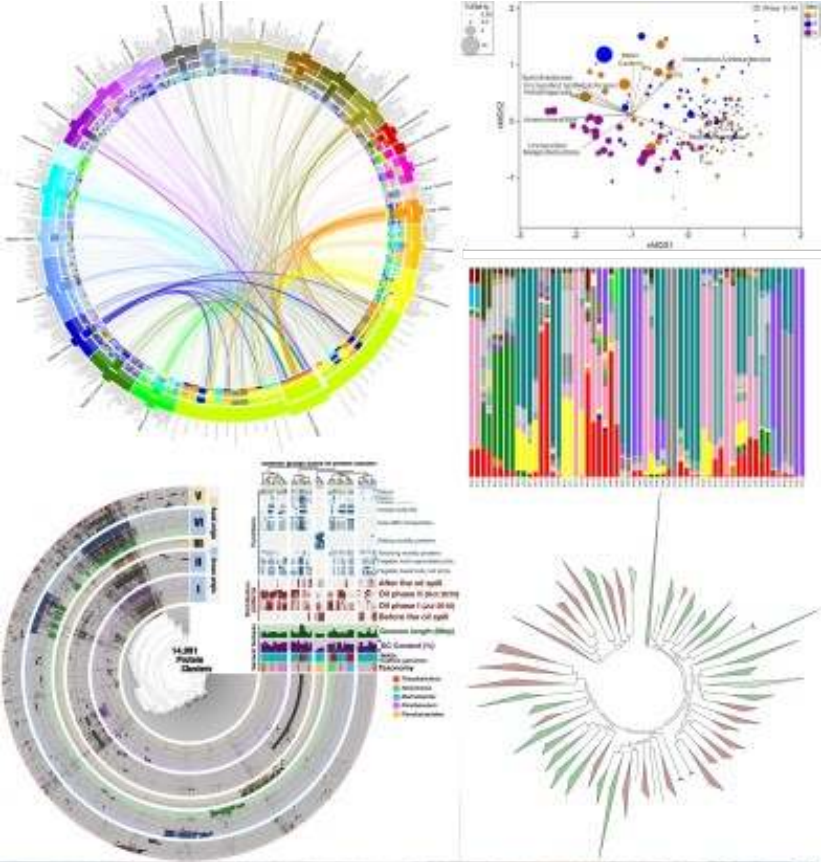


GIOVANNELLI LAB

DIPARTIMENTO DI BIOLOGIA

University of Naples Federico II

Environmental Metagenomic

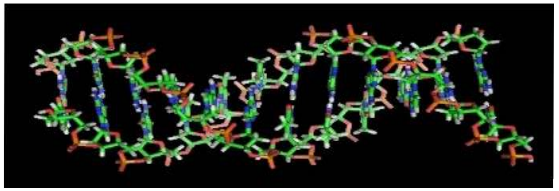
aa 2020-2021

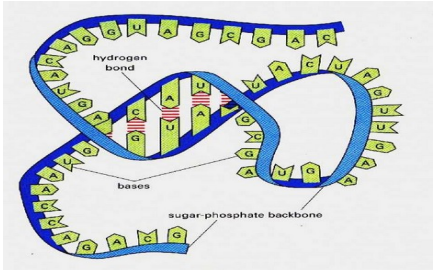
MICROBIAL GENOME ORGANIZATION

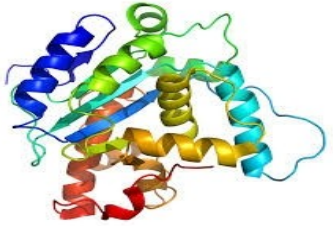
Donato Giovannelli

donato.giovannelli@unina.it • @d_giovannelli • http://dgiovannelli.github.io

Central Dogma in Molecular Biology







replication

DNA

GENES

transcription

RNA

TRANSCRIPTS

translation

Protein

PROTEIN

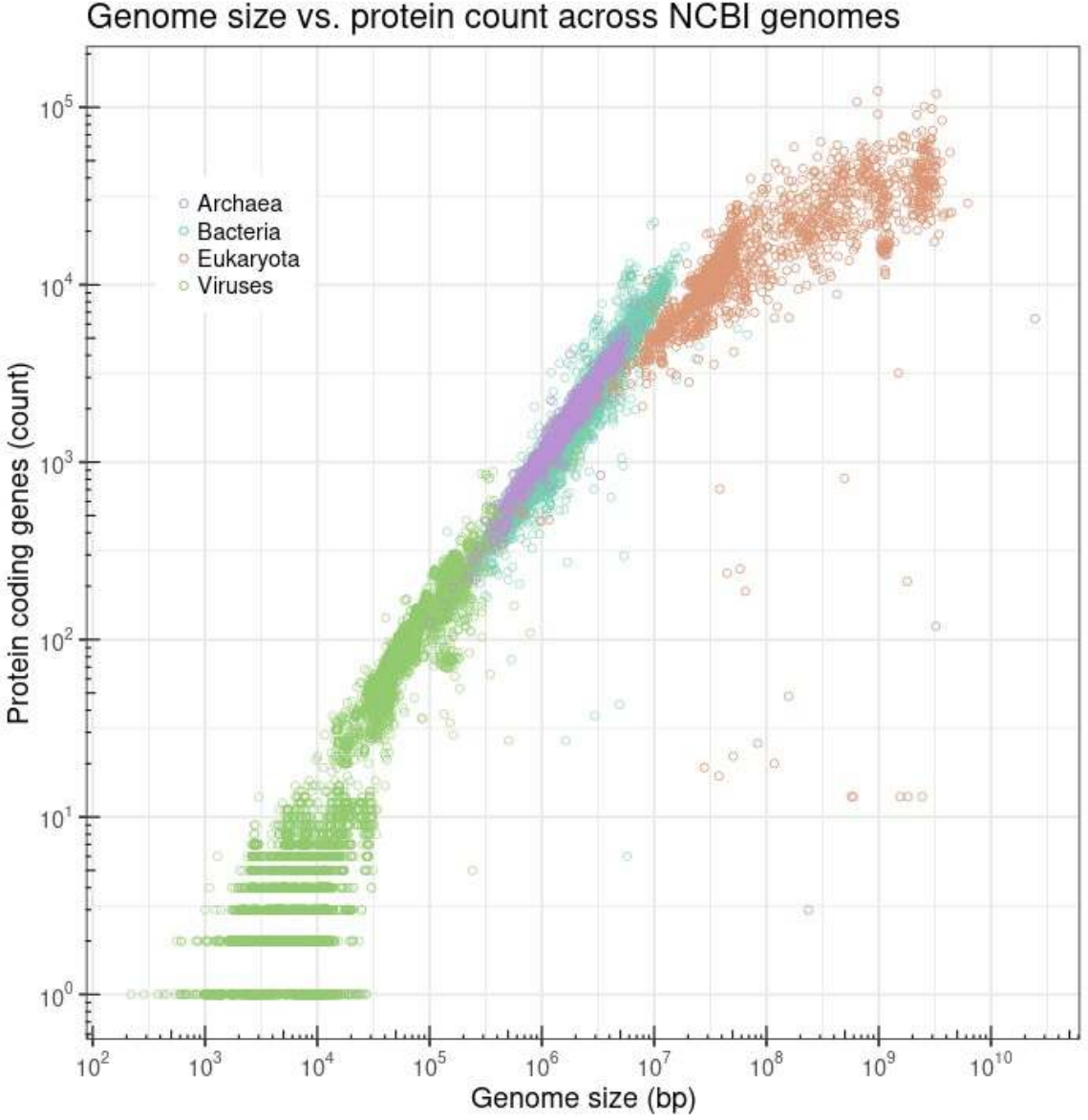
Prokaryotic Genomes

- Prokaryotic genomes range in size from those of large viruses to those of eukaryotic microbes
- Prokaryotic genomes are generally circular. One or more chromosome may be present
- One or more plasmid can be present (circular or linear), coding (often) for non essential genes conferring additional characteristics
- The gene density for 1kbp is around 1, and number of genome repeat is small. By comparison simple eukaryotes (*C. elegans*) has a gene density of 0.2, while humans have a gene density of 0.015
- Unlike prokaryotes, eukaryotic genomes contain a large fraction of noncoding DNA

Genome Size

Table 2.1 Examples of genomes for which a complete or draft sequence has been published

