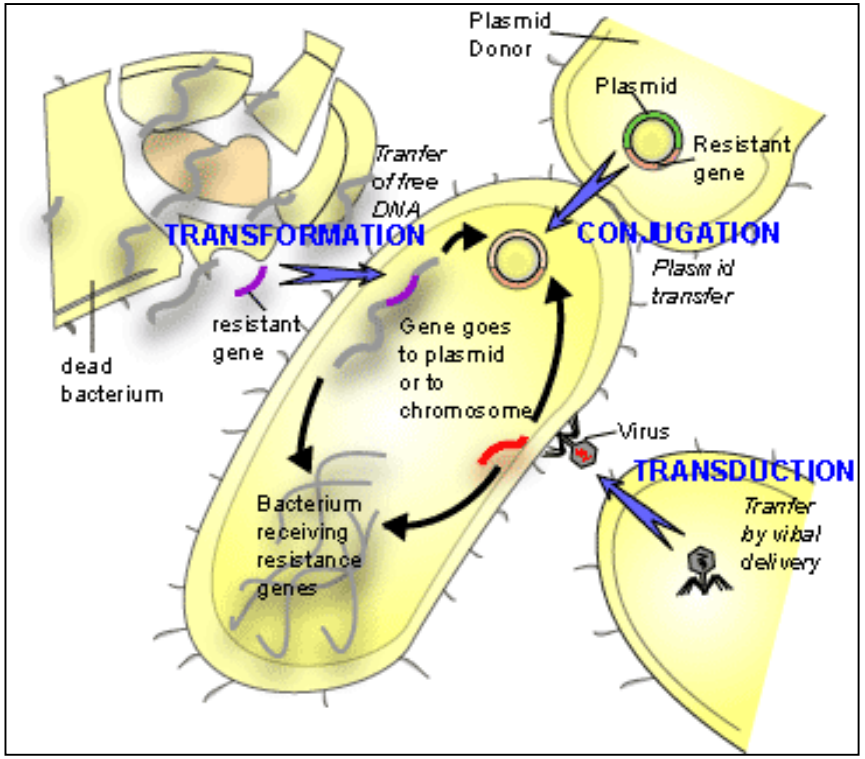
- Most bacteria are haploid which means that there is no such thing as dominance-recessive relationships among bacterial alleles
- Bacteria don't have sex in the animal/plant sense of sex (i.e., mating followed by recombination of whole genomes).
- Instead, bacteria acquire DNA from other bacteria through three distinct mechanisms:
  1. Transduction
  2. Transformation
  3. Conjugation
- This DNA may or may not then recombine into the recipient's genome.
- We use phrases like "Lateral" or "Horizontal" Gene Transfer to describe these sexual interactions
- Bacterial DNA is also subject to:
  5. Mutation
  6. Damage (not the same thing as mutation)
  7. Natural selection

Allele: One of a pair of genes at a single locus.

Lateral or Horizontal Gene Transfer (HGT)

- Lateral or horizontal gene transfer (HGT) is a process whereby genetic material contained in small packets of DNA can be transferred between individual bacteria.
- There are three possible mechanisms of HGT.
- These are:

1. Transduction,
2. Transformation, or
3. Conjugation.



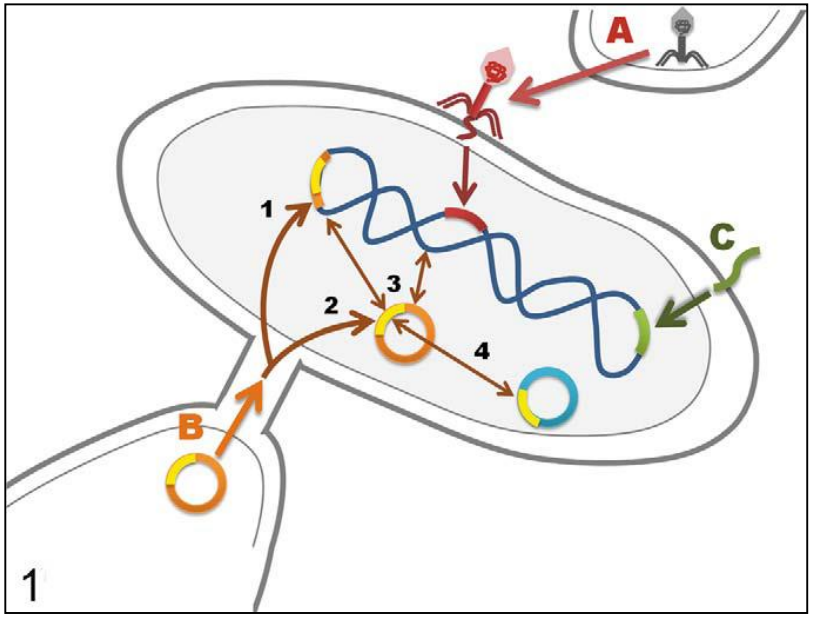
The bacterial plasmids

Mechanisms of DNA transfer between and within bacteria

A. Transduction: injection of DNA into a bacterium by a phage.

B. Conjugation: plasmid in a donor bacterium is transferred through a pilus into a recipient bacterium; plasmid may integrate into the chromosome (1) or remain in the cytoplasm (2); plasmid may be transferred between cytoplasmic and chromosomal locations (3); plasmid may exchange insertion sequences or transposons with other plasmids (4) or the chromosome.

C. Transformation: uptake of naked DNA from the environment.



Gyles and Boerlin,2014

Horizontal gene transfer vs. vertical transfer

1. **Lateral gene transfer (LGT) or Horizontal gene transfer (HGT):** any process in which an organism gets genetic material from another organism without being the offspring of that organism.
2. **Vertical transfer:** occurs when an organism gets genetic material from its ancestor, e.g., its parent or a species from which it has evolved.

Wikipedia,2012

Horizontal gene transfer

Known instances of HGT

- Antibiotic resistance genes on plasmids
- Insertion sequences
- Pathogenicity islands
- Toxin resistance genes on plasmids
- *Agrobacterium* Ti plasmid
- Viruses and viroids
- Organelle to nucleus transfers.

Horizontal gene transfer

How can you detect HGT's?

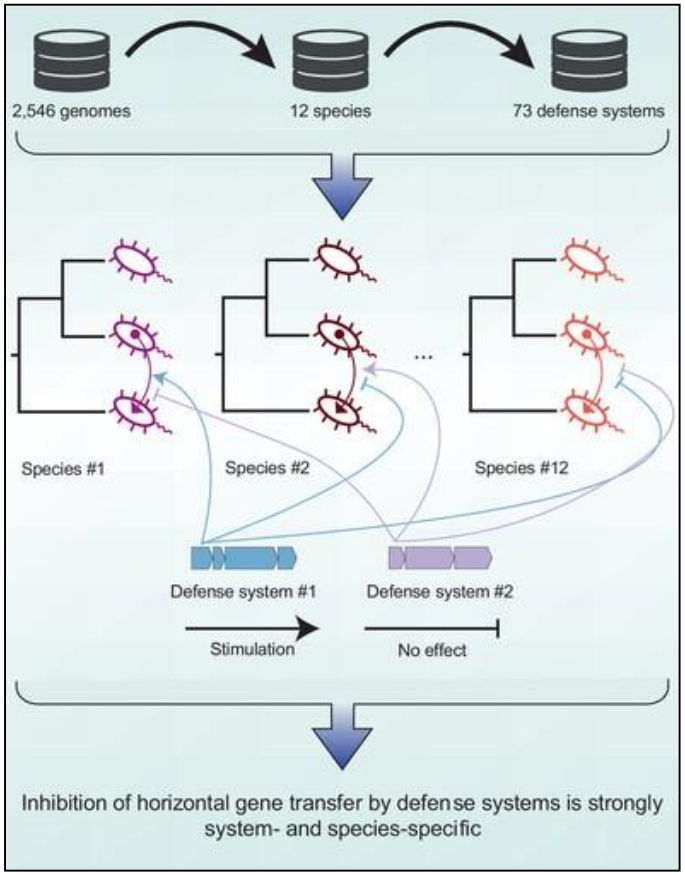
- DNA sequence information:
  - Phylogenetic trees
  - G+C Content
  - Codon bias
  - Sequences new to a genome will retain (for a while) the signatures of the donor genome and distinguished from ancestral DNA.

How can you detect HGT's?
G+C Content

- DNA is double stranded, G pairing with C.
- Measure the amount of G+C content in regions.
- If one region varies from most of the genome, than likely HGT.

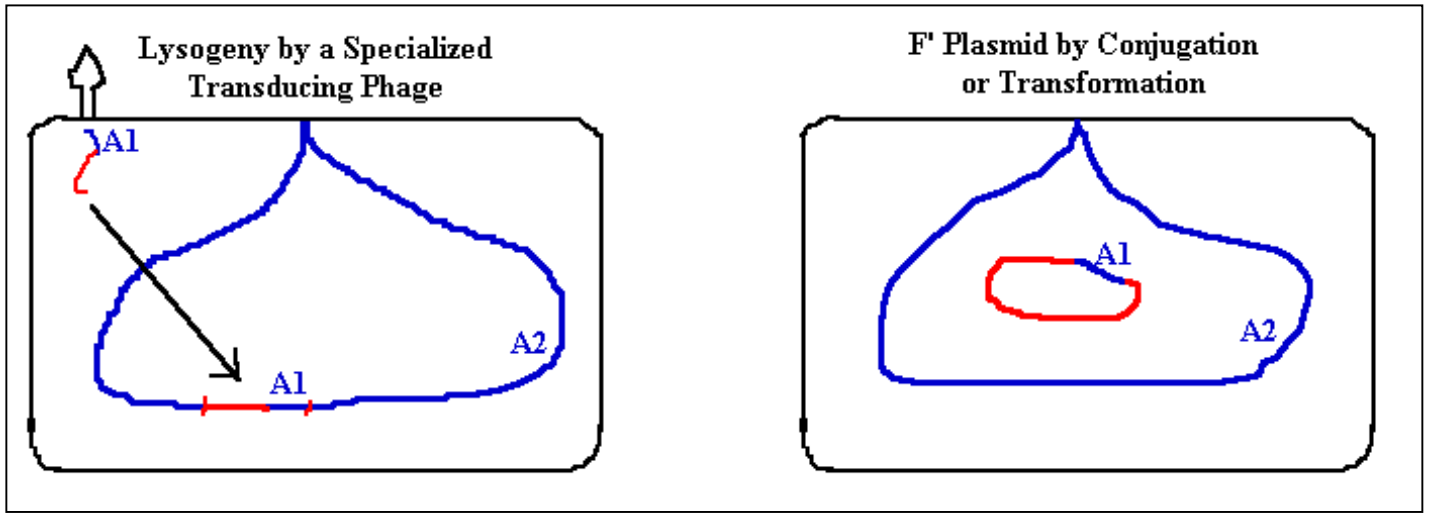
Defence systems and horizontal gene transfer in bacteria

- Horizontal gene transfer (HGT) plays a crucial role in the evolution of prokaryotes.
- This process is facilitated by mobile genetic elements (MGEs) but can be hindered by the activity of anti-MGE defence systems.
- Our research demonstrates that only a specific subset of these defence systems in certain bacterial taxa substantially reduces HGT.



Kogay *et al.*,2024

Comparison of
Transformation & Transduction



Two processes to create a partial diploid in bacteria are **specialized transduction** and **F' conjugation.**

1. Transformation

Competence

Bacterial transformation may be referred to as a stable genetic change brought about by the uptake of naked DNA (DNA without associated cells or proteins) to increase DNA quantity; Transformation will happen when bacteria are in a state of competence. i.e. in environmental conditions such as starvation and cell density.

Transformation

In bacteria and yeasts

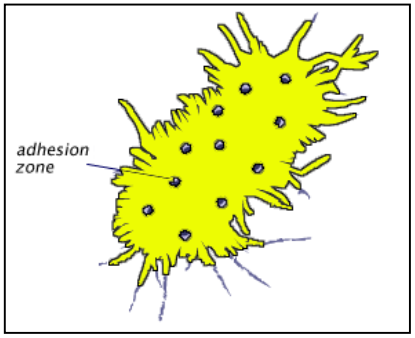
No need for cell to cell contact

- Transformation is the transfer of genetic information from a donor to a recipient using naked DNA without cell to cell contact.
- DNA uptake that facilitates recombination between strains.
- The recipient takes this DNA up.
- There is no requirement for cell to cell contact.
- The only limitation of this uptake is the restriction system.
- Bacteria use restriction enzymes (restriction endonucleases) to destroy the foreign DNA.

In eucaryotes other than yeast, transformation is called transfection.

Transformation
Competency

- Transformation is used to move DNA between bacteria, plants and animals.
- In bacteria, competency is an empirical matter.
- During rapid growth of *E. coli*, the cell membrane has hundreds of pores, called adhesion zones.
- **Transformation is made during this rapid growth, i.e. when these pores are present.**

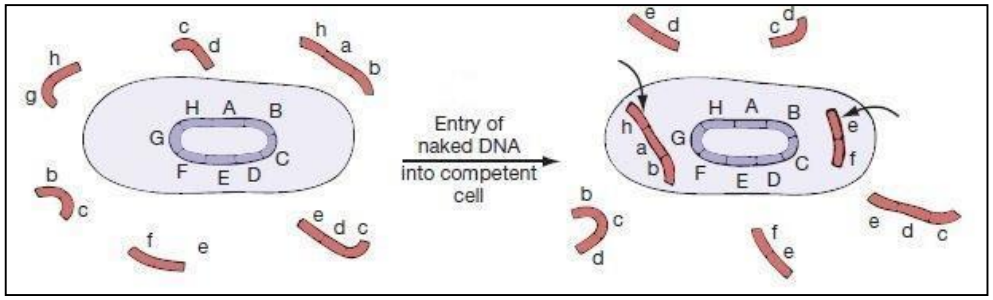


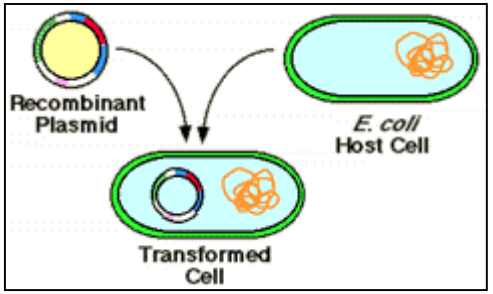
Transformation

- During this process, genes are transferred from one bacterium to another as naked DNA in solution.
- This process was first demonstrated in *Streptococcus pneumoniae*, and occurs naturally among a few genera of bacteria.
- This was first discovered in 1928 by Frederick Griffith.

Transformation

- It is the transfer of “naked” DNA after cell lysis.
- Free DNA may become available from dying or disintegrating bacteria.

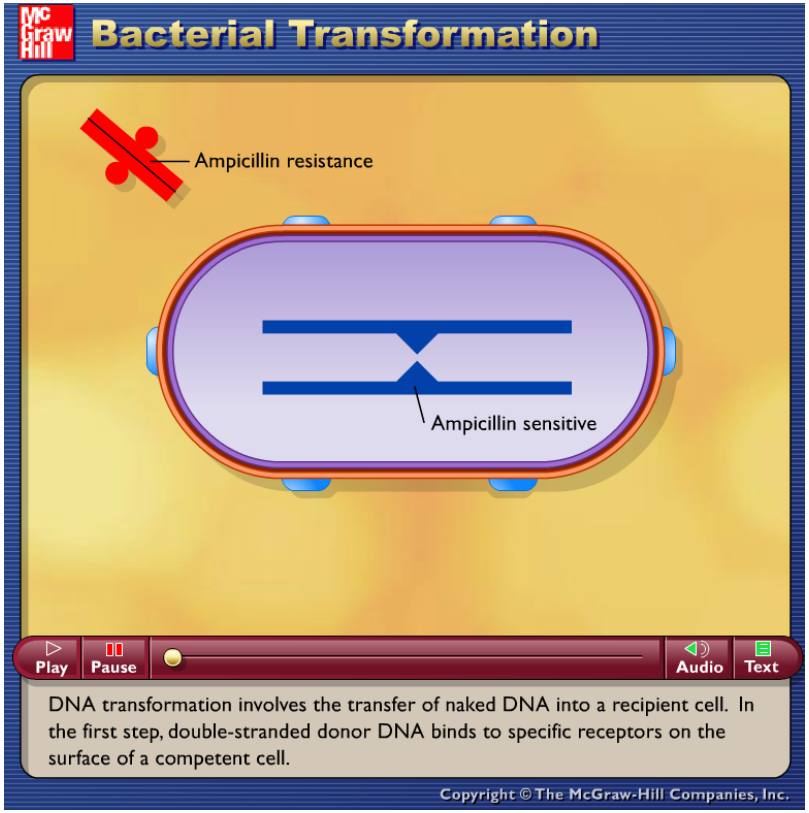




Transformation

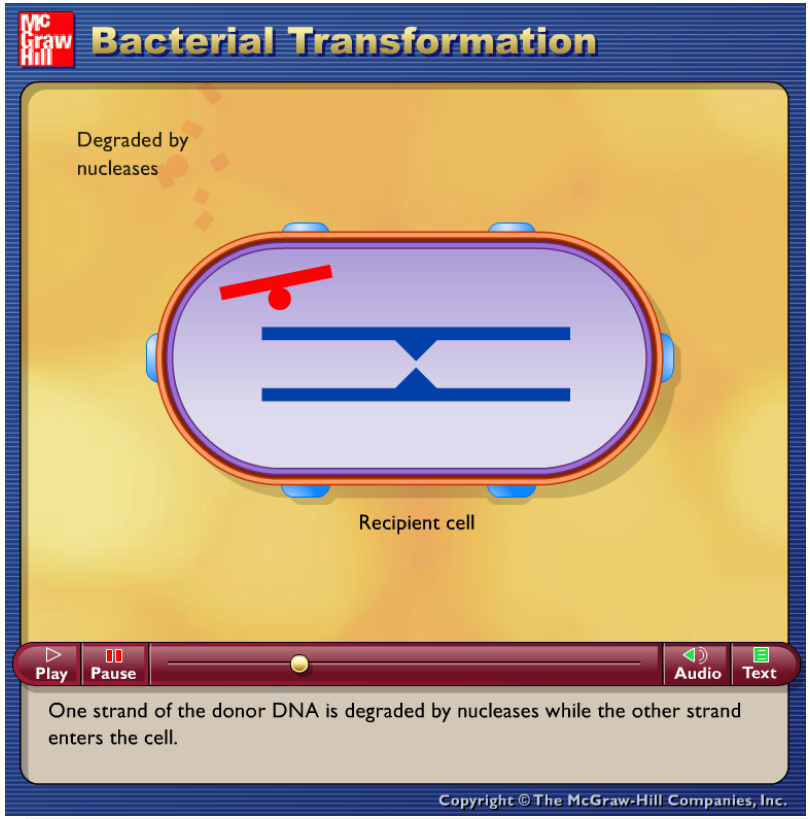
Ampicillin resistance gene transformation

- DNA from an ampicillin resistant bacterium is brought into an ampicillin sensitive organism by transformation.
- This involves the transfer of naked DNA into recipient cell.
- In the first step, double-strand DNA binds to specific on the surface of the competence cell.

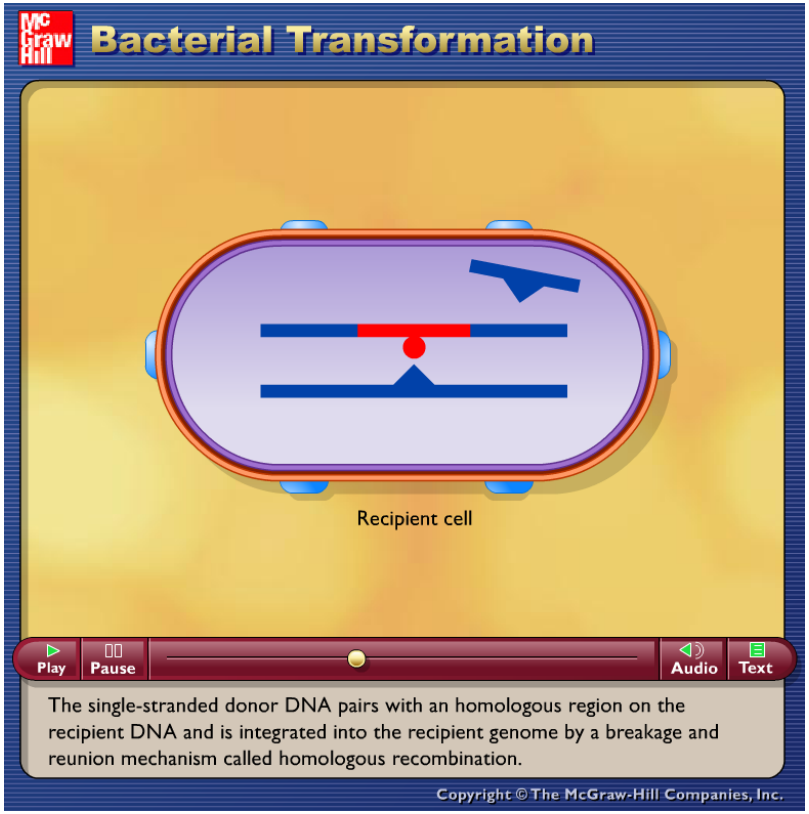


Transformation

Only single strand of donor DNA enter the recipient cell and the other is degraded by nuclease



One strand of the donor DNA is degraded by nuclease while the other strand enters the cell.

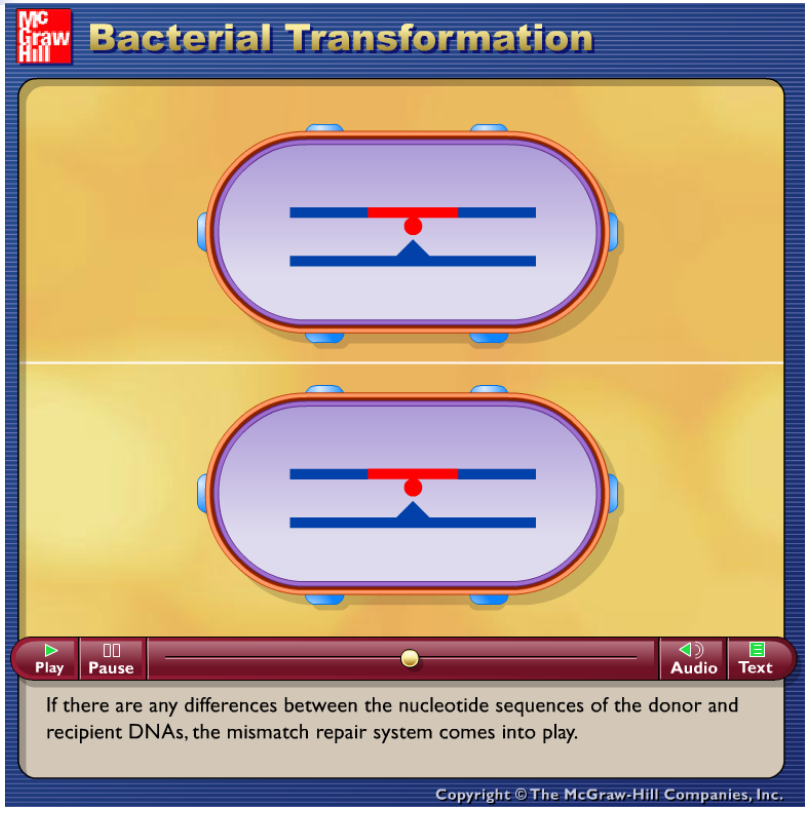


The single stranded donor DNA pair with homolgous region on recipient DNA and integrated into recipient genome.

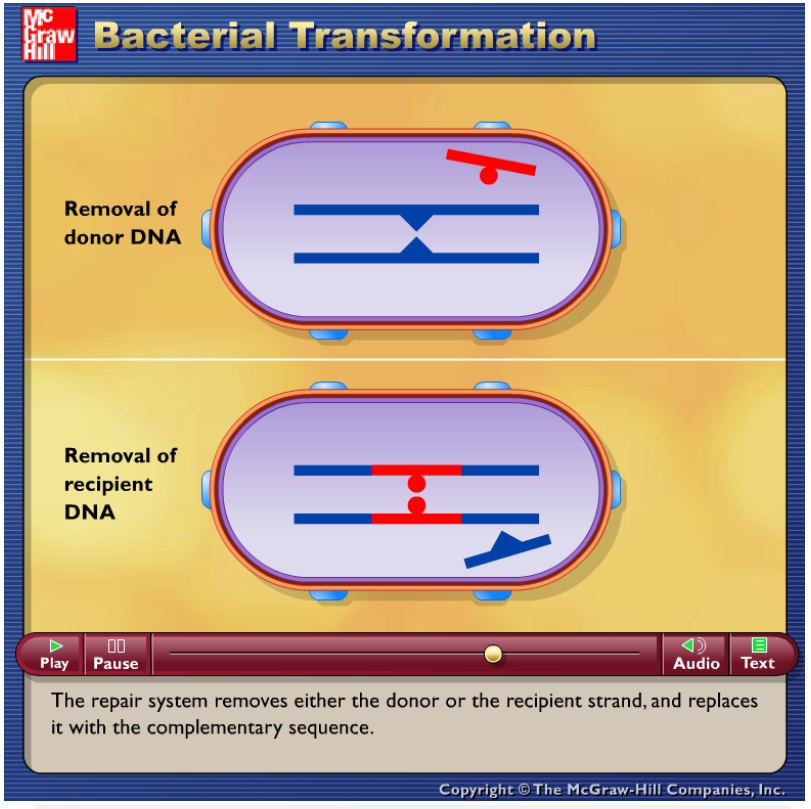
Transformation

Competency

The mismatch repair system will repair either the donor or the recipient strand



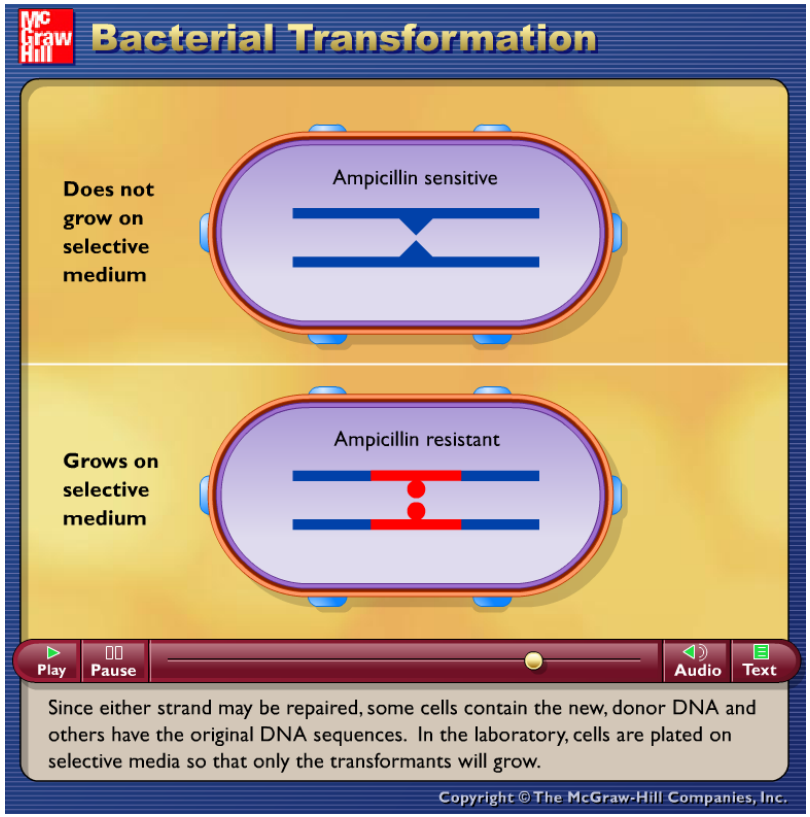
Mismatch sequence in sequence of recipient DNA. If there is any differences between the nucleotide sequences of donor and recipient DNAs, the mismatch repair system comes into play.

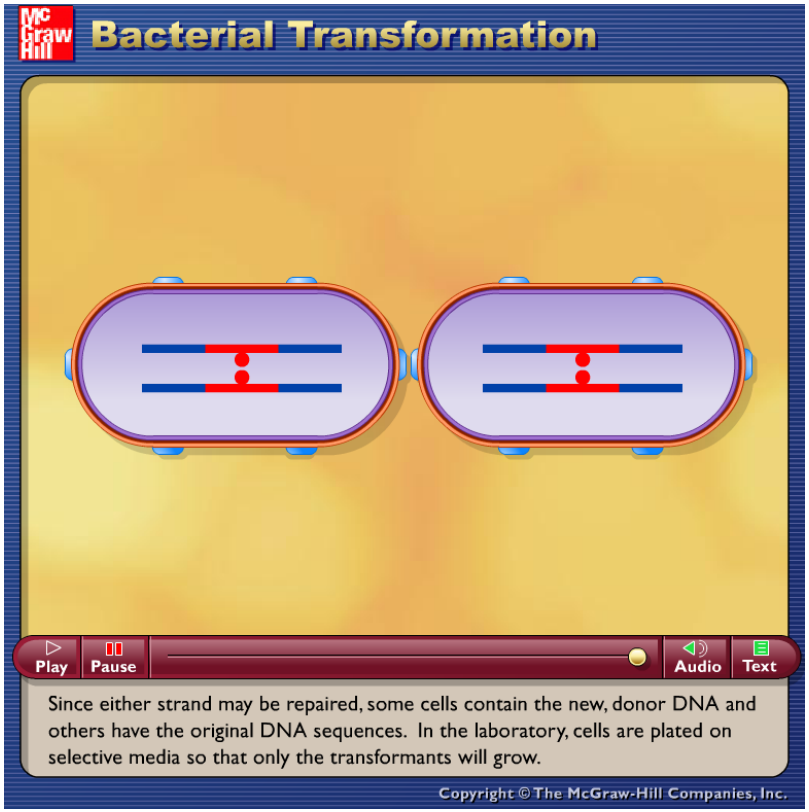


Mismatch sequence of donor (top) and recipient DNA (below) are removed.
The mismatch repair system removes either the donor or the recipient strand, and replaces it with complementary sequences.

Transformation

When DNA from an ampicillin resistant organism is transformed into an ampicillin sensitive organism, only the transformed cells will grow on selective media containing ampicillin





Only the transformed cells will grow on selective media containing ampicillin.
Ampicillin sensitive cells will not grow.

Transformation
Natural and experimental

1. Natural transformation (rare but dangerous):

- Many bacteria can acquire new genes by taking up DNA molecules (e.g., a plasmid) from their surrounding. While naturally transformable bacterial strains exist, these are sufficiently rare that induced transformation is more important to the geneticist.
- A captured DNA has to face restriction!

2. Experimental transformation:

- In the lab, DNA is introduced by one of the following methods:

I. Calcium chloride (followed by heat shock) method;

II. Electroporation;

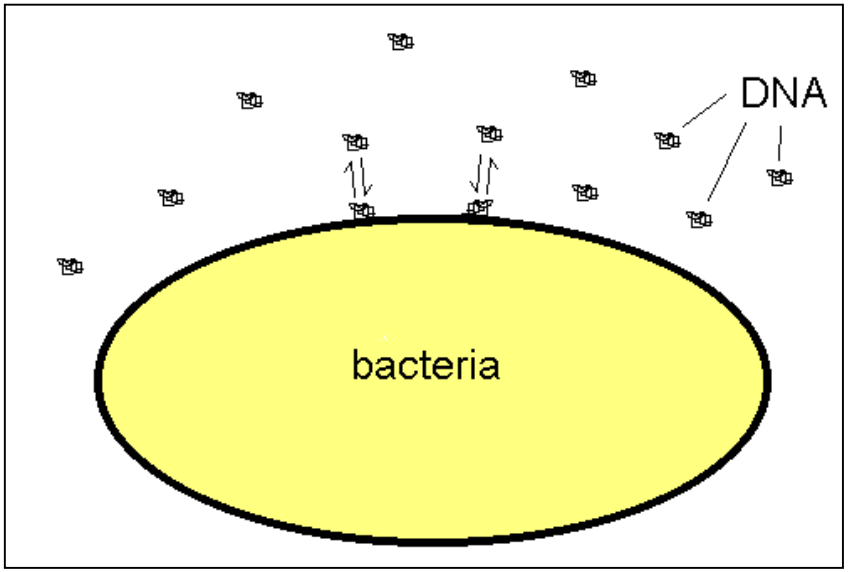
III. Liposomes.

Transformation

Artificial transformation

Calcium chloride method

- Some bacteria have specialized membrane proteins to bring DNA into their cells.
- Ca stimulates uptake of DNA into bacteria.



During the incubation on ice, DNA binds to the surface of the bacterium as a calcium-phosphate-DNA complex.

Artificial competency is typically induced using chemicals such as calcium chloride to increase the permeability of the cells.

Transformation

Artificial transformation

Calcium chloride method followed by heat shock

- **This step will be performed for you before you come into lab:**

1. Pick a single colony from a freshly grown plate of *E. coli* and disperse it in 100 ml of LB media in a 1 L flask. Incubate the culture at 37°C with vigorous shaking for approximately 3 hours. Cell density is monitored by determining $OD6_{00}$ and should be less than 10^8 cells/ml (log phase of growth - the most healthy bacteria).
2. Transfer 50 ml of this culture to a 50 ml conical tube and centrifuge at 2,000 rpm for 10 min.
3. Decant the supernatant into waste beaker (this must be sterilized before being dumped down the drain). Resuspend the pellet in 10 ml of ice cold 0.1 M $CaCl_2$. This is most easily done by resuspending in 1 ml, using the P1000 pipette and then adding 9.0 ml. Cut about 0.5 cm from the end of the blue tip before pulling *E. coli* through it, since the cells are fragile in this high calcium solution and may lyse if sheared. After they have been resuspended, centrifuge at 2,000 rpm for 10 min. (Sorvall HS-4 rotor).
4. Decant the supernatant into waste beaker (to be autoclaved later). Resuspend the pellet in 1.0 ml of ice cold 0.1 M $CaCl_2$.

Transformation

Artificial transformation

Calcium chloride method followed by heat shock

- **Bacterial Transformation:**

1. Pipette 200µl competent cells into each of 3 ice cold Eppendorf tubes. Label the tubes: Control, 1 ng, and 10 ng (1 ng is 10^{-3} ug, or 10^{-9} mg). The unknown plasmid is at a concentration of 1 ng/µl. A plasmid containing resistance to an antibiotic (usually ampicillin) is used as a vector. Add 1 ng of your unknown plasmid to one tube and 10 ng to the other. Place the tubes on ice for 30 min.
2. Put the tubes at 42°C for exactly 90 seconds. Return the cells to ice for 1-2 minutes.
3. Pipette the transformation mixtures onto labeled plates containing ampicillin and spread them around using a sterilized, bent glass rod spreader.
4. Place upside down in the 37°C incubator overnight.
5. 16-20 hours later, count the number of colonies on the plate with well-isolated colonies. Put parafilm around the edge of a plate and put it in a refrigerator for later use. Check the control plate to see that no colonies grew on it. Dispose of the plate and the control plate in the biohazard bag.

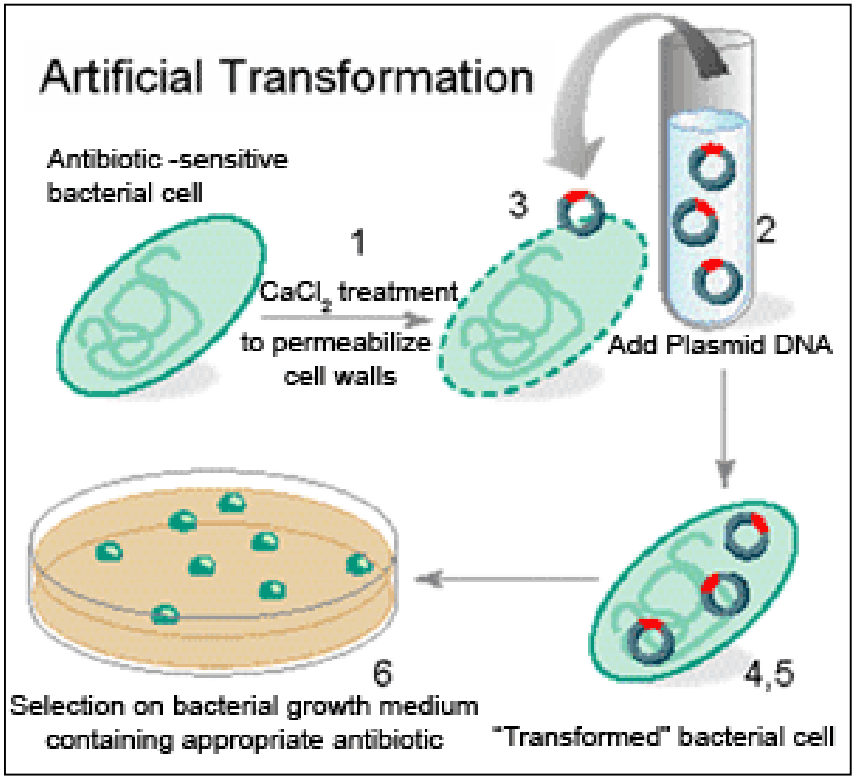
Determine the transformation efficiency the next day by counting the number of colonies. To calculate the competency of your cells, divide the number of colonies on your plate by the amount of DNA (in ug) you added to the transformation.

Transformation

Artificial transformation

Calcium chloride method followed by heat shock

1. Antibiotic sensitive bacterial cells e.g. *E.coli* have been treated with $CaCl_2$.
2. In a microtube, gently mix the plasmid DNA with the bacterium.
3. Incubate the mixture on ice for 20 minutes. Put the tubes at 42°C for exactly 90 seconds.
4. Transfer the bacterial suspension to LB agar with appropriate antibiotic.



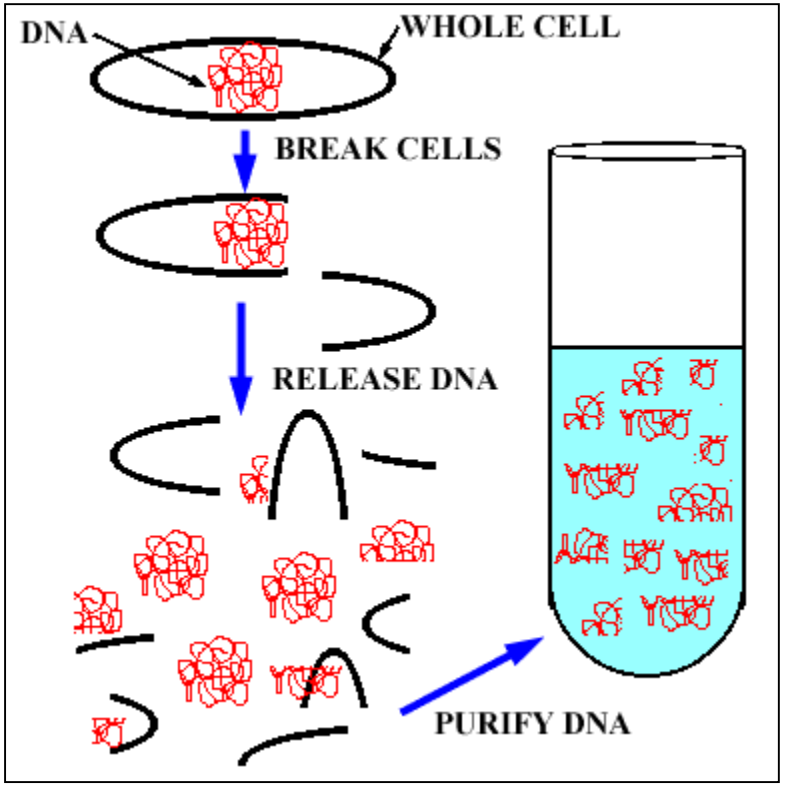
Note: In frozen state bacterial cell are not very "happy" = fragile.

Transformation

Transformation procedure

1. Isolation of cell-free or naked DNA

- The cells are broken and the DNA released.
- The cell-free DNA (Naked DNA), usually purified is subsequently isolated and collected.

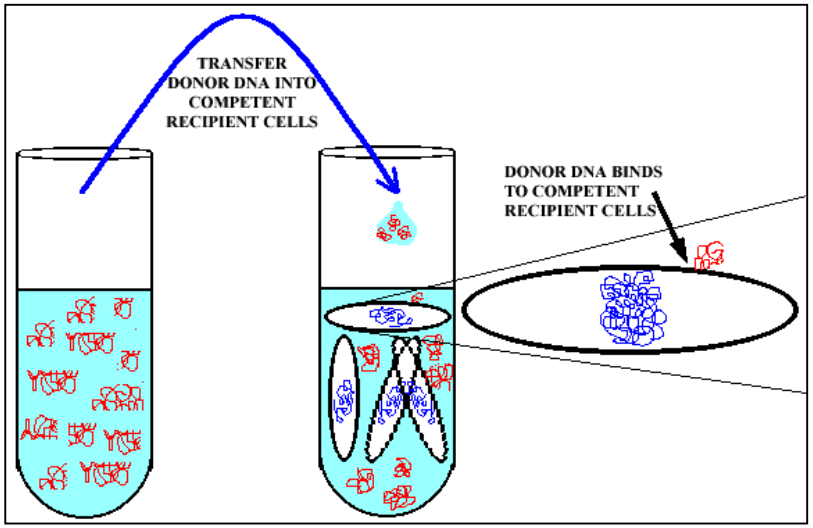


Transformation

Transformation procedure

2. Adding the naked DNA to recipient cells

- The naked donor DNA is incubated with the competent recipient cells to which it binds.