Species	Size of genome (Mb)	Approximate number of genes	References
Eukaryotes			
Arabidopsis thaliana (plant)	125	25 500	AGI (2000)
Caenorhabditis elegans (nematode worm)	97	19 000	CESC (1998)
Drosophila melanogaster (fruit fly)	180	13 600	Adams *et al.* (2000)
Homo sapiens (human)	3200	30 000–40 000	IHGSC (2001); Venter *et al.* (2001)
Saccharomyces cerevisiae (yeast)	12.1	5800	Goffeau *et al.* (1996)
Bacteria			
Escherichia coli K12	4.64	4400	Blattner *et al.* (1997)
Mycobacterium tuberculosis H37Rv	4.41	4000	Cole *et al.* (1998)
Mycoplasma genitalium	0.58	500	Fraser *et al.* (1995)
Pseudomonas aeruginosa PA01	6.26	5700	Stover *et al.* (2000)
Streptococcus pneumoniae	2.16	2300	Tettelin *et al.* (2001)
Vibrio cholerae El Tor N16961	4.03	4000	Heidelberg *et al.* (2000)
Yersinia pestis CO92	4.65	4100	Parkhill *et al.* (2001)
Archaea			
Archaeoglobus fulgidus	2.18	2500	Klenk *et al.* (1997)
Methanococcus jannaschii	1.66	1750	Bult *et al.* (1996)


Genome size vs. protein count across NCBI genomes
Archaea
Bacteria
Eukaryota
Viruses
Protein coding genes (count)
Genome size (bp)
10^0
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
10^{10}

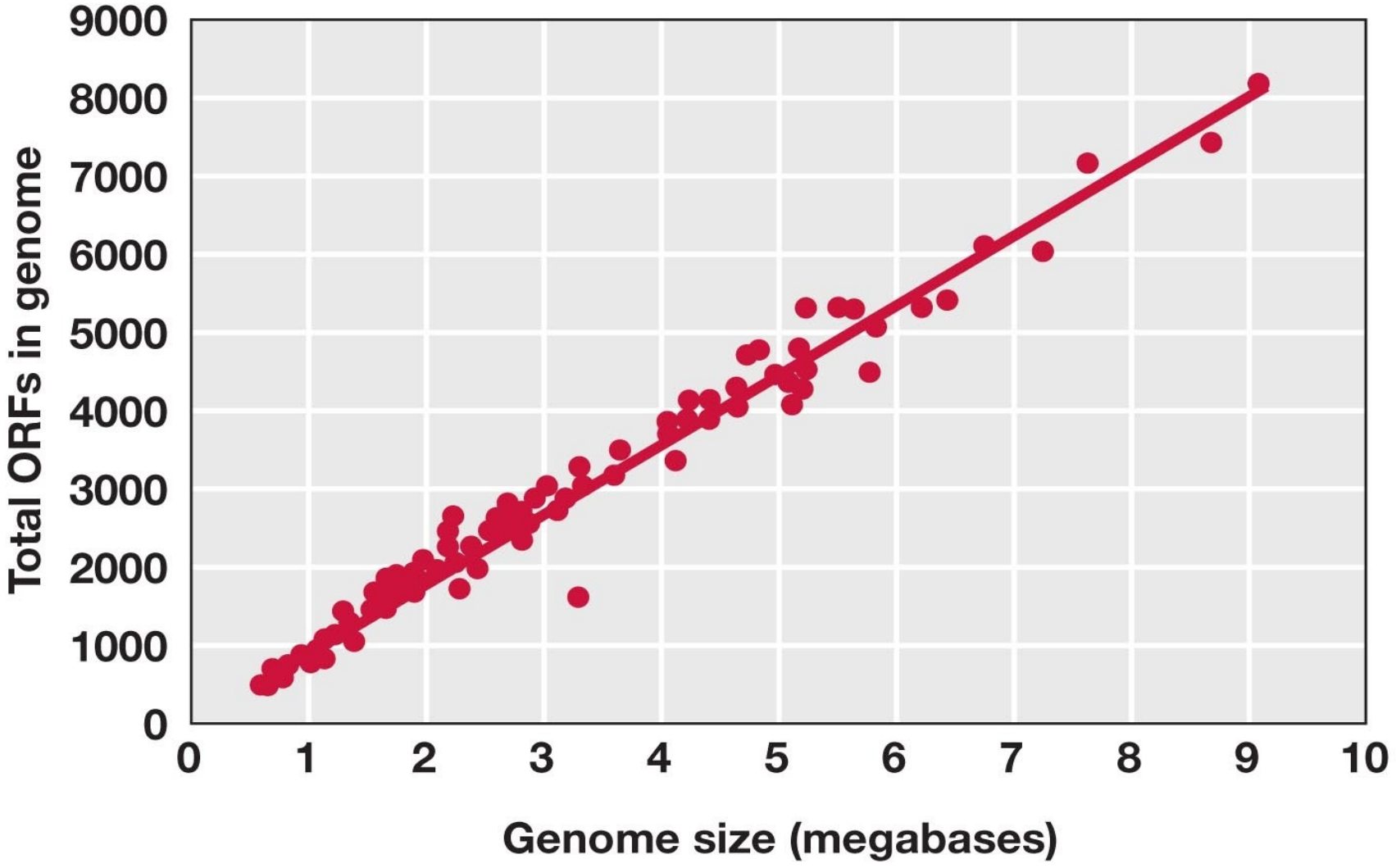
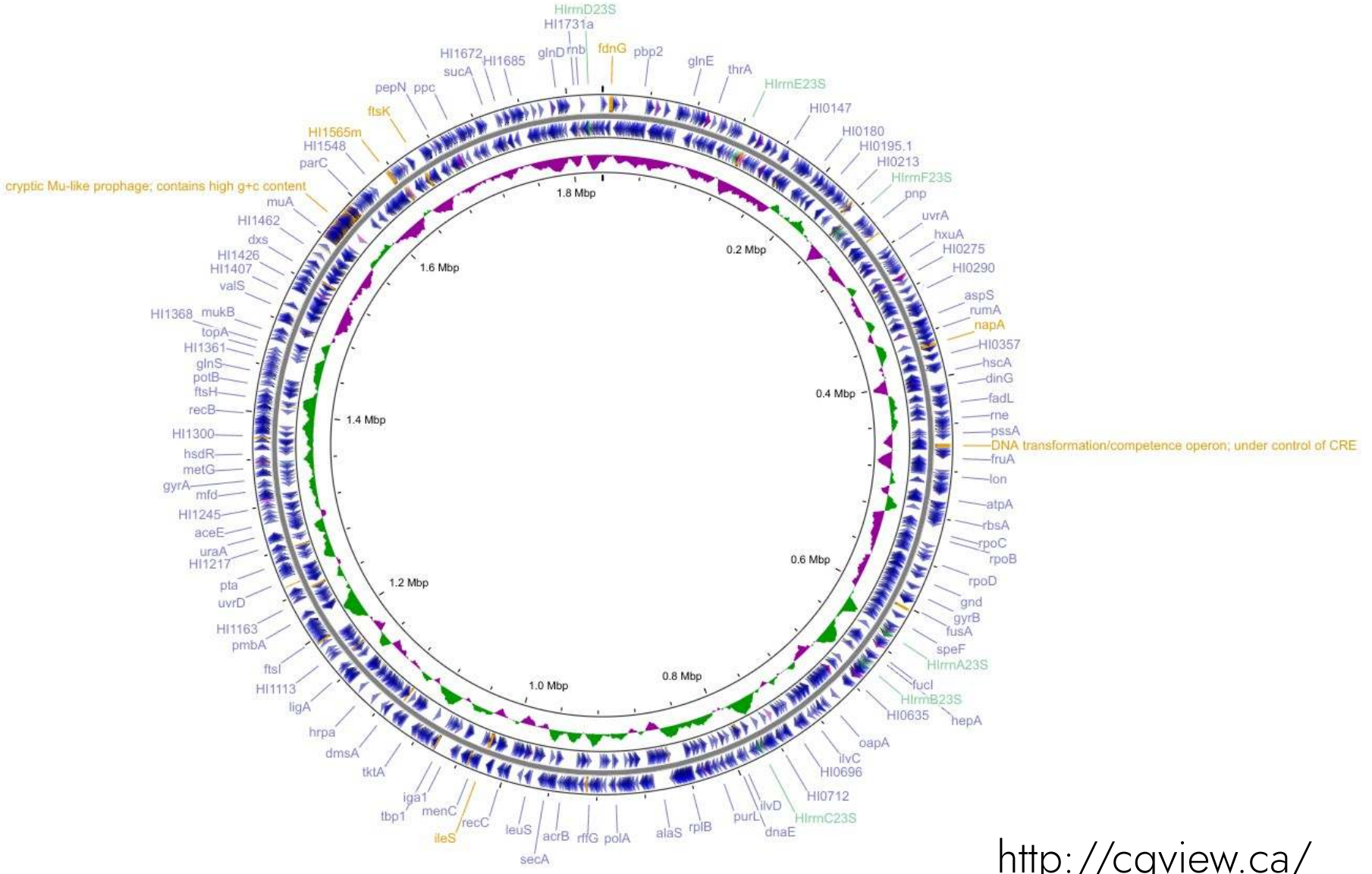

Total ORFs in genome
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
8
9
10
Genome size (megabases)