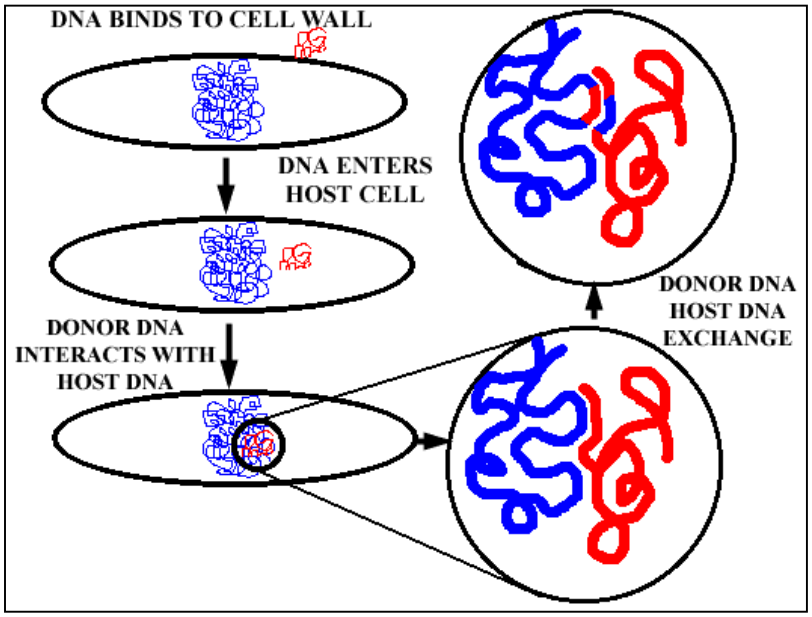


Transformation

Transformation procedure

3. Uptake and recombination of donor DNA

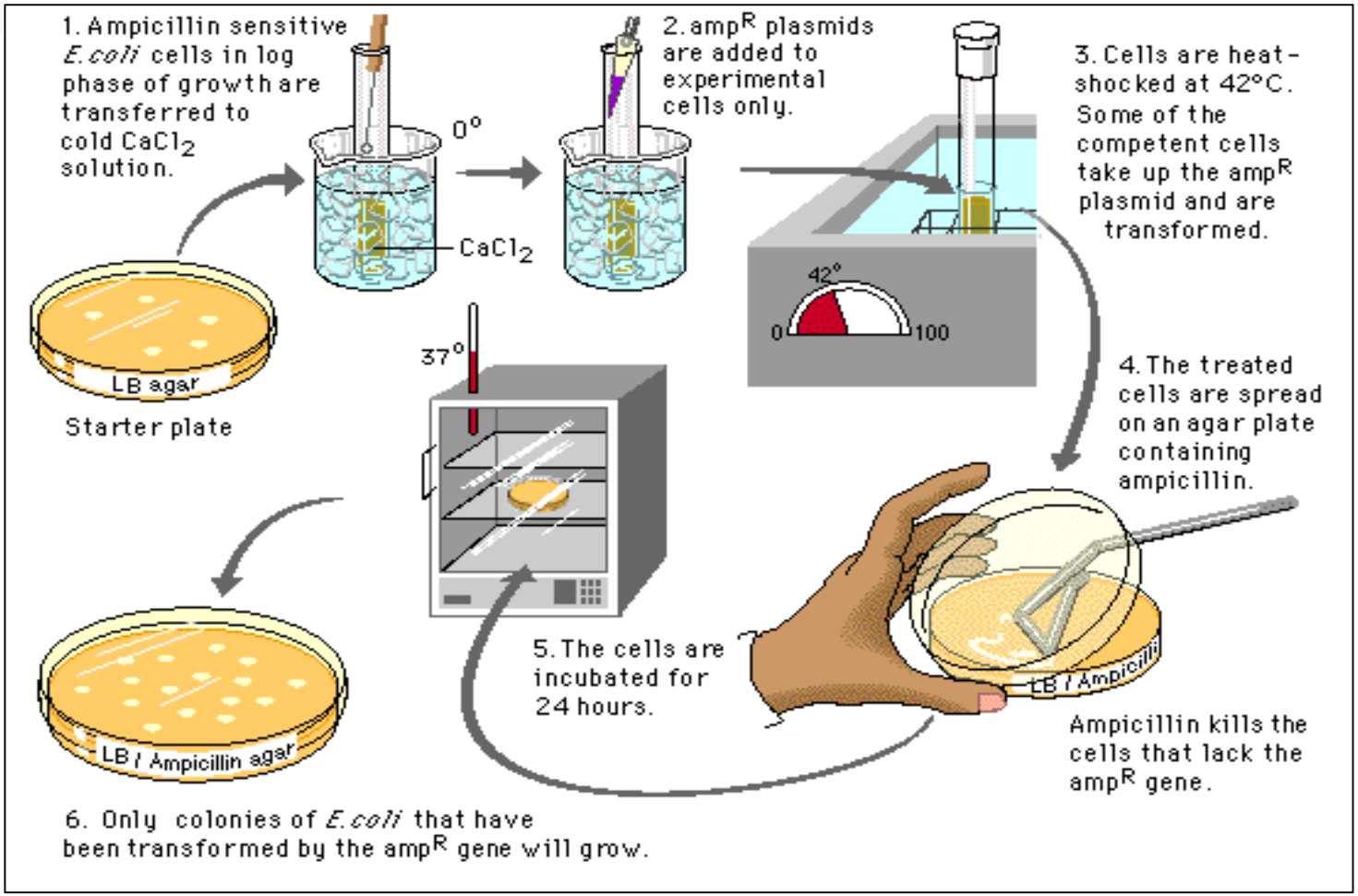
- The recipient cells take the donor DNA into their cytoplasm where it may exchange into the recipient's DNA or if it is a plasmid, it will replicate.



Plasmids can be used as vectors to carry foreign DNA into a cell. Once inside the cell, the plasmid is copied by the host cell's own DNA replication machinery.

Transformation Procedure

Calcium chloride method followed by heat shock



http://www.phschool.com/science/biology_place/labbench/lab6/concepts1.html

Transformation Procedure

Calcium chloride method followed by heat shock

- The purpose of this experiment is to transform the plasmid pGREEN into chemically competent *E. coli* cells.
- The selectable marker on the pGREEN plasmid is a beta-lactamase gene which confers resistance to the antibiotic ampicillin.
- The pGREEN plasmid also has a gene encoding a green fluorescent protein inserted into the cloning site.
- Recipient cells that are successfully transformed with the pGREEN plasmid will be ampicillin-resistant and will produce the green fluorescent protein, giving them a green coloration which is visible under ambient light conditions, and will fluoresce under UV light.

Cain *et al.*,2011

Transformation Procedure

Calcium chloride method followed by heat shock Continued

- **MEDIA NEEDED:**
- 2 LB plates containing Ampicillin (LB-amp)
- 2 LB plates
- LB broth
- Ice-cold Calcium chloride ($CaCl_2$)
- **CULTURES NEEDED:**
- *Escherichia coli* competent cells
- pGREEN plasmid DNA (5 ng/µl), on ice.

Cain *et al.*,2011

Transformation Procedure

Calcium chloride method followed by heat shock Continued

- **PROCEDURE:**

1. Label one of the LB-amp plates "Before" and the other LB-amp plate "After". Label one of the LB plates "Before" and the other LB plate "After".
2. Pre-warm all 4 plates and the small tube of LB broth at 37°C for 30 min.
3. Label an empty microfuge tube as "Tube A". Pipet 250 µl of ice-cold $CaCl_2$ into Tube A. Place Tube A on ice.
4. Use a sterile loop to transfer an isolated colony of *E. coli to* Tube A. Mix the bacteria with the $CaCl_2$ until no visible clumps remain in the tube. The $CaCl_2$ will increase the permeability of the bacterial cells, making them "competent" to take up DNA. Thus, they are now referred to as "competent cells". Put Tube A back on ice.
5. Using a sterile loop, streak a loopful of the competent cells from Tube A onto the LB plate labeled "Before".
6. Using a sterile loop, streak a loopful of the competent cells from Tube A onto the LBamp plate labeled "Before".
7. Pipet 10 µl of the pGREEN plasmid solution into Tube A, and mix by tapping gently.
8. Incubate Tube A on ice for 15 minutes.
9. Heat-shock the cells by placing Tube A in a 42°C water bath for 90 seconds. Place the tube immediately back on ice.
10. Add 250 µl of pre-warmed LB broth to Tube A. Incubate the tube at 37°C for 10 minutes.
11. Using a sterile disposable spreader, spread 100 µl of cells from Tube A on the pre-warmed LB-amp plate labeled "After".
12. Using a sterile loop, streak a loopful of the cells from Tube A onto the LB plate labeled "After".
13. Incubate all plates overnight at 37°C.

Cain *et al.*,2011

Transformation Procedure

Calcium chloride method followed by heat shock Continued

- **RESULTS:**

1. Examine the LB and LB-amp plates for growth and pigmentation, and record your results in the table below:

	Growth on LB	Growth on LB-amp	Growth on LB-amp
"Before"			
"After"			

2. Count the number of colonies on the LB-amp "After" plate.

   _____

3. Calculate the transformation efficiency:

Cain *et al.*,2011

Transformation Procedure

Calcium chloride method followed by heat shock Continued

- **Calculation of transformation efficiency**
- The success of a transformation experiment is determined by calculating the transformation efficiency, which is simply the number of transformants obtained per microgram of DNA.
- For example, if 2 ng of plasmid DNA was transformed into competent *E. coli*, and 237 transformants resulted, the transformation efficiency would be calculated as follows:
  - Step 1: Convert nanograms to micrograms

    2 ng = 0.002 µg
  - Step 2: Divide the # of transformants by the amount of DNA

    237 /0.002 µg = 118,500 transformants/µg DNA =

    1.19×10^5 transformants/µg

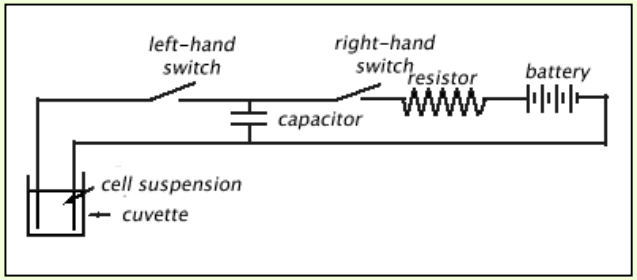
Cain *et al.*,2011

Transformation Procedure

Electroporation System

Electroporation method

- One way of physically introducing DNA into a cell is electroporation.
- The diagram shows an electrical circuit for a simple electroporation device.
- A solution of DNA fragments containing the gene of interest is added to the cuvette.
- The capacitor is charged by closing the right-hand switch. When the capacitor has been charged, the direct current pulse is discharged in the cuvette suspension by closing the left-hand switch.
- The DC pulse is thought both to disrupt temporarily the membrane and to "electrophorese" DNA into cells.



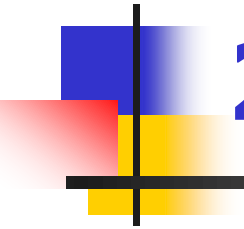
genemol.org

Transformation Procedure

Liposomes are lipid-bilayer bounded vesicles

Liposomes method

- To physically introduce DNA into cells you can use liposomes.
- Liposomes are lipid-bilayer bounded vesicles.
- Produced by hydrating lipids in aqueous solutions. If DNA is present in the solution, it becomes incorporated into the liposomes.
- Liposomes interact with cell membranes. The liposomal contents is transferred to the inside of the cell. Both membrane fusion and endocytosis have been implicated as mechanisms.
- Genes present in the transferred DNA can be expressed transiently.
- The transferred DNA may also integrate into chromosomes and cell lines containing the integrated gene may be selected.



2. Conjugation

Cell-cell contact

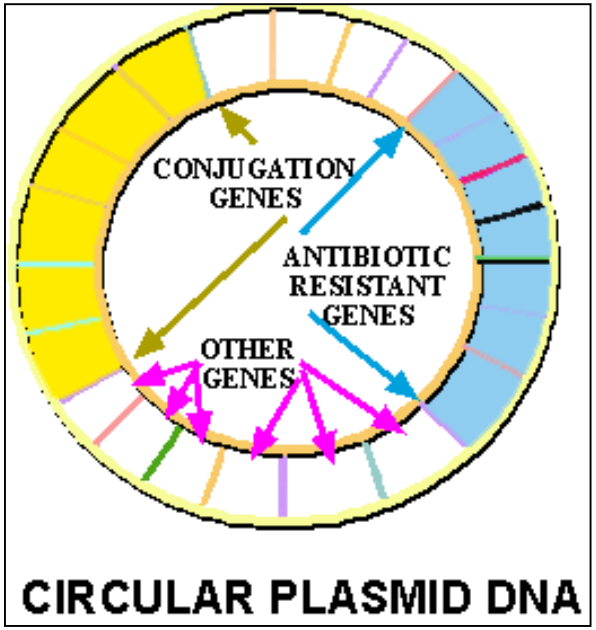
Conjugation

Transfer of plasmids from one cell to another

- This process requires contact between living cells (bacterial sex).
- One type of genetic donor cell (male) is an F^+; recipient cells (female) are F^-.
- F^+ cells contain plasmids called F factors; these plasmids are transferred to the F^- cells during conjugation.
- When the plasmid becomes incorporated into the chromosome, the cell is called an Hfr (high-frequency recombinant).
- During conjugation, an Hfr can transfer chromosomal DNA to an F^-.
- Usually, the Hfr chromosome breaks before it is fully transferred.

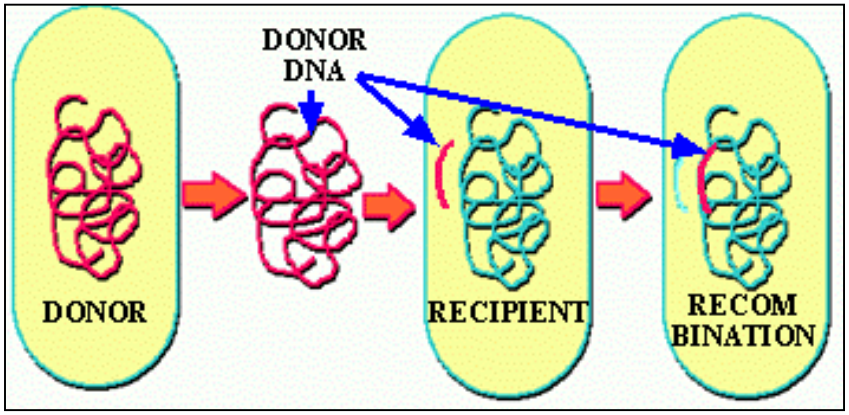
Fertility (F) plasmid

- A fertility plasmid carries the genes for conjugation as well as a number of other genes.
- In this figure the fertility plasmid also carries antibiotic resistant genes.



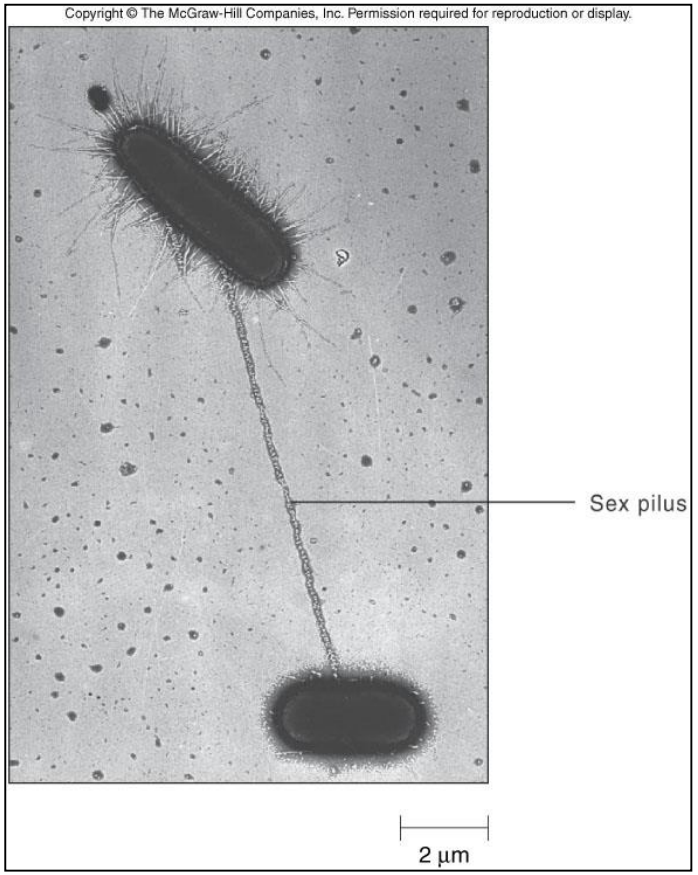
Conjugation
Bacterial exchange of DNA

- DNA from a donor cell is transferred to a recipient cell where it undergoes recombinational exchange, replacing one or more of the recipient's genes with those from the donor.



Conjugation

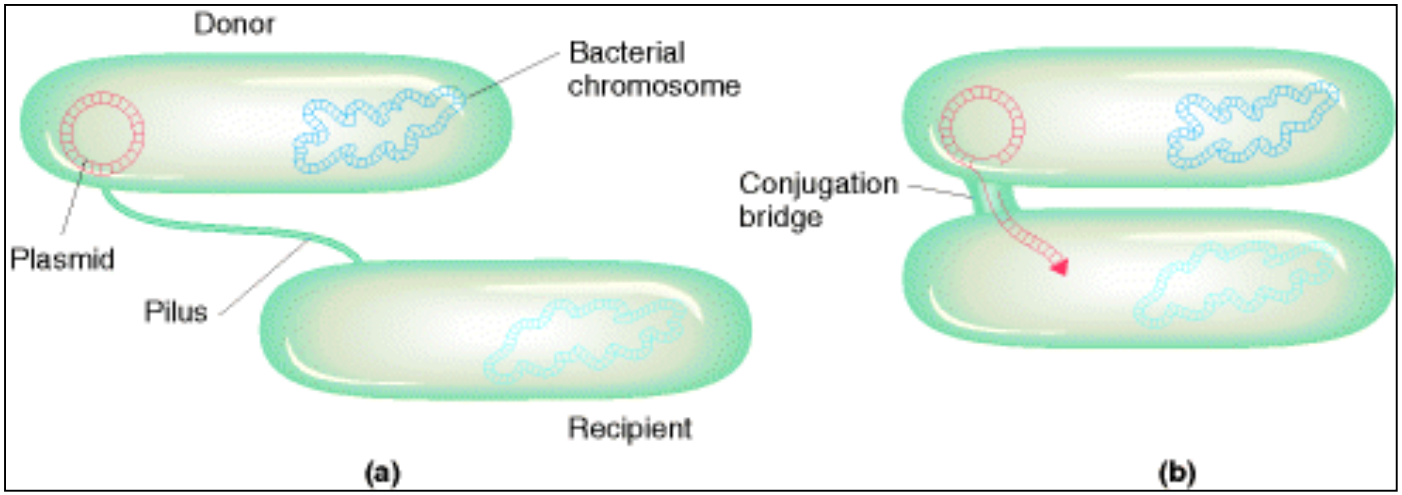
- The pili or sex pilus extends from one cell to its neighboring cell.
- Attachment allows the cells to be drawn close to one another.
- This results in the formation of a junction connecting the cytoplasm of one cell with the other.
- At this point small pieces of DNA (genetic material) known as plasmids can be exchanged between bacteria.



Conjugation

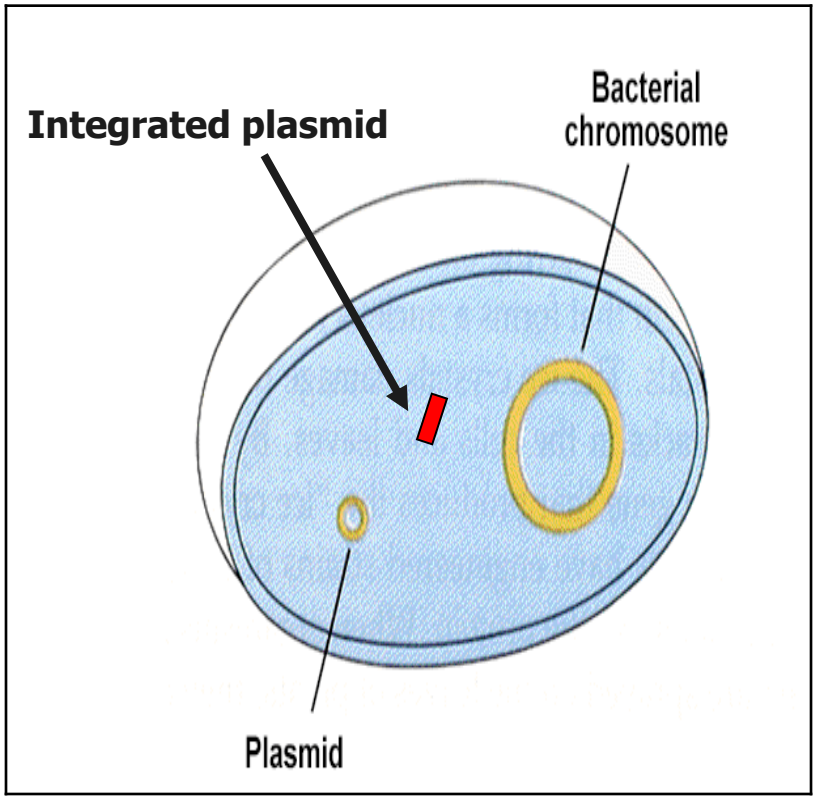
Transfer of F plasmid by conjugation

- Normally only the F^+ plasmid migrates but here the chromosome has been mobilised by integration giving the High Frequency Recombinant phenotype.



Conjugation

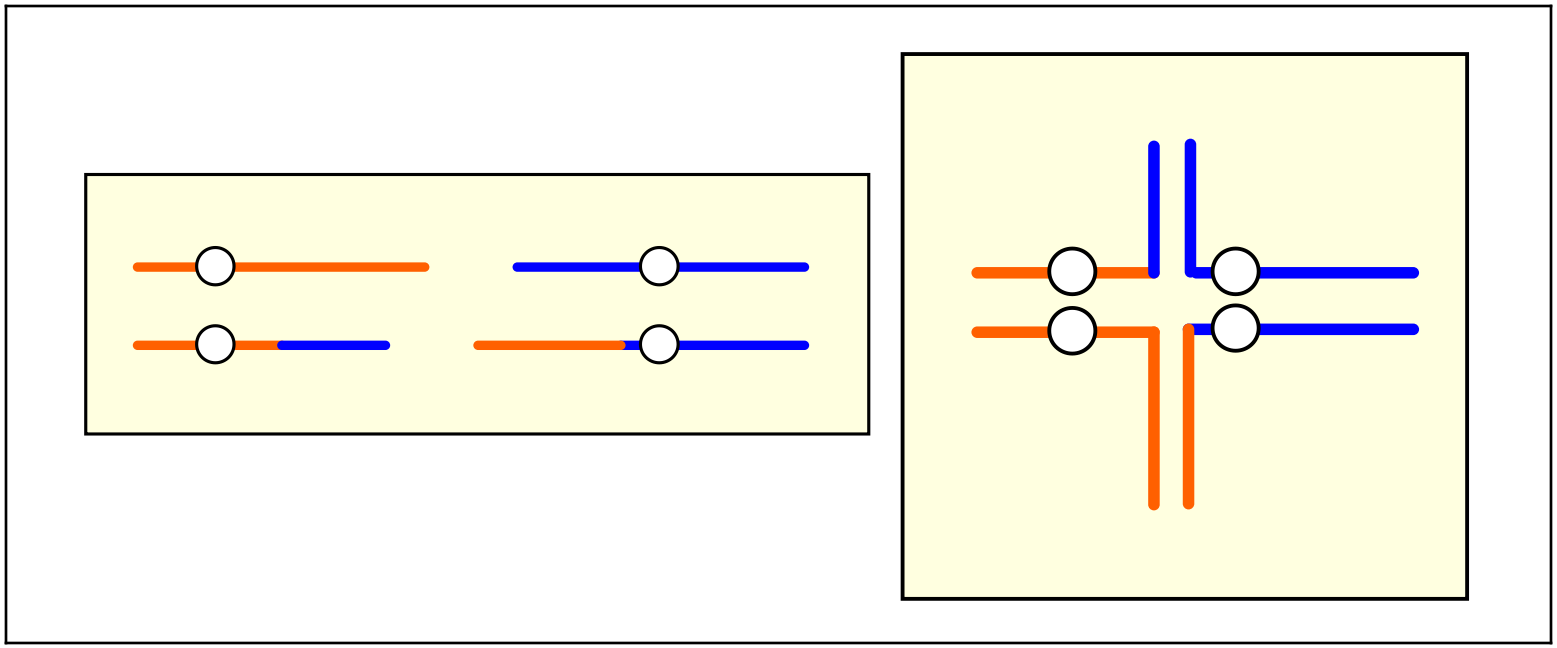
- Recall earlier that bacteria can contain extra DNA as plasmids.
- These plasmids can integrate into the bacterial chromosome as well.



Extrasomal plasmid

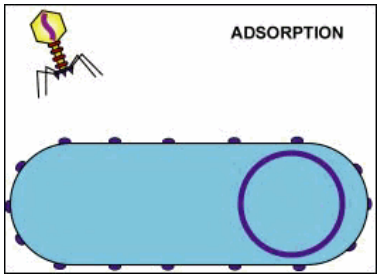
Conjugation

Genetic transfer occurred in one direction, not reciprocal



3. Transduction

Transducting particles/lysates



Transduction

Generalized and specialized transduction

- In this process, DNA is passed from one bacterium to another via a bacteriophage and is then incorporated into the recipient's DNA.

1. In generalized transduction, any bacterial genes can be transferred.
2. In specialized transduction, only genes adjacent to a prophage (lysogenic bacteriophage integrated into bacterial chromosome) can be transferred.

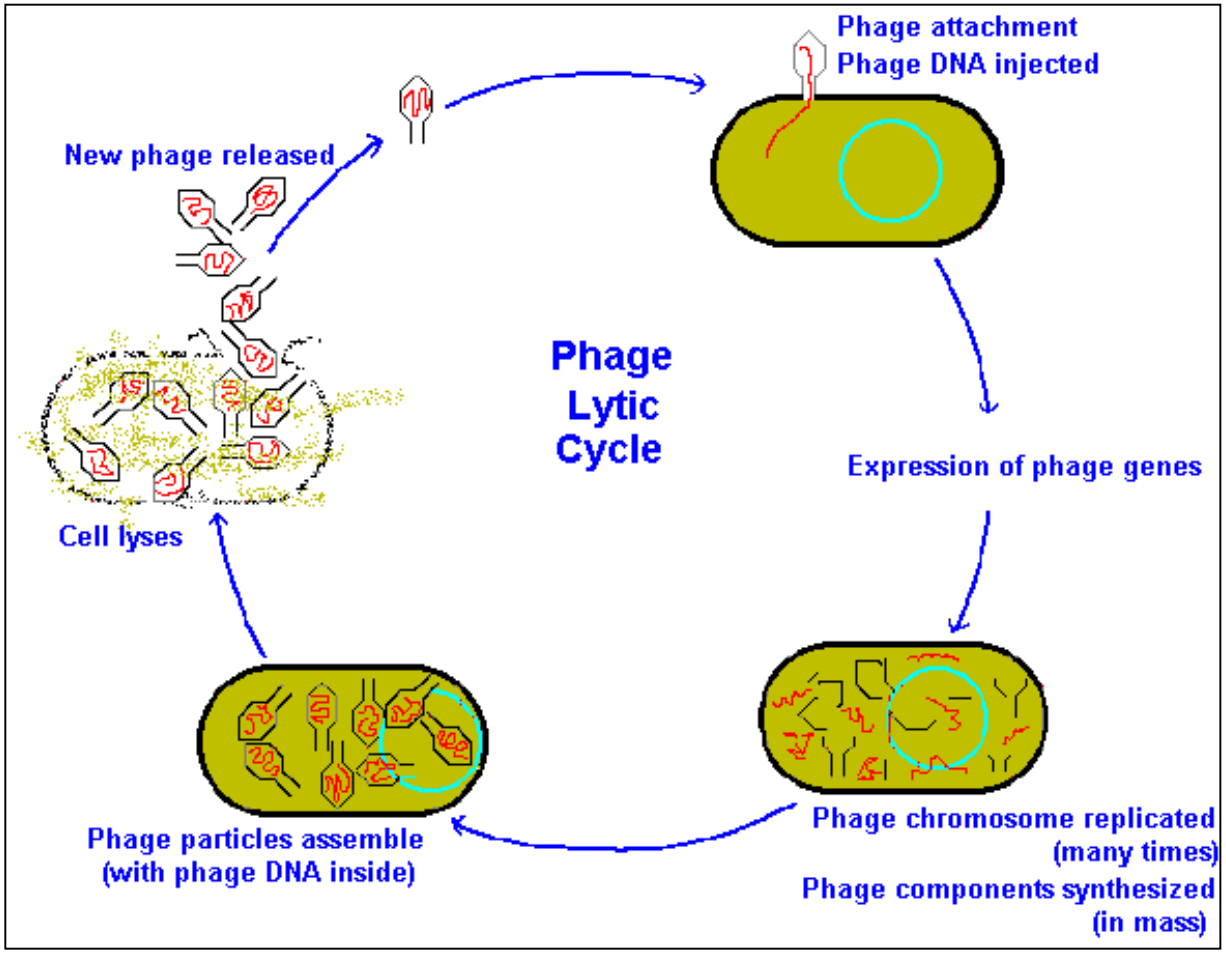
Only temperate phages, like λ are capable of specialized transduction. Both virulent and temperate phages may carry out generalized transduction. I am not aware that any RNA phage can carry out transduction (vocabulary).

Transduction
Specialized transduction

- Specialized transduction (also called restricted transduction) requires infection by temperate phage-host DNA adjacent to prophage is incorporated into phage genome.
- Prophages or plasmids integrated into an attachment site can be indicated by the name of the attachment site followed by a double colon and the phage genotype indicated in brackets.
- For example, *att*::[P22 *mnt*::Kan].

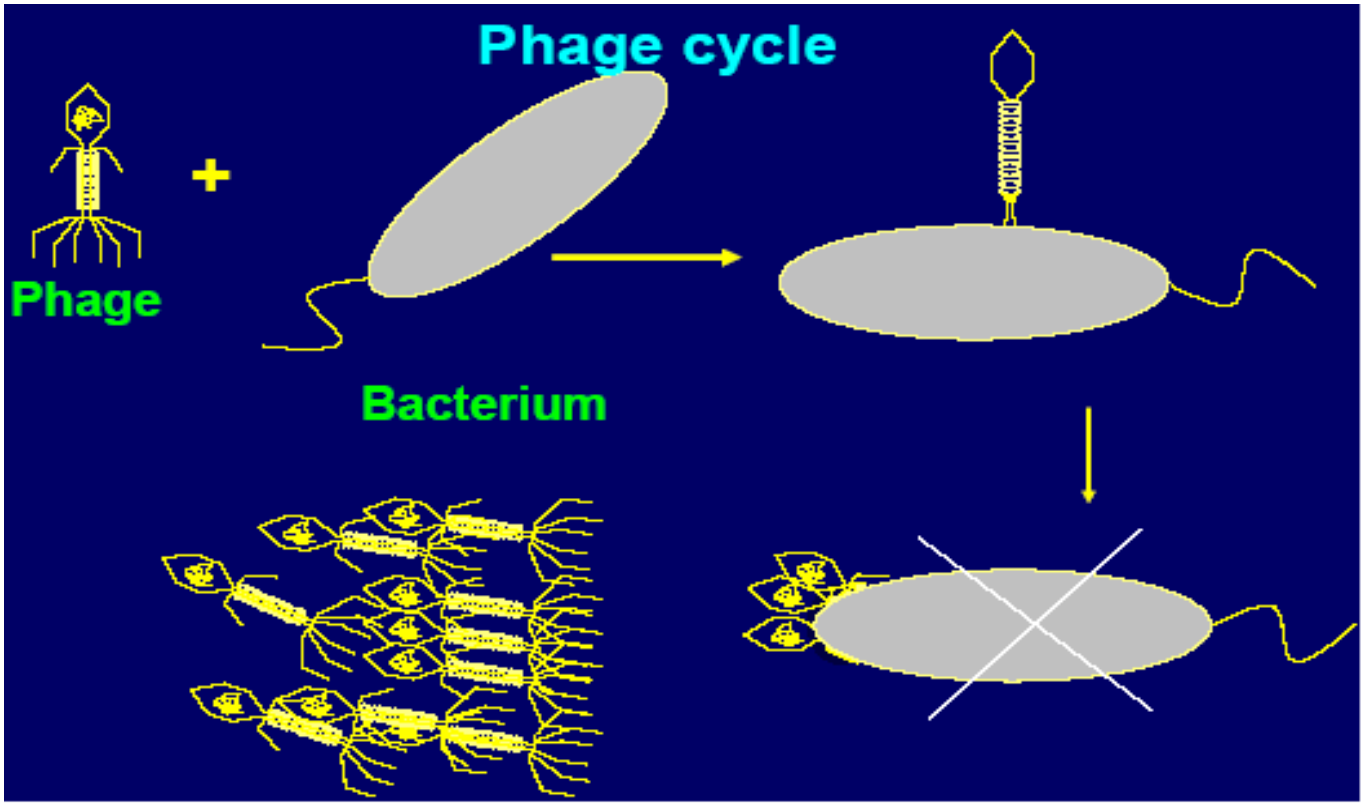
Transduction

Phage lytic cycle



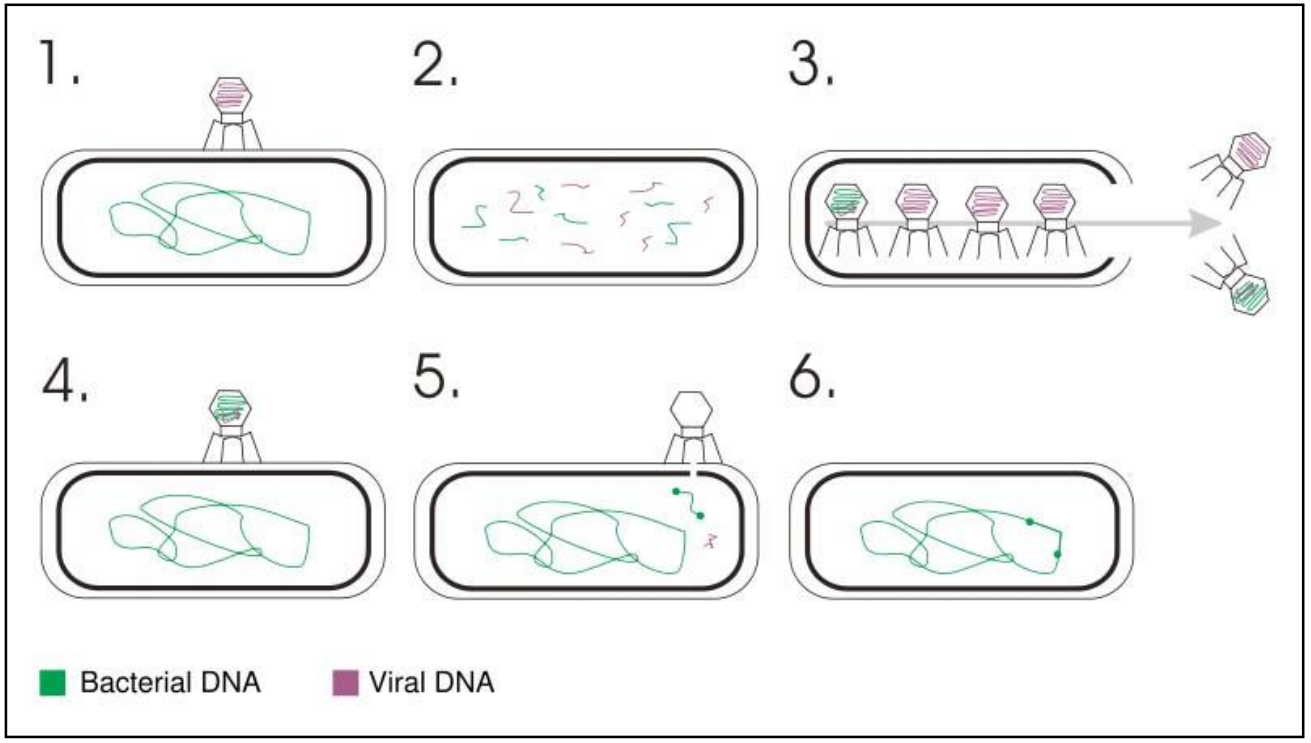
Phage cycle

Bacteriophage T4 tails fibers interact with its host (*Escherichia coli*) through membrane fusion



Mechanisms of genetic transfer
Transduction

- Viruses that infect bacteria; utilize syringe like mechanism to introduce genome.



Life Sciences-HHMI Outreach,2008

4. Mutation

- A mutation is a change in the base sequence of DNA; that change causes a change in the product coded for by the mutated gene.
- Mutations can be:

1. Harmful
2. Lethal
3. Helpful (beneficial)
4. Silent (neutral).

Mutation are often in microorganisms.

El-Safey Mohamed El-Safey; Johnson,2003

Mutation

Two classes of mutations

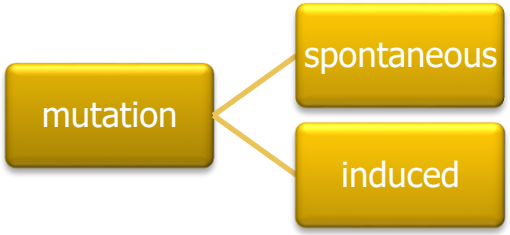
- A mutation may, or may not, affect the phenotype of the organism.
- Such a change may occur either:

1. Spontaneously (natural)mutations,
2. Induced mutations.

- The major source of variation within a bacterial species is mutation.
- In mutations, usually only a single gene changes at any one time.
- A bacterium that has mutated to resistance to an antibiotic (or other substance) is given the superscript "R".
- Thus, the phenotype ampicillin resistance is indicated as Amp^R.

Causes of mutations

Two classes of mutations



- Two classes of mutations are:

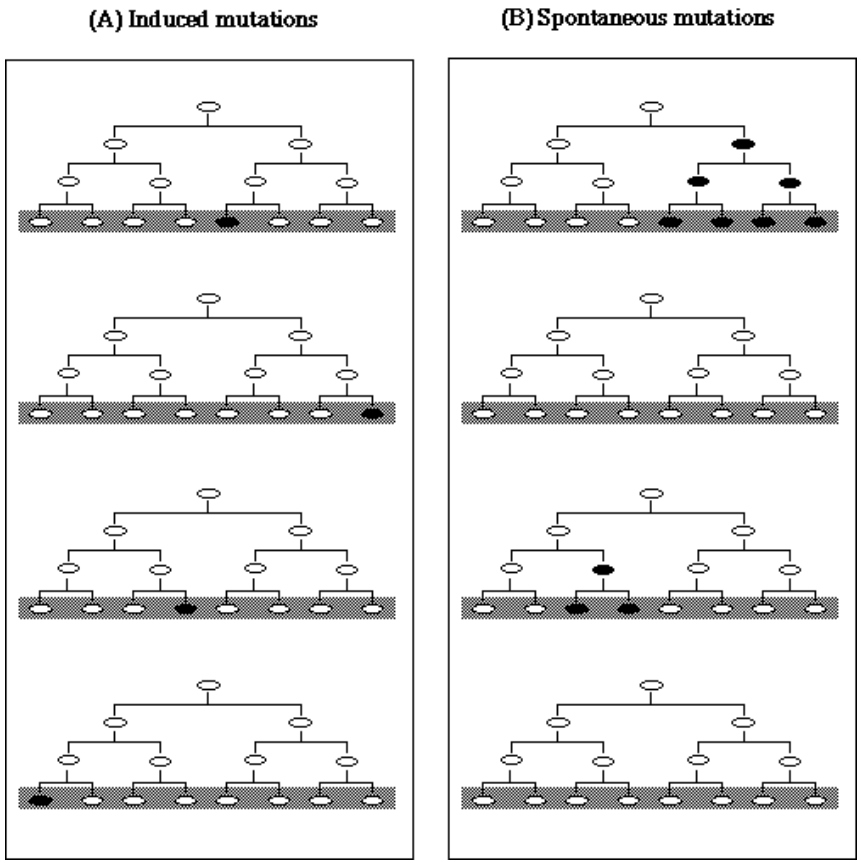
1. Spontaneous mutations (molecular decay),
2. Induced mutations caused by mutagens, agents that cause DNA to mutate.

- Spontaneous mutations occur without the presence of a mutagen.
- The spontaneous mutation rate varies from genome to genome.
- Mutagens are agents in the environment that cause permanent changes in DNA. E.g. X-rays, ultraviolet light, chemicals, etc.

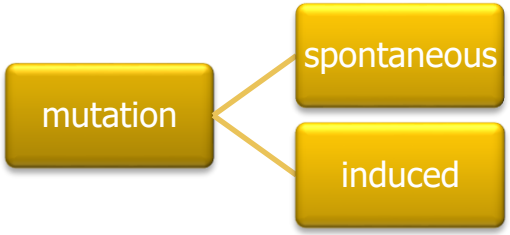
Causes of mutations

Spontaneous and induced mutations

- **Induced mutations:** Exposure to the toxic conditions might produce rare, resistant mutations in the population of bacteria.
- **Spontaneous mutations:** Spontaneous mutants might have occurred in the bacterial population prior to exposure to the toxic conditions, yielding resistant progeny cells.



Causes of mutations



1. Spontaneous mutations: A mutation without a known cause is called spontaneous mutations.

- A mutation that arises naturally and not as a result of exposure to mutagens.
- This type of mutation is not caused by any mutagens such as radiation or chemicals. Also called natural mutation.

Examples:

- Mistakes during DNA replication,
- Environmental mutagens.

2. Induced mutations: Caused by intentional exposure to a mutagen (radiation or chemicals).

Spontaneous mutations

In the presence of antibiotic concentrations

Gradient plate technique

- If an organism has the ability to produce mutant strains resistant to antibiotics, the nature of mutation, whether it is spontaneous or induced have to be tested.
- The spontaneous mutations due to resistance in antibiotics such as streptomycin are easily detected because they grow in the presence of antibiotic concentrations that inhibit the growth of normal bacteria.
- An excellent way to determine the ability of organisms to produce mutants that are resistant to antibiotic is to grow them on a gradient plate of a particular antibiotic.

Spontaneous mutations

In the presence of antibiotic concentrations

Gradient plate technique

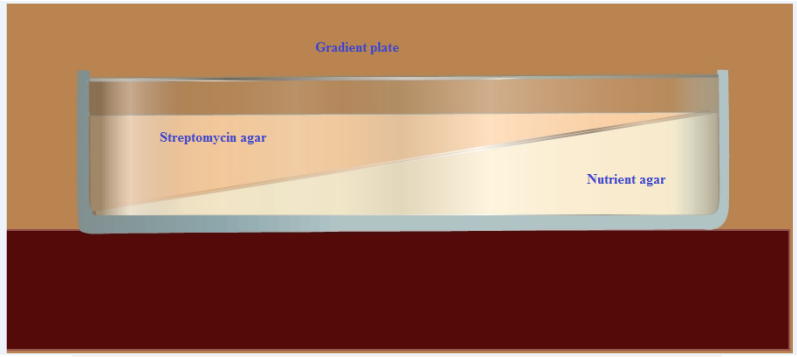
- A gradient plate is made by using streptomycin in the medium.
- E. coli, which is normally sensitive to streptomycin, will be spread over the surface of the plate and incubated for 24 to 72 hours.
- After incubation colonies will appear on both the gradients.
- The colonies develop in the high concentration are resistant to the action of streptomycin, and are considered as streptomycin resistant mutants.
- For isolation of antibiotic resistant of gram negative enteric bacteria, the antibiotics commonly used are Rifampicin, Streptomycin, and Erythromycin etc.

Spontaneous mutations

In the presence of antibiotic concentrations

Gradient plate technique

- The gradient plate consists of two wedges like layers of media: a bottom layer of plain nutrient agar and top layer of antibiotic with nutrient agar.
- The antibiotic in the top layer, diffuse into the bottom layer producing a gradient of antibiotic concentration from low to high.