Table 2.9 Examples of genome organization in prokaryotes

Genome organization			
Species	**DNA molecules**	**Size (Mb)**	**Number of genes**
Escherichia coli K-12	One circular molecule	4.639	4397
Vibrio cholerae El Tor N16961	Two circular molecules		
	Main chromosome	2.961	2770
	Megaplasmid	1.073	1115
Deinococcus radiodurans R1	Four circular molecules		
	Chromosome 1	2.649	2633
	Chromosome 2	0.412	369
	Megaplasmid	0.177	145
	Plasmid	0.046	40
Borrelia burgdorferi B31	seven or eight circular molecules, 11 linear molecules		
	Linear chromosome	0.911	853
	Circular plasmid cp9	0.009	12
	Circular plasmid cp26	0.026	29
	Circular plasmid cp32*	0.032	Not known
	Linear plasmid lp17	0.017	25
	Linear plasmid lp25	0.024	32
	Linear plasmid lp28-1	0.027	32
	Linear plasmid lp28-2	0.030	34
	Linear plasmid lp28-3	0.029	41
	Linear plasmid lp28-4	0.027	43
	Linear plasmid lp36	0.037	54
	Linear plasmid lp38	0.039	52
	Linear plasmid lp54	0.054	76
	Linear plasmid lp56	0.056	Not known

Table 2.8 Features of typical plasmids

Type of plasmid	Gene functions	Examples
Resistance	Antibiotic resistance	Rbk of *Escherichia coli* and other bacteria
Fertility	Conjugation and DNA transfer between bacteria	F of *E. coli*
Killer	Synthesis of toxins that kill other bacteria	Col of *E. coli*, for colicin production
Degradative	Enzymes for metabolism of unusual molecules	TOL of *Pseudomonas putida*, for toluene metabilism
Virulence	Pathogenicity	Ti of *Agrobacterium tumefaciens*, conferring the ability to cause crown gall disease on dicotyledonous plants

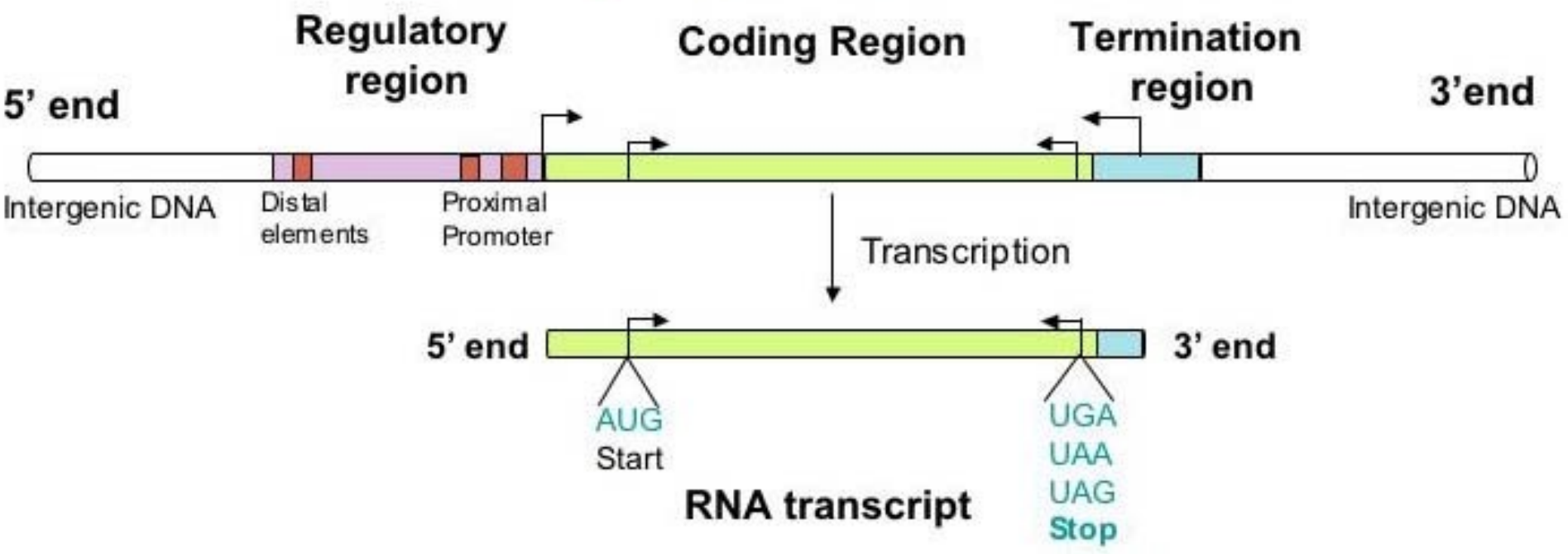


http://cgview.ca/

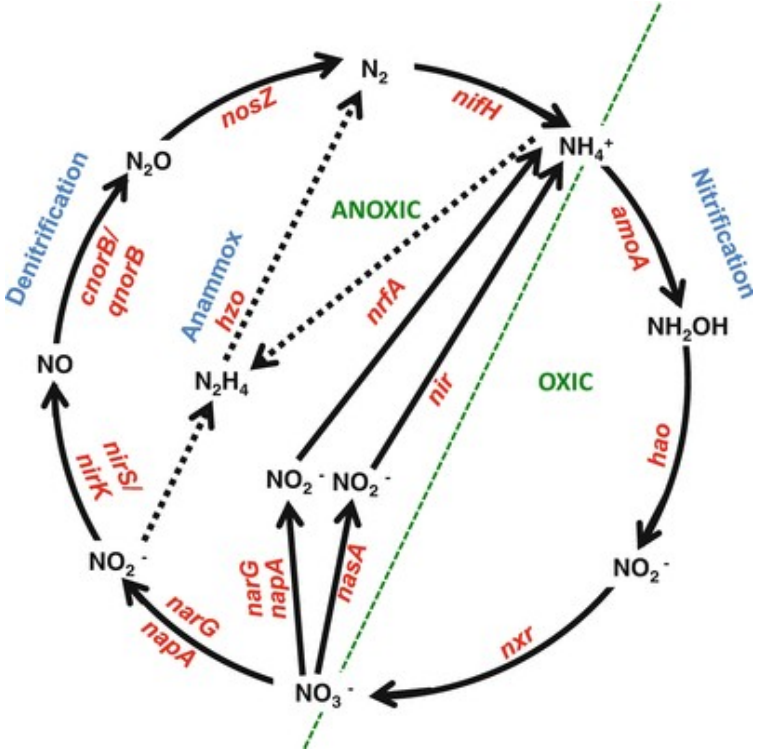
Prokaryotic genomes and the species concept

- Classical definition of species, defined during the 20th century, is based on evolutionary concepts and we now look on a species as a group of organisms that can interbreed with one another. The barrier to gene flow that is central to the species concept therefore does not hold with prokaryotes
- Early microbiologist described species in morphological terms, combining observations, staining and biochemical tests
- This type of classification was imprecise because many of the resulting species were made up of a variety of types with different properties
- Earlier molecular attempt to define prokaryotic species used 16S rRNA similarity as benchmark
- Genome projects have also confused our understanding of what constitutes a ‘species’ in the prokaryotic world.