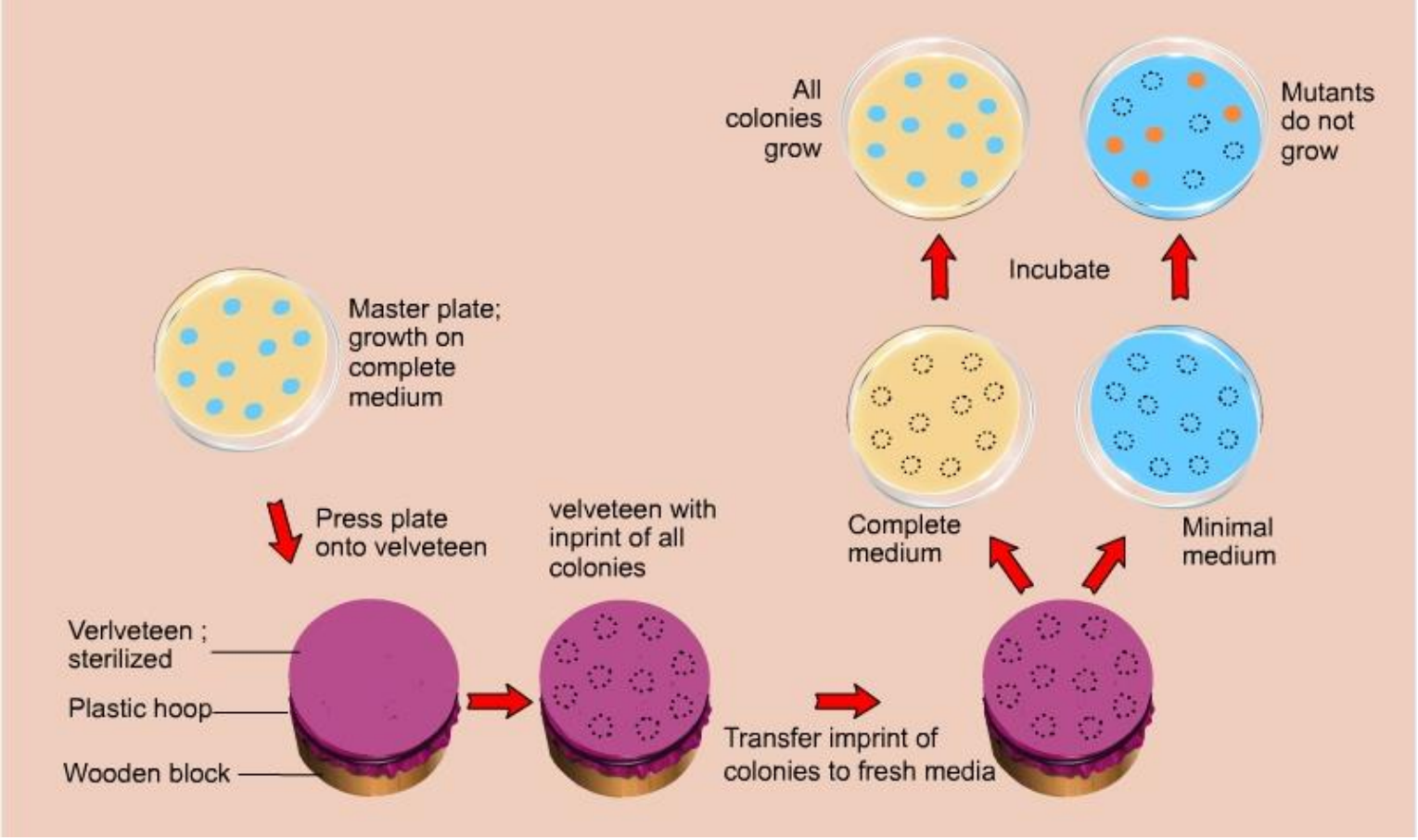
Gradient plate preparation steps

Amrita virtual lab.

Spontaneous mutations

In the presence of antibiotic concentrations

Replica plating technique



Amrita virtual lab.

Spontaneous mutation
Stability of mutants

- Streptomycin resistant strain of *Acidovorax citrulli* 30290^{str} was a spontaneous mutant (Weller and Saettler, 1978) obtained by plating train *Acidovorax citrulli* 30290 onto an ethanol bromcresol purple/brilliant blue R (EBB) agar (Schaad and Sechler, 1999) gradient plate containing 0 to 100 µg/ml streptomycin.
- The stability of strain 30290^{str} was confirmed by growing the bacterium for at least 30 generations in the absence of streptomycin (Feng *et al.*, unpublished).

Zaho *et al.*,2009

Colony phase variation

Mutation in *gacS* and *mutS* genes cause colony phase variation

- Colony phase variation is mediated by:

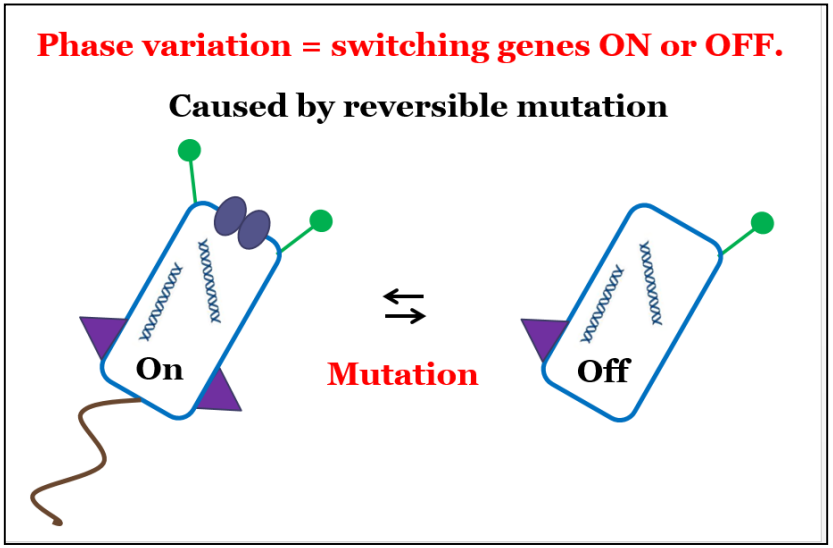
1. DNA mutations,
2. Reorganization, or
3. Modification.

- Mutations in the *mutS* gene increased the frequency of switching from phase I (opaque colonies) to phase II (translucent colonies) up to 1000-fold.
- Phase variation is based on structural changes at the DNA level and results in subpopulations of bacteria, as is often demonstrated by the presence of distinct morphological phases between colonies or within a colony.

Van den Broek *et al.*,2003

Phase variation
Mutation

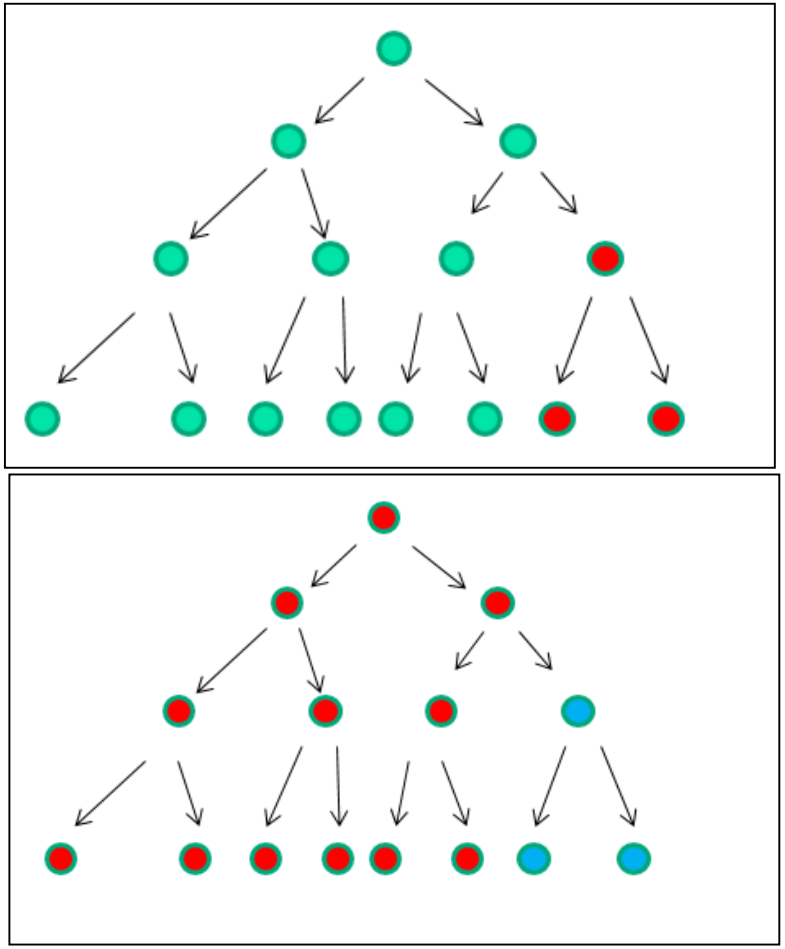
- Phase variation is the switching on and off of genes in response to mutation.
- These mutations are reversible, meaning genes can be switched on and off interchangeably.



Phase variation

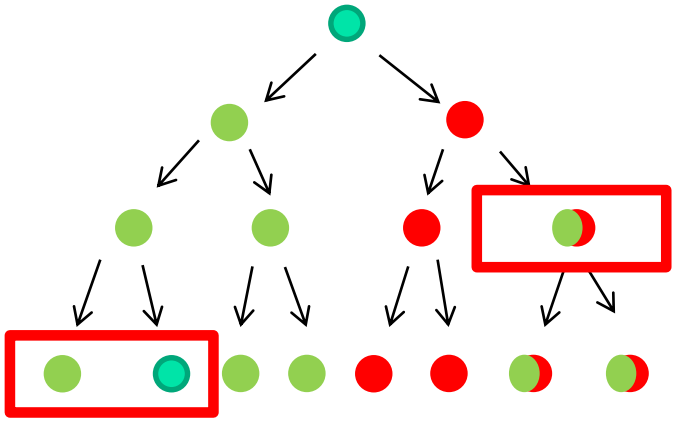
What does phase variation look like?

- Mutation is random.
- Like rolling a dice
- One cell will divide into two daughter cells
- With every division, there is a chance they will mutate (mutation rate)
- This mutation will be passed down to the daughter cells.
- It is a reversible process.



Phase variation

What does phase variation look like?

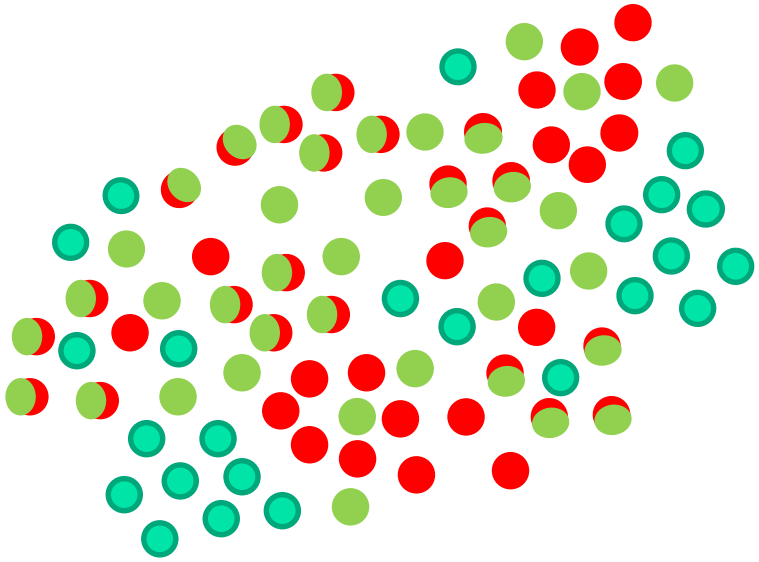


* Many bacteria have more than one phase variable gene
* Both genes can mutate at the same time

* Mutation isn't this quick in real bacteria.

Phase variation

What is the effect of phase variation on populations of cells?

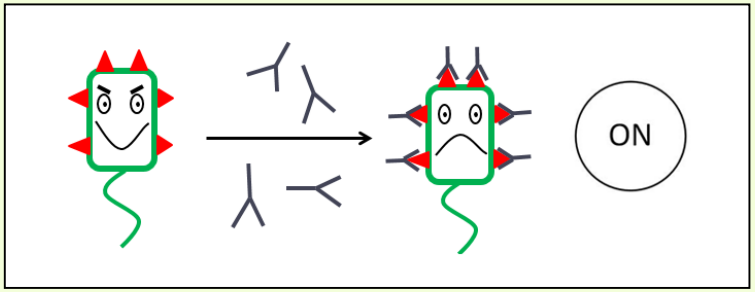


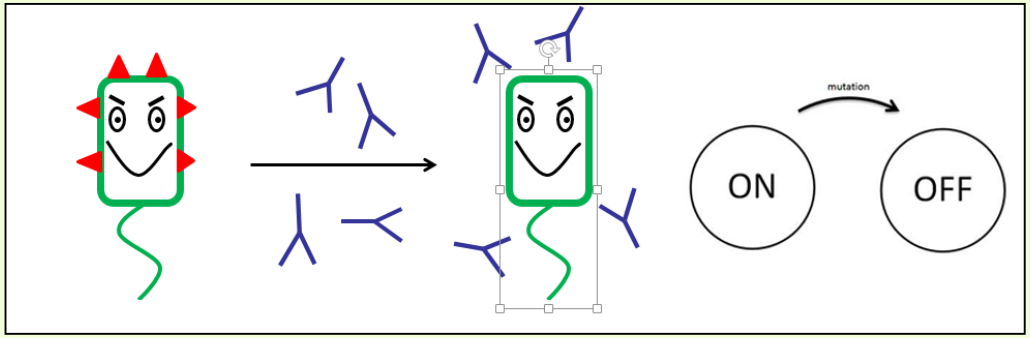
All of the different cells have the same DNA but have different antigens expressed.

Phase variation

How can phase variation benefit bacteria?

- Switching some genes off can help the bacteria to escape the immune system.





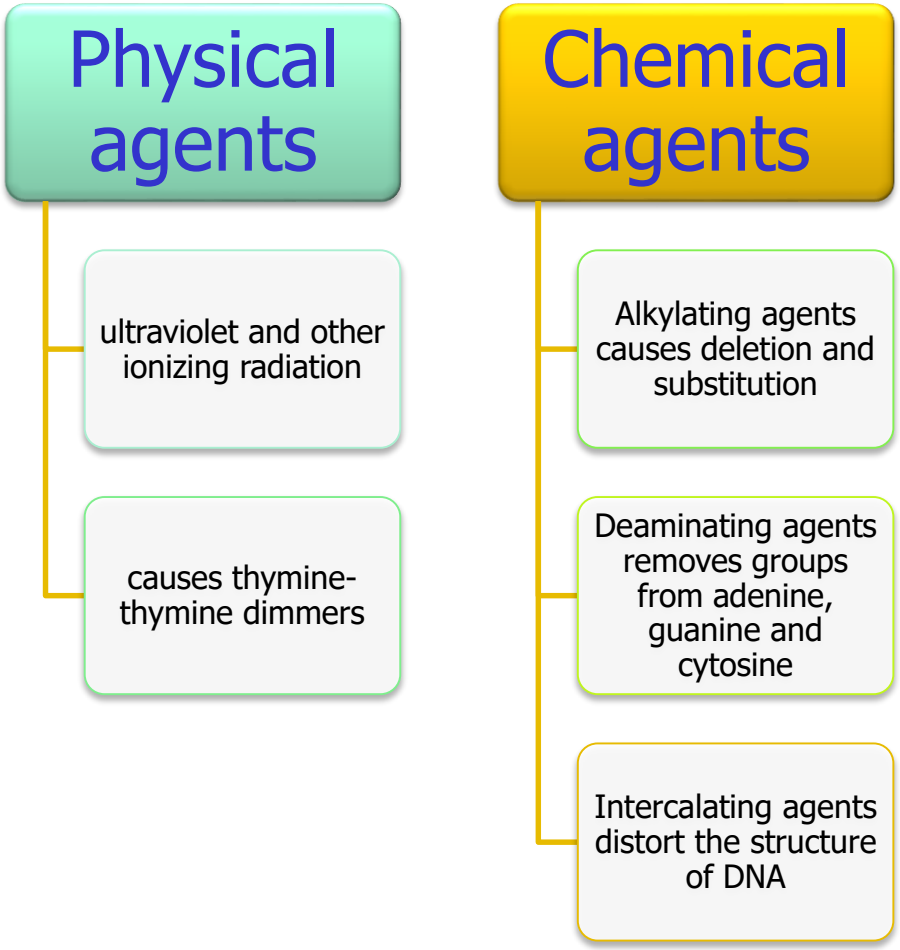
Causes of induced mutations

Chemicals

- Nitrosoguanidine (NTG),
- Hydroxylamine NH_2OH,
- Base analogous (e.g. BrdU), N-ethyl-N-nitrosourea (ENU)).
- Simple chemicals (e.g. acids),
- Alkylating agents (e.g. N-ethyl-N-nitrosourea, ENU).
- These agents can mutate both replicating and non-replicating DNA.
- In contrast, a base analog can only mutate the DNA when the analog is incorporated in replicating the DNA.
- Each of these classes of chemical mutagens has certain effects that then lead to transitions, transversions, or deletions.
- Methylating agents (e.g. liquid mutagen ethyl methanesulfonate, EMS)
- Polycyclic hydrocarbons (e.g. benzopyrenes found in internal combustion engine exhaust),
- DNA intercalating agents (e.g. ethidium bromide),
- DNA cross kinkers (e.g. platinum)
- Oxidative damage caused by oxygen (O) radicals.

Causes of induced mutations

Chemical and environmental agents



Causes of induced mutations

Chemical and environmental agents

- Chemical and environmental agents include alkylating agents, deaminating agents, and intercalating agents.

Alkylating agents

- deletion and substitution due to an apurinic site

Deaminating agents

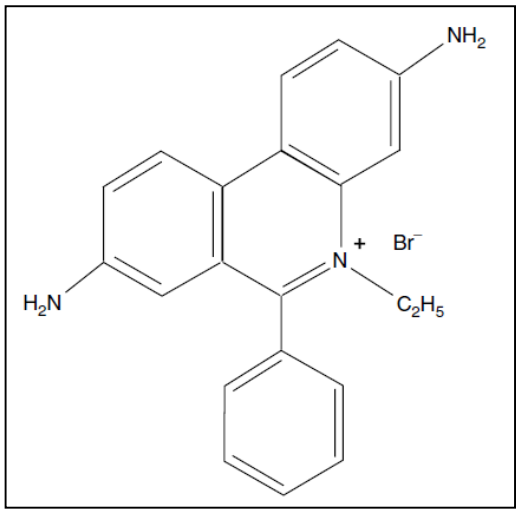
- removes groups from adenine, guanine and cytosine

Intercalating agent

- distort the structure of DNA by inserting itself in between adjacent base pairs.

Causes of induced mutations
Chemicals

- The intercalating agents, such as acridine orange and ethidium bromide, have a different mechanism of action.
- These molecules contain a flat ring structure which is capable of inserting (intercalating) into the core of the double helix between adjacent bases.
- The consequences of this are the addition (or sometimes deletion) of a single base when the DNA is replicated, giving rise to a frameshift mutation.
- These dyes (ethidium bromide in particular) are also much used in molecular biology in the detection of DNA, since the complex formed with DNA is fluorescent.



Ethidium bromide

Dale and Park,2004

Causes of induced mutations
Radiation

1. Short wave (X-rays, gamma rays): High penetration power; breaks DNA.
2. Non-ionizing (UV) longer wave: No penetrating power; forms thymine dimers.

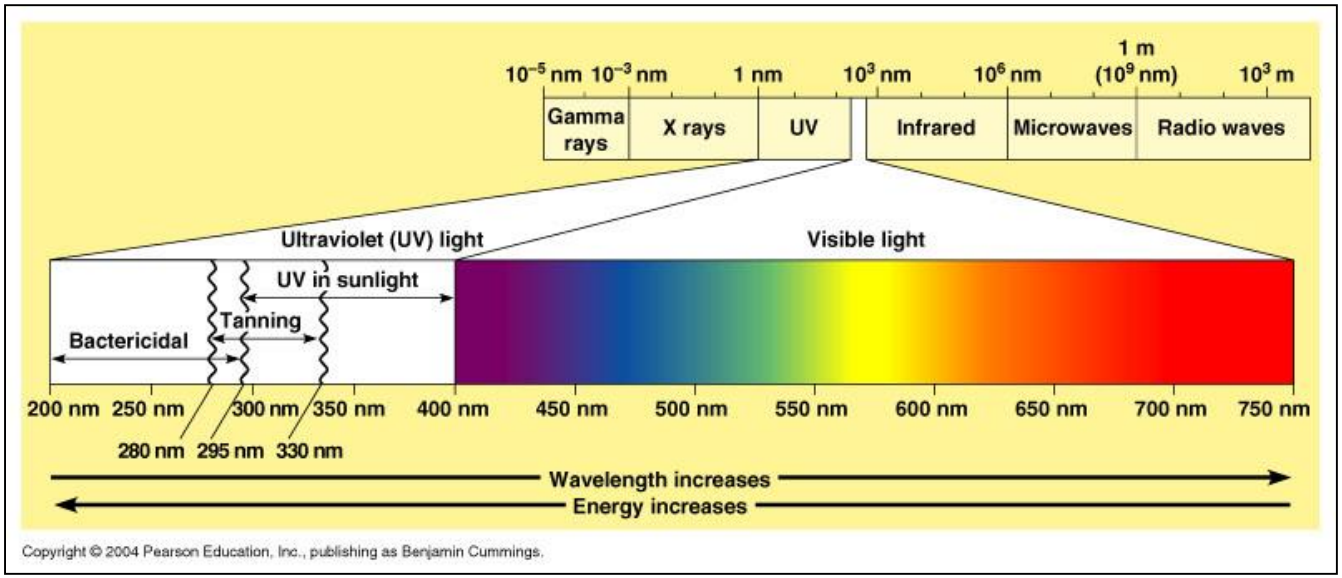
- Two nucleotide bases in DNA – Cytosine and thymine are most vulnerable to excitation that can change base-pairing properties.
- UV light can induce adjacent thymine bases in a DNA strand to pair with each other, as a bulky dimer.
- Organisms contain multiple repair systems.

Wikipedia,2008

Causes of induced mutations

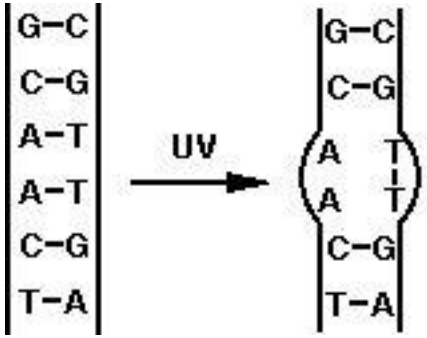
Radiation

- Ultraviolet radiation (nonionizing radiation) excites electrons to a higher energy level.
- DNA molecules are good absorbers of ultraviolet light, especially that with wavelengths in the 260 to 280 nm range.



Wikipedia,2008;…

Induced mutation
Radiation



- There are two forms of electromagnetic radiation that are mutagenic.

1. Ionizing radiation, such as gamma rays and X-rays, depending on the energy of the radiation, can create free radicals that result in problems ranging from point mutations to chromosome breaks.
2. Non ionizing radiation (UV) causes formation of pyrimidine dimers in the DNA molecule i.e. adjacent pyrimidine units are ionized forming highly reactive free radicals.

- These ionized pyrimidines interlink to form the dimers.
- Thus, the primary effect of UV on DNA is the creation of thymine dimers.
- Thymine dimers occur when two thymines are adjacent on a strand of DNA.

Induced mutation

Ultraviolet radiation

1. Changes the shape of the DNA in the cell.
2. Can cause problems during replication.
3. It is often used as a germicidal agent.

Effect of UV Radiation on Bacterial Growth

To study the effect of UV exposure on the growth of bacterial cells

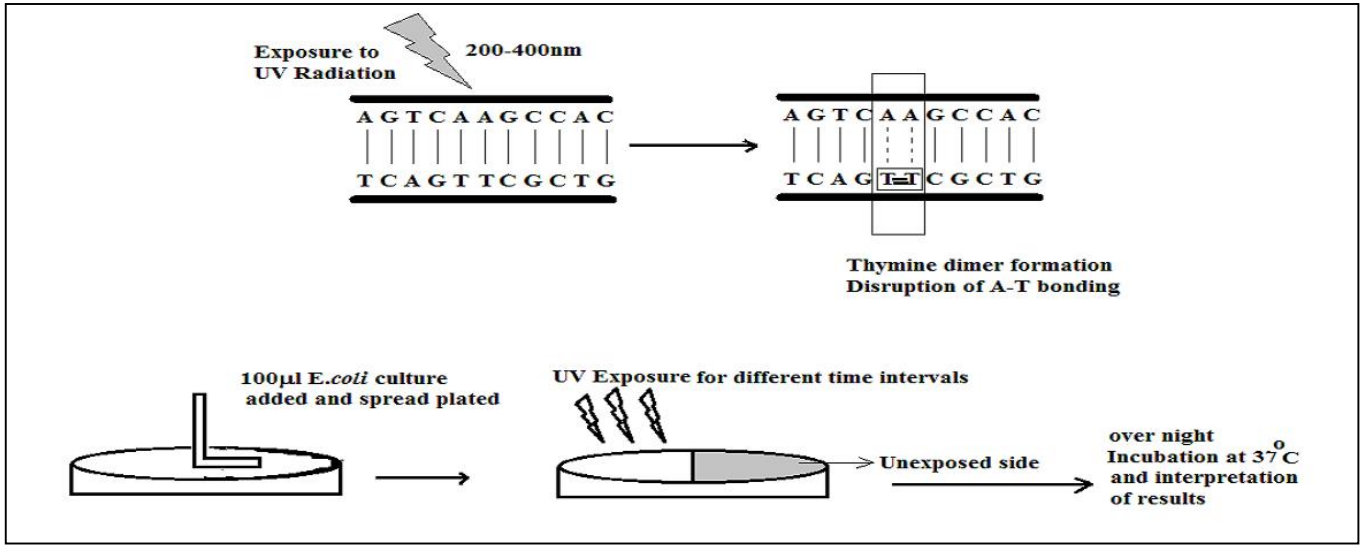
Protocol:

1. LB-nutrient agar plates are prepared.
2. Overnight culture of *E.coli* is taken and 0.1 ml of it was surface plated.
3. The plate is exposed to direct UV light for time intervals of 5, 10, 15, 30 minutes, by covering half of the plate with glass plate.
4. After exposure plates are incubated overnight at 37°C.
5. The number of bacterial colonies formed is counted in exposed and covered areas and percentage of reduction according to the respective time of exposure is determined using the given relation:

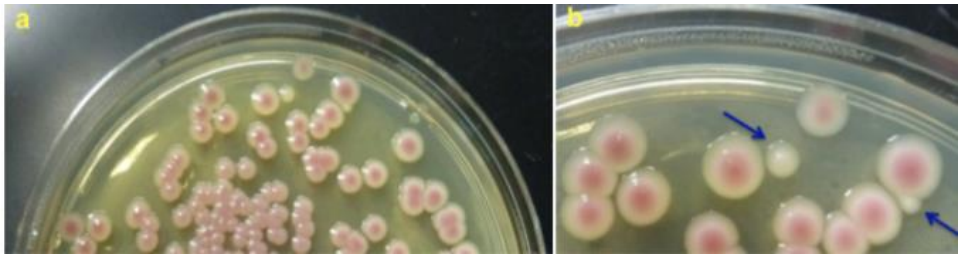
$$\% \text{ of Reduction} = \frac{\text{colonies in (covered area} - \text{exposed area)} \times 100\%}{\text{colonies in covered area}}$$

Molecular Cloning: A Laboratory Manual, 2nd ed., pp.1.25-1.28.

Effect of UV Radiation on Bacterial Growth (Continued)



Figures below show (a) a plate of *Serratia marcescens* after UV irradiation and part of the same plate enlarged (b). Arrows show two possible mutants.



Molecular Cloning: A Laboratory Manual, 2nd ed; Eby Bassiri

Causes of spontaneous mutations

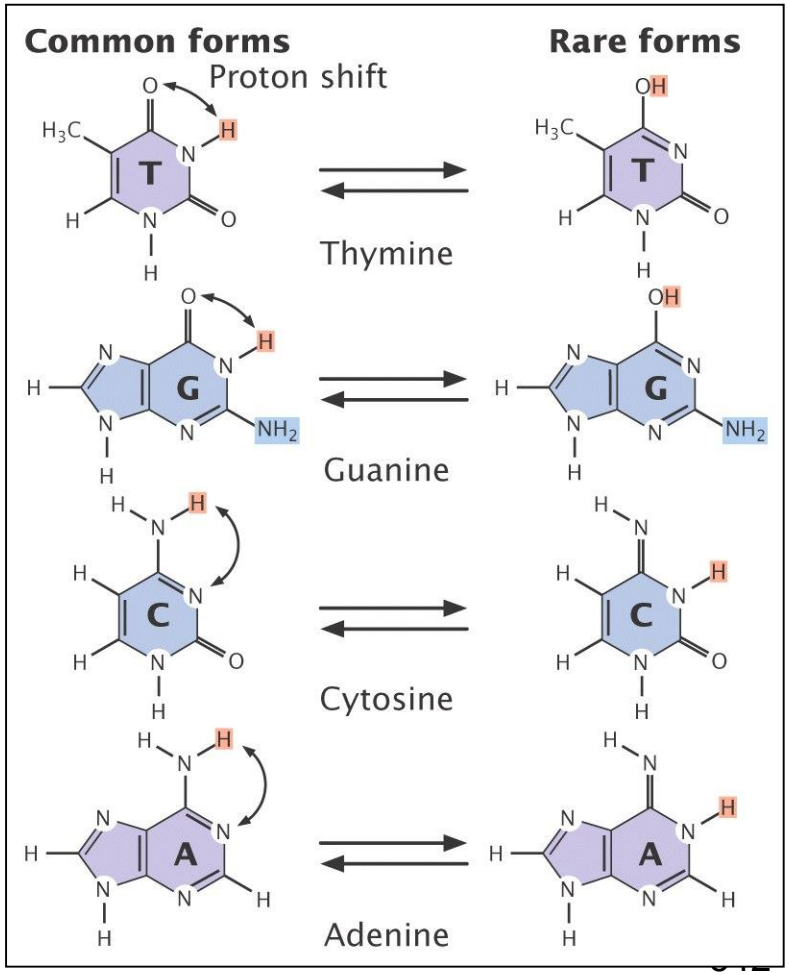
- Spontaneous mutations on the molecular level include:

1. Tautomerism: A base is changed by the repositioning of a hydrogen atom.
2. Depurination: Loss of a purine base (A or G) to form an apurinic site (AP site).
3. Deamination: Changes a normal base to an atypical base.

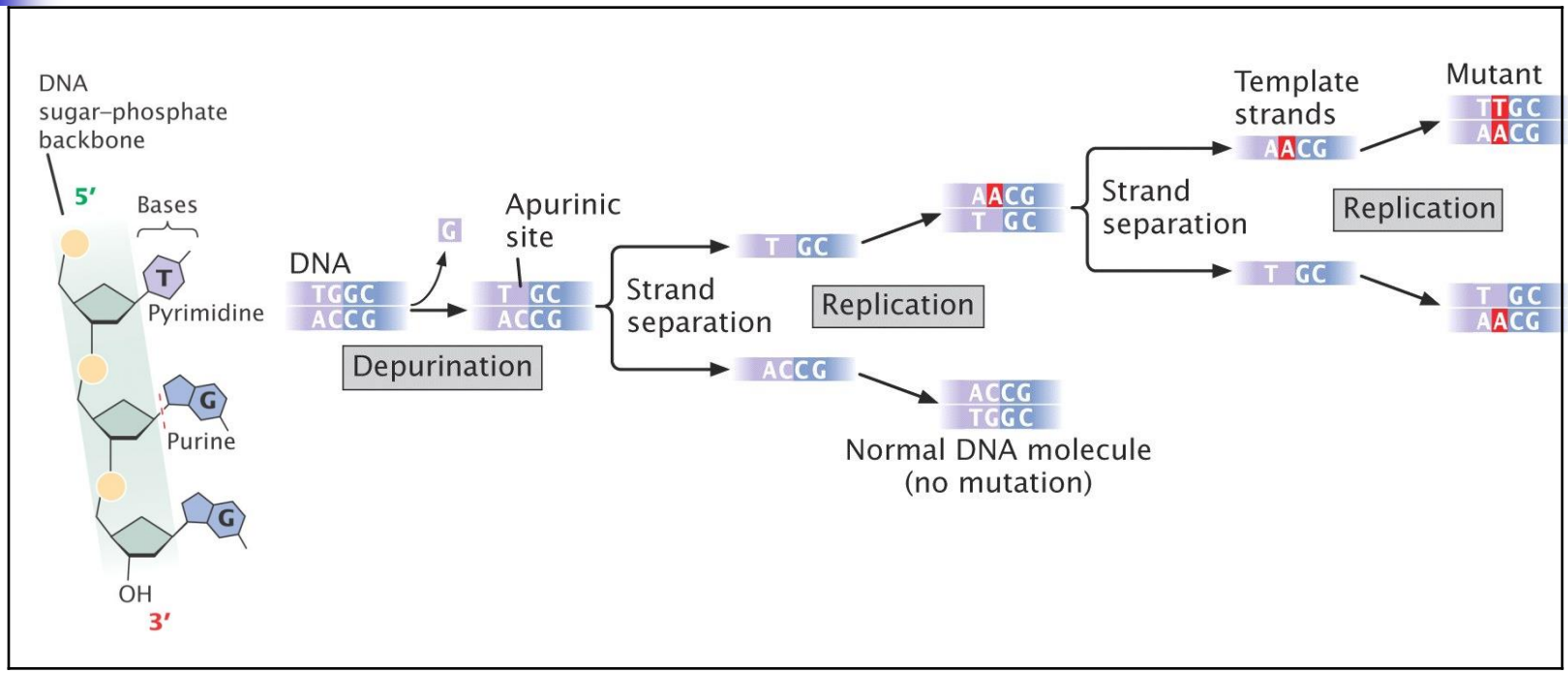
- Examples include C → U and A → HX (hypoxanthine), which can be corrected by DNA repair mechanisms; and 5MeC (5-methylcytosine) → T, which is less likely to be detected as a mutation because thymine is a normal DNA base.
- Transition – A purine changes to another purine, or a pyrimidine to a pyrimidine.
- Transversion – A purine becomes a pyrimidine, or vice versa.

Spontaneous mutation
Tautomerism

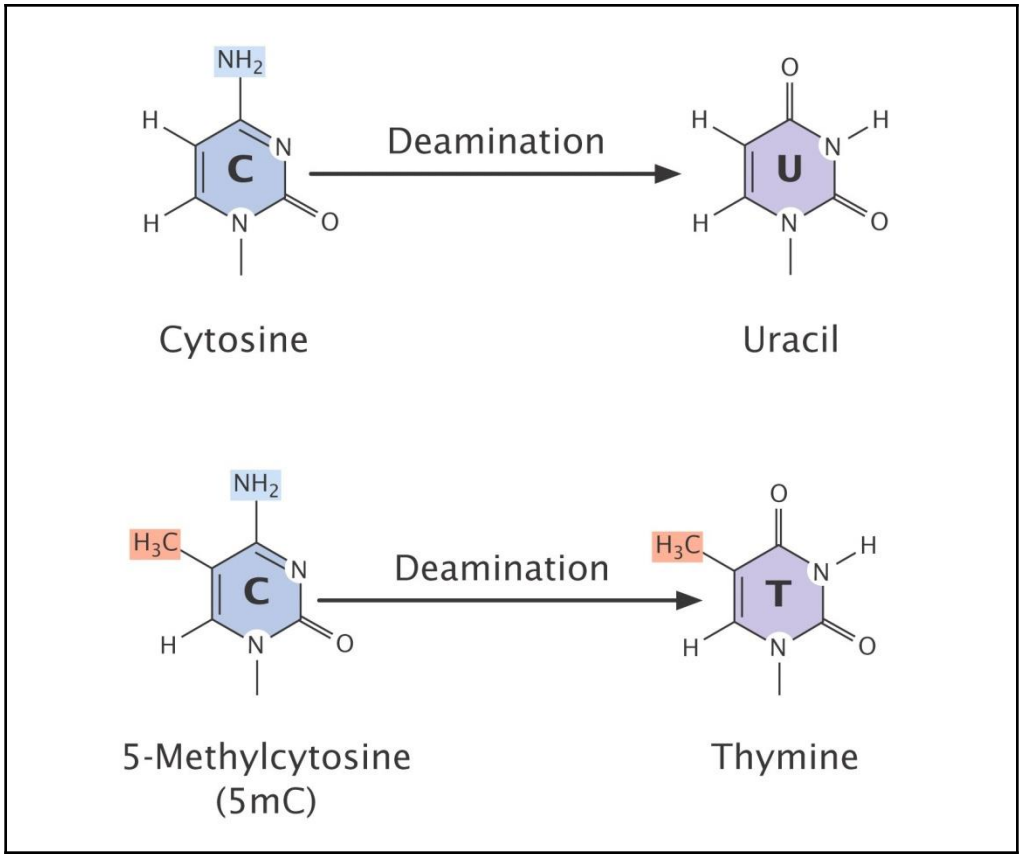
- Tautomeric shifts occur rarely in DNA bases and could lead to mismatched base pairing (incorporated error).



Spontaneous mutation
Depurination



Spontaneous mutation
Deamination



Types of mutations

1. Point mutation: Single base is substituted.
2. Missense mutation: Cause amino acid substitutions.
3. Nonsense mutation: Create stop codons. Base change changes single amino acid to stop codon.
4. Null or Knockout mutation: Mutation that totally inactivates a gene.
5. Deletion or insertion mutation: Change in number of bases making up a gene.
6. Frameshift mutation: One or a few base pairs are deleted or added to DNA (insertion or deletion of something other than multiples of three bases).
7. Reversion mutation: Mutated change back to that of wild type.

Types of mutations

Base substitution

GATCCGAGTATCGCAATTAGCA
GATCCGAGTGTCGCAATTAGCA

Deletion

GATCCGAGTATCGCAATTAGCA
GATCCGAGTAATTAGCA

Insertion

GATCCGAGTATCGCAATTAGCA
GATCCGAGTATCGCAGCATTAGCA

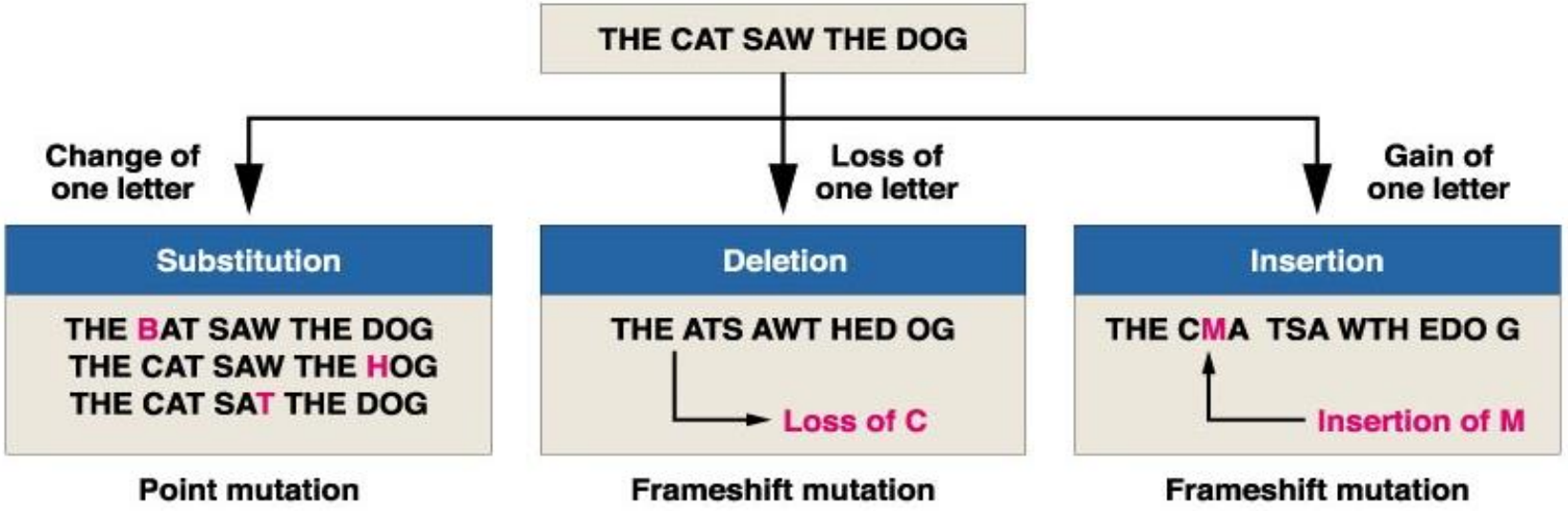
Duplication

GATCCGAGTATCGCAATTAGCA
GATCCGAGTATCTCGCAATTAGCA

Inversion

GATCCGAGTATCGCAATTAGCA
GATGCCAGTATCGCAATTAGCA

Three basic types of gene mutations



Point mutation

- A change in the single base pair of a nucleotide sequence is referred to as point mutation.
- Transition - A replaces G or C replaces T.
- Transversion - A replaces T or vice versa.

Consequences of point mutations

- Silent - Changed codon is for same amino acid.
- Missense - Changed codon is for different amino acid.
- Nonsense - Changed codon is a stop codon.
- Nitrous acid and 5-bromo -2-deoxyuricil are mutagens that yield point mutations.

Base substitutions

Transition / Transition

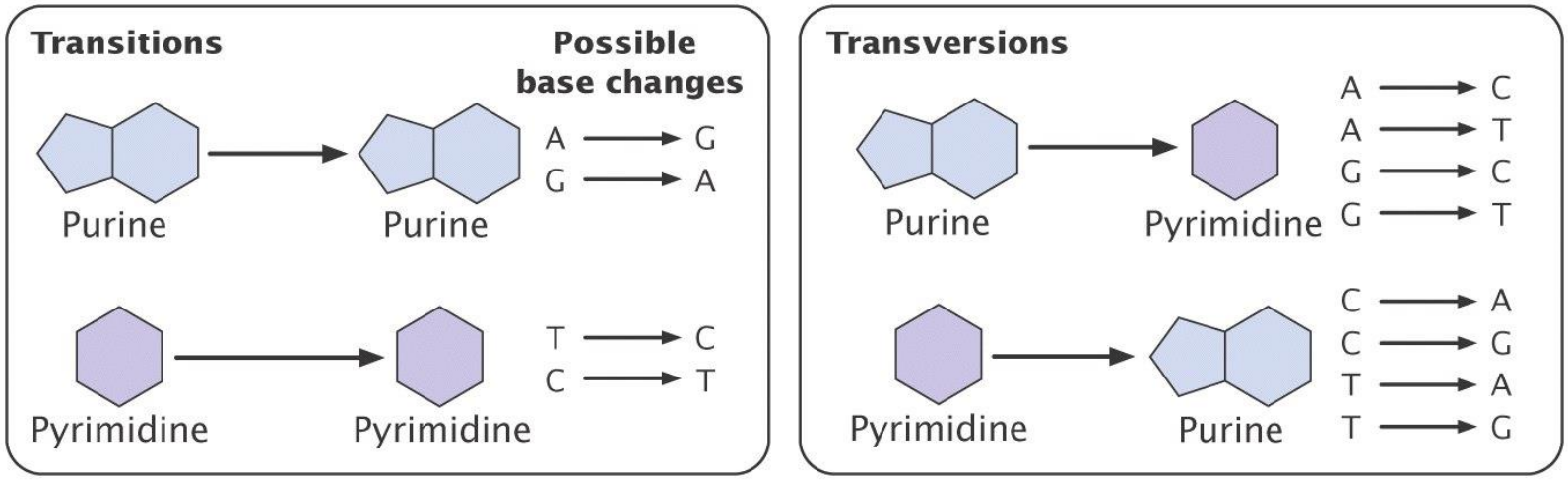
G·C ⇨A·T

transition

G·C ⇨T·A

transversion

Transition /Transversion



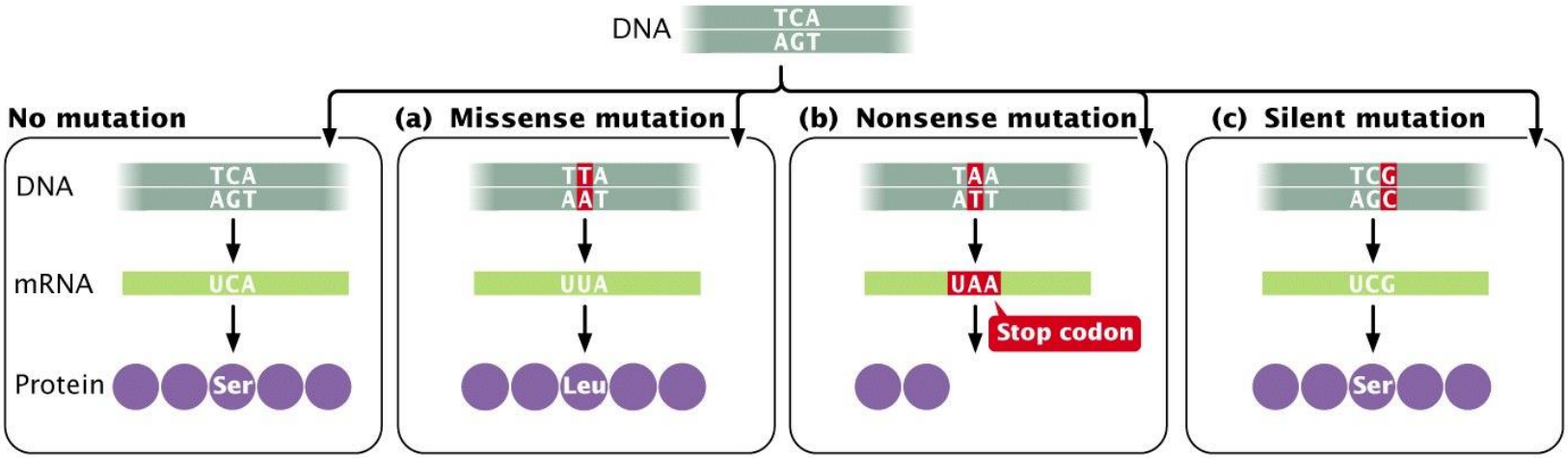
Silent mutation

- A silent mutation is a nucleotide substitution that alters a codon such that it still codes for the same amino acid.

DNA	RNA	Amino Acid
TCC	AGG	Arginine
GCC	CGG	Arginine

Consequences of point mutations

- Silent mutation:
- No phenotypic change;
- Degenerate code.



Missense mutation

Can result in significant changes in polypeptide

- A missense mutation is one that alters a codon so that is specifies a different amino acid in a protein (amino acid substitution).

THE CAT SAW THE BIG **D**OG

THE CAT SAW THE BIG **H**OG

This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.

Missense mutation

Normal b globin:

(G<u>A</u>G)

NH_2-val-his-leu-thr-pro-glu-glu…etc.

Sickle cell b globin:

(G<u>U</u>G)

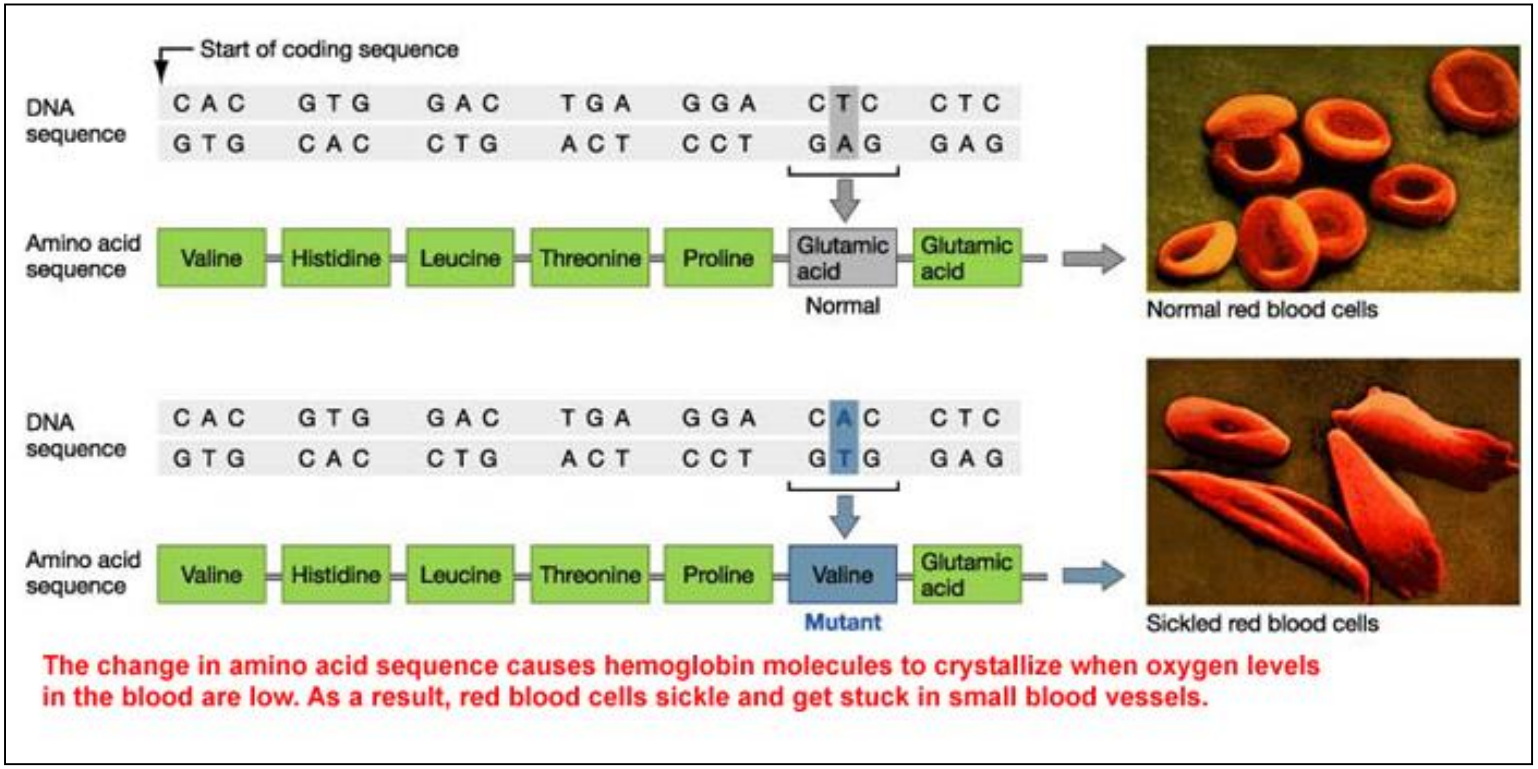
NH_2-val-his-leu-thr-pro-val-glu…etc.

Nonsense mutation

- A nonsense mutation converts an amino-acid-specifying codon to a stop codon.
- This causes premature termination of translation and results in a short, usually nonfunctional, protein.

Mutation

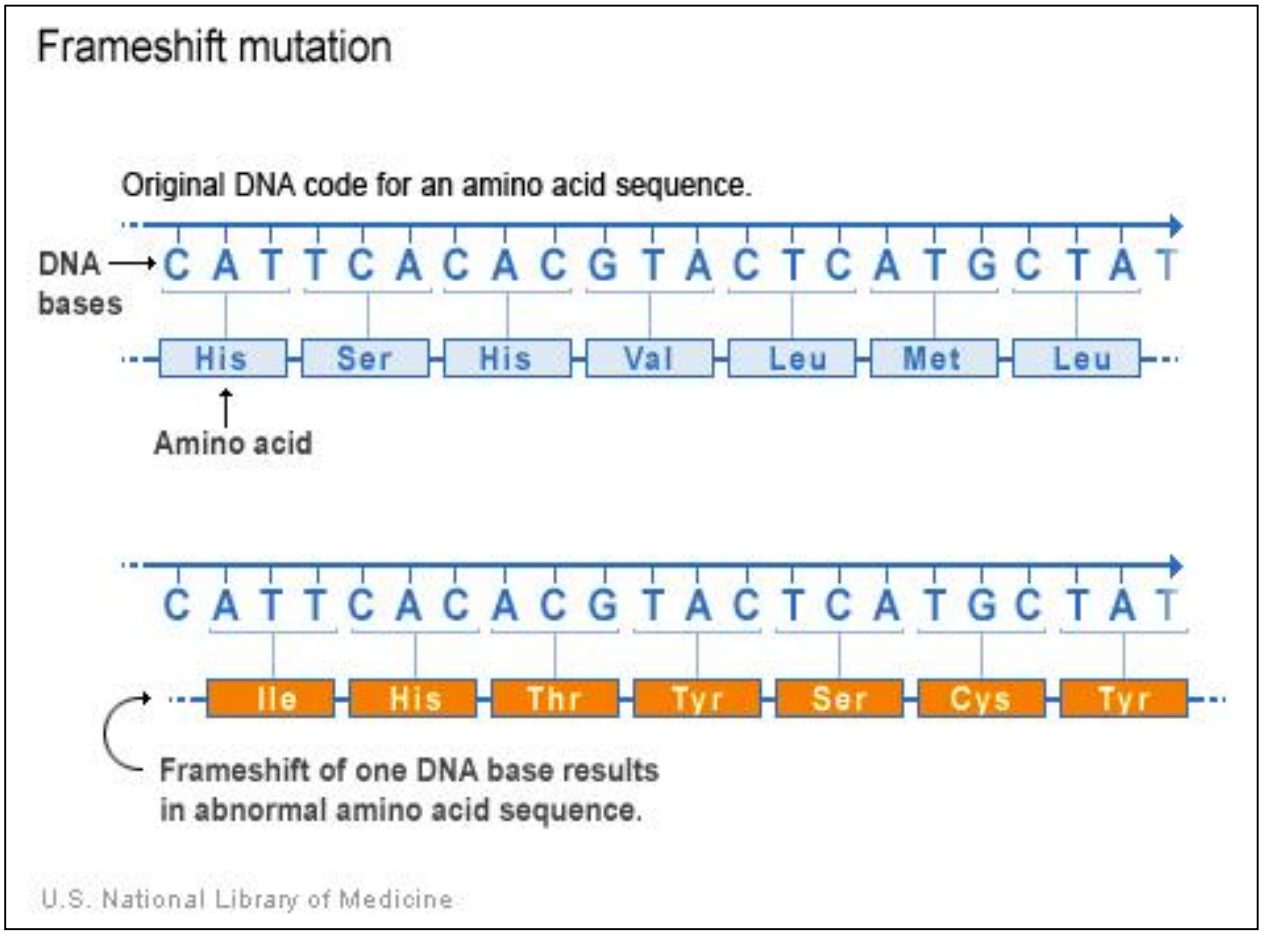
Change in primary amino acid sequence = defective protein- Sickle Cell



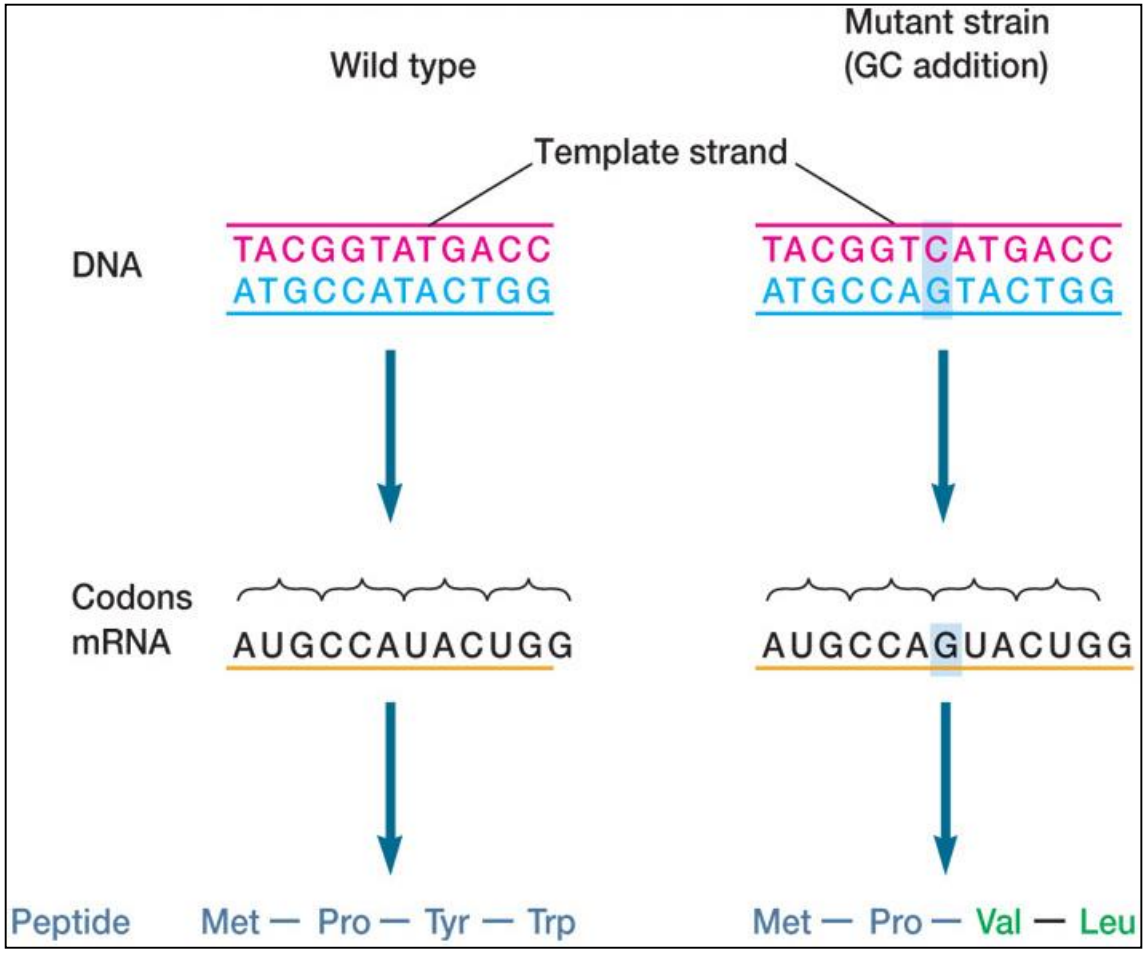
Frameshift mutation

- Frameshift mutation (also called a framing error or a reading frame shift) is a genetic mutation caused by indels (insertion or deletions) of a number of nucleotides that is not evenly divisible by three from a DNA sequence.
- In short, addition or deletion of one or more bases can result in a shift in the reading frame of codons.
- Transposon insertion or ethidium bromide can cause frameshifts.

Frameshift mutation



Frameshift mutation



What is the difference between mutation rate and mutation frequency?

- **Mutation Rate**: is the number of mutations per cell division.
- Because the number of cells in the population is so large, the number of cell divisions is approximately equal to the number of cells in the population (N).
- Therefore, the value of "h" can be determined by a fluctuation test.
- **Mutation Frequency**: is the ratio of mutants/total cells in the population.
- This is measured by plating out the cell on selective and non-selective media and counting the number of mutants per culture. This method is easier, but may show large fluctuations depending on when the first mutation arose in the population.

Frequency of mutation

Rates of mutation

- **Mutation rates are usually low:**
- In laboratory populations of growing *E. coli*, the rate of mutation per base pair is about 10^{-10} -10^{-11}.
- **High mutation rates:**
- 10,000-fold, 10^{-4} above spontaneous can be achieved under certain conditions in mutant *E. coli* bacteria.
- However, such populations die out eventually due to "error catastrophe".

Frequency of mutation

Rates of mutation

1. Mutation rate is the probability that a gene will mutate when a cell divides; the rate is expressed as 10 to a negative power.
2. Mutations occur randomly along a chromosome.
3. A low rate of spontaneous mutation is beneficial, providing genetic diversity needed for evolution.

Frequency of mutation
Calculation of mutation rates

- The mutation rate is the number of mutations per cell division. Because the cell population is so large, the number of cell divisions is approximately equal to the number of cells in the population (N).

$$a = \frac{h}{N}$$

- In the example shown above, if h was determined from a fluctuation test with 10^7 cells per tube, then the mutation rate would be:

$$a = \frac{h}{N} = \frac{0.6 \text{ mutations / tube}}{10^7 \text{ cell divisions / tube}} = 6 \times 10^{-8} \text{ mutations / cell divisions}$$

Frequency of mutation
Calculation of mutation rates

- Assume a mutation rate of u = 1 out of a million spores per generation or 1 x 10^{-6}.
- This means that, on average, in a population of one million individuals (spores, bacterial cells, or virus particles), you can expect to find one mutant for any given locus per generation.
- In a population of 10 million individuals, you would expect to find 10 mutants for any locus.
- And in a population of 1 billion individuals, you expect to find 1000 mutants for any locus.

The American Phytopathological Society,2018

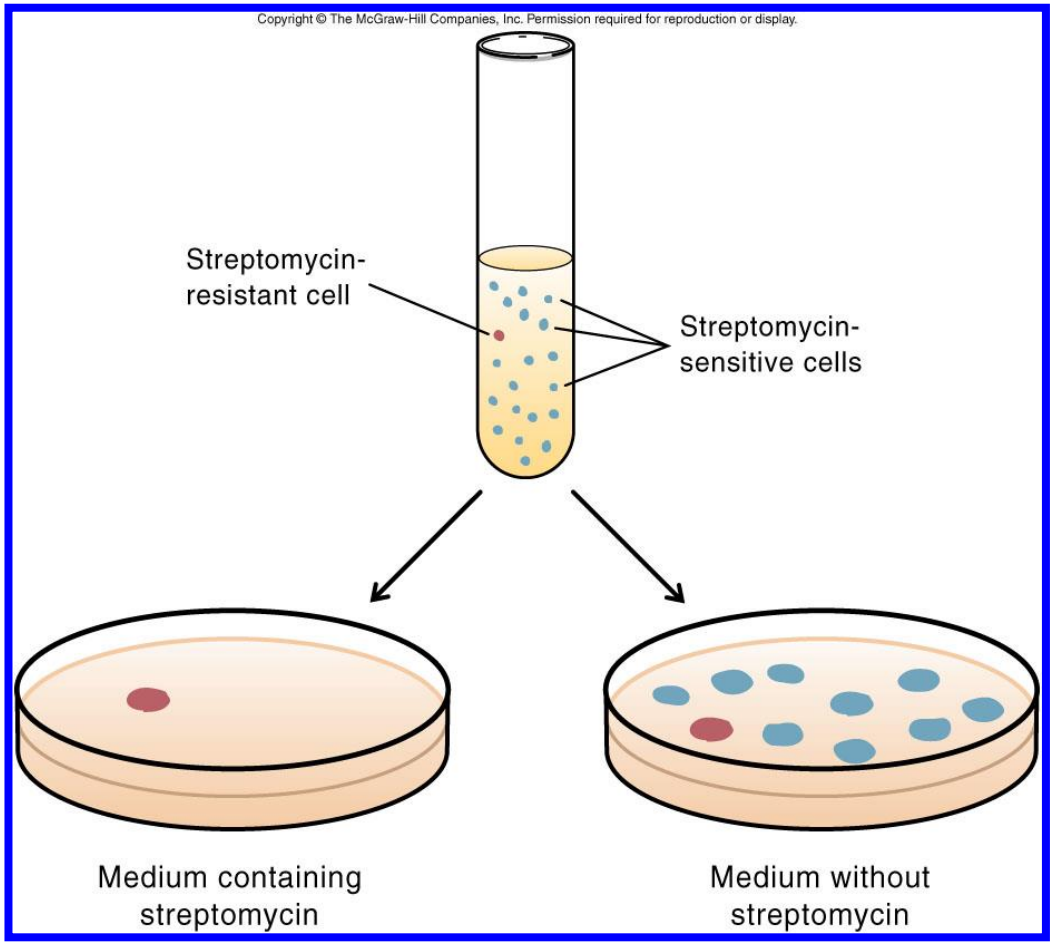
Identifying mutants

- Mutants can be detected by selecting or testing for an altered phenotype.

1. Positive selection involves the selection of mutant cells and rejection of nonmutated cells.
2. Replica plating is used for negative selection- to detect, for example, auxotrophs that have nutritional requirements not possessed by the parent (nonmutated) cell.

Mutation

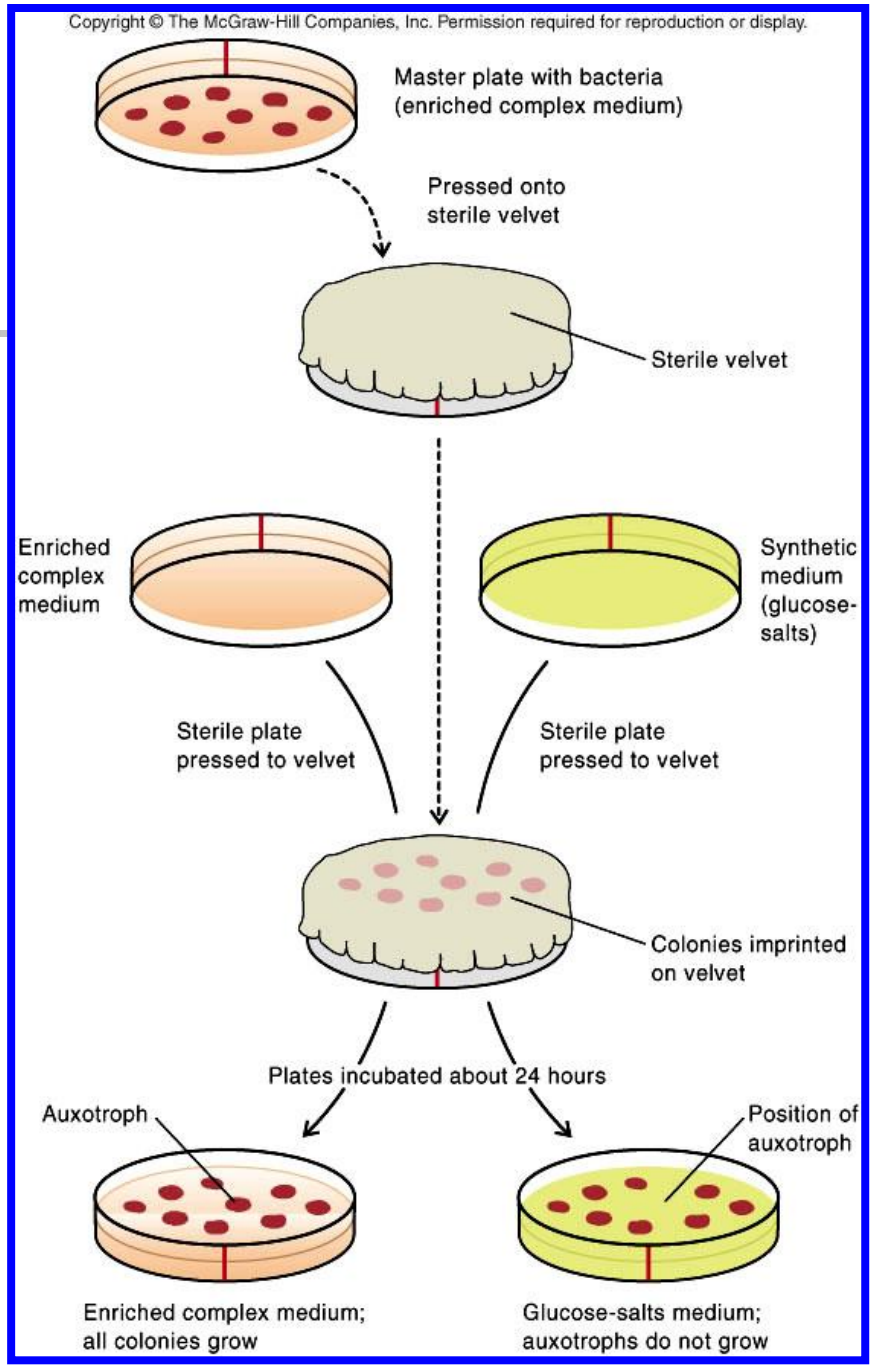
Direct selection for mutants



Mutation

Indirect selection for mutants

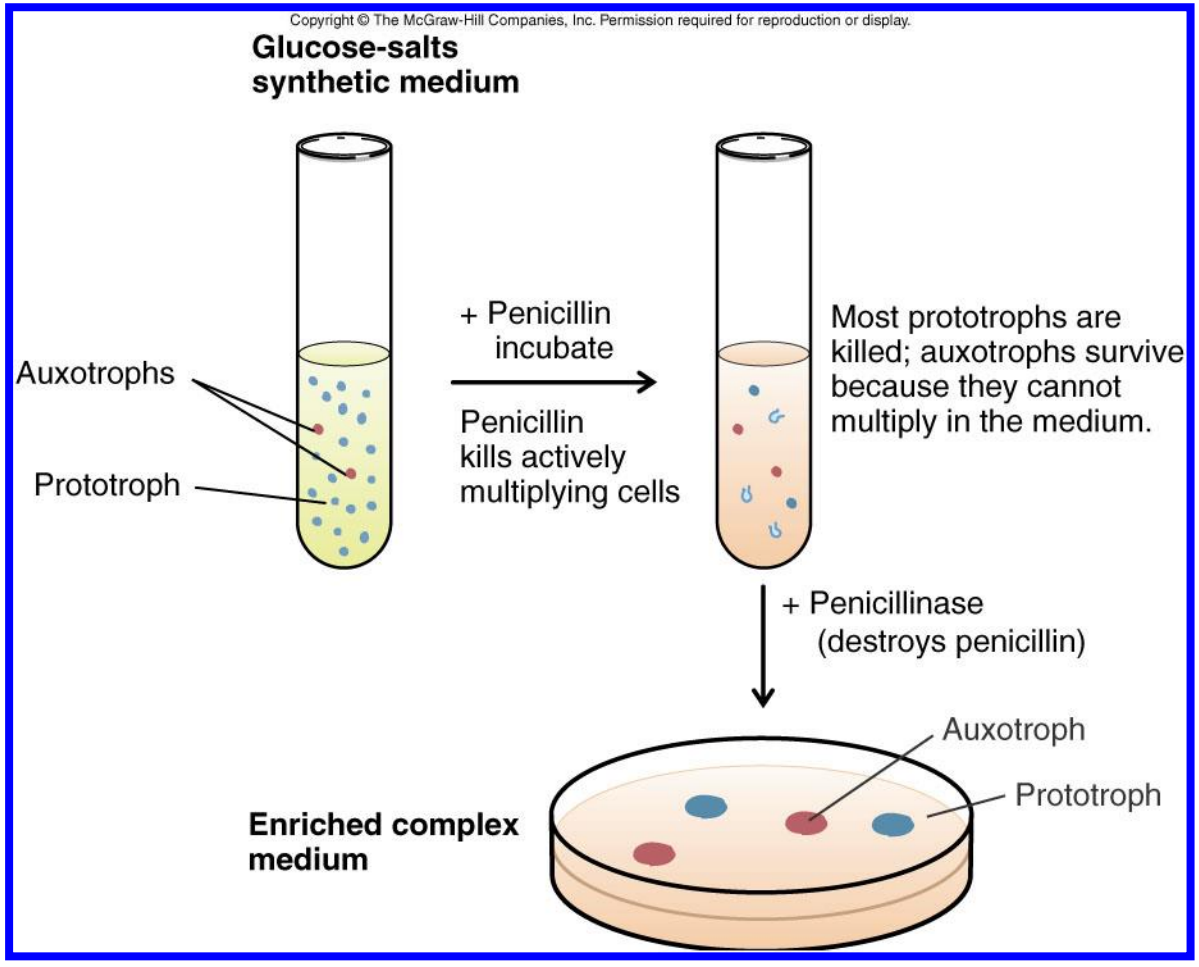
Replica plating



Mutation

Indirect selection for mutants

Penicillin enrichment



Mutation

Indirect selection

Lethal mutation

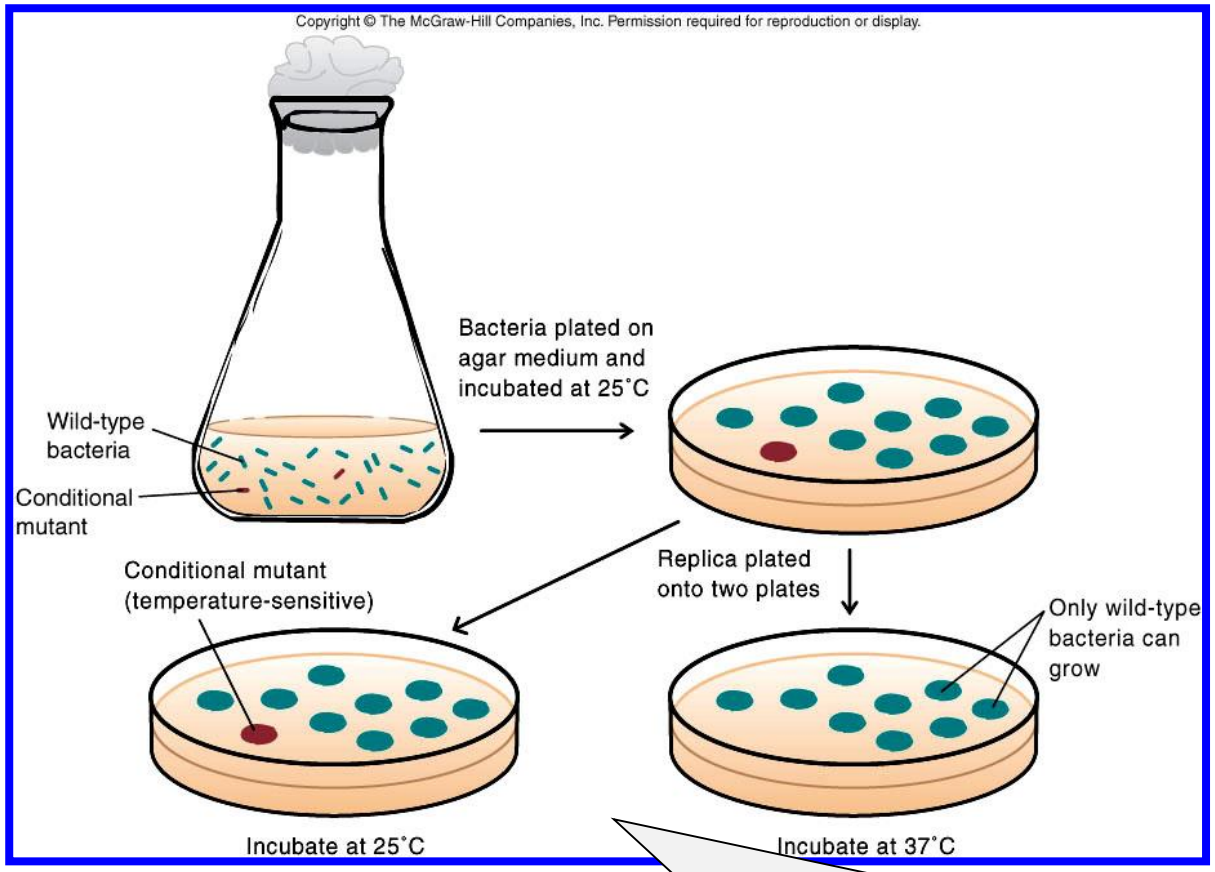
- Mutation which affects vital functions resulting in the death of the organism – (i.e. nonviable mutation).

1. A conditional lethal mutant may be able to live under certain conditions–permissive (not restricted, free) conditions.
2. Commonest type of conditional mutant is the temperature sensitive (t_s) mutant which is able to live at the permissive temperature of 35°C but not at the restrictive temp (39°C).

Mutation

Indirect selection

Lethal mutation



This is one example of isolation of mutants carrying conditionally lethal mutations found in essential genes.

Repair of mutated DNA

1. Nucleotide excision repair can enzymatically repair mutational damage to DNA by cutting out and replacing the damaged portion.
2. Photoreactivation enzymes (light-repair enzymes; photolyases) can repair thymine dimers caused by exposure to ultraviolet radiation in the presence of visible light.

Part II
Recombinant DNA technology

Isolate, cut and recombine DNA from different species

Restriction enzymes ⟹ Cloning vectors ⟹ DNA ligases

Background

- In 1978, Genentech made the first important contribution of recombinant DNA technology to medicine.
- The company succeeded in cloning the human insulin gene into a bacterial plasmid vector that allowed the human insulin protein to be grown up in *E. coli*.

Recombinant DNA technology
Basic tools

- The basic tools, which required to produce recombinant DNA molecules are:

1. Restriction enzymes: Cut DNA at unique restriction sites.
2. Cloning vectors (plasmids or viruses): Vehicles that transport genes into host cells for cloning.
3. DNA ligases: Join the ends of two DNA fragments (joining of sugar to phosphate in DNA fragments).

Recombinant DNA technology
Common steps

- Introduction of recombinant DNA molecule into an appropriate host cell via transformation or transduction.
  - Each cell receiving rDNA = Clone.
  - May have thousands of copies of rDNA molecules/cell after DNA replication.
  - As host cell divides, rDNA partitioned into daughter cells.

Recombinant DNA technology
Common steps

- DNA molecules are digested with enzymes called restriction endonucleases which reduces the size of the fragments → Renders them more manageable for cloning purposes.
- These products of digestion are inserted into a DNA molecule called a vector → Enables desired fragment to be replicated in cell culture to very high levels in a given cell (copy #).

The vector carrying the gene of interest, allows the foreign gene to be replicated, transcribed, and translated into a protein by bacteria (usually *E. coli*).

Recombinant DNA technology

- Production of a unique DNA molecule by joining together two or more DNA fragments not normally associated with each other.
- DNA fragments are usually derived from different biological sources.
- Therefore, recombinant DNA technology is a set of tools that allows molecular biologists to create novel DNA molecules that do not exist in nature, and to produce large quantities of such recombinant DNAs for research and therapeutic purposes.

Wikipedia

Recombinant DNA technology

- A body of techniques for cutting and splicing together different pieces of DNA.
- When segments of foreign DNA are transferred into another cell or organism, the protein for which they code may be produced along with substances coded for by the native genetic material of the cell or organism.
- These cells become "factories" for the production of the protein coded for by the inserted DNA.

Applications of recombinant DNA technology

- Pharmaceuticals (e.g., insulin)
- Specialized bacteria (e.g., for sewage treatment)
- Agricultural products (e.g., herbicide-resistant crops)
- Oligonucleotide drugs
- Genetic testing
- Gene therapy
- Gene mapping
- DNA fingerprinting

In the fields genetics and evolutionary computation, a locus (plural loci) is the specific location of a gene or DNA sequence on a chromosome.
The ordered list of loci known for a particular genome is called a genetic map.
Genetic mapping is the process of determining the locus for a particular biological trait.

Recombinant DNA technology

Escherichia coli: An ideal bacterium for recombinant DNA technology

- *Escherichia coli* is an ideal organism for growing up recombinant ideal organism DNA molecules because:

1. It is easy to culture;
2. Easily grown through asexual cycles in as little as 20 minutes;
3. Generally has just one gene per trait;
4. Can have colonies representing over 100 individuals displayed on a single Petri plate;
5. Can participate in gene exchange by conjugation, transformation, and transduction;
6. Serves as host to numerous viruses and plasmids.
7. As it turns out, plasmids (as well as viruses) make excellent vectors for carrying foreign genes into bacteria.

1. Restriction Endonucleases

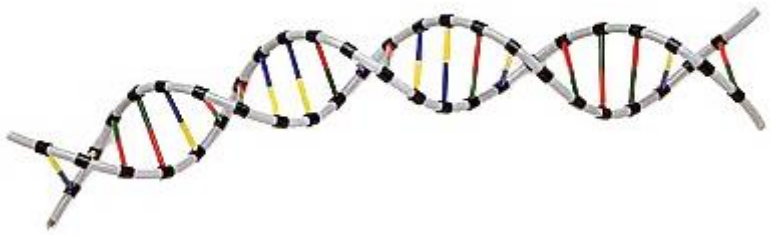
Thermophilic bacteria growth curves and a particular condition for restriction enzymes production.

Restriction Endonucleases
Molecular scissors

- The major tools of recombinant DNA technology are restriction enzymes (nucleases).
  - First discovered in the late 1960s.
  - They work by cutting up the foreign DNA, a process called restriction.
  - Most restriction enzymes are very specific.
  - Recognizing short, specific nucleotide sequences in DNA molecules and cutting at specific points within these sequences.
  - There are hundreds of restriction enzymes and more than 150 different recognition sequences.

Restriction Endonucleases

Molecular scissors



- Restriction Enzymes scan the DNA sequence.
- Find a very specific set of nucleotides.
- Make a specific cut.

Restriction Endonucleases

Molecular scissors

Restriction endonucleases

- Recognize specific base sequences in DNA
- Cut DNA at those recognition sites

Restriction Enzymes

- Nucleases are further described by addition of the prefix "endo" or "exo" to the name:

1. Endonuclease applies to sequence specific nucleases that break nucleic acid chains somewhere in the interior, rather than at the ends, of the molecule.
2. Exonucleases function by removing nucleotides from the ends of the molecule.

Restriction Endonucleases

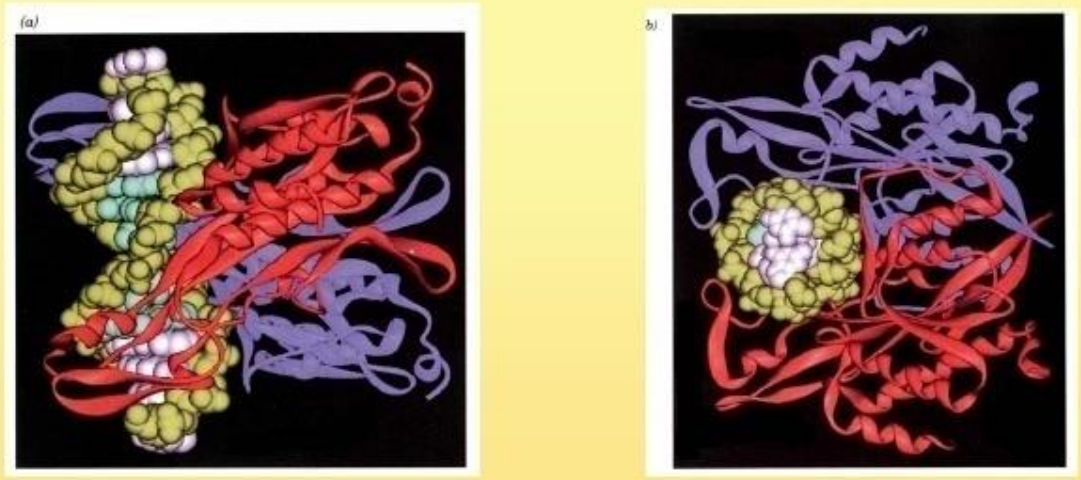
Molecular scissors

- Over 3000 have been identified.
- More than 600 available commercially.
- Routinely used for DNA modification and manipulation in laboratories.
- http://en.wikipedia.org/wiki/Restriction_enzyme

A restriction endonuclease
Structure

Structure of a restriction endonuclease

- Restriction enzymes are usually dimers of identical subunits, analogous to the symmetry of their binding sites in DNA.



Restriction Endonucleases
Molecular scissors

- The first endonucleases discovered was from *Escherichia coli* EcoRI.
- Named after the organism from which they were derived:
  - EcoRI from *Escherichia coli*
  - BamHI from *Bacillus amyloliquefaciens*
- Protect bacteria from bacteriophage infection.
  - Restricts viral replication
- Bacterium protects it's own DNA by methylating those specific sequence motifs.

Motif: A sequence of amino acids or nucleotides that perform a particular role and is often conserved in other species or molecules.

Types of Restriction Enzymes

Four classes of restriction endonucleases

- There are four classes of restriction endonucleases: types I, II, III and IV.
- All types of enzymes recognise specific short DNA sequences and carry out the endonucleolytic cleavage of DNA to give specific double-stranded fragments with terminal 5'-phosphates.
- They differ in their:

1. recognition sequence,
2. subunit composition,
3. cleavage position, and
4. cofactor requirements.

Types of Restriction Enzymes

Type II enzymes, which are predominantly used in biotechnology

- **Type I:** Recognize specific sequences but then track along DNA (~1000-5000 bases) before cutting one of the strands and releasing a number of nucleotides (~75) where the cut is made. A second molecule of the endonuclease is required to cut the 2nd strand of the DNA.
- **Type II:** Recognize a specific target sequence in DNA, and then break the DNA (both strands), within the recognition site.
- **Type III:** Intermediate properties between type I and type II. Break both DNA strands at a defined distance from a recognition site.
- **Type IV:** target methylated DNA.

Types of Restriction Enzymes

Some RE types

TYPES AND ACTIVITIES OF RESTRICTION ENZYMES

Type I
Cleaves DNA at random sites far from its recognition sequence

Type II
Cleaves DNA at defined positions close to or within its recognition sequence

Type IIG
Cleaves outside its recognition sequence with both REase and MTase enzymatic activities in the same protein

Type IIP
Cleaves symmetric targets and cleavage sites

Type IIS
Recognizes asymmetric sequences

Type III
Cleaves outside its recognition sequence and require two sequences in opposite orientations within the same DNA

Type IV
Cleaves modified (e.g., methylated) DNA

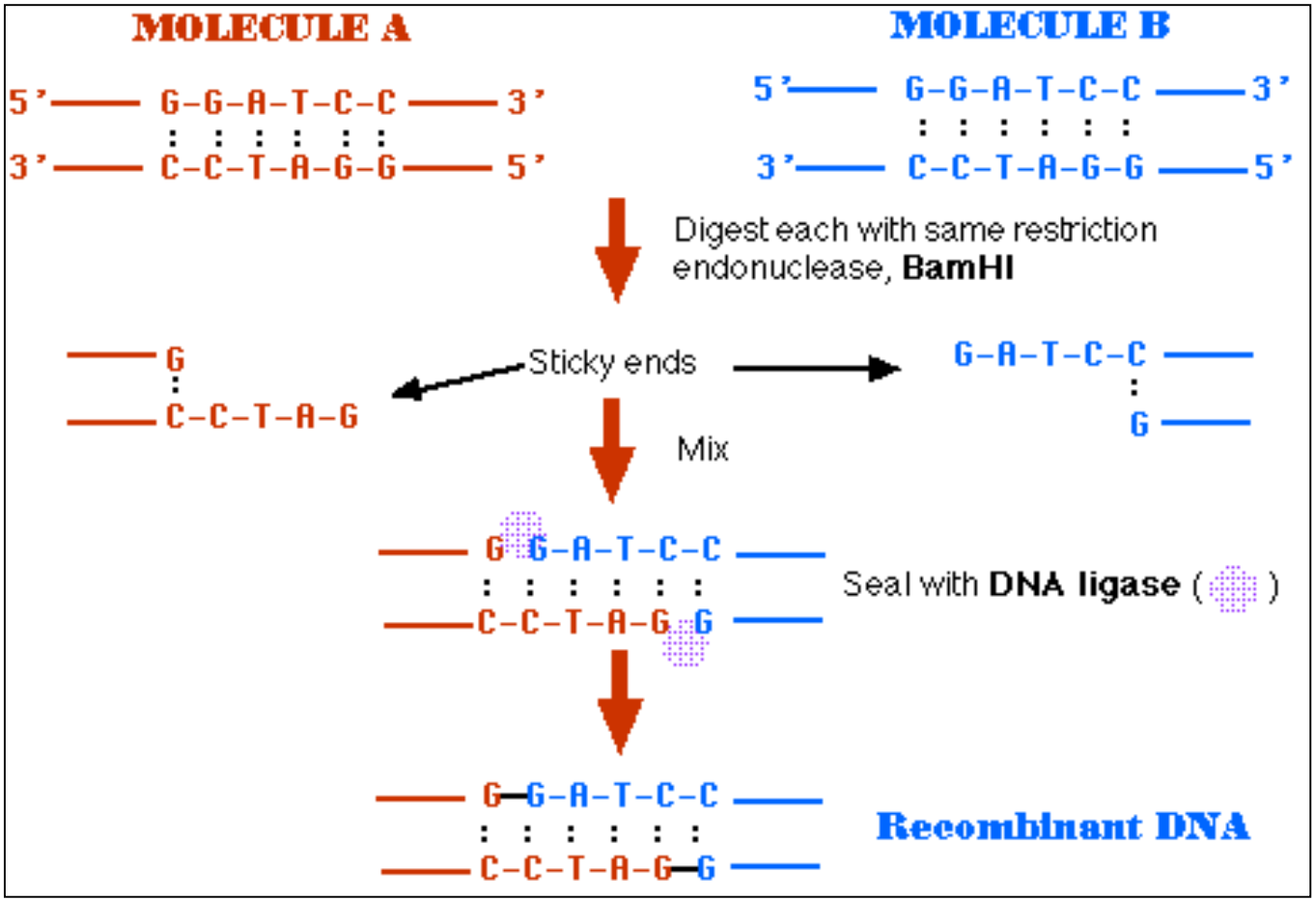
Restriction Endonucleases

Mode of action

- A restriction endonuclease functions by "scanning" the length of a DNA molecule.
- Once it encounters its particular specific sequence (recognition site), it will bond to the DNA molecule and makes one cut in each of the two sugar-phosphate backbones of the double helix.
- Once the cuts have been made, the DNA molecule will break into fragments.