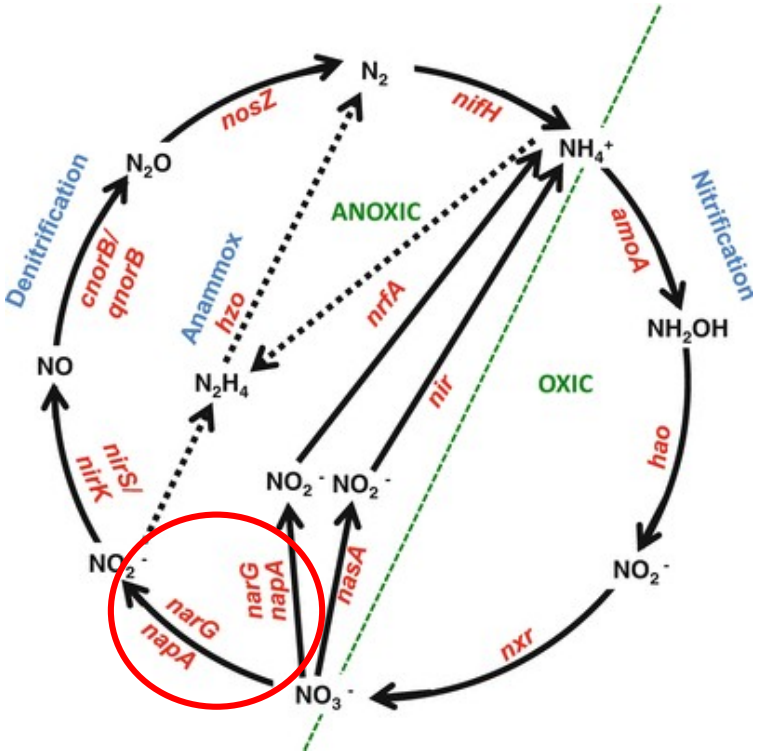
Gene organization in Prokaryotes



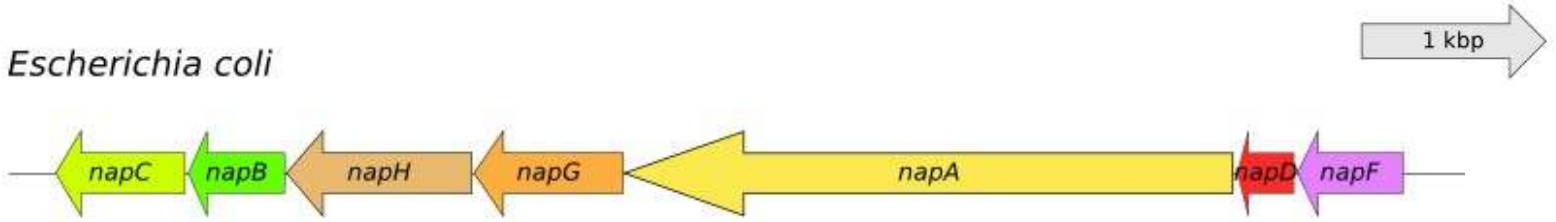
Visualizing genes organization

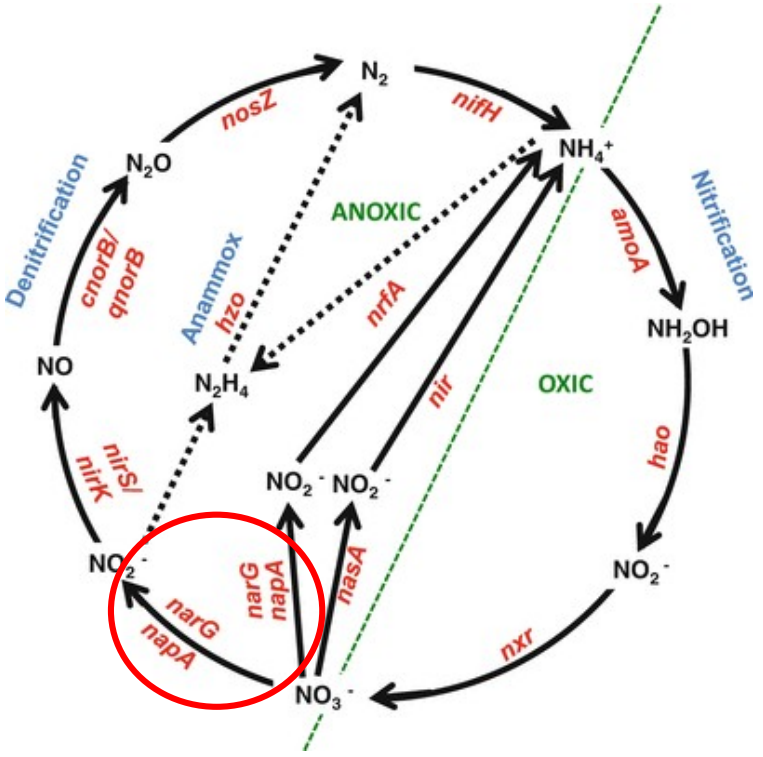


Visualizing genes organization

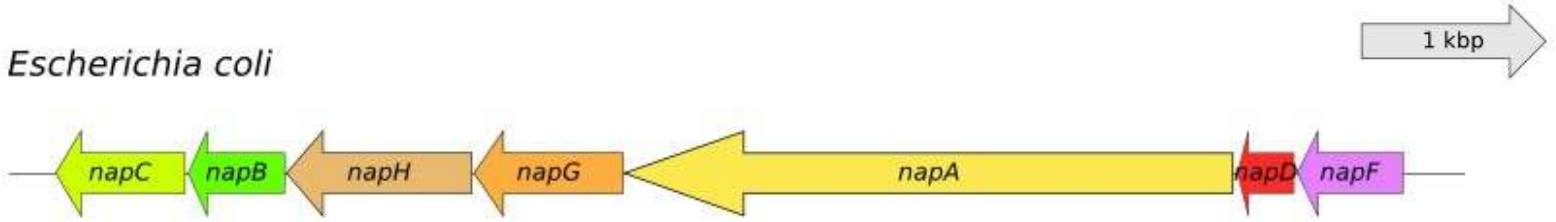


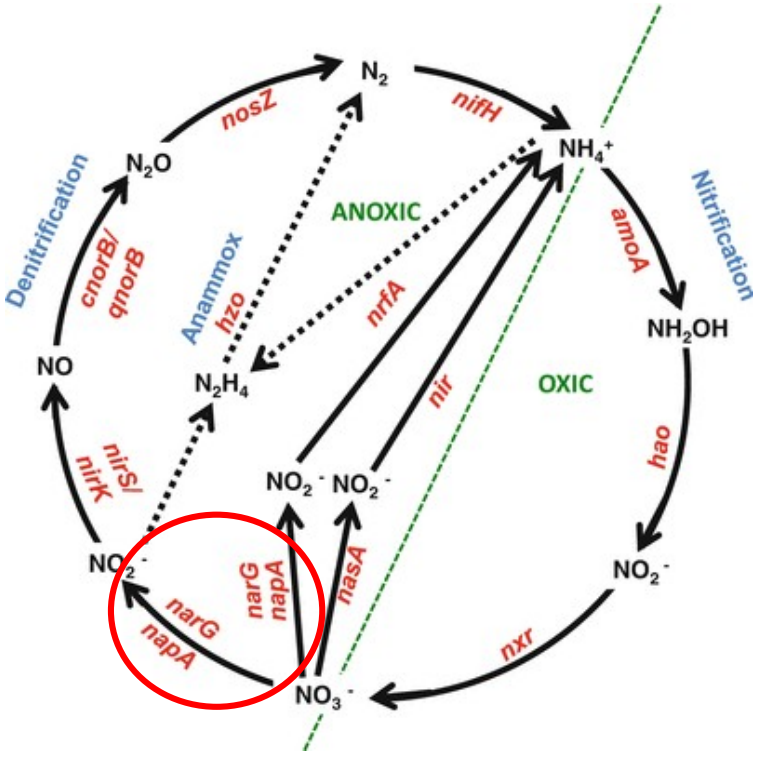
Visualizing genes organization

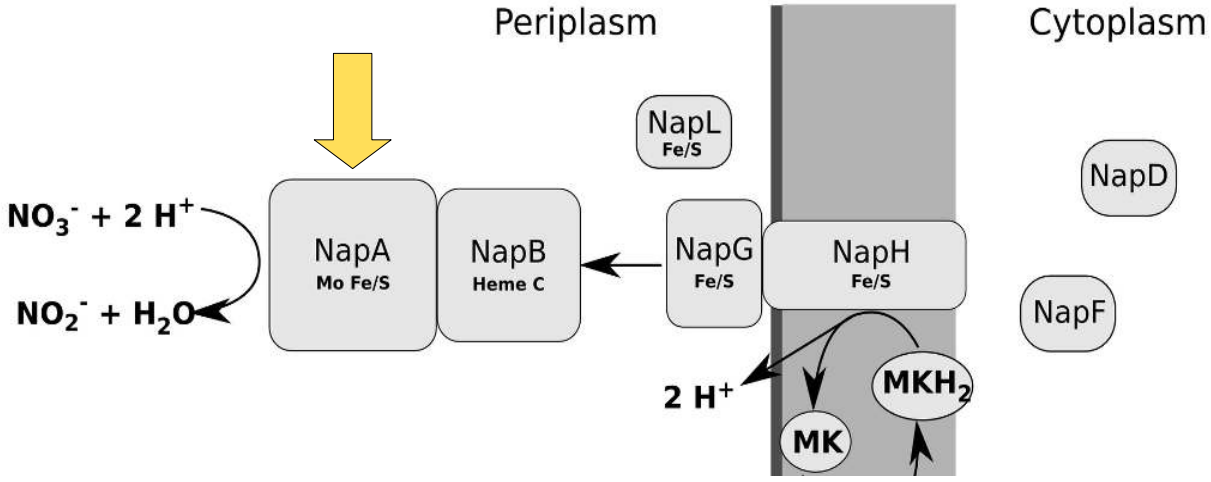




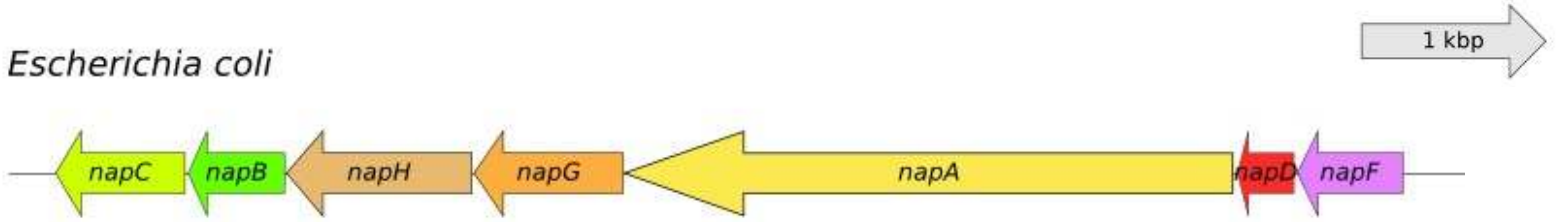
Visualizing genes organization

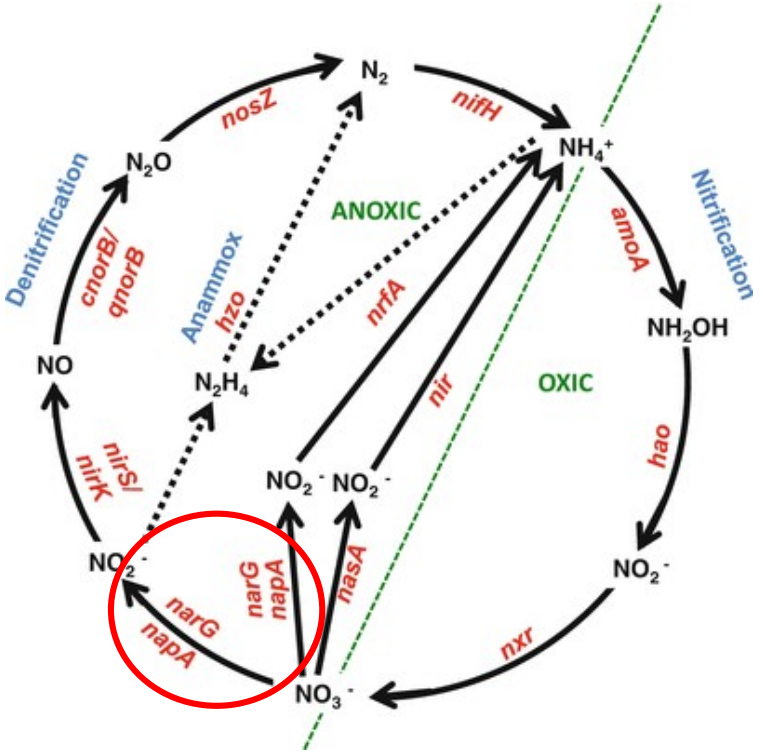


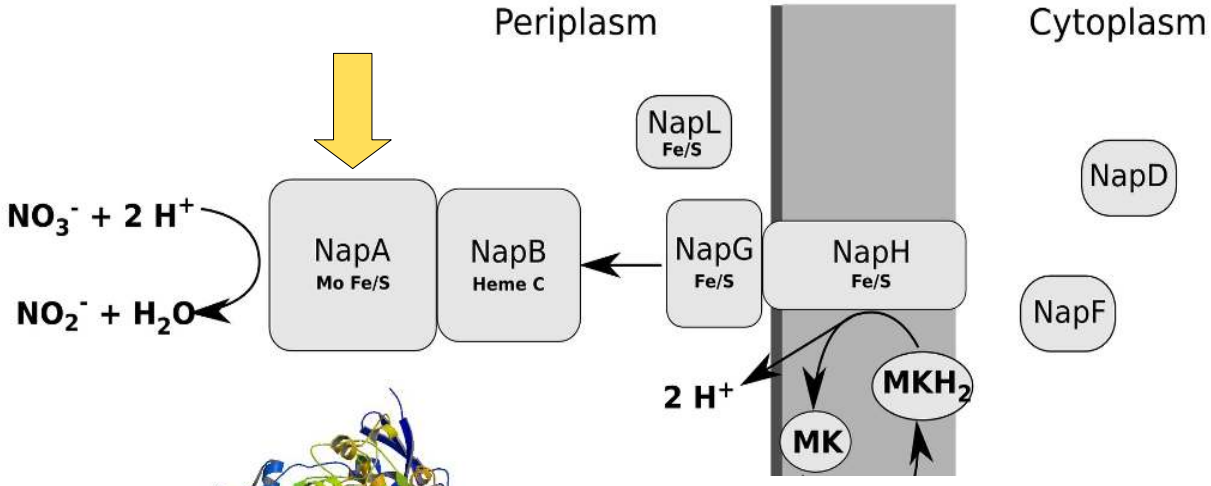


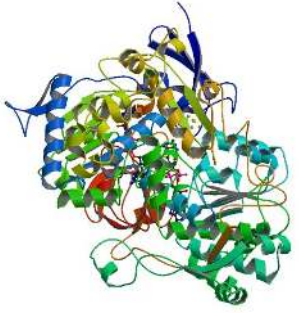