Restriction Endonucleases

Blunt and Sticky ends

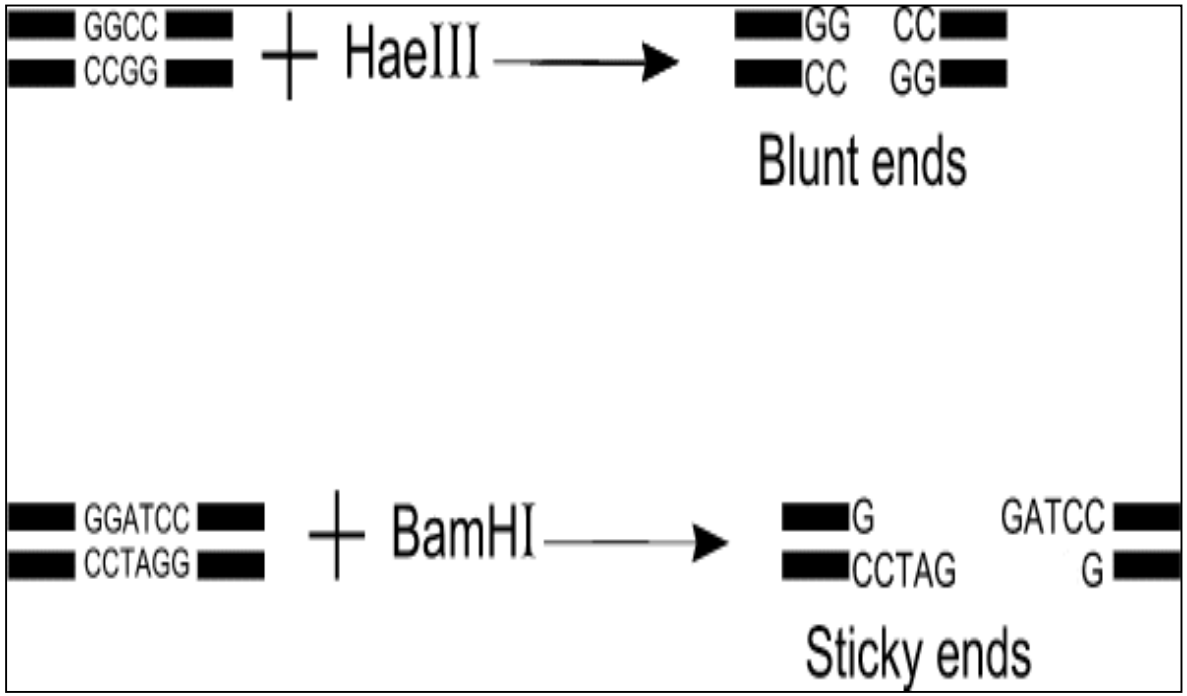


P. Mclaughlin

Restriction Endonucleases

Blunt and Sticky ends

HaeIII and BamHI



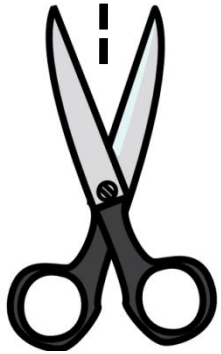
Restriction Endonucleases

Blunt ends

HaeIII

- Once the recognition site was found HaeIII could go to work cutting (cleaving) the DNA.

5’ TGACGGGTTCGAGGCCAG 3’
3’ ACTGCCCAAGGTCCGGTC 5’



Restriction Endonucleases

Blunt ends

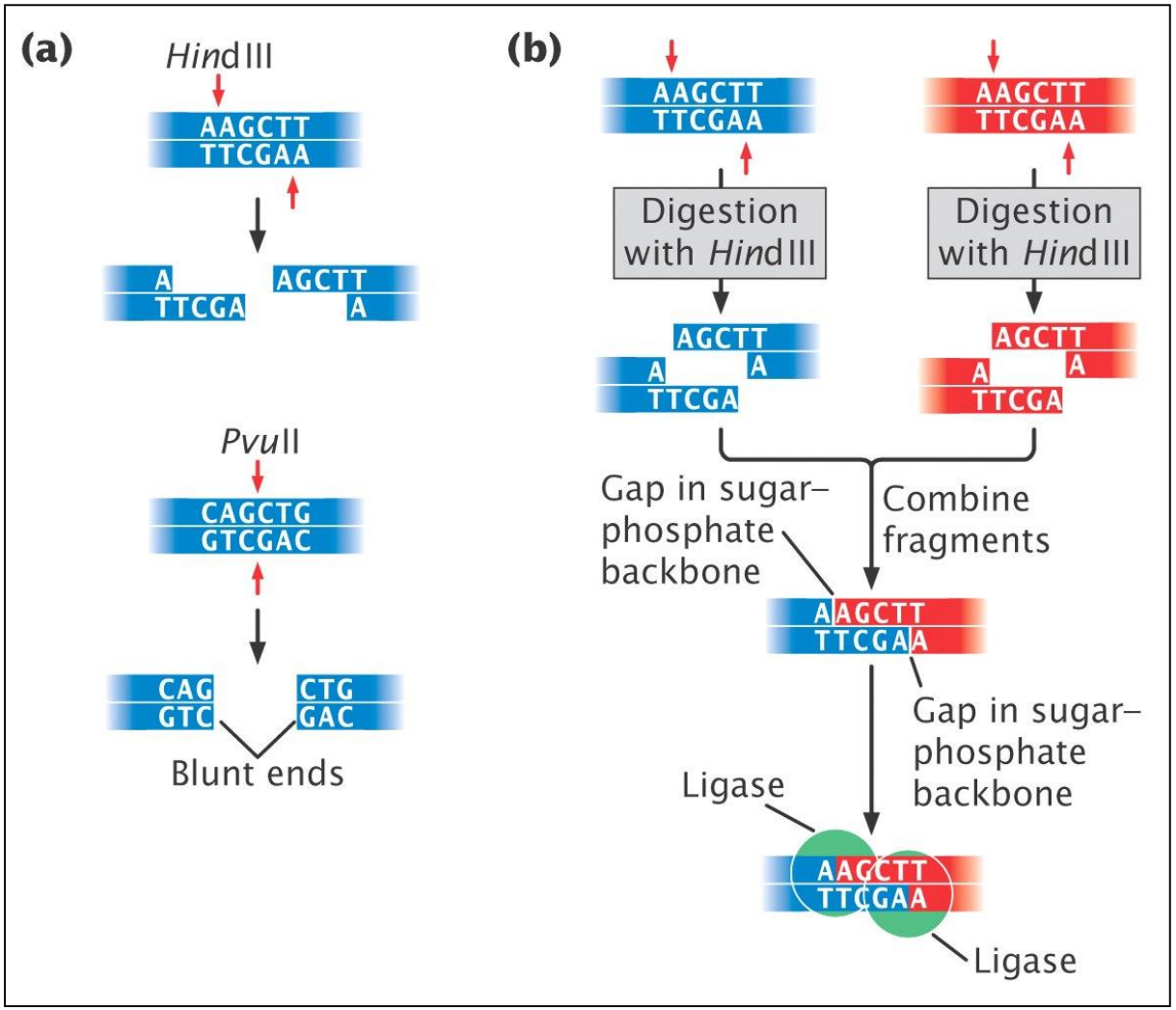
HaeIII

- These cuts produce what scientists call "blunt ends".

5' TGACGGGTTCGAGG	CCAG 3'
3' ACTGCCCAAGGTCC	GGTC 5'

Products generated by restriction enzymes

HindIII and PvuII



Restriction enzymes

How restriction enzymes are named

- The restriction enzymes are named by the following convention:

1. The first letter of the name comes from the genus of the organism they are isolated from
2. The second two letters of the name come from the species of the organisms from which they are isolated
3. The next letter refers to the name of the strain.
4. Roman number at end refers to the order in which enzyme was isolated from bacteria. E.g. EcoRI:

- E= Escherichia (genus),
- co= coli (species),
- R= RY13 (strain) and
- 1 (first identified).

Restriction Enzymes

Characterisitcs of some common type II enzymes

Table 18.2 Characteristics of some common type II restriction enzymes used in recombinant DNA technology

Enzyme	Microorganism From Which Enzyme Is Isolated	Recognition Sequence	Type of Fragment End Produced
*Bam*HI	*Bacillus amyloliquefaciens*	5′–G↓GATCC–3′ 3′–CCTAG↑G–3′	Cohesive
*Cof*I	*Clostridium formicoaceticum*	5′–GCG↓C–3′ 3′–C↑GCG–5′	Cohesive
*Dra*I	*Deinococcus radiophilus*	5′–TTT↓AAA–3′ 3′–AAA↑TTT–5′	Blunt
*Eco*RI	*Escherichia coli*	5′–G↓AATTC–3′ 3′–CTTAA↑G–5′	Cohesive
*Eco*RII	*Escherichia coli*	5′–↓CCAGG–3′ 3′–GGTCC↑–5′	Cohesive
*Hae*III	*Haemophilus aegyptius*	5′–GG↓CC–3′ 3′–CC↑GG–5′	Blunt

Type II restriction endonucleases

Type II restriction enzyme

EcoRI

- Many Type II restriction endonucleases recognize palindromic sequences i.e. symmetrical sequences which read in the same order of nucleotide bases on each strand of DNA (always read 5' to 3').

- For example, *Eco*RI recognizes the sequence:

  5'-G A A T T C-3'

  3'-C T T A A G-5'

- The high specificity for their recognition site means that DNA (target sequence and cloning vector) will always be cut reproducibly into defined fragments (important for molecular cloning).

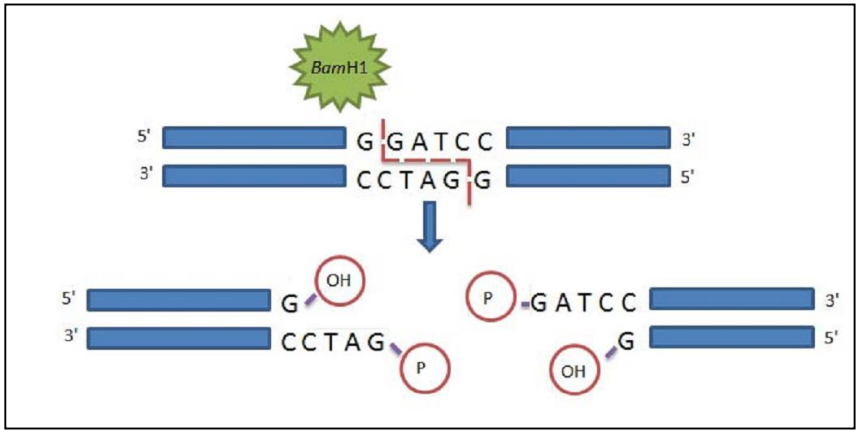
The DNA sequence ACCTAGGT is palindromic because its nucleotide-by-nucleotide complement is TGGATCCA. Reading 5' to 3' forward on one strand matches the sequence reading backward 5' to 3' on the complementary strand.

Restriction Enzyme Digestion

Type II restriction enzyme

Bam H1 enzyme

- Every restriction enzyme has a specific restriction sit e at which it cuts a DNA molecule. For example restriction sequence for *Bam* HI is GGATCC (type II restriction enzyme).
- The most abundantly used restriction enzymes are type II restriction enzymes which cleave at specific restriction site only.



Restriction Enzyme Digestion

BamH1 enzyme

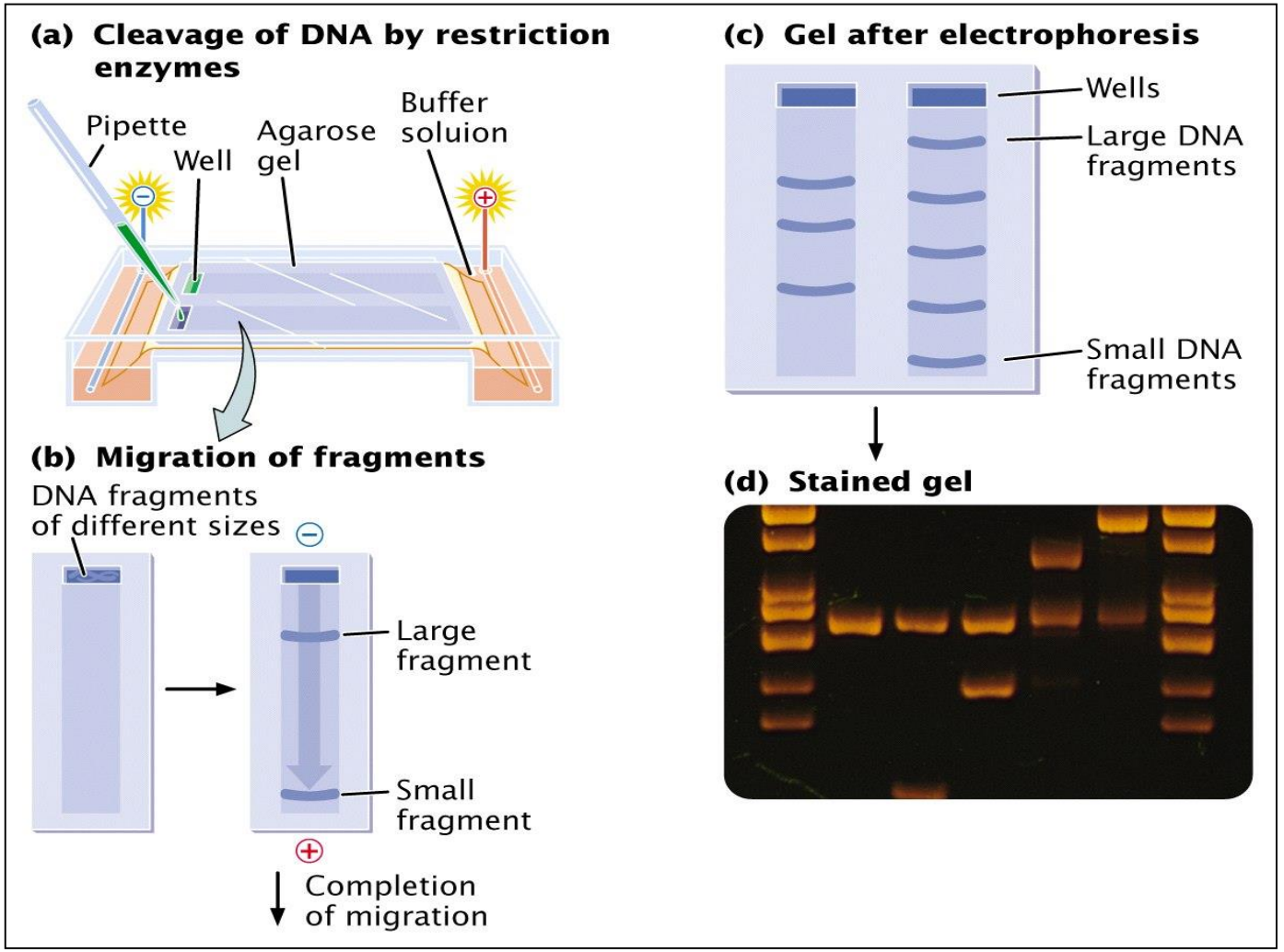
Procedure

- Take 1.5 µg of PUC18 DNA (10µl) in a fresh eppendorf.
- To this, add 11.5 µl of sterile water followed by 5 µl of 10X buffer.
- Add 1.5 µl of *BamH1* enzyme (1 units) and incubate the mixture at 37˚C for 2 hrs.
- Prepare 0.7% agarose gel and load the samples including 1 Kb DNA ladder, undigested pUC18 DNA and *BamH1* digested PUC18 DNA.
- Run the gel at 100 V for 1 hr.
- Visualize the gel under UV illuminator.
- 10ul of the sample and 2µl of the dye were mixed.
- Laod 10µl of this into the gel.

Restriction enzyme Digestion

Electrophoresis

Viewing DNA fragments



Frequency of cutting

- Because of their restriction site specificity, the restriction endonucleases cut DNA into fragments whose average length is determined by the number of base pairs in the restriction site (and to a lesser extent by the ratio of bases in the DNA).
- For DNA that has equal amounts of all four bases, each base has a probability of 1/4 at any particular position in the DNA strand.

Frequency of cutting

- For a restriction site of 4 base pairs, the probability of random occurrence of that sequence is (1/4)(1/4)(1/4)(1/4) = 1/256.
- For 6 base pairs, the probability is 1/4,096, and for 8 base pairs it is 1/65,536.

Frequency of cutting

Frequency of cutting

Average distance between cuts is:

$$4^n$$

where "*n*" is number of bp's in recognition site.

- 4-base cutter: 4^4 = 256 bp
- 5-base cutter: 4^5 = 1,024 bp
- 6-base cutter: 4^6 = 4,096 bp
- 8-base cutter: 4^8 = 65,536 bp

•>2000 different restriction enzymes, > 200 Δ rec. sequences.

2. Cloning DNA

DNA Vectors for Gene Cloning

In DNA cloning, the plasmid or phage chromosome used to carry the cloned DNA segment

Cloning DNA

From the Greek - klon, a twig

- Inserting DNA into a plasmid - Joining with ligase.
- Plasmid mixed with host bacteria and enters cell by transformation.
- Screen bacteria to find cells with correct gene for cloning.
- Genes cloned as host cell divides.

Whole genome sequencing (WGS)

BAC-by-BAC (clone-by-clone shotgun strategy) and Whole Genome Shotgun (WGS) sequencing

- To read the DNA, the chromosomes are cut into tiny pieces, each of which is read individually. When all the segments have been read they are assembled in the correct order.
- There are essentially two ways to sequence a genome:

1. BAC-by-BAC strategy (also known as hierarchical shotgun sequencing or a clone-by-clone shotgun strategy);
2. Whole Genome Shotgun (WGS) sequencing.

See also Bacterial Diagnosis-Part I and Bacterial Management-Part I files.

Trivedi,2000;..

Cloning DNA

Sequence Libraries

Genomic library vs. cDNA library

- To find or isolate a gene or promoter you need a library.
- Library is a collection of sequences.
- Libraries are of two general types:

1. Genomic library – goal is to have all the sequence information in the genome represented in the library.
2. cDNA library - cDNA (complementary DNA) is a very selective library and contains only expressed genomic information (only exons).

Cloning DNA

Sequence Libraries

Genomic library vs. cDNA library

- **Genomic library:**
- Organism specific and represents the whole genetic material.
- Contain expressed genes, non-expressed genes, exons and introns, promoter and terminator regions, but mostly repetitive DNA.
- **cDNA library:**
- Sequences are obtained by the conversion of mRNA to cDNA.
- Contains coding sequences including 5′ and 3′ untranslated regions.
- no intergenic regions.

Genomic libraries

Gene Library

- A gene library is a collection of host cells that contain all of the genomic DNA of the source organism.
- A gene library is also called a clone bank, or a gene bank.

1. Genomic libraries

- A genomic library consists of a population of *E. coli* cells, each of which contains a vector with a different random fragment of the genome of the organism that donated the DNA.

Genomic libraries

- Collections of cloned sequences.
- Can represent:

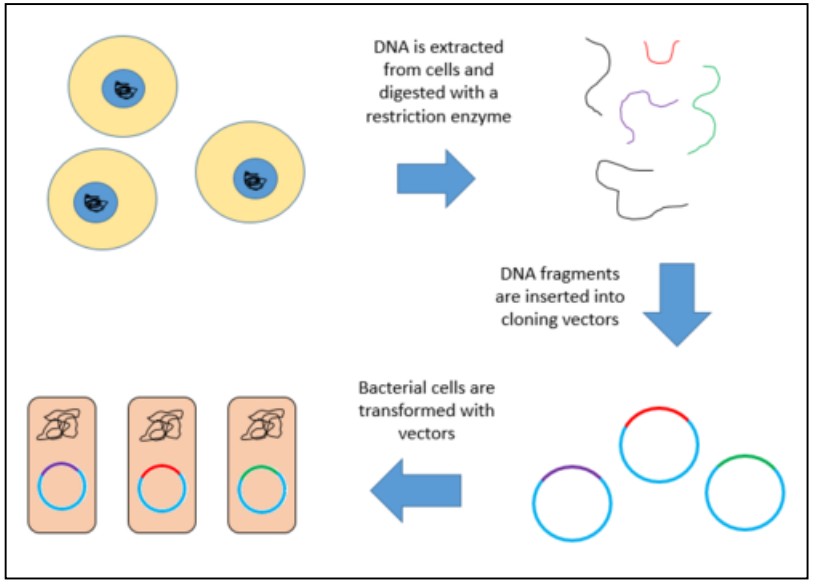
1. an entire genome,
2. a single chromosome, or
3. the set of genes expressed by a particular cell type.

Creating a gene library

- The source organism's genomic DNA is digested (cut) into clonable elements and inserted into host cells.
- Conditions of the digestion reaction are set to give a partial, not a complete, digestion.

Create a genomic library

1. Extract and purify DNA.
2. Digest the DNA with a restriction enzyme.
3. Insert the fragments of DNA into vectors that were cut with the same restriction enzyme.



Whole genome sequencing (WGS)

Making a BAC library

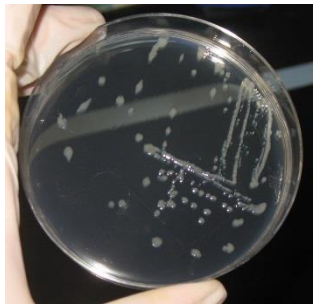
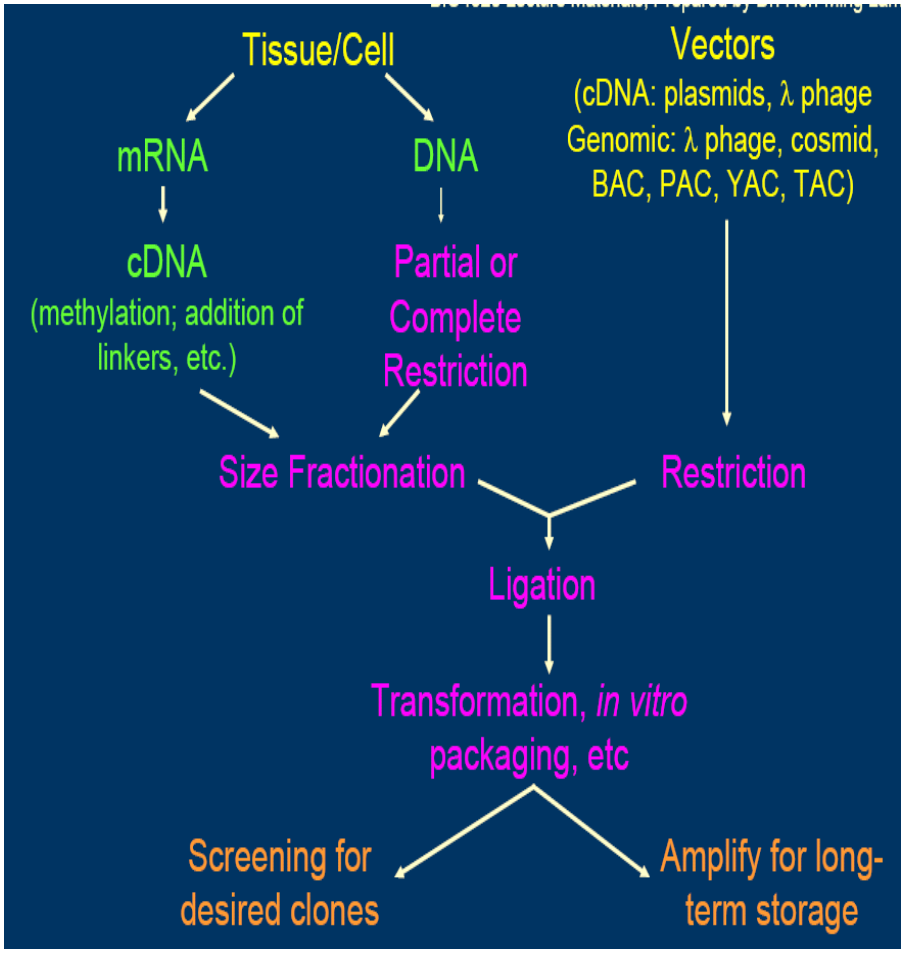


Plate growing *Escherichia coli* colonies

- Four key stages in the process of creating a BAC library:

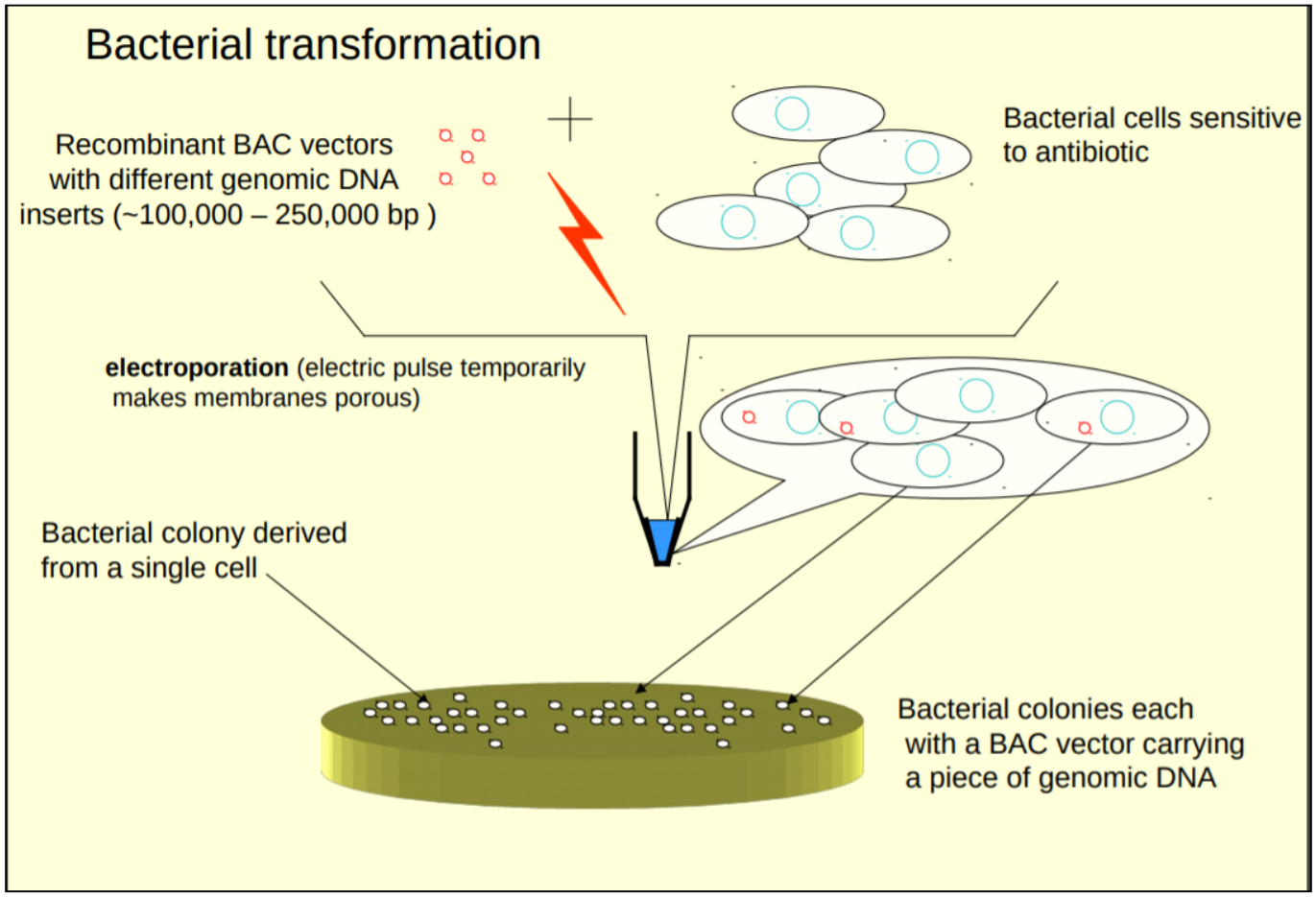
1. Extracting DNA from white blood cells and using restriction enzymes to "cut up" the DNA into smaller fragments.
2. Gel electrophoresis to separate and select DNA fragments by size.
3. Creating BAC clones by inserting DNA fragments into vectors and transferring them into bacterial cells.
4. Selection of the BAC clones for the library.

Create a genomic library

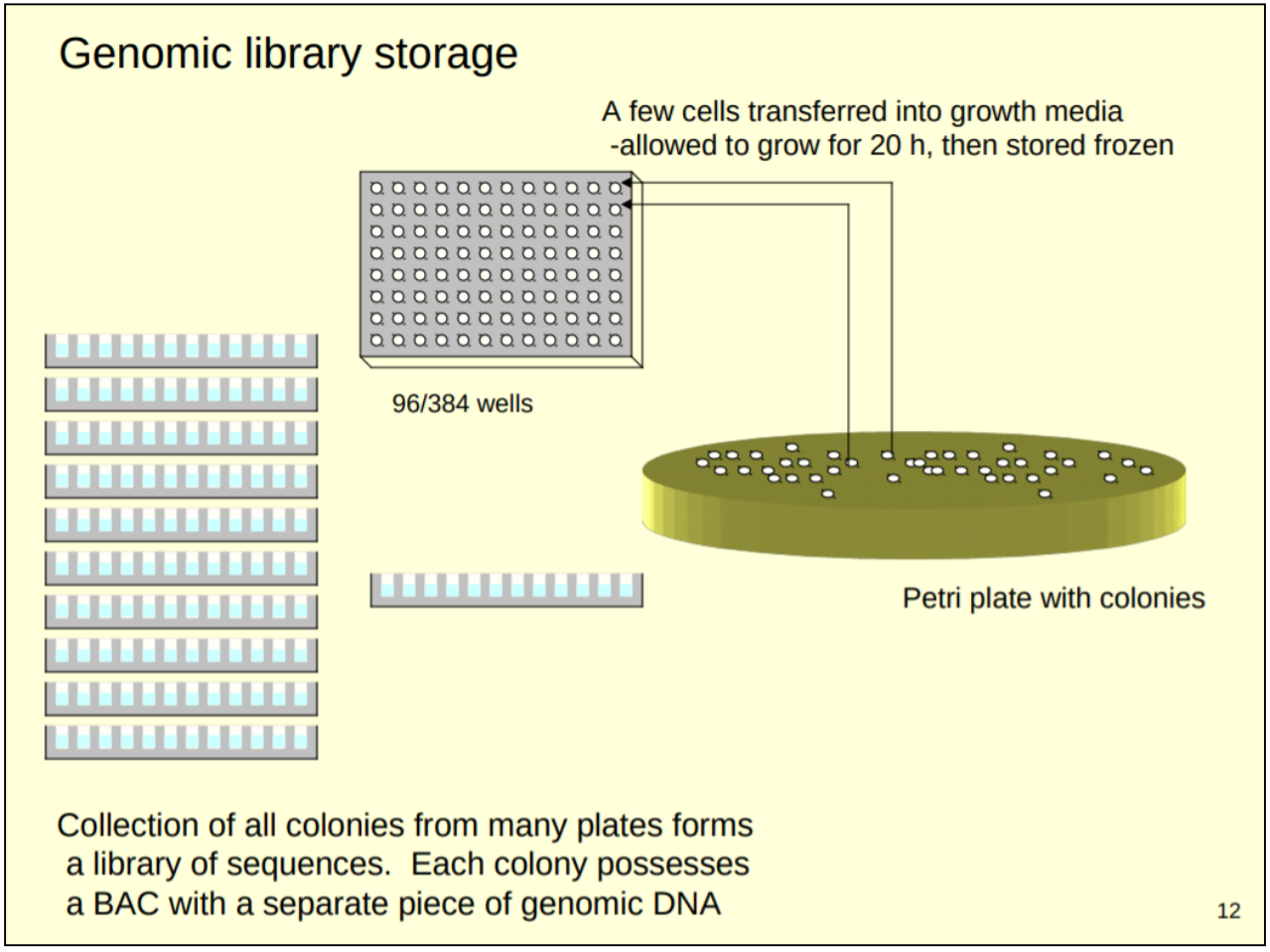


Tal H Hasson

Create a genomic library
Making a BAC library

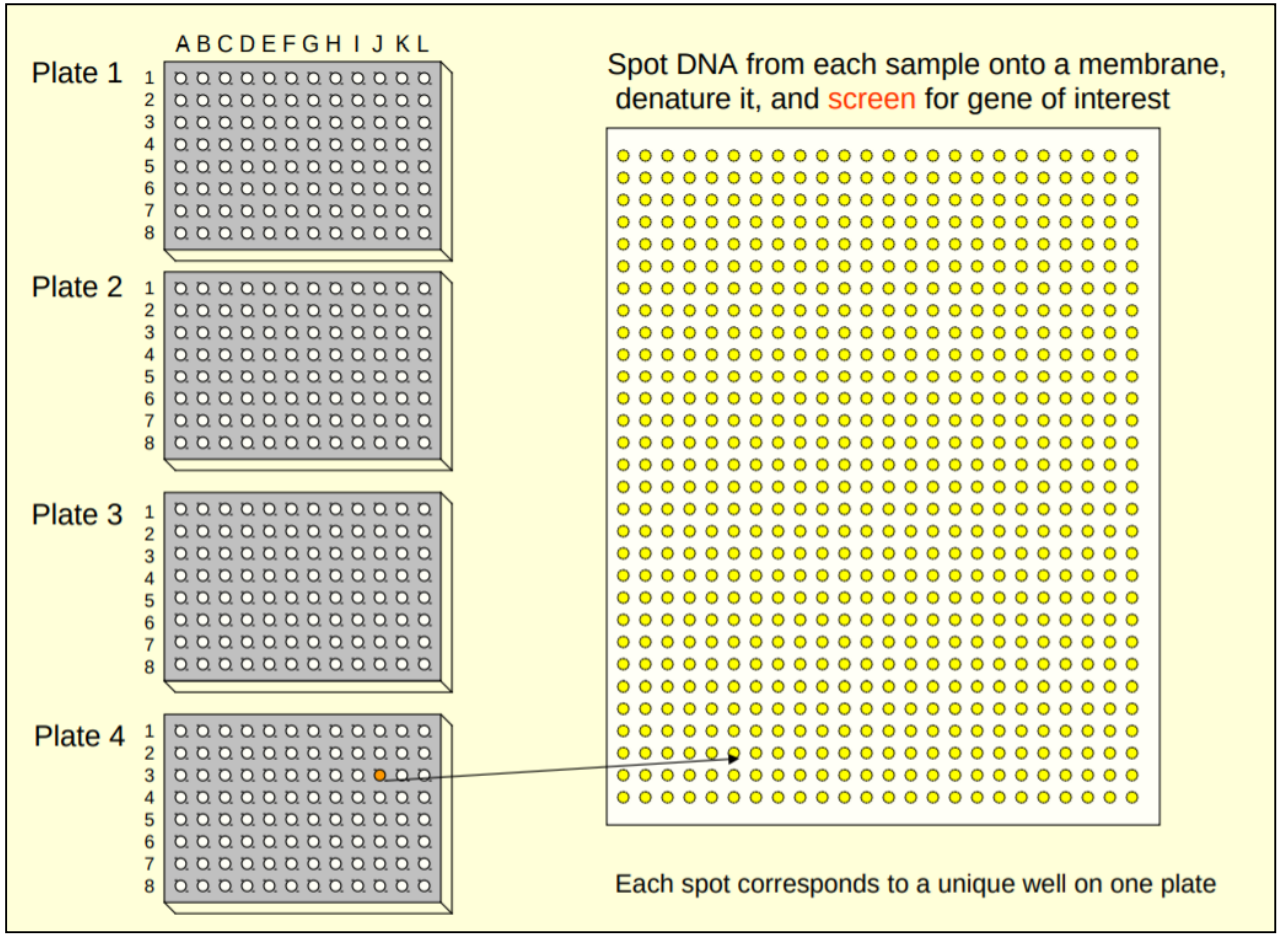


Create a genomic library
Making a BAC library



Create a genomic library

Making a BAC library



PLNT2530 (2018)

Create a genomic library

Making a BAC library

Uses

Genomic libraries

Uses

-Mapping genomes by finding clones representing contiguous pieces

-Providing clones for sequencing entire genomes

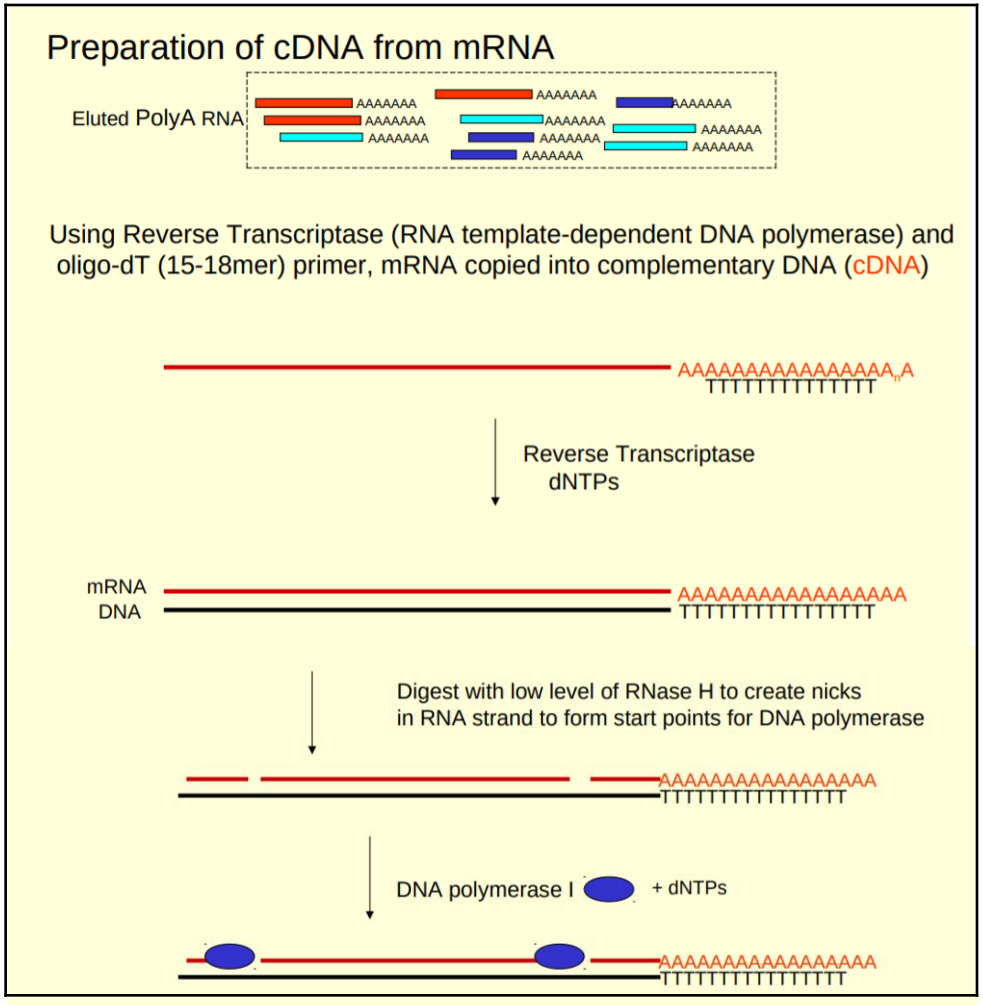
-Finding promoters for a gene when you have the coding sequence

2. Create a cDNA library

- Made from mRNA by reverse transcription of mRNA into complementary DNA (cDNA).
- Since transcribed sequences represents only small percentage of genome, only the small percentage of the genome that is transcribed will appear in the cDNA library.
- Thus a cDNA Library is far more selective if we are looking for a specific gene than is a Genomic Library.

2. Create a cDNA library

Preparation of cDNA from mRNA



2. Create a cDNA library

Packing the cDNA

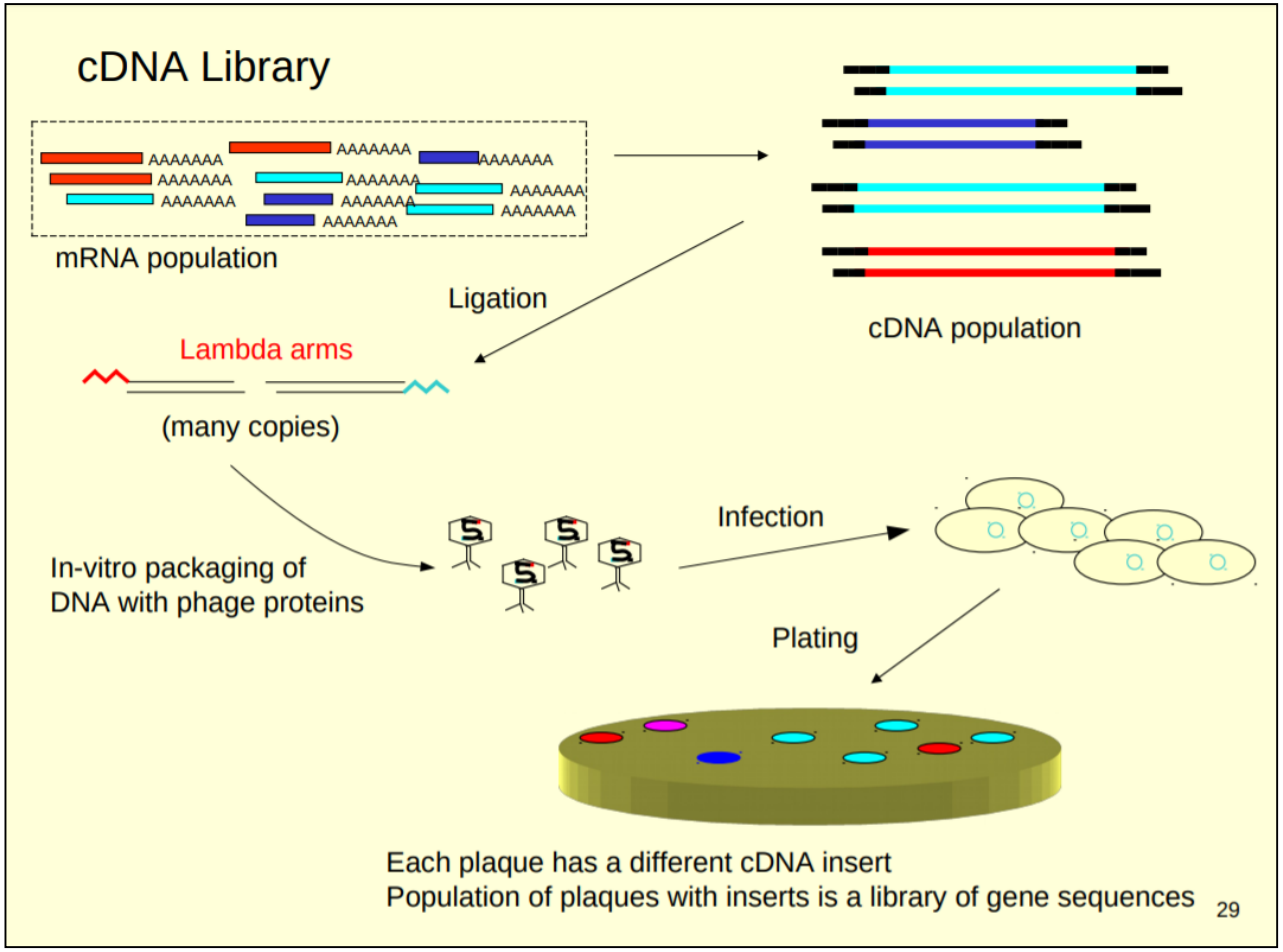
cDNA can be stored in plasmids or phages

- The final step is to ligate the sticky ends of the cDNA with the λ-phage arms that have complementary sticky ends, thereby inserting the Double strand cDNA into the vector.

2. Create a cDNA library

Preparation of cDNA from mRNA

cDNA can be stored in plasmids or phages



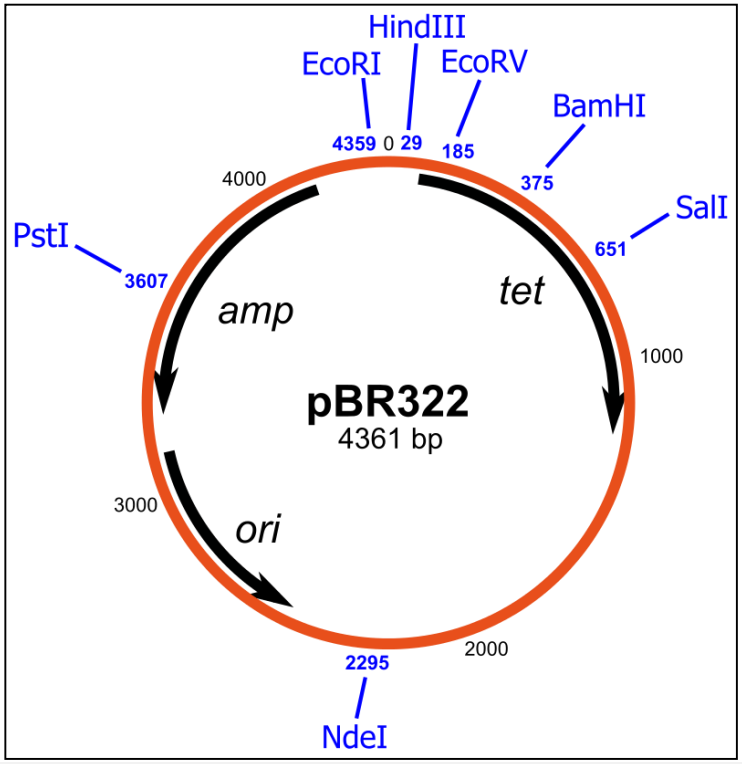
Cloning vectors
Reasons for a vector

- Allow for transport of individual fragments
  - Can get into its host
- Allow for replication of individual fragments
  - Can replicate in its host
- Allow for purification of individual fragments
  - Can extract it from its host

Cloning vectors

Properties of cloning vectors

- Able to independently replicate themselves (and the introduced DNA).
- Contain a number of unique restriction endonuclease recognition sites.
- Carry a selectable marker (e.g., an antibiotic resistance gene).
- Easy to recover from the host cell.



Schematic representation of the pBR322 plasmid, one of the first plasmids widely used as a cloning vector.

Plasmid vectors are typically engineered to contain two things: (1) a selectable marker such as an antibiotic-resistance gene to allow selection of transformed cells and (2) a multiple cloning site where a gene of interest can be inserted.

Cloning vectors

Main classification of vectors

On the basis of host cell used

- After construction of a recombinant DNA these can be introduced into a host cell.
- So depending on the host cell the vectors are designed and constructed.
- All the parts of the vectors must be functionally compatible with the host.
- For example, if we are making a vector for a bacterial host, it must have a suitable origin of replication which will be functional in a bacterial cell.

Choice of vector

Main classification of vectors

On the basis of host cell used

- Depends on nature of protocol or experiment,
- Type of host cell to accommodate rDNA.
  - Prokaryotic
  - Eukaryotic
- These include:

1. vectors for bacteria
2. vectors for yeast
3. vectors for animals
4. vectors for plants

http://www.biotechnologynotes.com;...

Cloning vectors

Main types of vectors

Vector	Vector Host	Insertion size
Plasmids	Bacteria	Up to 15 kb
Bacteria virus (lambda)	Bacteria	Up to 25 kb
Cosmid (engineered vectors that combine characteristics of both plasmids and phage)	Bacteria	30-45 kb
BAC (bacterial artificial chromosome)	Bacteria	100-500 kb
YAC (yeast artificial chromosome)	Yeast	250-1000 kb

Cloning vectors

Main classification of vectors

Vectors for Bacteria

- Vectors are those DNA molecules that can carry a foreign DNA fragment when inserted into it.
- Vectors are also known as vehicle DNAs.
- Based on the nature and sources, the vectors are grouped into:

1. bacterial plasmids,
2. bacteriophages,
3. Cosmids,
4. Fosmid, and
5. Phagemids.

Cloning vectors

Main classification of vectors

Vectors for Bacteria

- These are special bacterial origin of replication and antibiotic resistance selectable markers.
- Bacteria support different kinds of vectors:

1. plasmid vectors (e.g. pBR 324),
2. bacteriophages vectors (e.g. phage lambda(λ),
3. Cosmids (e.g. SuperCos1),
4. Fosmid (e.g. pFOS1),
5. Phagemids (e.g. pUC118, 119 and 120),

Cloning vectors

Main classification of vectors

Vectors for Bacteria

Cloning vectors	Natural occurrence	Size (Kb)*	Selective marker**
Plasmids			
pACYC 177	*Escherichia coli*	3.7	Amp^r, Kan^r
pBR 322	*E. coli*	4.0	Amp^r, tct^r
pBR 324	*E. coli*	8.3,	Amp^r, tet^r, El imm.
pMB9	*E. coli*	5.8	Tcf
pRK 646	*E. coli*	3.4	Amp^r
pC194	*Staphylococcus aureus*	3.6	Ery^r
pSA 0501	*S. aureus*	4.2	Str^r
pBS 161-1	*Bacillus subtilis*	3.65	Tet^r
pWWO	*Pseudomonas putida*	117	Kan^r
Cosmids			
pJC 74	Derived plasmid from colicinogenic factor E1 (Col E1) in *E.coli*	16	Amp^1, El imm.
pJC 720	do	24	El umn./Rif^r
pHC 79	Derivative of pBR 322, widely used E. coli cloning vectors	6	Amp^r, Tet^r
Viruses			
Phage $M13^+$	*E. coli*	6.4	-
Phage /	*E. coli*	4.9	-

Vectors for Bacteria

1. Plasmid vector

- Plasmids are the extrachromosomal, self-replicating, and double stranded closed and circular DNA molecules present in the bacterial cell.
- Naturally occurring plasmids can be modified by *in vitro* techniques.
- Cohen *et al.*, 1973 for the first time reported the cloning DNA by using plasmid as vector.

Vectors for Bacteria
Plasmid vector

- A plasmid can be considered a suitable cloning vehicle if it possesses the following features:

1. It can be really isolated from the cells,
2. It possesses a single restriction site for one or more restriction enzyme (s),
3. Insertion of a linear molecule at one of these sites does not alter its replication properties,
4. It can be reintroduced into a bacterial cell and cells carrying the plasmid with or without the insert can be selected or identified (Bernard and Helinski, 1980),
5. They do not occur free in nature but are found in bacterial cells.

Vectors for Bacteria
Plasmid vector

- **All commercial cloning plasmids have at least 3 components in common:**

1. An origin of replication (Ori):

- This DNA sequence ensures that the plasmid will be recognized by the bacterial replication machinery and replicated along with the bacterial chromosome.

2. A multiple cloning site:

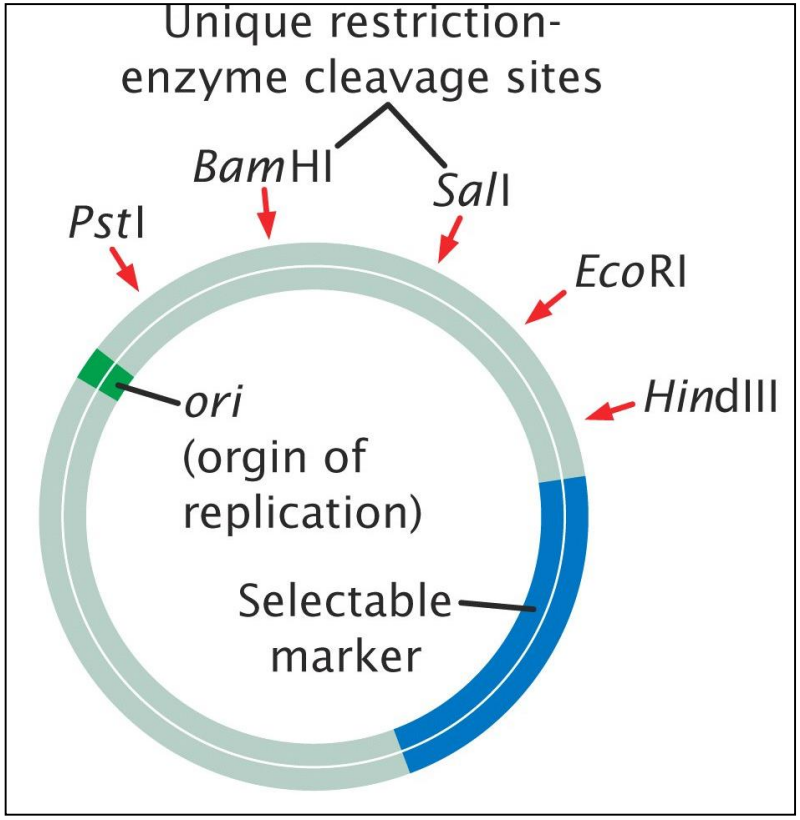
- Restriction enzymes will only cut DNA at specific sites.
- For example, the restriction enzyme "Eco RI" cuts DNA only at the sequence: 5'-GAATTC-3'.

3. A selectable marker:

- The plasmid must carry a selectable marker that allows only those cells that actually pick up the plasmid to be selected for against the background of all the cells that did not.

Vectors for Bacteria
Plasmid vector

- Unique restriction enzyme cleavage sites
- Antibiotic resistance genes with retained bacterial promoter:
- *bla* - ampicillin
- *aad* - spectinomycin, streptomycin
- *nptIII* - kanamycin, neomycin, amikacin
- Present in plant transformants from both biolistic and *Agrobacterium* transformation.

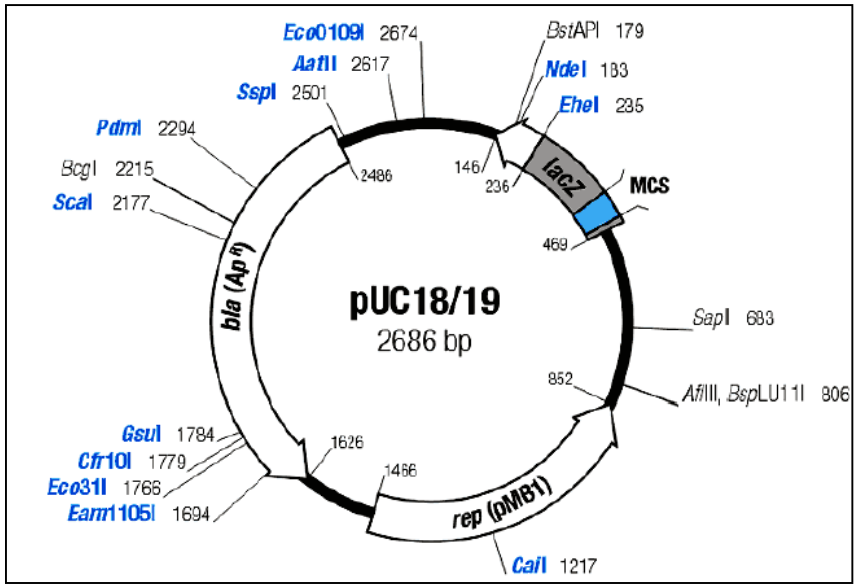


Vectors for Bacteria

Plasmid vector

pUC18

- pUC18: Derivative of pBR322.
- Advantages over pBR322:
- Smaller – so can accommodate larger DNA fragments during cloning (5-10kbp)
- Higher copy # per cell (500 per cell = 5-10x more than pBR322)
- Multiple cloning sites clustered in same location = "polylinker".



Natural vs. commercial plasmid vectors

- There are many plasmid vectors that are commercially available.
- All are derived from plasmids that were originally isolated from bacterial cells in nature.
- However, commercial cloning plasmids differ from natural plasmids because they have been modified in clever ways that make them particularly useful for the creation of recombinant DNA molecules.

Vectors for Bacteria

2. Bacteriophage/phage vectors

- A bacteriophage, or phage for short, is a virus that infects bacteria.
- Like other types of viruses, bacteriophages vary a lot in their shape and genetic material.
- Phage genomes can consist of either DNA or RNA, and can contain as few as four genes or as many as several hundred.

Vectors for Bacteria

Bacteriophage/phage Vs. plasmid vectors

- Plasmid vectors can carry up to 15 kb of inserted DNA, Whereas
- Phage vectors can carry inserts of 23 kb.

Vectors for Bacteria

Bacteriophage/phage vectors

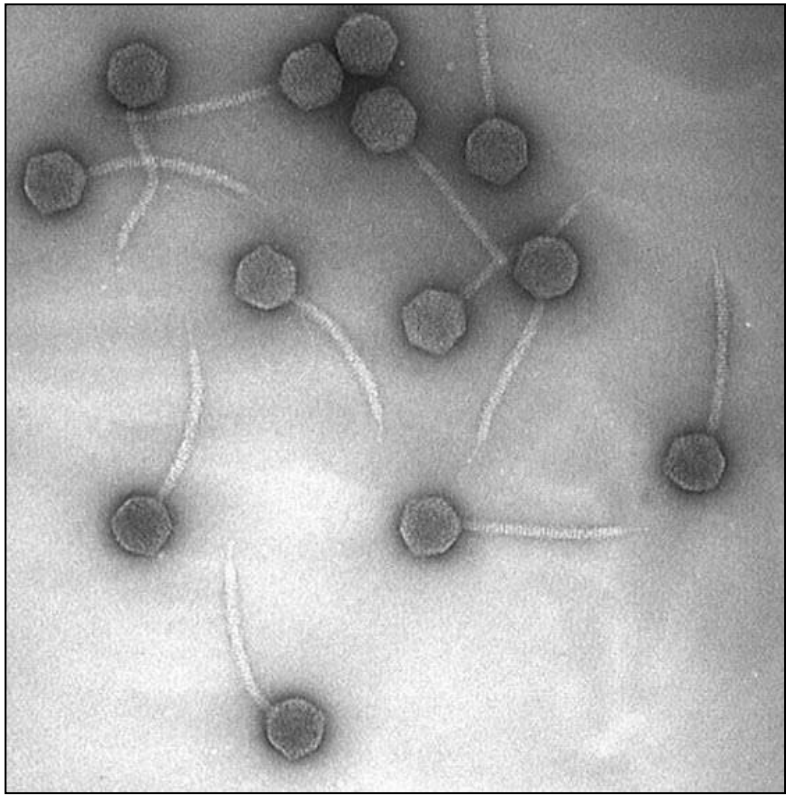
Lambda (λ) Phage Vector

- Bacteriophage lambda (λ) infects *E. coli*.
- Double-stranded, linear DNA vector – suitable for library construction.
- Can accommodate large segments of foreign DNA.
- Central 1/3 = "stuffer" fragment
  - Can be substituted with any DNA fragment of similar size without affecting ability of lambda to package itself and infect *E. coli*.
  - Accommodates ~15 kbp of foreign DNA
  - Foreign DNA is ligated to Left and Right Arms of lambda→ Then either:
    - Transfected into *E. coli* as naked DNA, or
    - Packaged *in vitro* by combining with phage protein components (heads and tails) (more efficient, but labor intensive and expensive).

Bacteriophage/phage vectors

Lambda (λ) Phage Vector

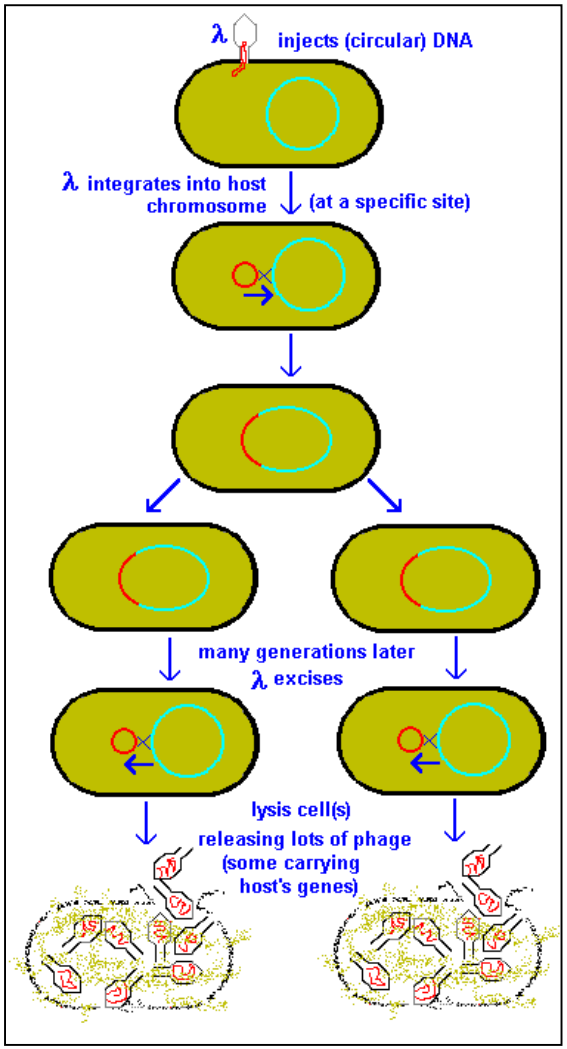
- The lambda phage, also called *Enterobacteria phage* λ and colphage λ, is a type of temperate bacteriophage or bacterial virus that infects the *Escherichia coli* (*E. coli*) species of bacteria.
- The virus may be housed in the genome of its host via lysogeny.



Bacteriophage/phage vectors

Lambda (λ) Phage Vector

- Lytic cycle:
- Bacteriophage lambda (λ)s frequently used by molecular geneticists because most of lamda's 49,000 bases can be replaced with foreign DNA.
- So it can act as a vehicle for transferring DNA.



Vectors for Bacteria

3. Cosmid vectors

- Cosmids are hybrids of λ phages and plasmids.
  - Lambda components include the cos part of this hybrid that are necessary for insertion into phage heads.
  - Plasmid DNA components include ORI + Antibiotic resistance gene.
- Very large inserts can be accommodated by cosmids.
- Cosmids can carry inserts up to 44 kb.
- Cosmids are cloning vectors that can replicate autonomously like a plasmid and be packaged into a phage.

Vectors for Bacteria

3. Cosmid vectors

- Cosmid vectors can accommodate up to 42 kb of DNA.
- They are multicopy plasmids in *E. coli*, facilitating DNA isolation and *in vitro* manipulation.
- However, deletion of inserted DNA sometimes occurs with cosmid clones, perhaps indicative of sequences that are deleterious to *E. coli*, particularly at high copy number.
- SuperCos1 (Agilent), is a novel, 7.9-kb cosmid vector which was used to make an ordered *S. coelicolor* cosmid library has proved to be a very effective vector, but its derivatives require modification before they can be transferred to and stably maintained in a *Streptomyces* host.

Vectors for Bacteria

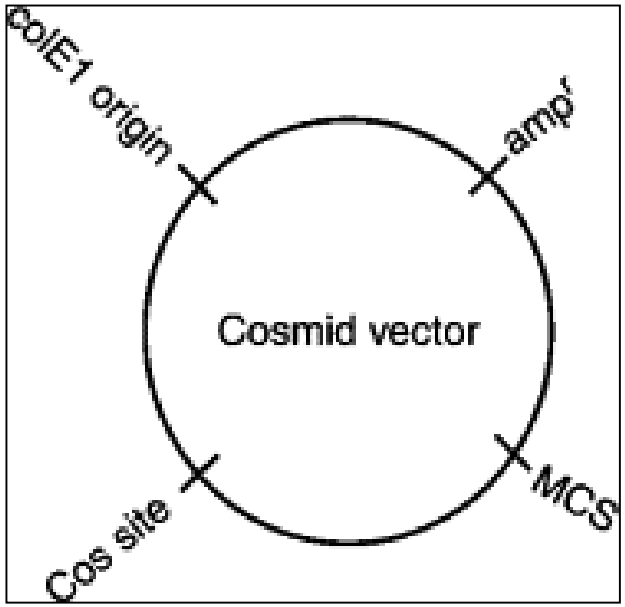
Cosmid vectors

Essential features of a cosmid vector

- Constructed by combining:

1. plasmid origin of replication,
2. selectable marker gene, and
3. bacteriophage (cos site) for genetic elements.

- Cloning capacity up to 45 kbp.
- Used for constructing genomic library.



ScienceDirect,2019

Vectors for Bacteria

4. Fosmid vectors

- A new cloning vector with a low copy number was developed based on F-factor plasmid and this is called the fosmid vector.
- Thus, fosmids are similar to cosmids but are based on the bacterial F-plasmid.
- The cloning vector is limited, as a host (usually *E. coli*) can only contain one fosmid molecule.
- Fosmids can hold DNA inserts of up to 40 kb in size; often the source of the insert is random genomic DNA.

ScienceDirect,2019;..

Vectors for Bacteria

4. Fosmid vectors

- Fosmids are used for genomic library construction of complex genomes.
- A fosmid library is prepared by extracting the genomic DNA from the target organism and cloning it into the fosmid vector.
- A fosmid is a low copy number plasmid constructed by combining the features of F-factor plasmid and the cos site.
- These libraries are stably maintained as they are low in copy number, i.e. one copy per cell.
- The fosmid vector library is used to check the quality of the sequence generated from genome projects.

ScienceDirect,2019

Vectors for Bacteria
Fosmid vectors

- pFOS1 was the first developed fosmid vector.
- The main application of a fosmid vector is that a stable genomic library can be constructed and maintained for longer.
- Fosmid libraries are useful in filling gaps in the contig assembly.
- They are also used for detecting deletions, insertion and rearrangements, etc.
- The size of the fosmid vector is ~8 kbp.
- It has parB and parA genes to maintain the copy number of the fosmid.
- A kanamycin-resistant gene or an ampicillin-resistant gene is used as a selectable marker.

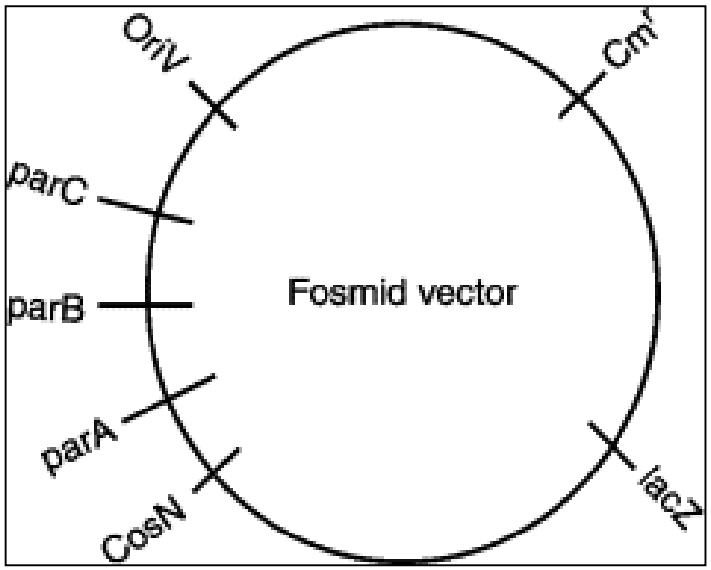
ScienceDirect,2019

Vectors for Bacteria

Fosmid vectors

Essential features of a fosmid vector

1. *E. coli* -based cloning vector.
2. Cloning capacity is up to 100 kbp.
3. Able to maintain foreign DNA stably over generations.
4. Low copy number vector.



ScienceDirect,2019

Vectors for Bacteria

5. Phagemid vectors

- A phagemid is a DNA-based cloning vector, which has both bacteriophage and plasmid properties.
- Phagemid is a phage whose genome contains a plasmid that can be removed during the infection of a host with a second, helper phage.
- These vectors carry, in addition to the origin of plasmid replication, an origin of replication derived from bacteriophage.

Vectors for Bacteria

5. Phagemid vectors

- We call them phagemids ('phage' from M13 bacteriophage and 'mid' from plasmid).
- Although M13 vectors are very useful for the production of single-stranded versions of cloned genes they do have one disadvantage.
- There is a limit to the size of DNA fragment that can be cloned with an M13 vector, with 1500bp generally being looked on as the maximum capacity.
- To get around this problem a number of novel vectors have been constructed which are the hybrids of plasmids and M13 vectors.

Vectors for Bacteria

Phagemid vectors

Certain advantages of Phagemid over phage vectors

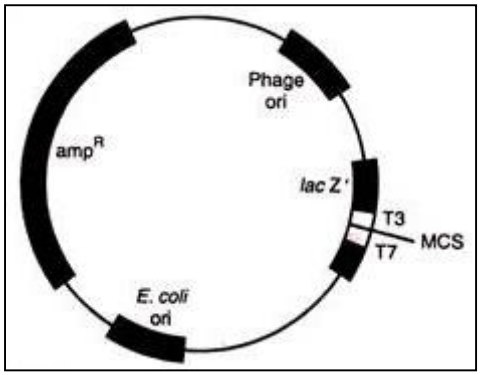
1. The carrying capacity of phagemid is higher than phage vectors.
2. Phagemid has higher efficiency in transformation than phage vectors.
3. Phagemids are genetically more stable than recombinant phage vectors.
4. Phagemids can be exploited to generate single stranded DNA template for sequencing purposes.
5. Single stranded phagemid vectors inside the phage can be targeted for site-directed mutagenesis.
6. Single stranded vectors can be used to generate hybridization probes for mRNA or cDNA.

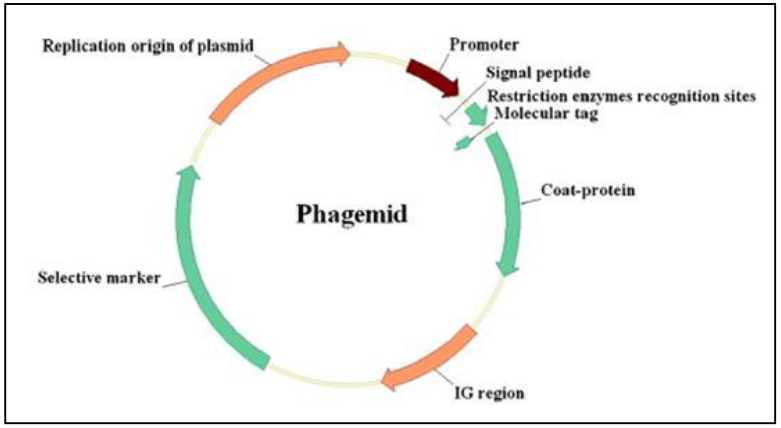
Vectors for Bacteria
Phagemid vectors

- **A typical phagemid has following parts:**

1. Phage M13 origin of replication.
2. A portion of lac Z' gene driven by lac promoter.
3. A multiple cloning site (MCS) with lac Z' gene.
4. Phage T7 and T3 promoter sequences flanking the MCS sequences.
5. ColE1 origin of replication.
6. amp^R resistant gene.

The main advantage of the phagemid system is that it can be used to provide single-or double-stranded material without any re-cloning.





Phagemid vector for phage display.
Examples of phagemids are the vectors pUC118, 119 and 120.

Qi *et al.*,2012;..

T-A cloning vectors

T-vectors

TA cloning (also known as rapid cloning or T cloning)

- A method for direct cloning a PCR product, by the T-vector technique. This is cheap and easy way to clone PCR products with A 3' overhangs.
- **Making the T-vector:**
- Digest 5 µg pBluescript II with EcoRV (pronounced "eco R five"). EcoRV is a type II restriction endonuclease (blunt cutter).
- Heat kill the enzyme or gel purify using Qiagen column elute in 50µl EB.
- Add 10 µl of 10X PCR buffer, 2 µl of 100 mM dTTP, 37.0 µl of distilled water and 1.0 µl of Taq DNA polymerase. Incubate at 72 C for 2 hours.
- Purify the T-vector by phenol/chloroform extraction and ethanol precipitation or purify over a Qiagen column. Resuspend/elute the prepared T-vector in 100 µl of water or TE. We use about 2- 5 µl in a ligation reaction.
- **Cloning the PCR product:**
- Purify the PCR product over a Qiagen column (Gel purify if necessary). Elute in 50 µl EB. Use 5-10 µl of this PCR product in a ligation reaction e.g.
- 10 µl PCR insert
- 5 µl T-Vector
- 4 µl 5X Ligase buffer
- 1 µl T4 Ligase
- Ligate 15 min RT(rapid ligation kit) to overnight (15C).
- Transform into *E. coli* (XL1Blue).
- Plate on Amp (75µg/ml) X-gal (spread 60ul of 2% solution on plate) Tet (15µg/ml) IPTG (0.1-1mM)
- Pick white colonies to prep.

T-A cloning vectors

T-vectors

TA cloning (also known as rapid cloning or T cloning)

- In genetics, pBluescript (pBS) or pBluescript II is a commercially available phagemid, derived from the PUC19 plasmid, containing several useful sequences for use in cloning with bacteriophage.
- The sequences include a multiple cloning site sequence (MCS), antibiotic resistance sequence to ampicillin and an *E. coli* and f1 helper phage origin of replication.

T-A cloning vectors

T-vectors

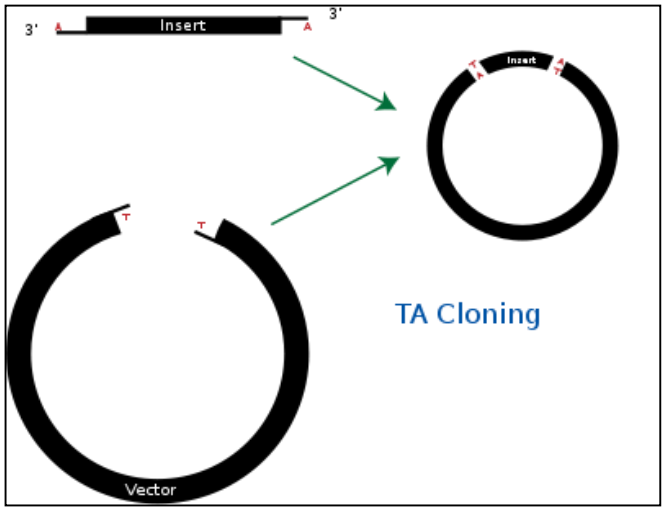
TA cloning (also known as rapid cloning or T cloning)

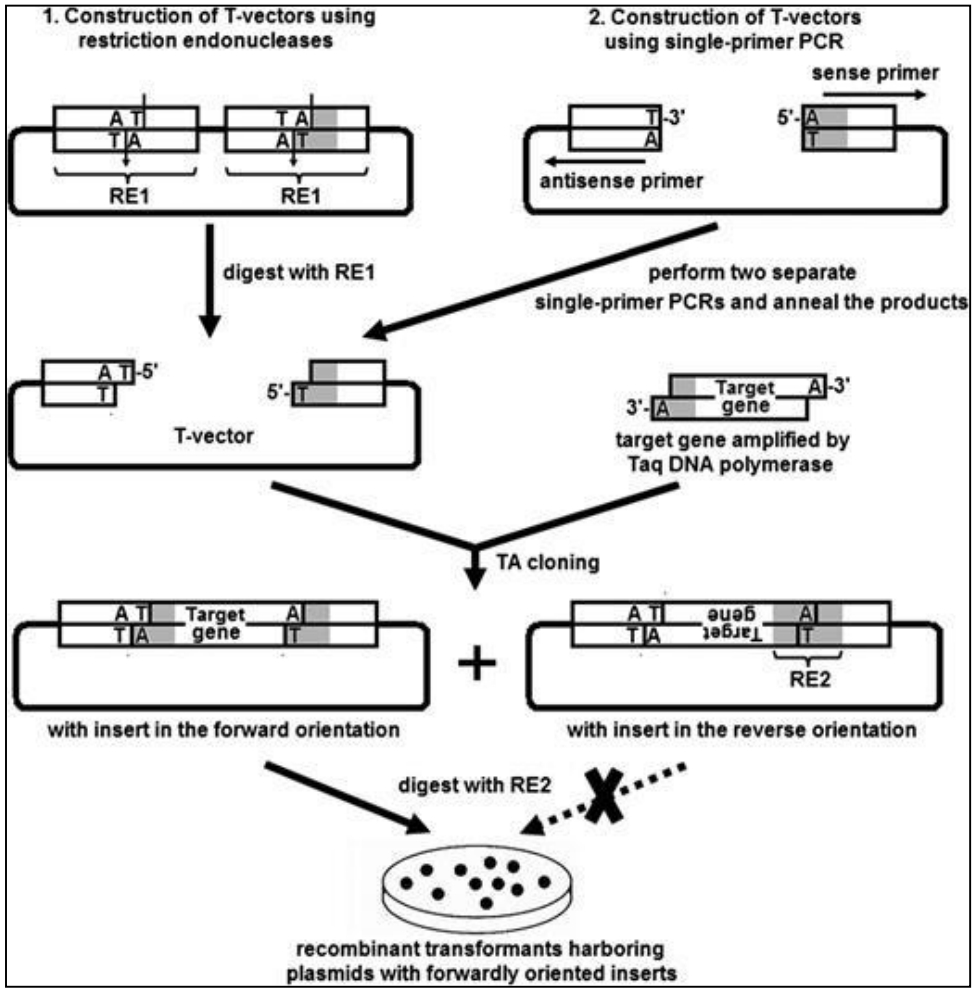
- TA cloning is a simple and convenient method of subcloning polymerase chain reaction (PCR) products.
- the PCR product can be directly cloned into a linearized cloning vector that have single base 3'-T overhangs on each end.
- "TA" is short for "thymine" and "adenine."
- This cloning technique utilizes the ability of thymine to hybridize to adenine in the presence of ligases.
- **Restriction enzymes are not used**, unlike the traditional subcloning method.
- Instead, PCR products are amplified using Taq polymerases enzymes.

T-A cloning vectors

T-vectors

TA cloning (also known as rapid cloning or T cloning)





Wikepdia, Yao *et al.*,2016

Cloning vectors

Yeast Vectors

- All the yeast vectors can be divided into three types:

1. Yeast cloning vectors (or Yeast plasmid vectors);
2. Yeast expression vectors;
3. Yeast artificial chromosomes (YAC).

Cloning vectors
Shuttle vectors

- Shuttle vectors are those which can multiply into two different unrelated species.
- Shuttle reactors are designed to replicate in the cells of two species, as they contain two origins of replication, one appropriate for each species as well as genes that are required for replication and not supplied by the host cell, i.e., it is self-sufficient with the process of its replication.

Cloning vectors
Shuttle vectors

- The shuttle vectors are of following types:

1. **Eukaryotic – Prokaryotic Shuttle Vectors:**

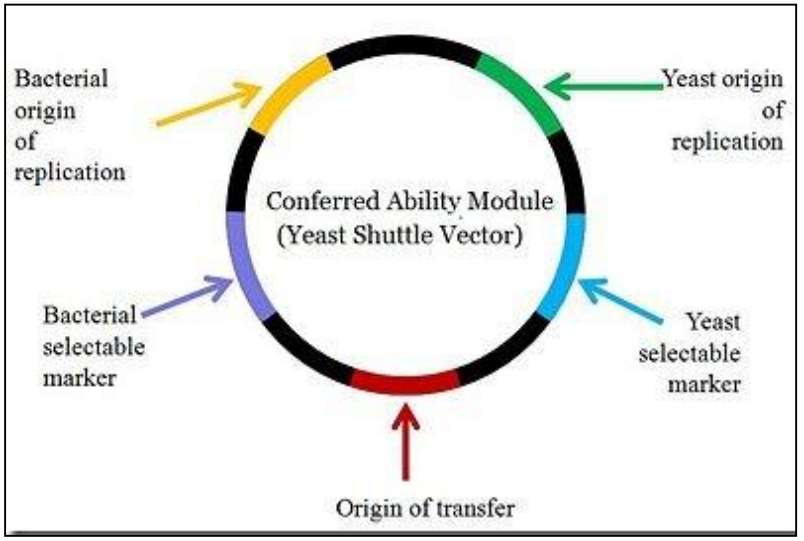
- Vectors that can propagate in eukaryotes and prokaryotes. e.g., YEp vectors can be propagated in yeast (fungi) as well as in *E. coli* (bacteria).

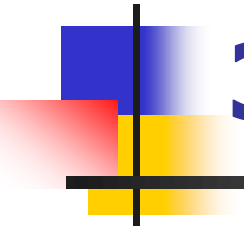
2. **Prokaryotic – Prokaryotic Shuttle Vectors:**

- Vectors that can be propagated in two unrelated prokaryotic host cells, e.g., RSF1010 vectors can be propagated both in bacteria as well in spirochetes.

Cloning vectors
Shuttle vectors

- Example of Shuttle Vector:
- pHV14, pEB10, pHP3, etc. replicate both in *Bacillus subtilis* and *E.coli*.
- pJDB219 is another shuttle vector that can replicate in *E. coli* and Yeast.
- pJDB219 is another shuttle vector that can replicate in *E. coli* and Yeast (*Saccharomyces cerevisiae*).





3. DNA ligation

Ligation of DNA fragments

Principles of cloning, vectors and cloning strategies

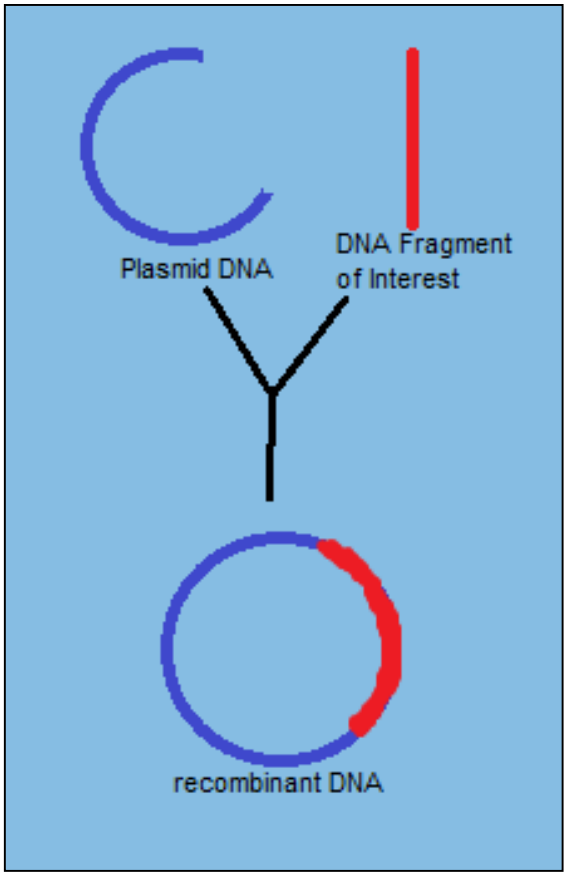
Ligation of DNA fragments
Principle

- The basic strategy in molecular cloning is to insert a DNA fragment of interest (a segment of DNA) into a DNA molecule (called a vector) that is capable of independent replication in a host cell.
- The result is a recombinant molecule composed of the DNA insert linked to vector DNA sequences.
- Construction of these recombinant DNA molecules is dependent on the ability to covalently seal single stranded nicks in DNA.
- This process is accomplished both *in vivo* and *in vitro* by the enzyme DNA ligase.
- DNA ligation is the process of joining together two DNA molecules ends (either from the same or different molecules).

Ligation of DNA fragments
Plasmid cloning strategy

- Involves five steps:

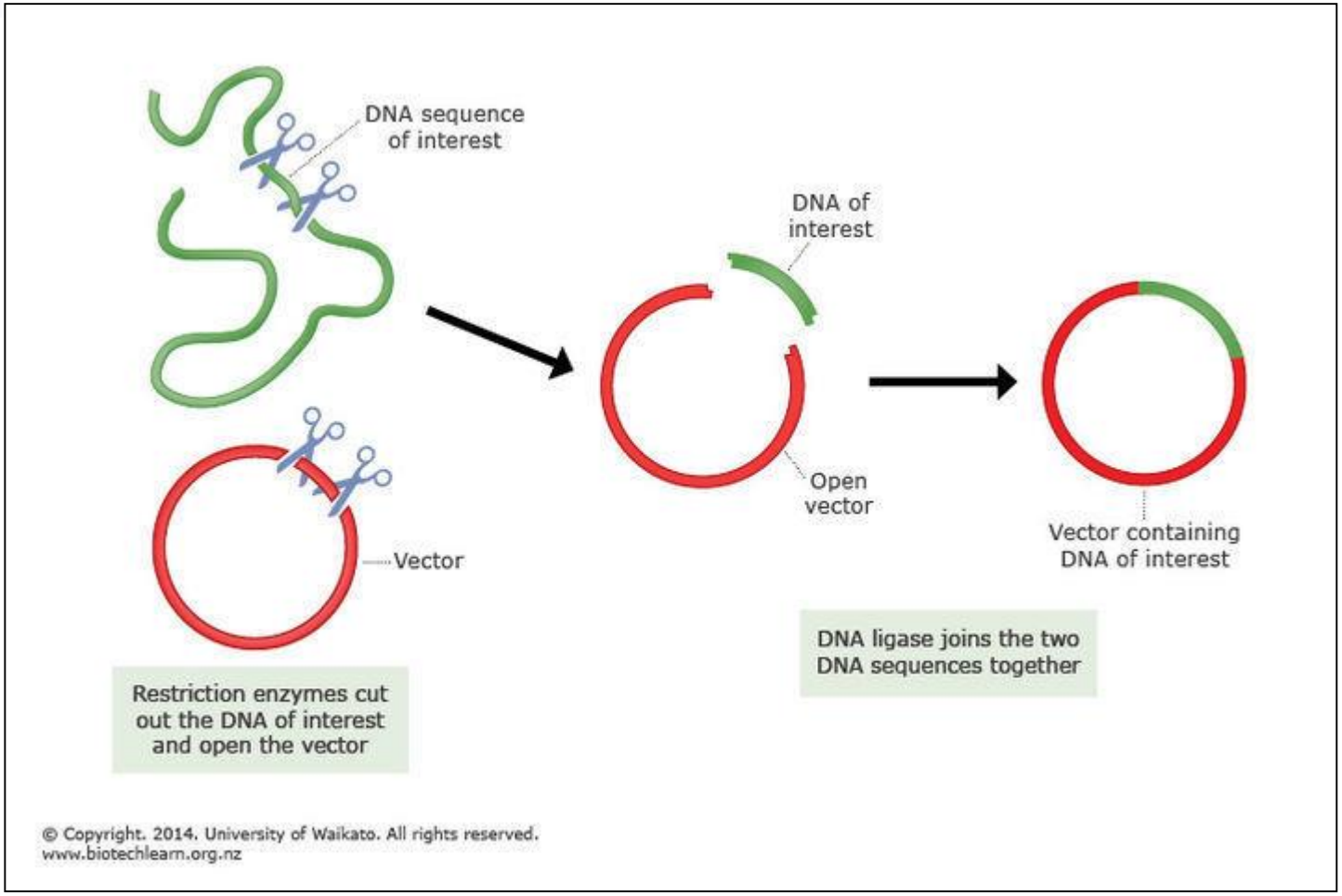
1. Enzyme restriction digest of DNA sample.
2. Enzyme restriction digest of DNA plasmid vector.
3. Ligation of DNA sample products and plasmid vector.
4. Transformation with the ligation products.
5. Growth on agar plates with selection for antibiotic resistance.