Visualizing genes organization









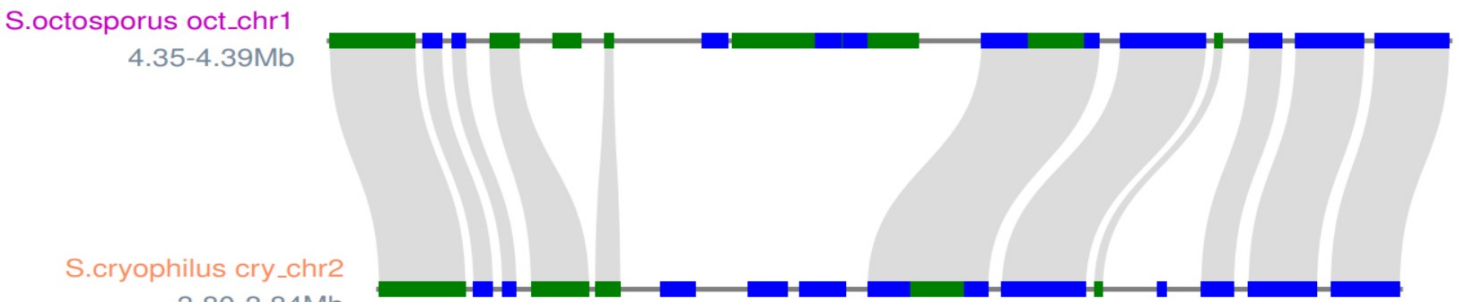
Sequence Homology

Sequence similarity: Defines the degree of similarity between two sequences. Similarity is calculated by different algorithms assigning a different weight to gaps

Sequence homology: assumes that the similarity observed is derived from common ancestry

Conserved regions: portion of a gene that do not change significantly when comparing homologs

Synteny: area of a gene/genome that is conserved in organization across different orgsnisms



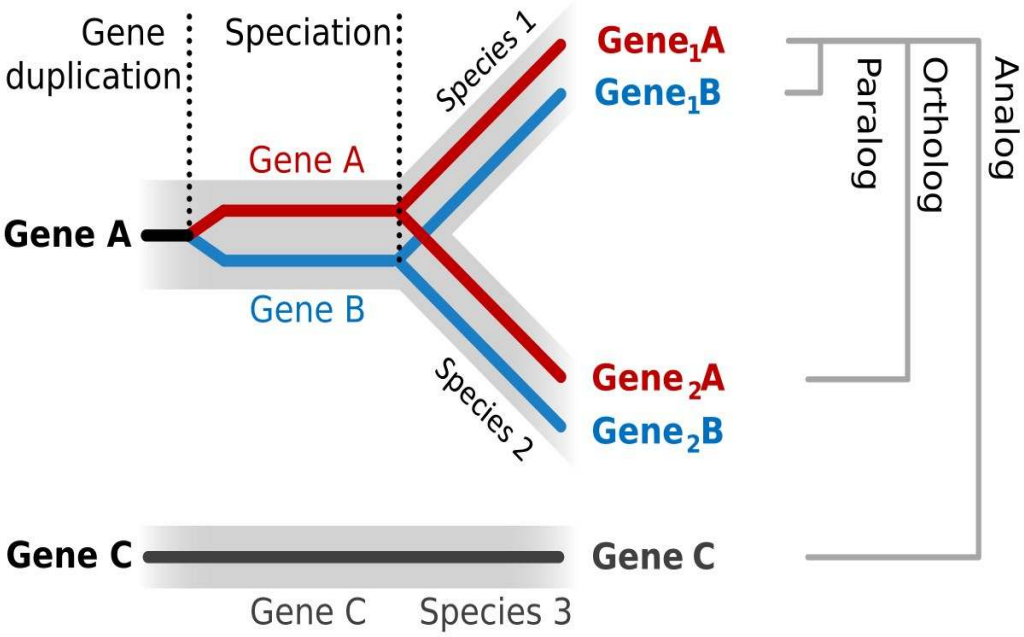
Gene Families, Duplications, and Deletions