Ralph Gibson;..

Ligation of DNA fragments

Procedure

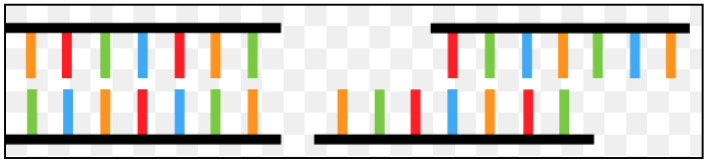


Ligation of DNA fragments

DNA ligase

Sticky ends (cohesive ends) or blunt ends

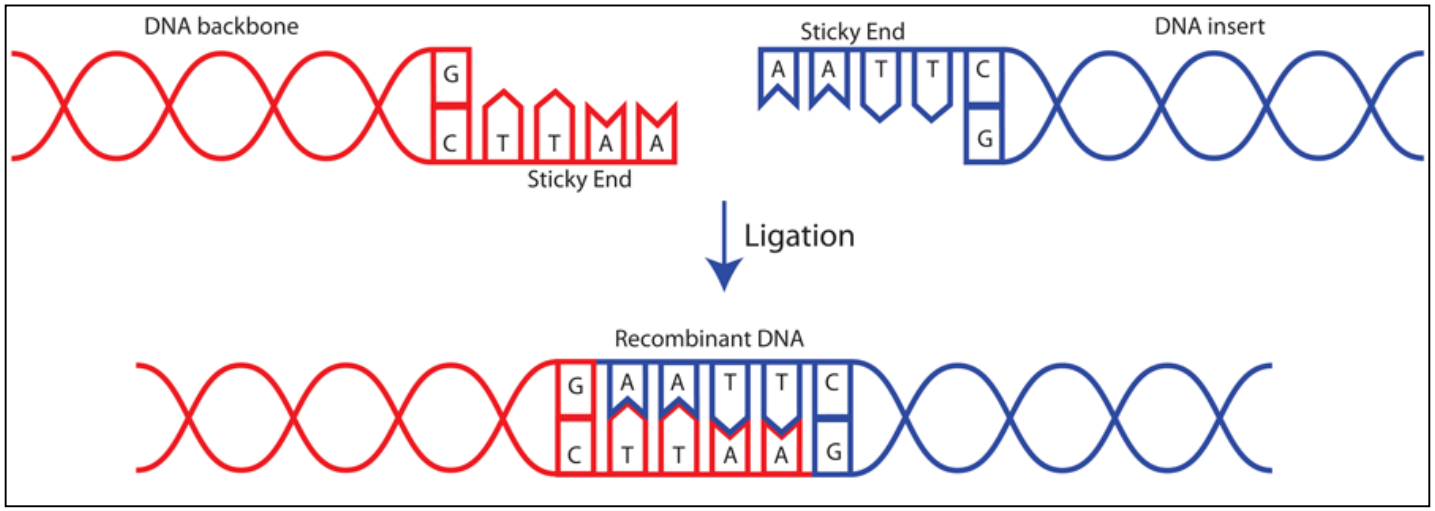
- DNA ends refer to the properties of the end of DNA molecules, which may be sticky ends (cohesive ends), blunt ends or in other forms.
- The enzyme that joins the DNA fragments is called DNA ligases.
- The enzyme extensively used in joining DNA fragments is T_4 DNA ligase.
- The ligase joins both cohesive end as well as blunt ended DNA.
- It is a single polypeptide with a M.W of 68,000 Dalton requiring ATP as energy source.
- The maximal activity pH range is 7.5-8.0.
- The enzyme exhibits 40% of its activity at pH 6.9 and 65% at pH 8.3.



Ligation of DNA fragments

DNA ligase

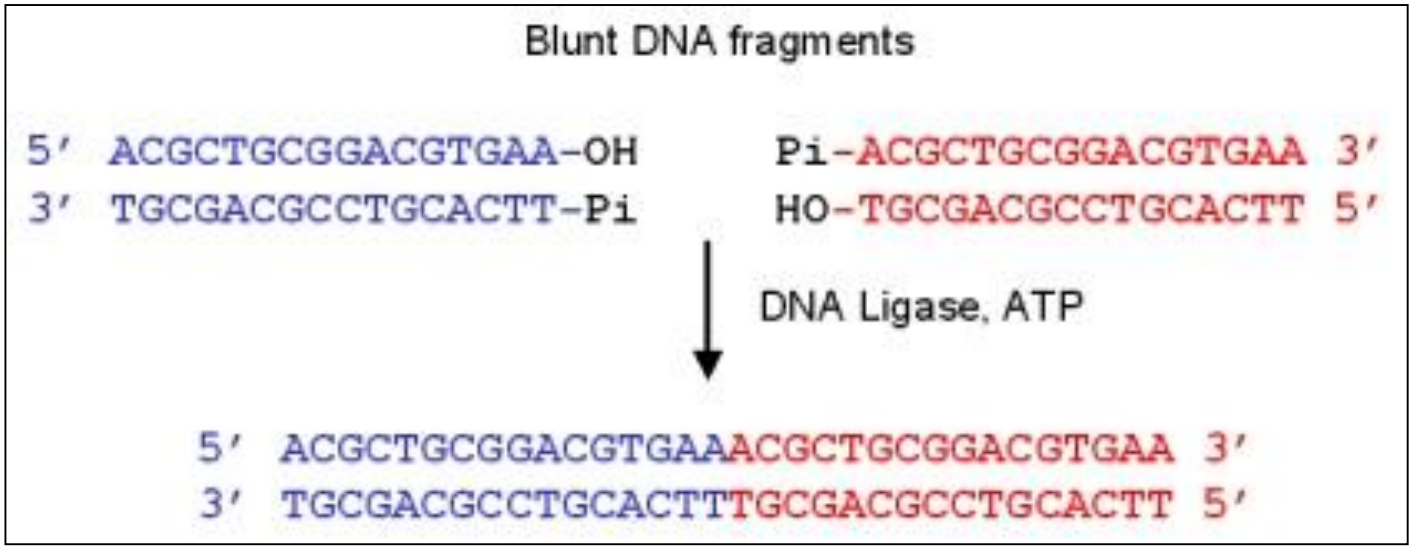
Sticky ends (cohesive ends)



Ligation of DNA fragments

DNA ligase

Blunt ends



Blaber,2001

Ligation of DNA fragments
Procedure

- **Materials Required For Restriction Enzyme Digestion:**
- pUC18 DNA
- *Bam*H1 enzyme
- 10X buffer
- 1Kb Ladder
- Sterile water
- Agarose
- 6X loading dye
- 1.5 ml Sterile Vials
- Ethidium Bromide
- 1X TAE buffer

- **Procedure:**
- Three separate vials are taken and are labelled as reaction, +ve control and –ve control. (details in the next slide)

Materials	Reaction	+ve control	-ve control
Double digested DNA (EcoR1/hind3)	1.5µl	-	1.5µl
10X instant Ligation buffer	1µl	1µl	1µl
T4 DNA Ligase	1µl	1µl	-
Water	6.5µl	8µl	7.5µl
Total	10µl	10µl	10µl

Ligation of DNA fragments
Procedure

- Three separate vials are taken and are labelled as reaction, +ve control and –ve control.
- 1.5µl of PCR product of DNA fragment is added to reaction and –ve control vials only.
- 1 µl of 10X Ligation buffer is added to each of the three vials.
- 1 µl of T4 DNA Ligase (1 U) is added to reaction and +ve control vials only.
- 6.5 µl, 8 µl, 7.5 µl of water is added to reaction, +ve control and –ve control vials, respectively.
- The total volume in each of the vials is 10 µl. Incubate for 1 hr at 37°C.
- The prepared mixtures can be analyzed in bacterial transformation in bacterial cells or they can be analyzed onto agarose gel.

Ligation of DNA fragments

Procedure

Blue-white colony screening method

- DNA of interest is ligated into a vector.
- The vector is then inserted into a competent host cell viable for transformation, which are then grown in the presence of X-gal.

1. Cells transformed with vectors containing recombinant DNA will produce white colonies;
2. Cells transformed with non-recombinant plasmids (i.e. only the vector) grow into blue colonies.

- This method of screening is usually performed using a suitable bacterial strain, but other organisms such as yeast may also be used.

Wikipedia,2018

Ligation of DNA fragments

Procedure

Blue-white colony screening method

- Blue-white screening is a rapid and efficient technique for the identification of recombinant bacteria.
- It relies on the activity of β-galactosidase, an enzyme occurring in *E. coli*, which cleaves lactose into glucose and galactose.

1. If in clones with no recombinant DNA (wild type), β-galactosidase is produced, X-gal is hydrolyzed to form 5-bromo-4-chloro-indoxyl, which spontaneously dimerizes to produce an insoluble blue pigment called 5,5′-dibromo-4,4′-dichloro-indigo(appear blue colonies).
2. While the recombinant ones appear white (colonies).

Ligation of DNA fragments

Procedure

Blue-white colony screening method

- Isopropyl β-D-1-thiogalactopyranoside (IPTG) is used along with X-gal for blue-white screening.
- IPTG is a non-metabolizable analog of galactose that induces the expression of lacZ gene.
- It should be noted that IPTG is not a substrate for β-galactosidase but only an inducer.
- For visual screening purposes, chromogenic substrate like X-gal is required.

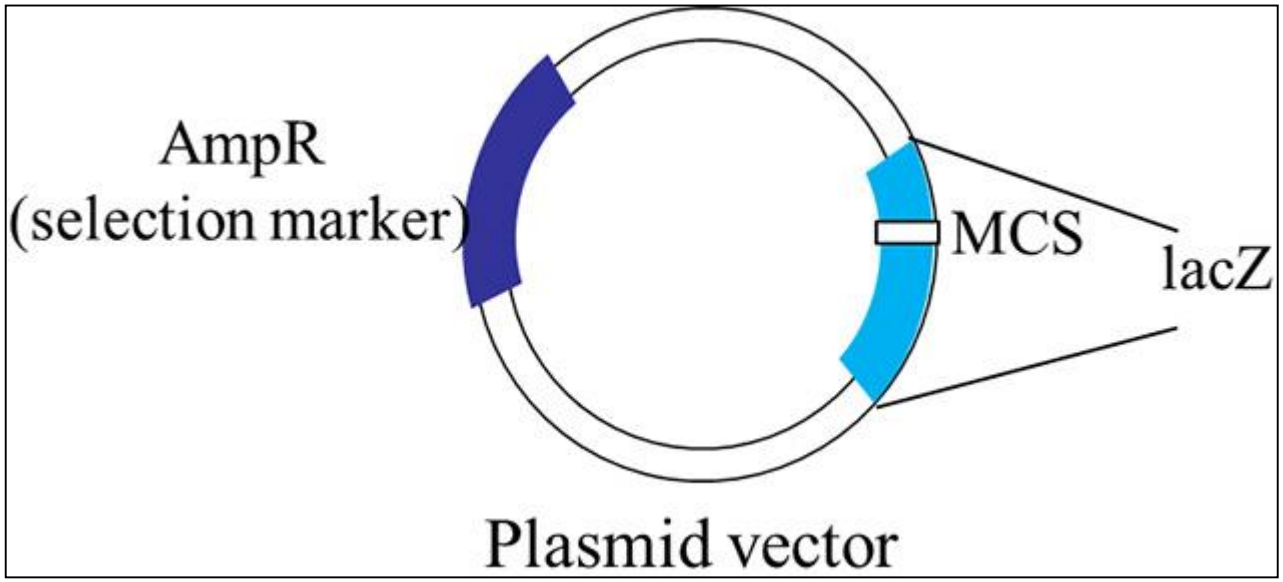
Wikipedia,2018

Ligation of DNA fragments

Procedure

Blue-white screening

- A schematic representation of a typical plasmid vector that can be used for blue-white screening.

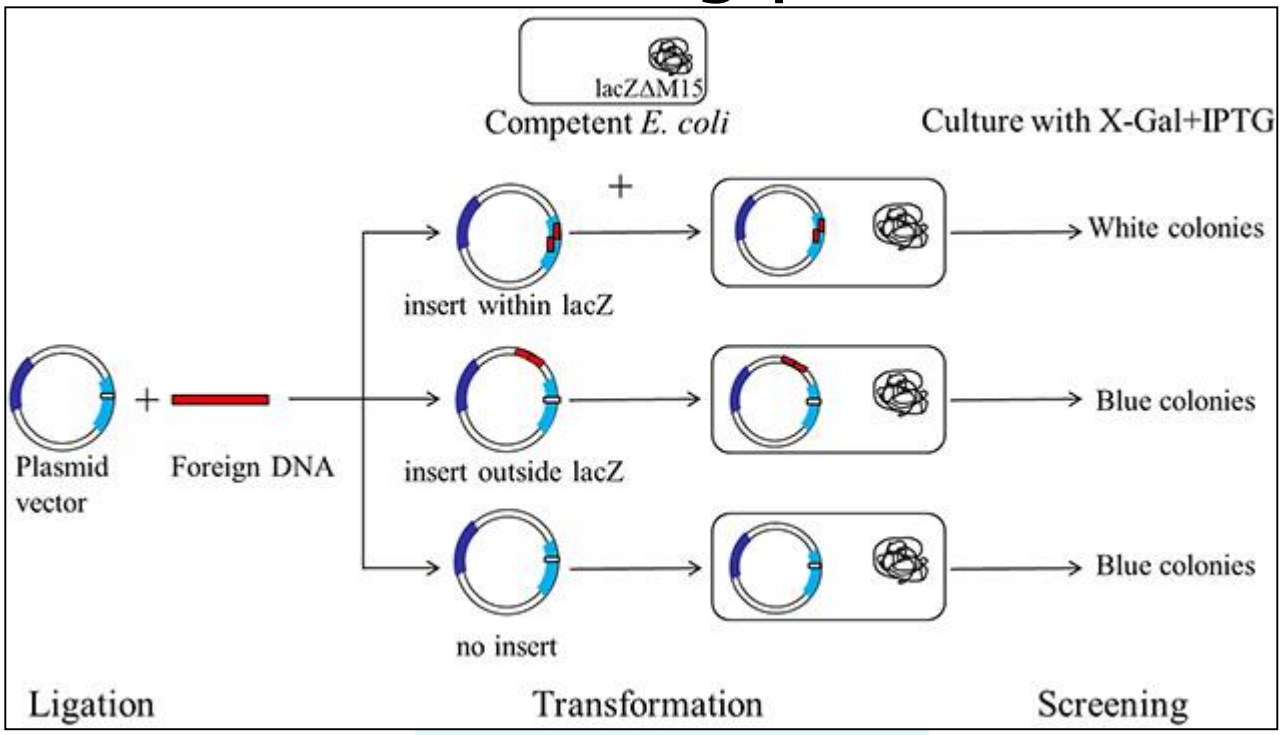


Ligation of DNA fragments

Procedure

Blue-white screening

- A schematic representation of a typical blue-white screening procedure.



Ligation of DNA fragments

Procedure

Blue-white colony screening method

- An LB agar plate showing the result of a blue white screen.

1. white colonies: transformed cells with vectors containing recombinant DNA.
2. blue colonies: transformed cells with non-recombinant plasmids.

Blue-white color selection of recombinant bacteria using a chromogenic substrate known as X-gal (5-bromo-4-chloro-indoxyl β-D-galactoside).

Wikipedia,2018

Advantages of genetic transfer

A. Production of a transgenic plants resistant to pests/diseases

Transgenic Crops

- Plants are genetically engineered to:

1. Resist the effects of weed killers (herbicides).
2. Tolerate salt in the soil.
3. Resist insect pests (produce *Bt* toxins from *Bacillus thuringensis*).
4. Produce pharmaceuticals and vaccines.

- Therefore, most of the transgenic crop varieties currently grown by farmers are either:

1. Herbicide tolerant or
2. Insect pest-resistant.

Transgenic plants
Resistant to pests

1. ISAAA, Pub. No. 26, 2002; 2. Field trials;
3. Makhathini Flats

Developing countries & small holders benefit from biotech

Bt cotton delivers economic advantages over conventional cotton[1]

	China	India[2]	South Africa[3]	Mexico
Farm Size	0.5Ha	2Ha	< 3Ha	20Ha
Increased Yield	5~10%	40%	25%	3~20%
Reduced Insecticide	50~67%	50%	32%	50%
Positive Economic Return	$360 - ~550/Ha	$75 - ~200/Ha	$50/Ha	$45 - ~600/Ha

Increased income and time savings

Newell-McGlouhalin,2005

Transgenic plants
Resistant to pests

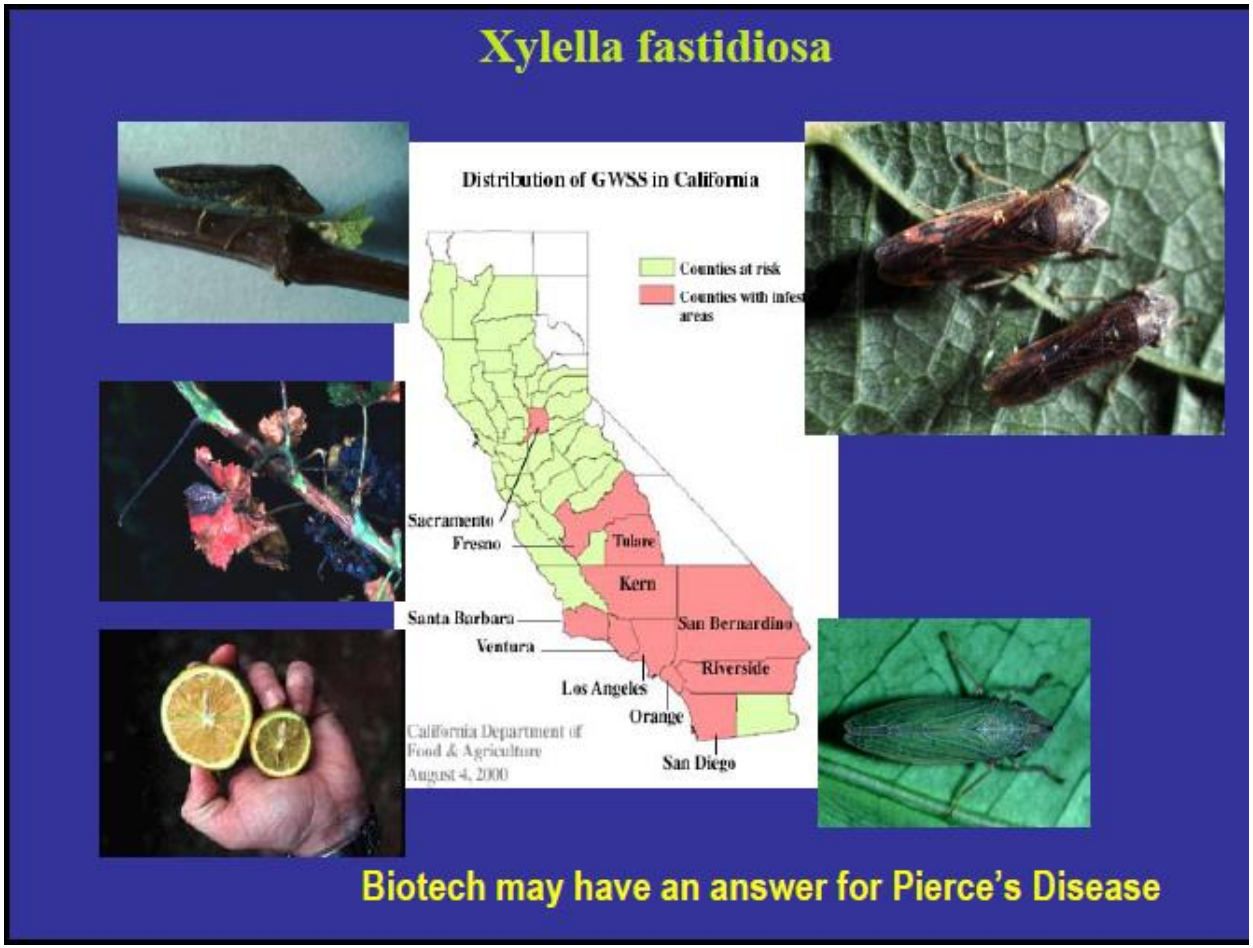
Promising Early results with Bt corn in the Philippines

- Bt corn approved 2002
- 49,400 acres planted
- 5,000 farmers using Bt corn
- Income increased by 34% for farmers planting Bt corn

- Up to 90% reduction in mycotoxin fungi that produce fumonisins
- **Effects on animals: feed refusal, short-term illness, reproductive dysfunction, death.**
- **Human effects: suppression immune system, reproductive dysfunction, cancer, death.**

Newell-McGlouhalin,2005

Transgenic plants
Resistant to diseseas

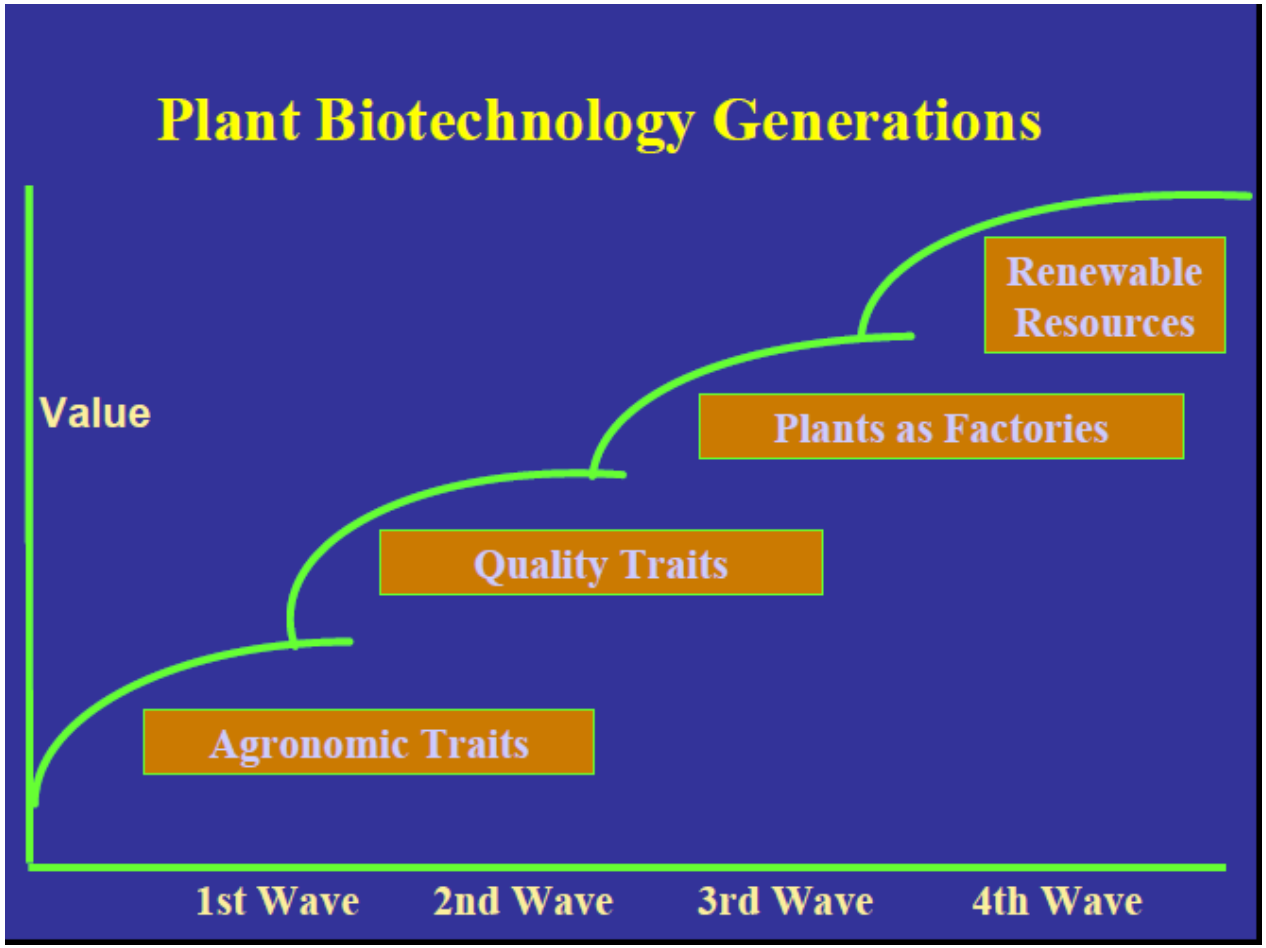


Newell-McGlouhalin,2005

Advantages of genetic transfer

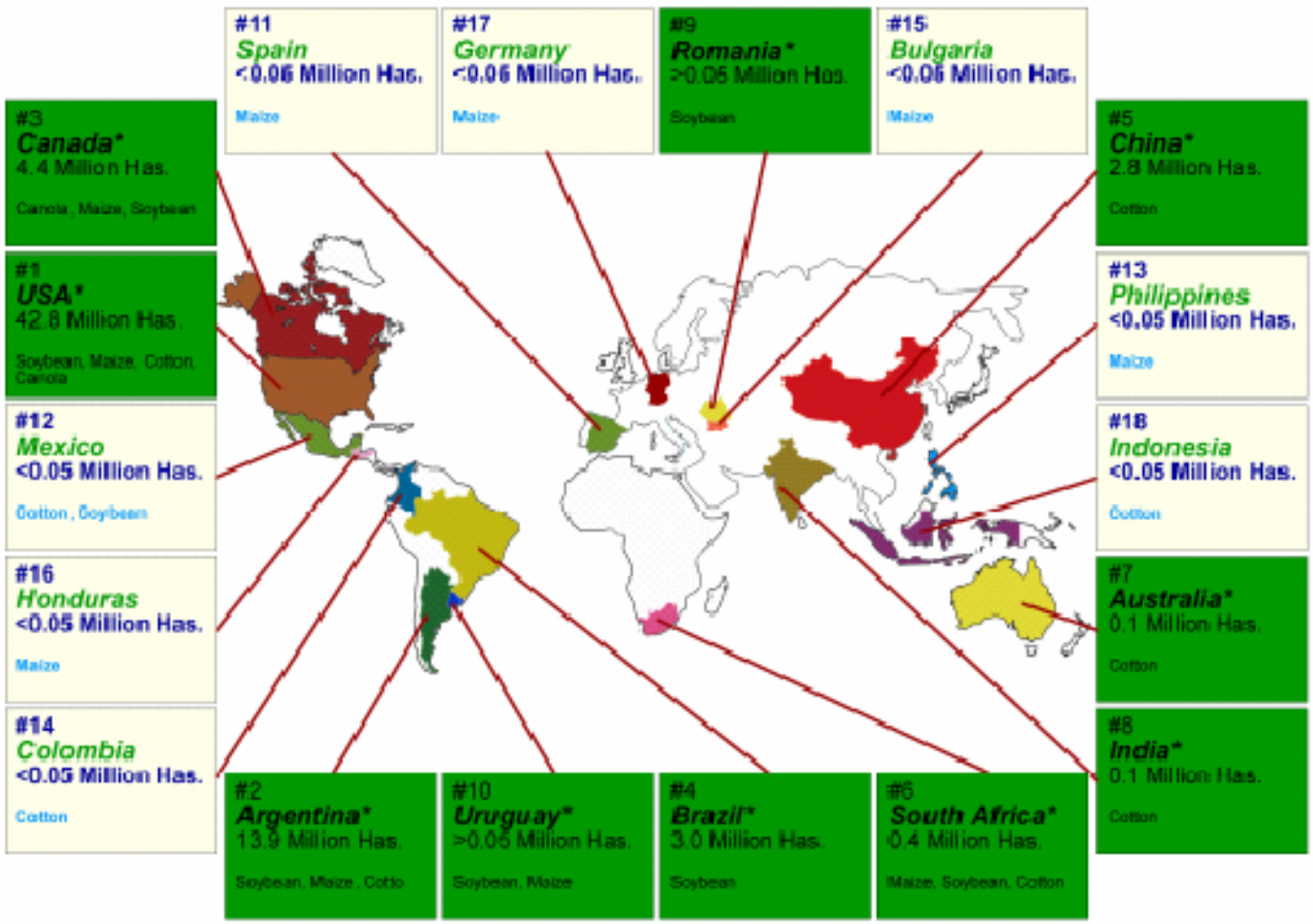
B. Production of a transgenic plants with high agronomic traits

Plant biotechnology



Newell-McGlouhalin,2005

GM Crop Countries, 2003



■ * 10 countries growing 50,000 hectares or more of GM crops.

2001 US impacts	YIELD	NET VALUE	PESTICIDE
US TOTAL	+3,785 m lb	+ $1,459 m	- 45.6 m lb

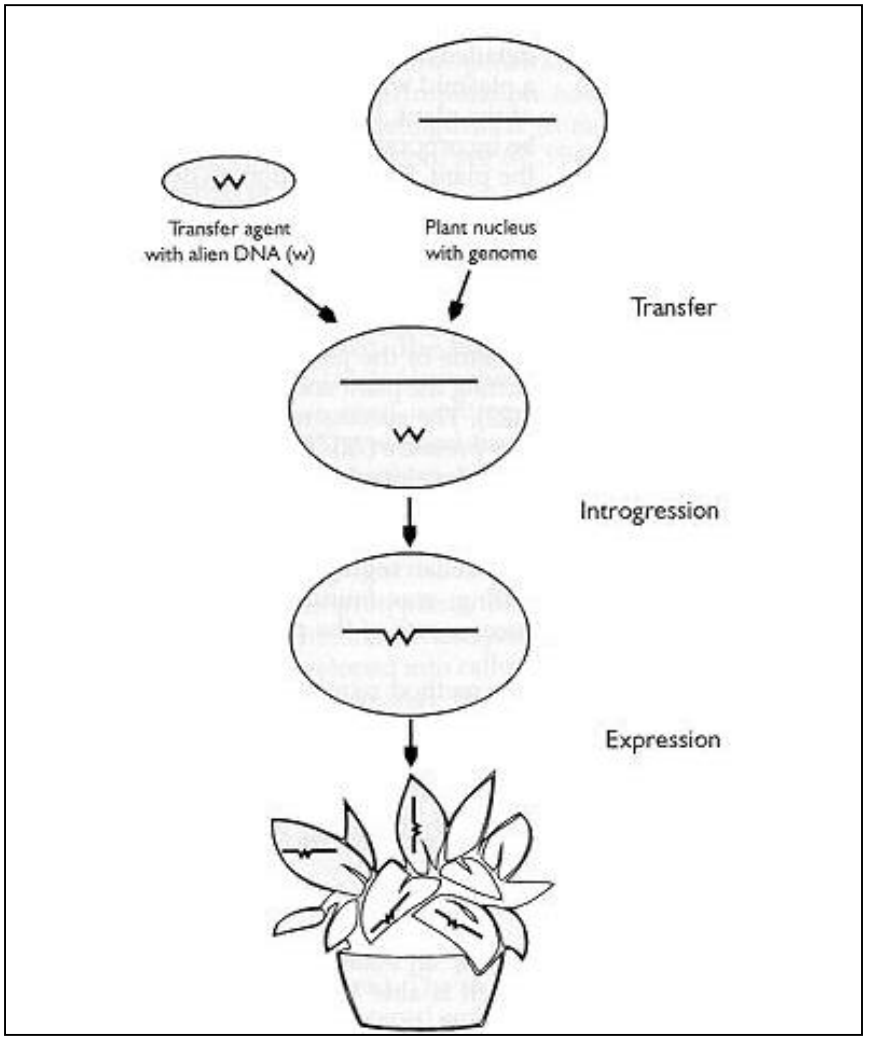
Source: C. James, ISAAA Brief 30, 2003

Newell-McGlouhalin,2005

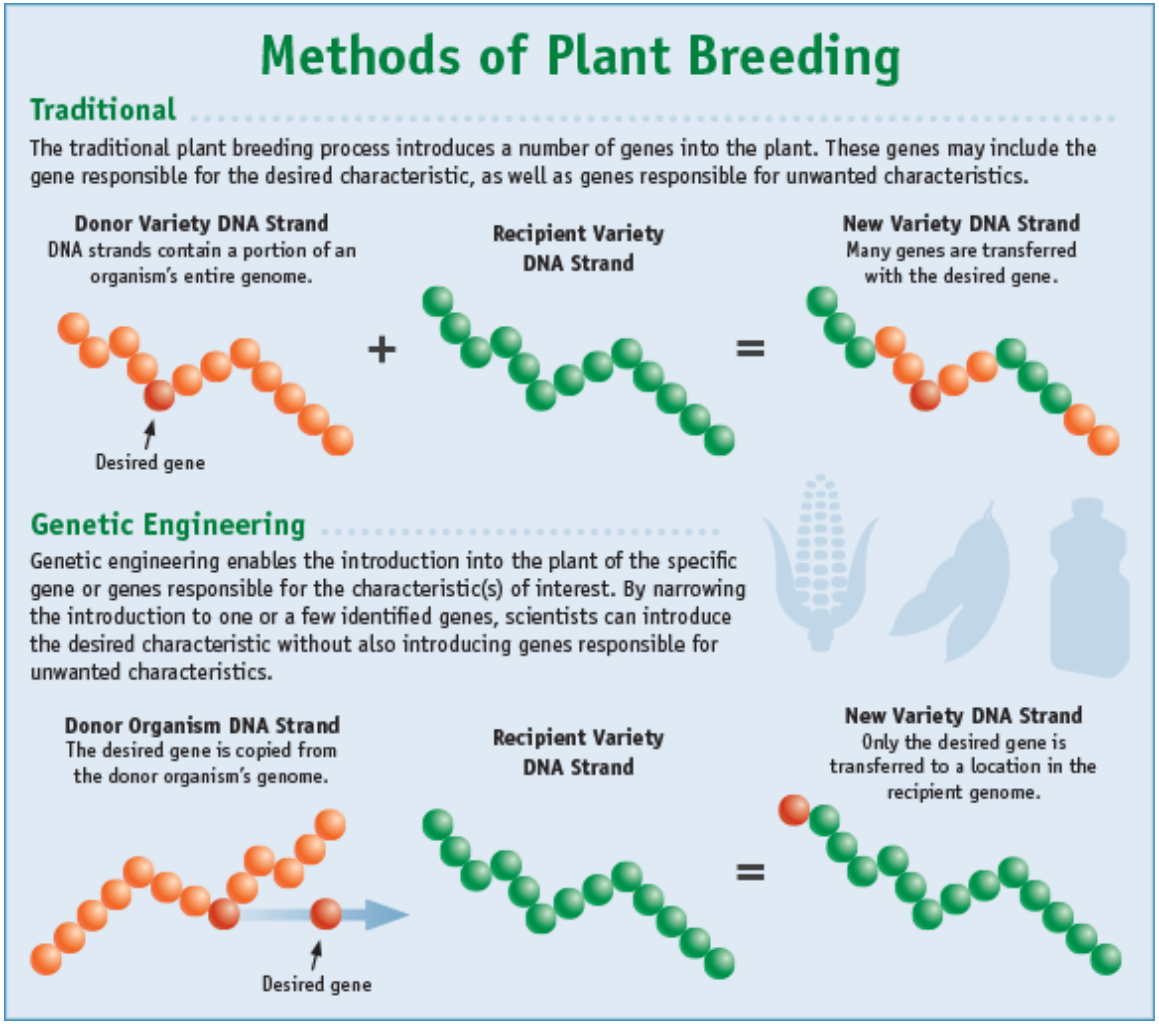
Techniques of recombinant DNA and biotechnology

- **Designing genes for insertion.**
- **Methods to introduce (insert) foreign genes into plant cell genome.**

Transgenic plant with align DNA

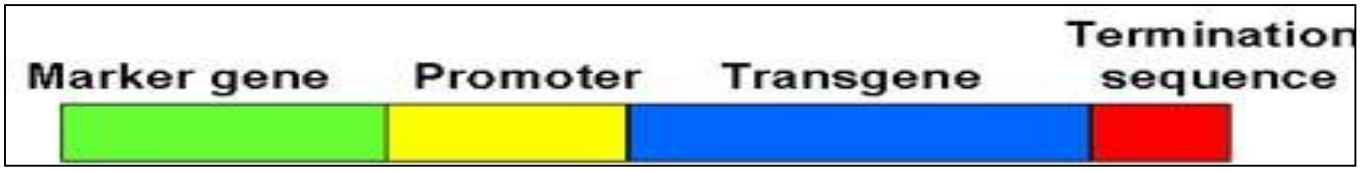


Transgenic plant with desired gene



Designing genes for insertion

- Once a gene has been isolated and cloned (amplified in a bacterial vector), it must undergo several modifications before it can be effectively inserted into a plant.



Simplified representation of a constructed transgene containing necessary components for successful integration and expression.

Designing Genes for Insertion

1. A promoter sequence must be added for the gene to be correctly expressed (i.e., translated into a protein product).
2. The termination sequence signals to the cellular machinery that the end of the gene sequence has been reached.
3. A selectable marker gene is added to the gene "construct" in order to identify plant cells or tissues that have successfully integrated the transgene.

Transformation methods

DNA insertion methods

- DNA must be introduced into plant cells.
- Direct:

1. Chemical method
2. Electrical method
3. Physical methods

- Indirect:

1. *Agrobacterium tumefaciens*

Transformation methods

Physical (direct) and *Agrobacterium*-mediated transformations

Methods of delivery of DNA into the cell

1. Direct methods

-protoplast microinjection
-particle bombardment
-protoplast polyethyleneglycol (PEG) method
-protoplast electroporation
-silicon carbide fibers

2. Agrobacterium -mediated transformation

Agrobacterium tumefaciens
Agrobacterium rhizogenes (Hairy Root)

Transformation methods

Physical (direct) and *Agrobacterium*-mediated transformations

Types of delivery systems

Naked DNA

Biolistics

- SiC fibres
- Protoplasts
- Electroporation
- Pollen

Vectored

- *Agrobacterium*
- Viruses

Transformation methods

Physical (direct) and *Agrobacterium*-mediated transformations

1. **The "Gene Gun" method** (also known as microprojectile bombardment or biolistics):

- This technique, which is shown and explained in the animated demo section of this web site, has been especially useful in transforming monocot species like corn and rice.

2. **The *Agrobacterium* method:**

- Transformation via *Agrobacterium* has been successfully practiced in dicots (broadleaf plants like soybeans and tomatoes) for many years, but only recently has it been effective in monocots (grasses and their relatives).
- In general, the *Agrobacterium* method is considered preferable to the gene gun, because of the greater frequency of single-site insertions of the foreign DNA, making it easier to monitor.

Transformation methods

Physical Methods

1. Particle bombardment
2. Microinjection
3. Silicon Carbide whiskers

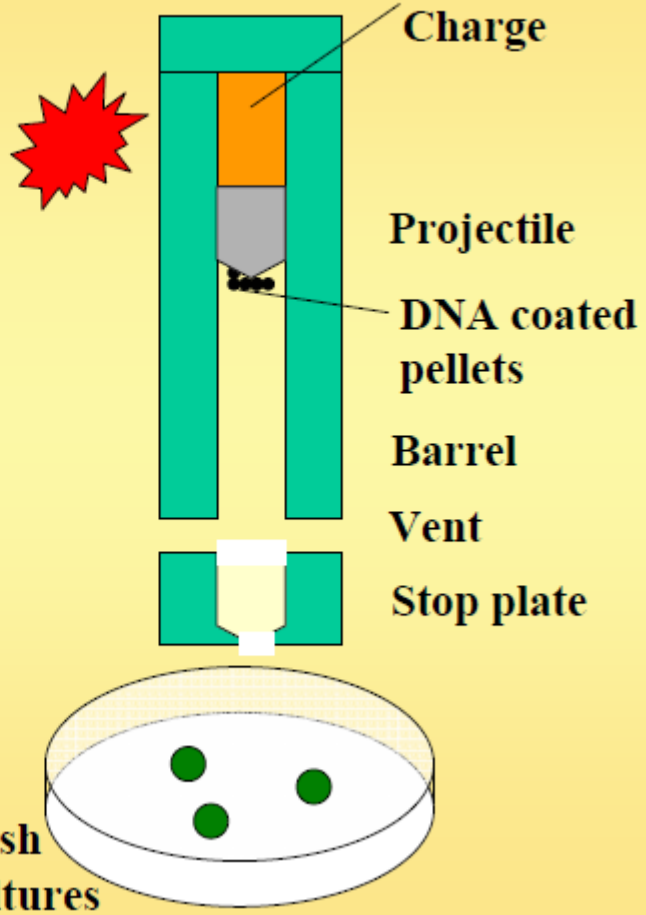
1. The "Gene Gun" method

Microprojectile bombardment

Biolistics Microprojectile Bombardment

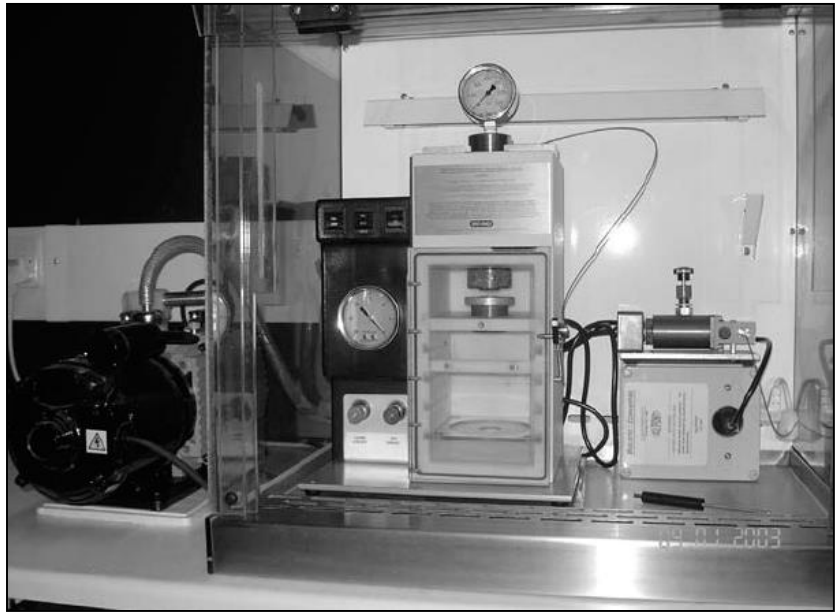
- DNA coated on pellets is forced down the barrel of a 'Particle Gun' by an explosive charge

- Shot with particle gun through cell wall of plant cells



The "Gene Gun" method
Microprojectile bombardment

- A gene gun used to transform plants by firing gold particles (~ 1 μ diameter) coated with DNA at plant tissue.
- The DNA contains the gene of interest and a selectable marker allowing only transformed cells to proliferate.
- These may then be regenerated into intact plants.



Strange,2003

The "Gene Gun" method

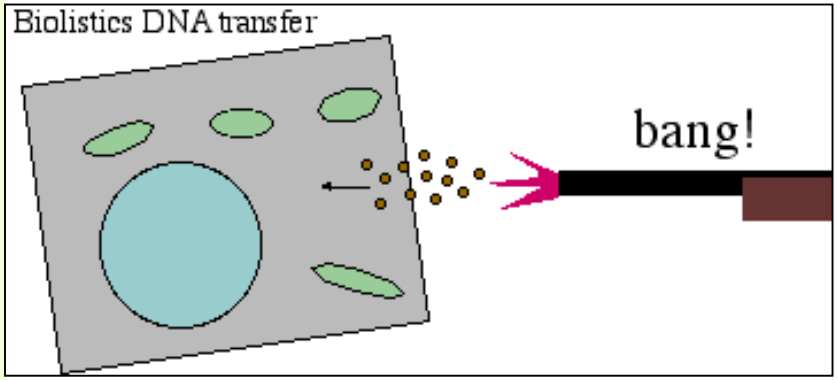
Microprojectile bombardment

- Uses a 'gene gun'
- DNA is coated onto gold (or tungsten) particles (inert) gold is propelled by helium into plant cells.
- if DNA goes into the nucleus it can be integrated into the plant chromosomes.
- Cells can be regenerated to whole plants.

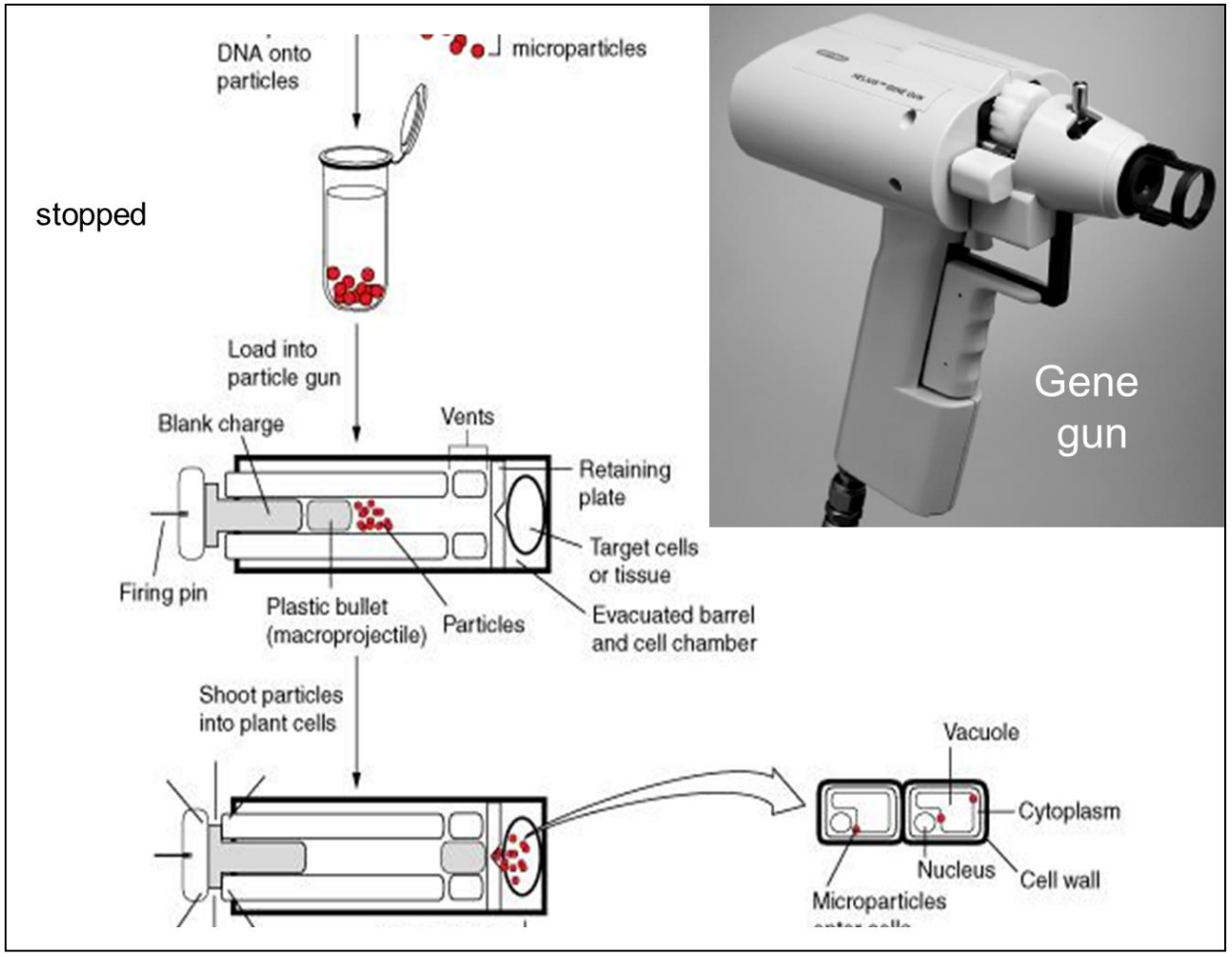
Gene Gun

The "Gene Gun" method
Microprojectile bombardment

- In the "biolistic" (a cross between biology and ballistics) or "gene gun" method, microscopic gold beads are coated with the gene of interest and shot into the plant cell with a pulse of helium.
- Once inside the cell, the gene comes off the bead and integrates into the cell's genome.



A gene gun
Particle gun



The "Gene Gun" method

Drawbacks with gene gun

- Here gold particles, coated with DNA, are fired into recipient plants.
- Although expression of foreign DNA is often obtained by this technique, it is usually transient and is frequently quickly lost as the DNA is seldom integrated into the host genome.

2. Microinjection

- Most direct way to introduce foreign DNA into the nucleus.
- Achieved by electromechanically operated devices that control the insertion of fine glass needles into the nuclei of individuals cells, culture induced embryo, protoplast.
- Labour intensive and slow.
- Transformation frequency is very high, typically up to ca. 30%.

3. Silicon Carbide Whiskers

- Silicon carbide forms long, needle like crystals.
- Cells are vortex mixed in the present of whiskers and DNA.
- DNA can be introduced in the cells following penetration by the whiskers.

The *Agrobacterium* method
In direct method

- The tumor-inducing principle of *Agrobacterium* resides in the **Ti plasmid**.
- Ti plasmids can be divided into four groups, according to the types of opine (novel derivatives of arginine) that are made:
- Nopaline,
- Octopine,
- Agropine,
- Ri plasmids.

The *Agrobacterium* method
In direct method

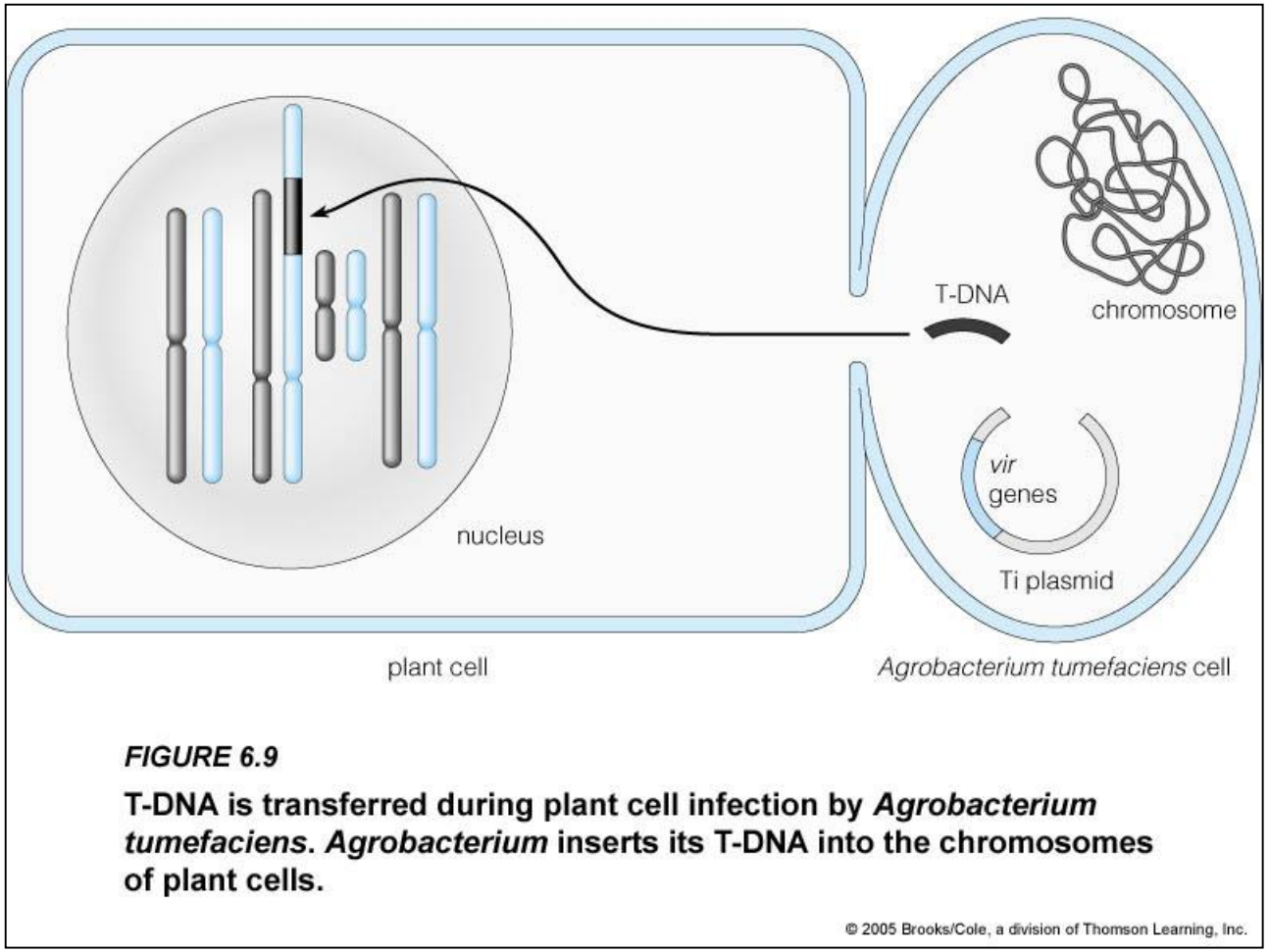
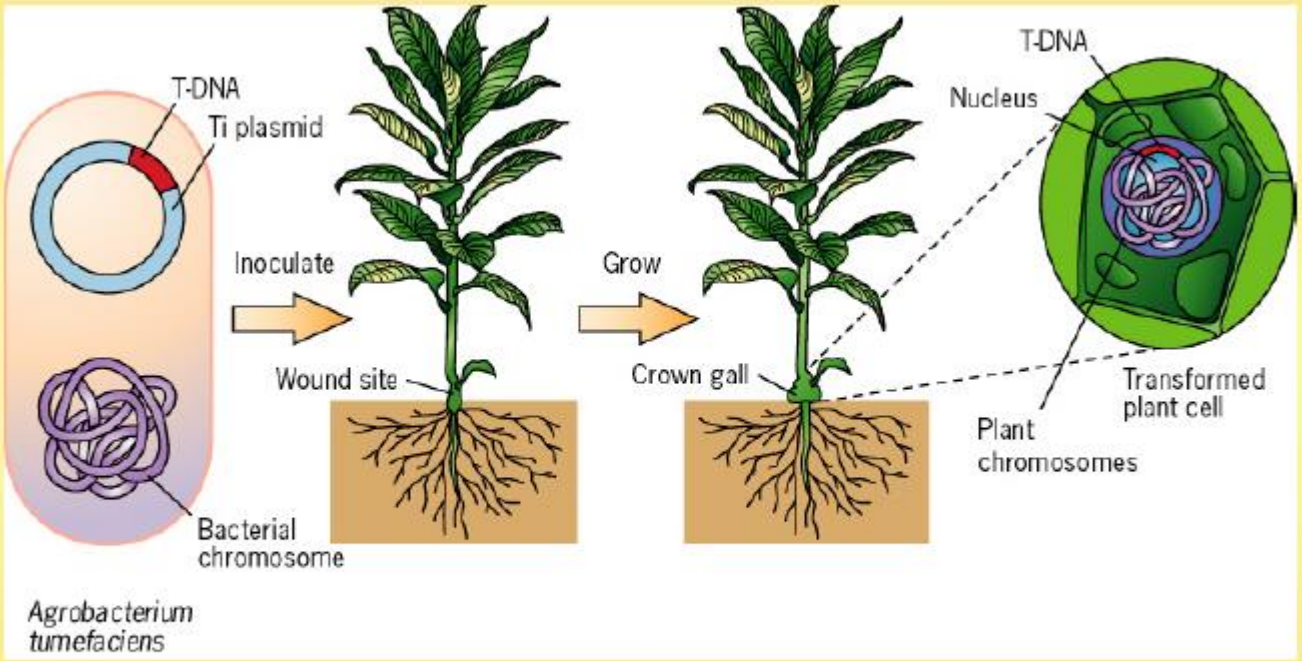


FIGURE 6.9

T-DNA is transferred during plant cell infection by *Agrobacterium tumefaciens*. *Agrobacterium* inserts its T-DNA into the chromosomes of plant cells.

© 2005 Brooks/Cole, a division of Thomson Learning, Inc.

The *Agrobacterium* method
In direct method



Transformation of plant cells by *Agrobacterium tumefaciens* harboring a wild-type Ti plasmid.

Ti plasmids

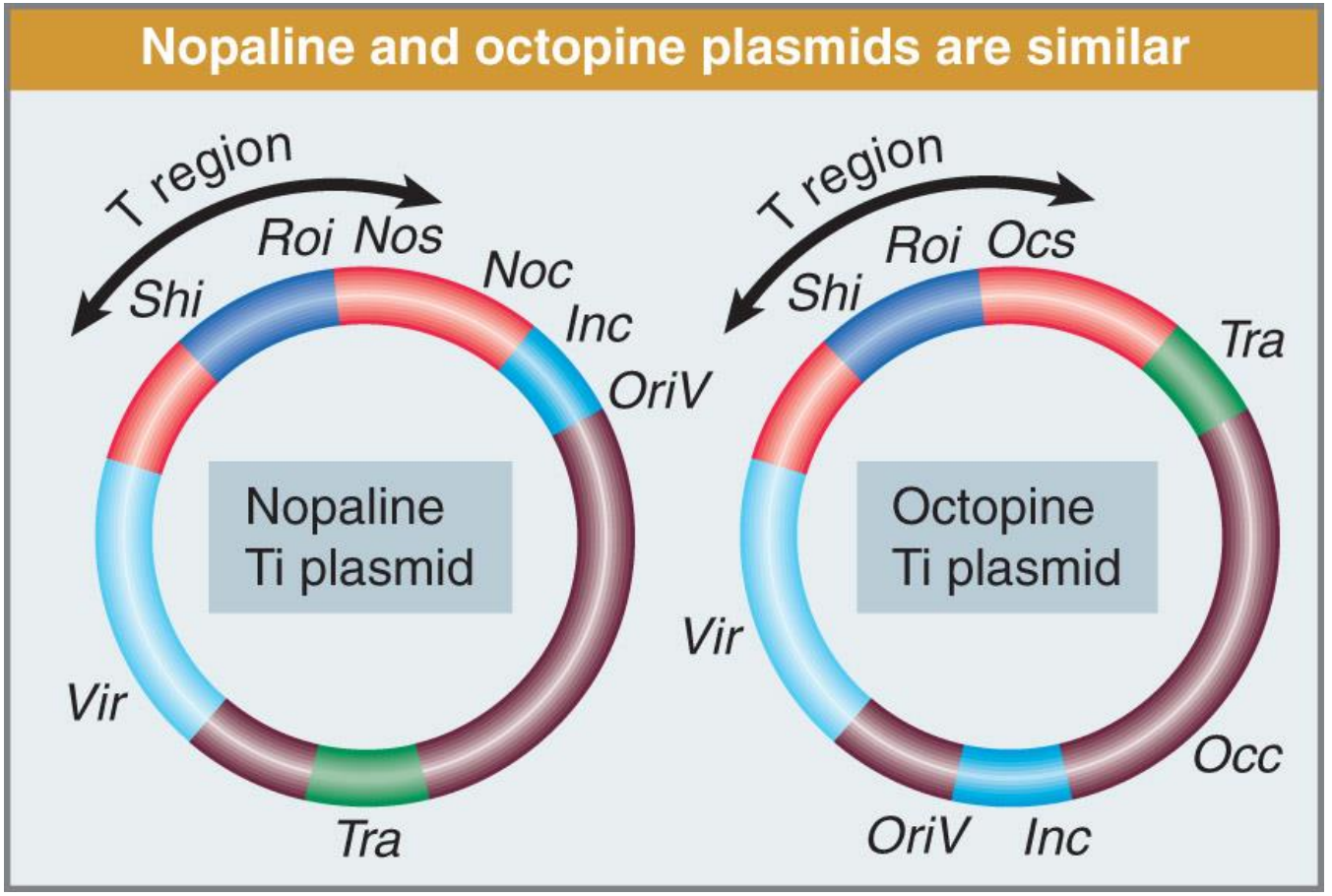
Carry genes involved in both plant and bacterial functions

Ti genes function in bacteria and in plants		
Locus	Function	Ti Plasmid
vir	DNA transfer into plant	all
shi	shoot induction	all
roi	root induction	all
nos	nopaline synthesis	nopaline
noc	nopaline catabolism	nopaline
ocs	octopine synthesis	octopine
occ	octopine catabolism	octopine
tra	bacterial transfer genes	all
Inc	incompatibility genes	all
oriV	origin for replication	all

Genes IX,2008

Ti plasmids

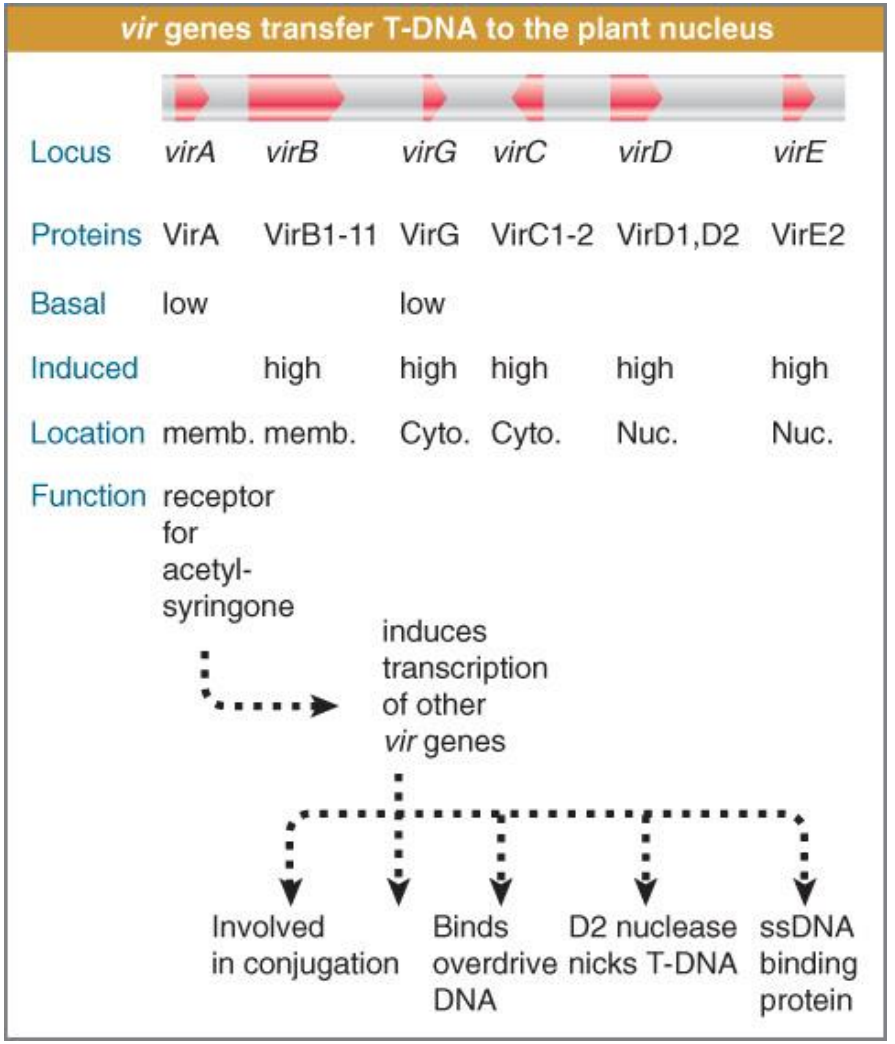
Nopaline and octopine Ti plasmids carry a variety of genes, including T-regions that have overlapping functions



Genes IX,2008

Ti plasmids

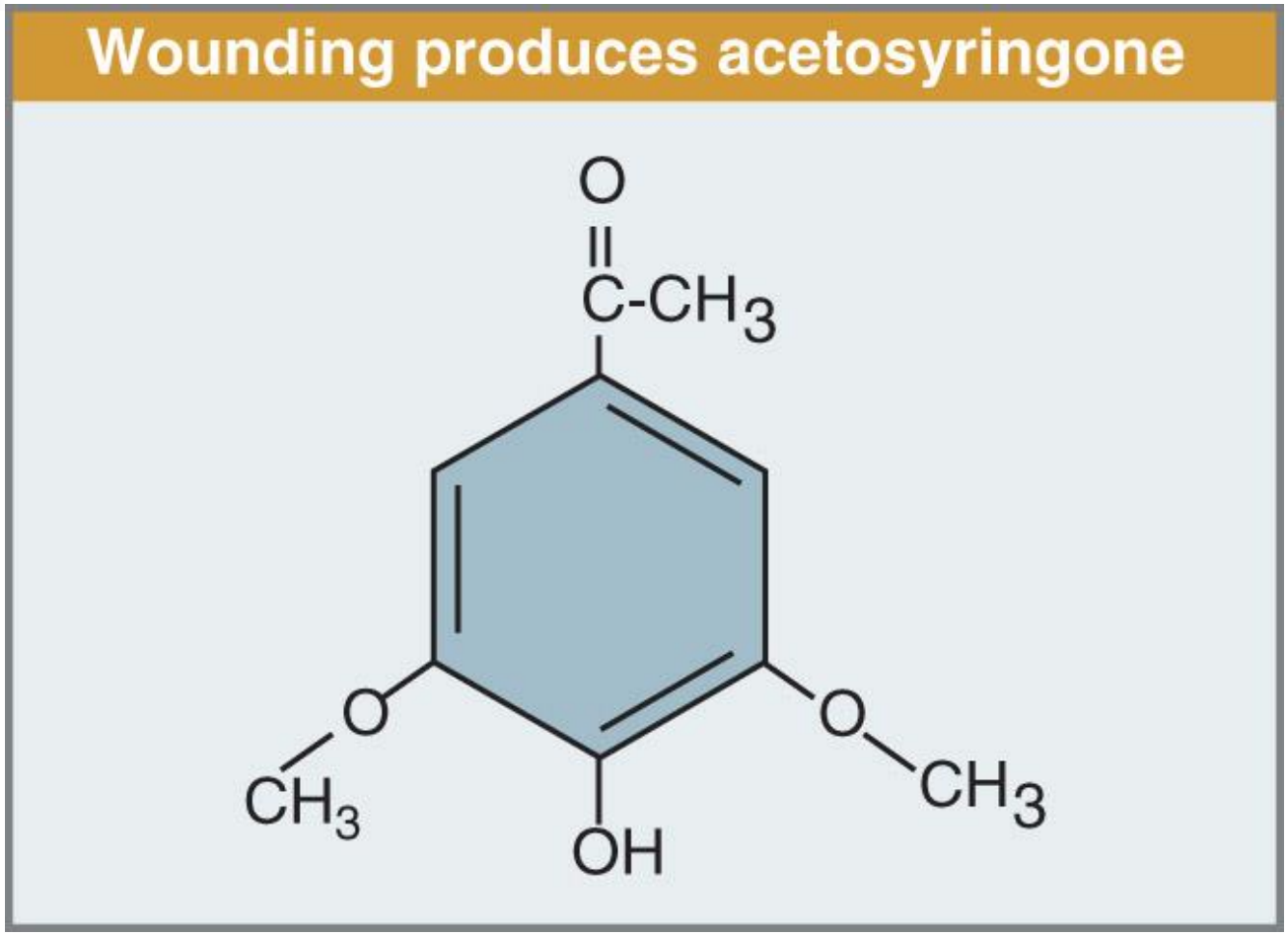
The *vir* region of the Ti plasmid has six loci that are responsible for transferring T-DNA to an infected plant



Genes IX,2008

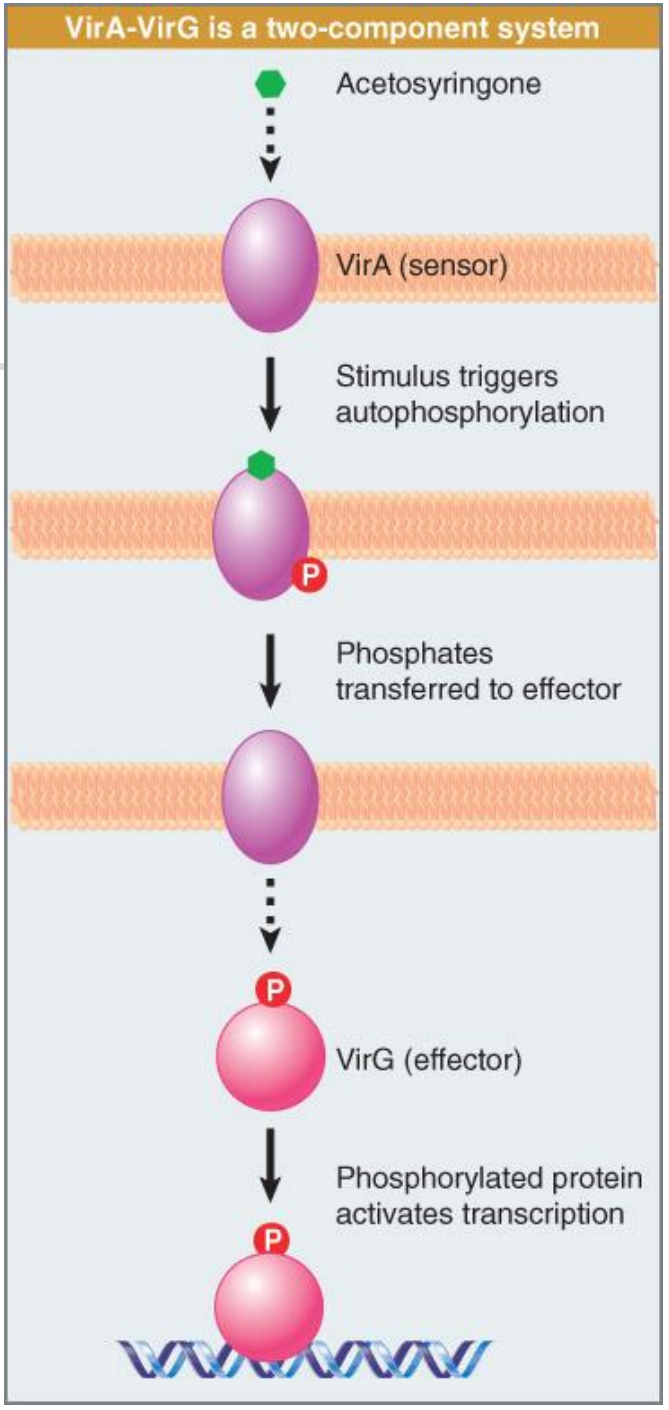
Ti plasmids

Acetosyringone (4-acetyl-2,6-dimethoxyphenol) is produced by *N. tabacum* upon wounding and induces transfer of T-DNA from *Agrobacterium*



Genes IX,2008

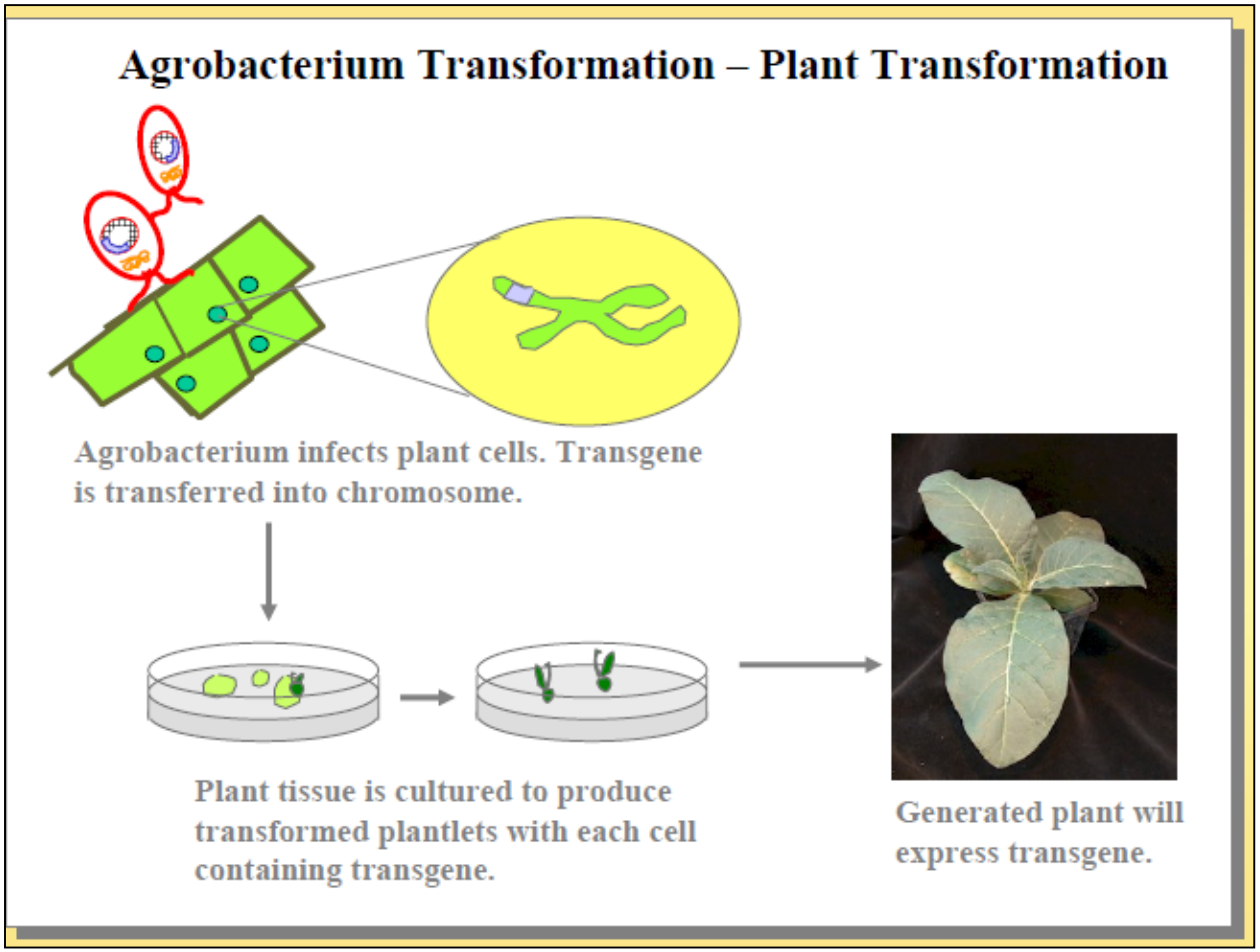
The two-component system of VirA-VirG responds to phenolic signals by activating transcription of target genes.



Genes IX,2008

The *Agrobacterium* method

Indirect method



Disadvantages of genetic transfer

GE (Genetically Engineered) or GMO foods

Disadvantages of genetic engineering (GE) products

- Health
- Environment
- Social and Cultural
- Economic

Disadvantages of genetic engineering (GE) products

- Unknown long-term effects: GE products are so new in the market that their long term effects are unknown.
- Most GE foods are tested in the United States using "Substantial equivalence" that is criticized by many opponents of GE.
- Allergens and toxins: GE products could produce new toxic substances (toxins) or allergens (substances that produce an allergic reaction in people). For example, some GM soybean plants. New viruses and bacteria: Genes could possibly transfer from one species to another.
- These genes could create new viruses and bacteria with properties that present drugs cannot deal with.
- Superbugs: Genes could possibly transfer from one species to another.
- If an antibiotic resistant gene were to transfer from an organism into a disease-creating bacteria, then an antibiotic-resistant bug (bacteria) would be created.

Disadvantages of GE

Gene flow

- Bacteria and fungi are capable of capturing and using genetic material from their surroundings (for example, from decaying plant matter or microorganisms).
- GM material, plants and decaying material, could possibly transfer their genes to other organisms such as bacteria; these genes could then be taken up other organisms.
- The effect of the absorbed gene in these other species is unknown.
- Biochemical instability: In a living organism, genes do not always work on a "one gene, one trait" basis.
- Genes interact with each other and also with their immediate environment.
- The introduction of new genetic material may create unforeseen interactions with other genes that may produce toxins or allergens.

Disadvantages of GE

Gene flow

- Antibiotic resistance: The process of genetic modification uses selection markers to indicate which organisms have taken up the required gene.
- A commonly used marker is an antibiotic resistance gene.
- If gene transfer occurs between species then bacteria and viruses may gain antibiotic resistance creating "Superbugs".
- Decreased biodiversity: With the advantages promoted for GM crops, certain cultivars are being planted to the exclusion of other cultivars.
- This will lower the number of genetic variations in our environment (the diversity of our biosphere or biodiversity).

Disadvantages of GE

- Cultural values: When transgenic species are formed, genes are transferred from one species to another.
- Various beliefs (religious, cultural, social) believe that species are separate entities and have their own spirits.
- Mixing of these spirits is unacceptable.
- Certain religions prohibit the eating of certain types of meat.
- With GM food, can there be a guarantee that this tomato does not contain, for example, a pig gene.

Disadvantages of GE

- Loss of markets: Products in the global economy must meet market needs.
- A proportion of consumers actively purchase non-GM produce.
- Any company or country that does not meet those needs could lose markets or market share.
- For example, in the United Kingdom, some supermarket chains have decided to accept only non-GE foods.
- This response to consumer pressure creates a market for non-GM products.
- Trade barriers: Concerns over GM foods (particularly in Europe) have led to legislation covering GM foods and their safety.
- This legislation could be used by countries to create trade barriers - that the imported food was genetically modified and so was a risk to the people's health and safety.

General terms and abbreviations

- **Alleles:** A variant of the DNA sequence at a given locus is called an allele. A **locus** (plural **loci**) itself is the specific location of a gene or DNA sequence on a chromosome.
- Or one of a number of alternative forms of a gene that can occupy a given genetic locus on a chromosome.
- As individuals carry two copies of each gene, one on each pair of chromosomes, they may have identical (homozygous) or different (heterozygous) alleles.
- **bp base pairs:** Measure of size of piece of DNA.
- **Biolistic:** A method (biological ballistics) of transecting cells by bombarding them with microprojectiles coated with DNA.
- **Bioinformatics:** The genome sequence consists only of millions of DNA bases in a defined order. Bioinformatics is the use of computers to obtain biologically interesting information from the sequence data.
- This starts with identifying genes, primarily by searching for open reading frames (ORFs) – although in practice it is more complex than that.
- The gene sequences are then automatically translated to determine the protein sequences they encode.

General terms and abbreviations

- **cDNA:** Complementary DNA, a double-stranded DNA molecule prepared in vitro by copying an RNA molecule back into DNA using reverse transcriptase. The RNA component of the resulting RNA-DNA hybrid is then destroyed by alkali, and the complementary strand to the remaining DNA strand synthesized by DNA polymerase. The resulting double-stranded DNA can be used for cloning and analysis.
- **cDNA library:** a very selective library which is designed so that it will have only the coding sequences of expressed genes represented.
- **Coda:** A duplex DNA sequence complementary to an RNA molecule of interest, carried in a cloning vector.
- **Constitutive expression:** Genes that are expressed whenever protein synthesis occurs, usually housekeeping genes (central to metabolism).
- **DAPI 4´,6-diamidino-2-phenylidine:** A fluorescent DNA binding dye; used to label chromosomes.
- **Downregulation:** repression, or suppression – decrease the rate of gene transcription.
- **Down regulation reduction:** Not cessation, in the effect being studied, e.g. gene expression.
- **Downstream:** 1. Location of a motif or domain in a gene nearer the 3' end of the sequence than a reference site 2. later reactions in a biochemical cascade or pathway.

General terms and abbreviations

- **FISH:** Fluorescence in situ hybridization.
- **Flanking region**: The DNA sequences extending on either side of a specific locus or gene.
- **5' flanking region:** A region of DNA that is adjacent to the 5' end of the gene. The 5' flanking region contains the promoter, and may contain enhancers or other protein binding sites. It is the region of DNA that is not transcribed into RNA.
- **3' flanking region:** A region of DNA which is NOT copied into the mature mRNA, but which is present adjacent to 3' end of the gene.
- **Genomics**: Sequencing and molecular characterization of genomes.
- **Genomic library:** Contains a collection of recombinant DNA clones that collectively constitute the genome of the organism (see DNA library).
- **GFP:** Green fluorescent protein; used as a marker for the protein product of an expressed gene.
- **Gene editing** (or genome editing) is the insertion, deletion or replacement of DNA at a specific site in the genome of an organism or cell.
- **Helicases:** A GTP- or ATP-dependent enzyme that can unwind a nucleic acid duplex.

General terms and abbreviations

- **Hfr**: High frequency of recombination. A strain of bacteria that has incorporated an F factor into its chromosome and can then transfer the chromosome during conjugation. In *Escherichia coli*, a cell having its fertility factor integrated into the bacterial chromosome; a donor (male) cell.
- **Housekeeping gene:** Is typically a constitutive gene that is transcribed at a relatively constant level.
- **Inducible expression:** Genes that are normally turned off but are expressed under certain conditions, e.g. to use new carbon/energy source (catabolism).
- **kb kilobase**: 1000 base pairs; measure of size of piece of DNA.
- **-mer:** Denotes number of units in a molecule; e.g. 12-mer oligonucleotide indicates a molecule with 12 nucleotides.
- **Metagenomics:** the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics (See the file microbial ecology).
- **Methylation:** (DNA or protein methylation) denotes the addition of a methyl group to a substrate or the substitution of an atom or group by a methyl group. Methylation can be involved in modification of heavy metals, regulation of gene expression, regulation of protein function, and RNA processing. Protein methylation is one type of post-translational modification.

General terms and abbreviations

- **NADPH (Nicotinamide adenine dinucleotide phosphate):** is a coenzyme used in anabolic reactions, such as lipid and nucleic synthesis, which require NADPH as a reducing agent.
- **Null mutation:** A mutation that results in complete absence of function.
- **Okazaki fragments:** The pieces of DNA produced during the formation of the lagging strand that are later ligated together to form a complete DNA.
- **Operon:** A genetic unit containing several genes coding for the enzymes for a metabolic pathway.
- The genes are clustered and are transcribed together into a polycistronic mRNA (a mRNA molecule that contains the code for more than one protein or enzyme).
- The transcription of each operon is initiated at a promoter region and controlled by a neighbouring regulatory gene.
- This regulatory gene specifies a regulatory protein (a repressor or apoinducer) that binds to the operator sequence in the operon to either prevent or allow its transcription.
- **Operator (gene):** Region of DNA found in many bacterial operons to which the repressor (or apoinducer) binds, thus preventing (or allowing) transcription of the operon.
- ***oriC*** **:** *Escherichia coli* origin of DNA replication.

General terms and abbreviations

- **Promoter:** A region of a DNA molecule to which an RNA polymerase binds and initiates transcription.
- **Quorum sensing:** Control of gene expression based on bacterial density (i.e., existence of a quorum).
- **Open reading frame (ORF):**
- A nucleic acid sequence (DNA or RNA sequence) with a reading frame that contains nonstop codons; it can therefore potentially be translated into a polypeptide.
- The region of the nucleotide sequences from the start codon (ATG) to the stop codon is called the open reading frame.
- Each sequence contains all the information to produce a particular protein.
- **Repressible expression:** Genes that are normally expressed, but are turned off under certain circumstances, e.g., when sufficient quantities of a factor (such as an amino acid) is present (anabolism).
- **Ribozymes:** RNA molecules that act as enzymes; molecular scissors that cut RNA; catalyze fundamental biological processes.
- **Single nucleotide polymorphisms (SNP):** Individual variations in sequence.

General terms and abbreviations

- **Transposable element:** A DNA sequence that can move from one chromosomal site to another.
- **Transposon:** A class of transposable element with characteristic flanking sequences.
- **Transposons (The simplest transposons or ISs):** are insertion sequences (IS) that contain only a single gene.
- **Transposons (The larger transposons or Tns):** composite transposons which may have various genes.
- **Transposition:** Involves the movement of a DNA sequence to another position on the same or different DNA molecule.
- **Upstream** identifies sequences in the opposite direction from expression; for example, the bacterial promoter is upstream of the transcription unit, the initiation codon is upstream of the coding region.
- **Upregulation:** activation, or promotion – increase the rate of gene transcription.

Selected References

- Abedone, S.T. Power point presentations and lecture notes. The Ohio State University. www.phage.org. abedon.1@osu.edu.
- Agroresearch,2001. Disadvantages of genetically modified organisms.
- Bacterial genetics, 93303 bact gen class. ppt. Power point presentation.
- BIO 184 Laboratory Manual Page 38 CSU. Sacramento Updated: 12/19/2005 EXPERIMENT 3: Creating recombinant DNA molecules.
- Chapter 4. Gene=Enzyme, RNA Transcription, Genetic Code. Biology Dept. Kenyon College, BIOL 114.
- Dyer, D. 1999. Bacterial Genetics.
- Farabee, M.J. 2001. Control of gene expression, Use for educational purposes is encouraged.
- HFSP Workshop Reports. 2000. Cell Division and the Replicon. Senior editor: Jennifer S. Altman, Strasbourg.
- Hobden, J. A. and Jeffery A. Hobden,2006. Bacterial Genetics. PDF
- Lake Michigan College. Principles of Biology I (111). Revised November 12, 2003.
- LECTURE 4: DNA replication and genome structure.
- LECTURE 4, Mar 2000. Bacterial genetics. Lecture aims.
- Love, J. 2003. Teacher's Study Guide for Lesson Twenty One. Prokaryotic Genetics.
- Microbiology Unit 3. Bacterial Genetics and Viruses.

Selected References

- Poole, A. 2006. LUCA, the Last Universal Common Ancestor of all life. Stockholm Univ., Sweden.
- Singleton, P. 2004. Bacteria in biology, biotechnology and medicine. 6th ed., Wiley, 559 pp.
- Replication of DNA. htm. How is cellular DNA copied?
- Sherry Fuller-Espie 2003. Recombinant DNA Technology. Cabrini College. The genetic code. Htm. What songs do the "notes of DNA" dictate?
- VIRTUE Newsletter: Science.