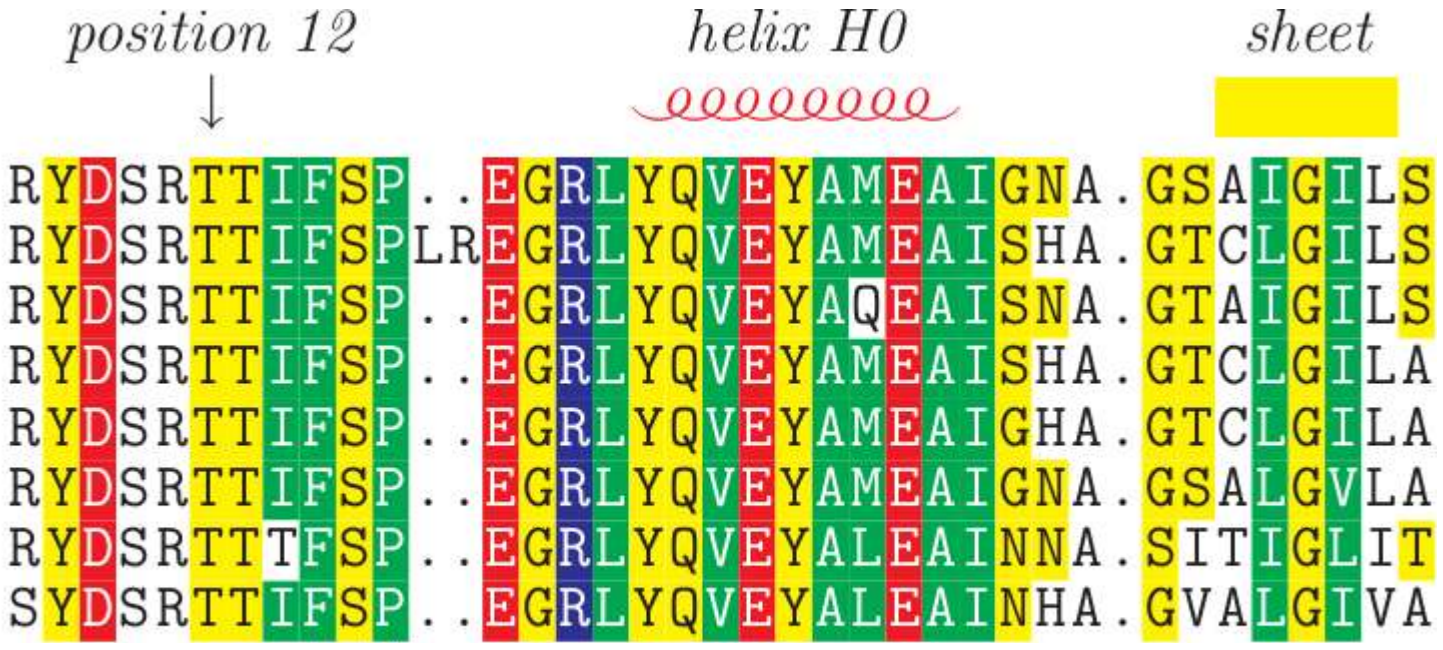
Homologs: related sequence that have common genetic ancestry

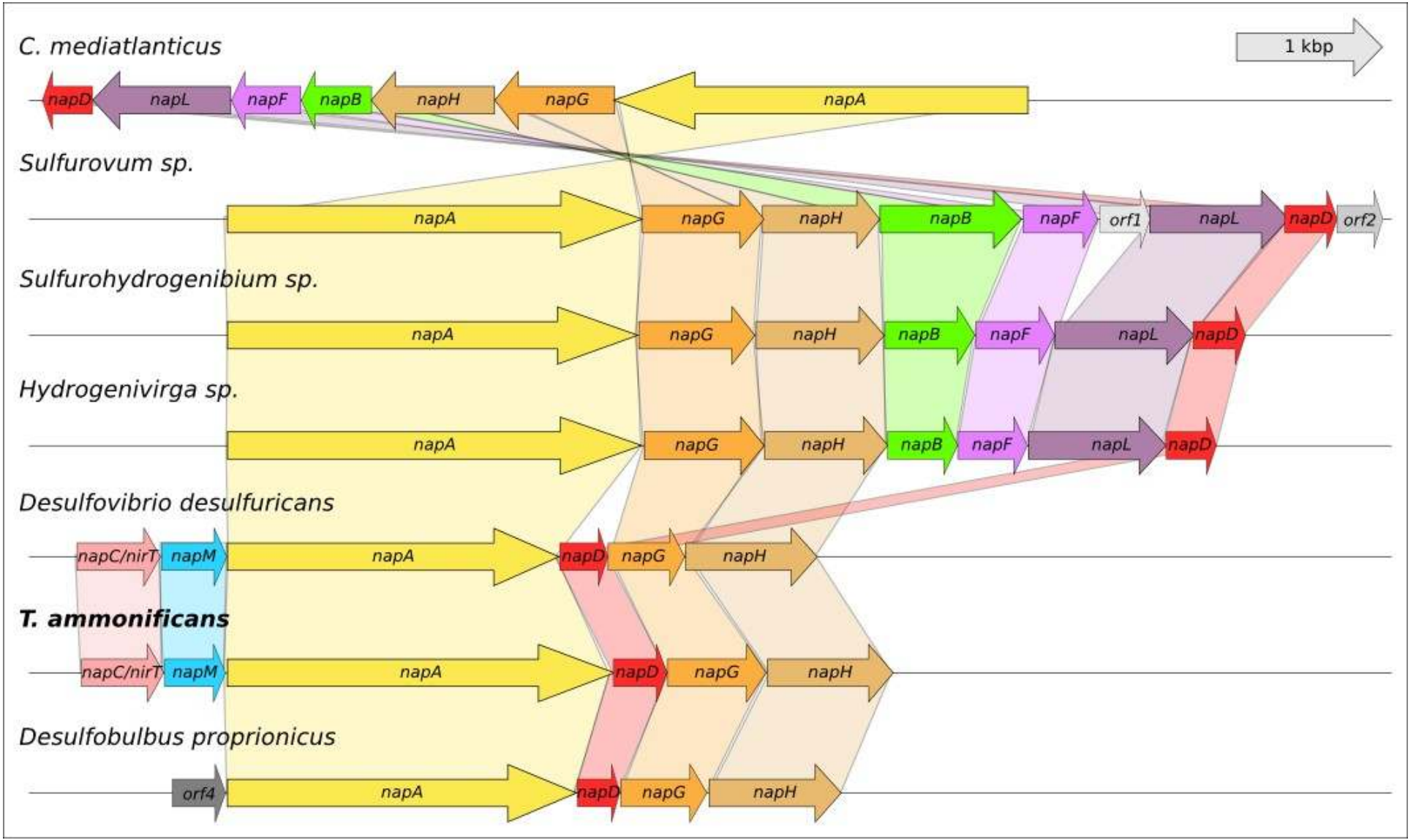
Paralogs: genes within an organism whose similarity to one or more genes in the same organism is the result of gene duplication

Gene families: groups of homologous genes

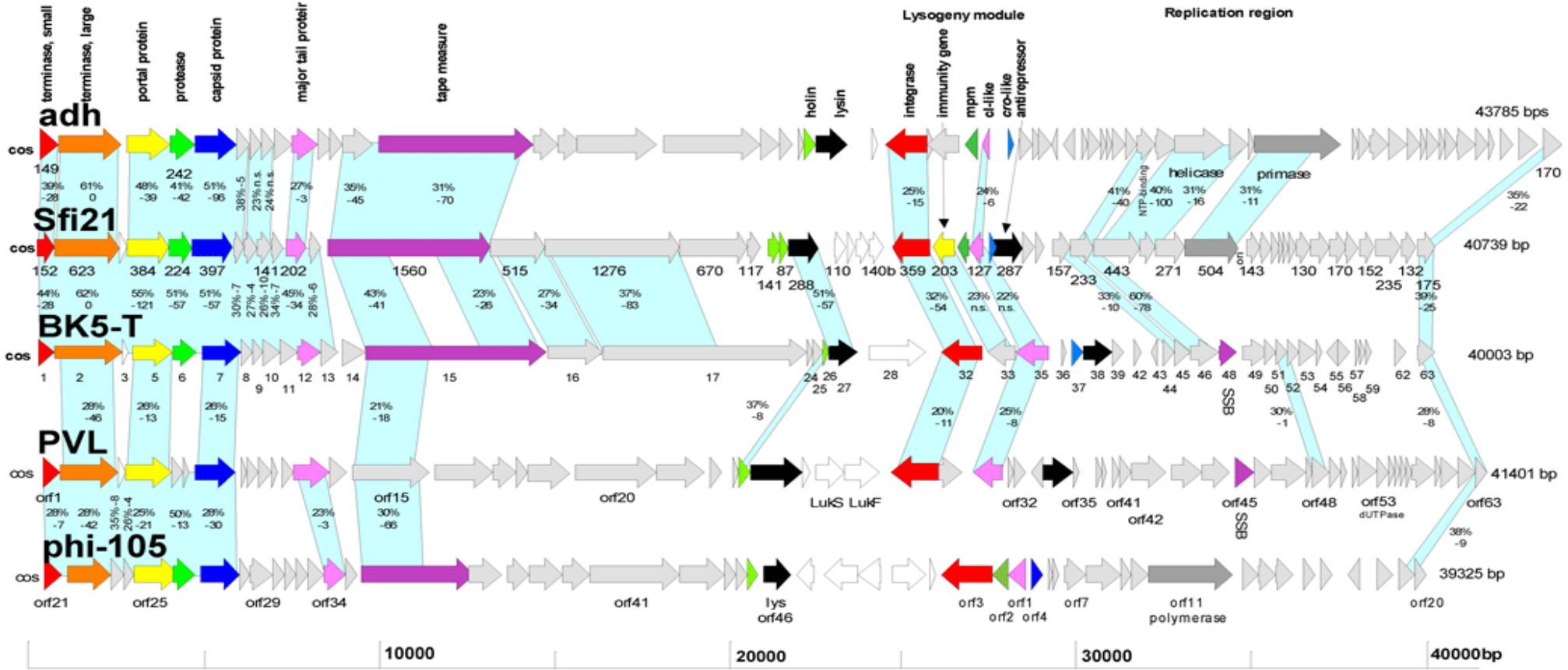
Orthologs: genes found in one organism that are similar to those in another organism but differ because of speciation




position 12
helix H0
sheet
RYDSRTTIFSP..EGRLYQVEYAMEAIGNA.GSAIGILS
RYDSRTTIFSPLREGRLYQVEYAMEAISHA.GTCLGILS
RYDSRTTIFSP..EGRLYQVEYAQEAISNA.GTAIGILS
RYDSRTTIFSP..EGRLYQVEYAMEAISHA.GTCLGILA
RYDSRTTIFSP..EGRLYQVEYAMEAIGHA.GTCLGILA
RYDSRTTIFSP..EGRLYQVEYAMEAIGNA.GSALGVLA
RYDSRTTTFSP..EGRLYQVEYALEAINNA.SITIGLIT
SYDSRTTIFSP..EGRLYQVEYALEAINHA.GVALGIVA


C. mediatlanticus
1 kbp
napD
napL
napF
napB
napH
napG
napA
Sulfurovum sp.
napA
napG
napH
napB
napF
orf1
napL
napD
orf2
Sulfurohydrogenibium sp.
napA
napG
napH
napB
napF
napL
napD
Hydrogenivirga sp.
napA
napG
napH
napB
napF
napL
napD
Desulfovibrio desulfuricans
napC/nirT
napM
napA
napD
napG
napH
T. ammonificans
napC/nirT
napM
napA
napD
napG
napH
Desulfobulbus proprionicus
orf4
napA
napD
napG
napH

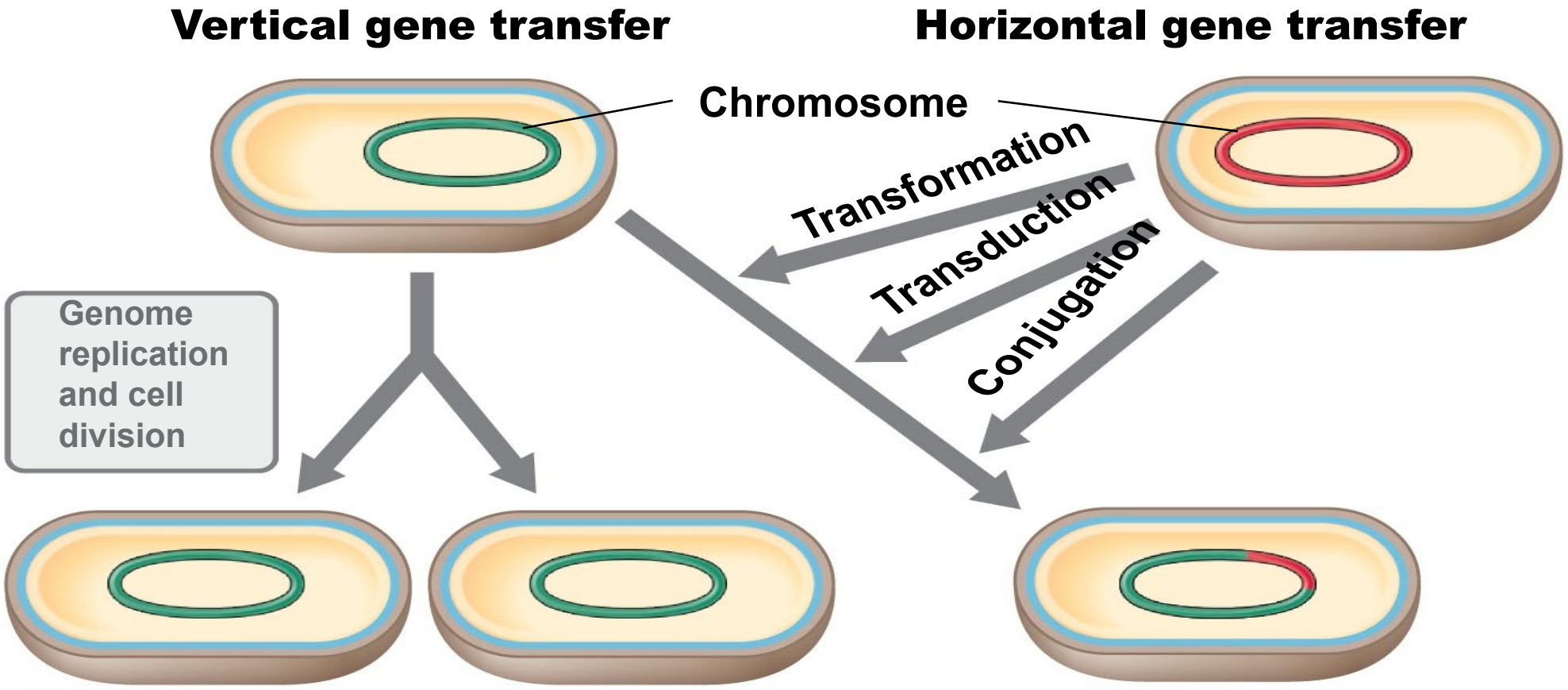
Viral Genomes

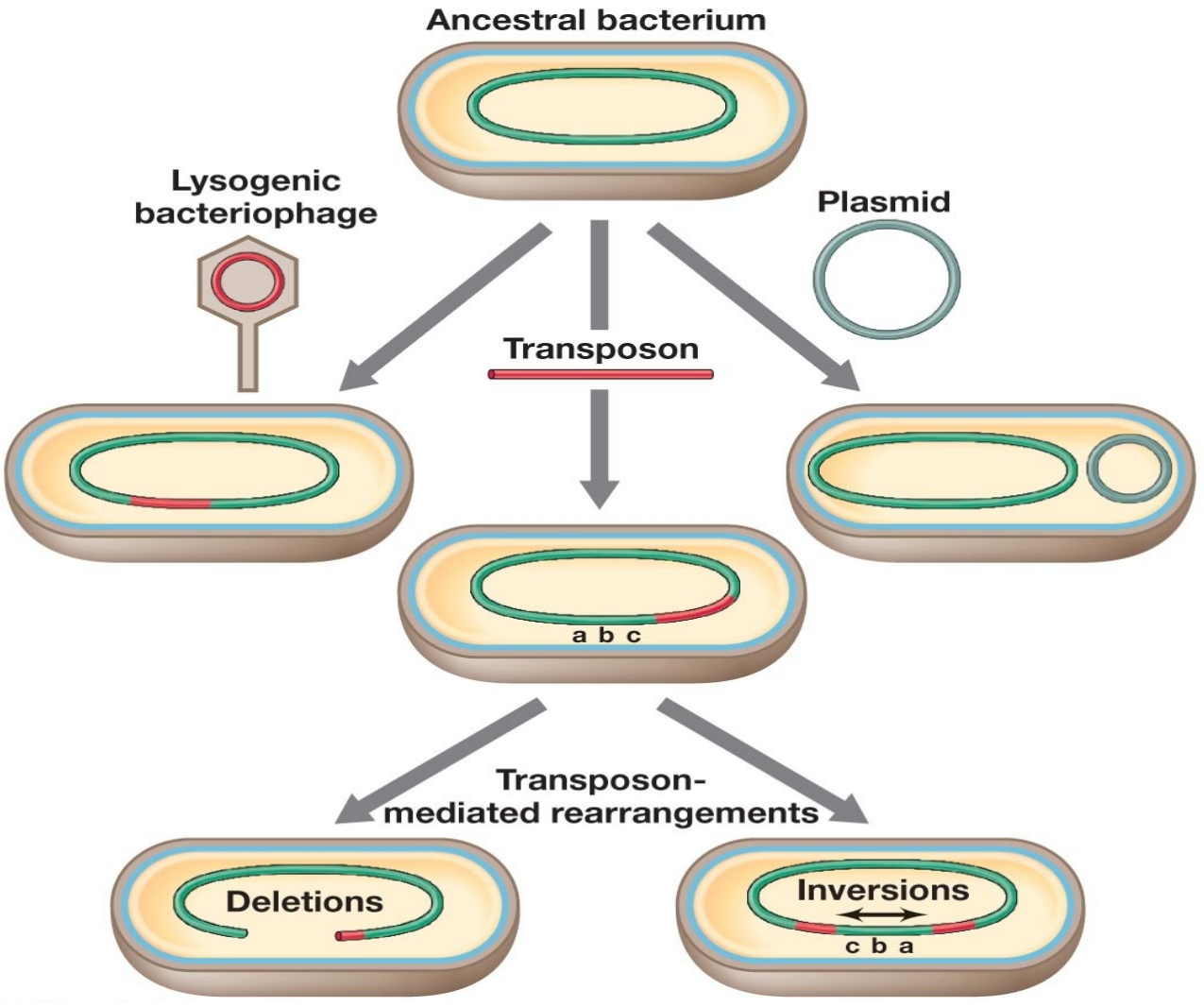


20,000-100,000 bp in size (30-100 genes) on average

Horizontal Gene Transfer

- *Horizontal gene transfer* is the transfer of genetic information between organisms, as opposed to vertical inheritance from parental organism(s)
- It is considered one of the major mode of prokaryotic evolution
- May be extensive in nature
- It is a major confounding mechanism when studying prokaryotic evolution


Vertical gene transfer
Horizontal gene transfer
Chromosome
Transformation
Transduction
Conjugation
Genome replication and cell division


Ancestral bacterium
Lysogenic bacteriophage
Plasmid
Transposon
a b c
Transposon-mediated rearrangements
Deletions
Inversions
c b a

Horizontal Gene Transfer Impact

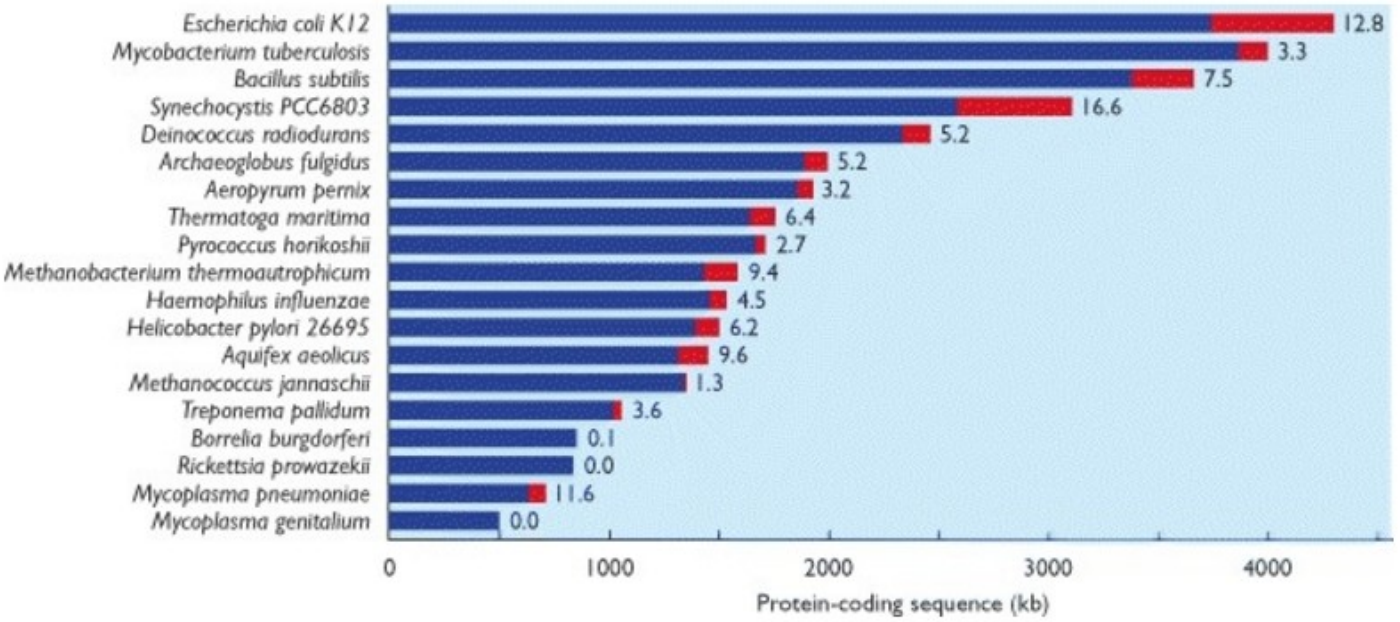
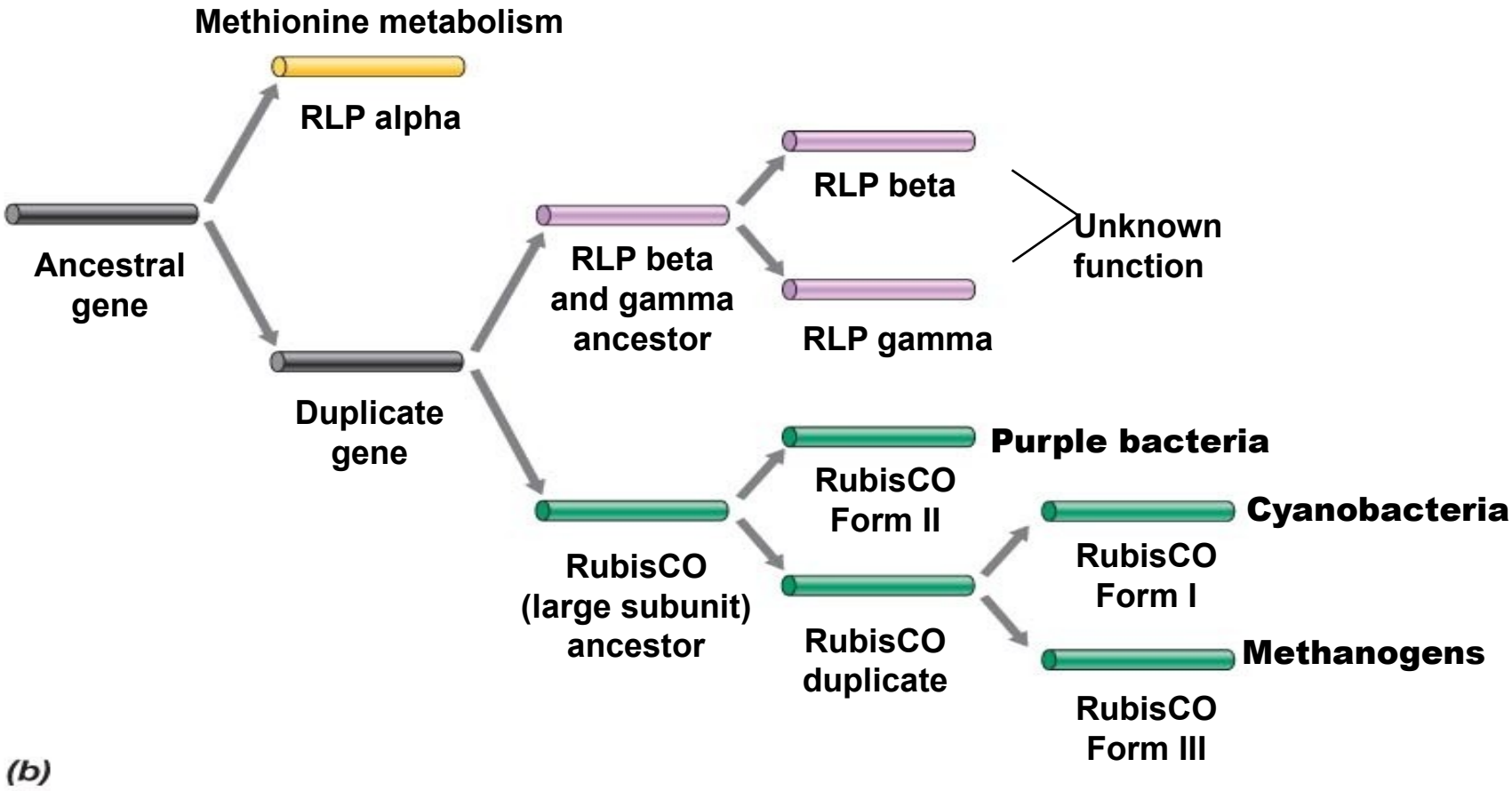
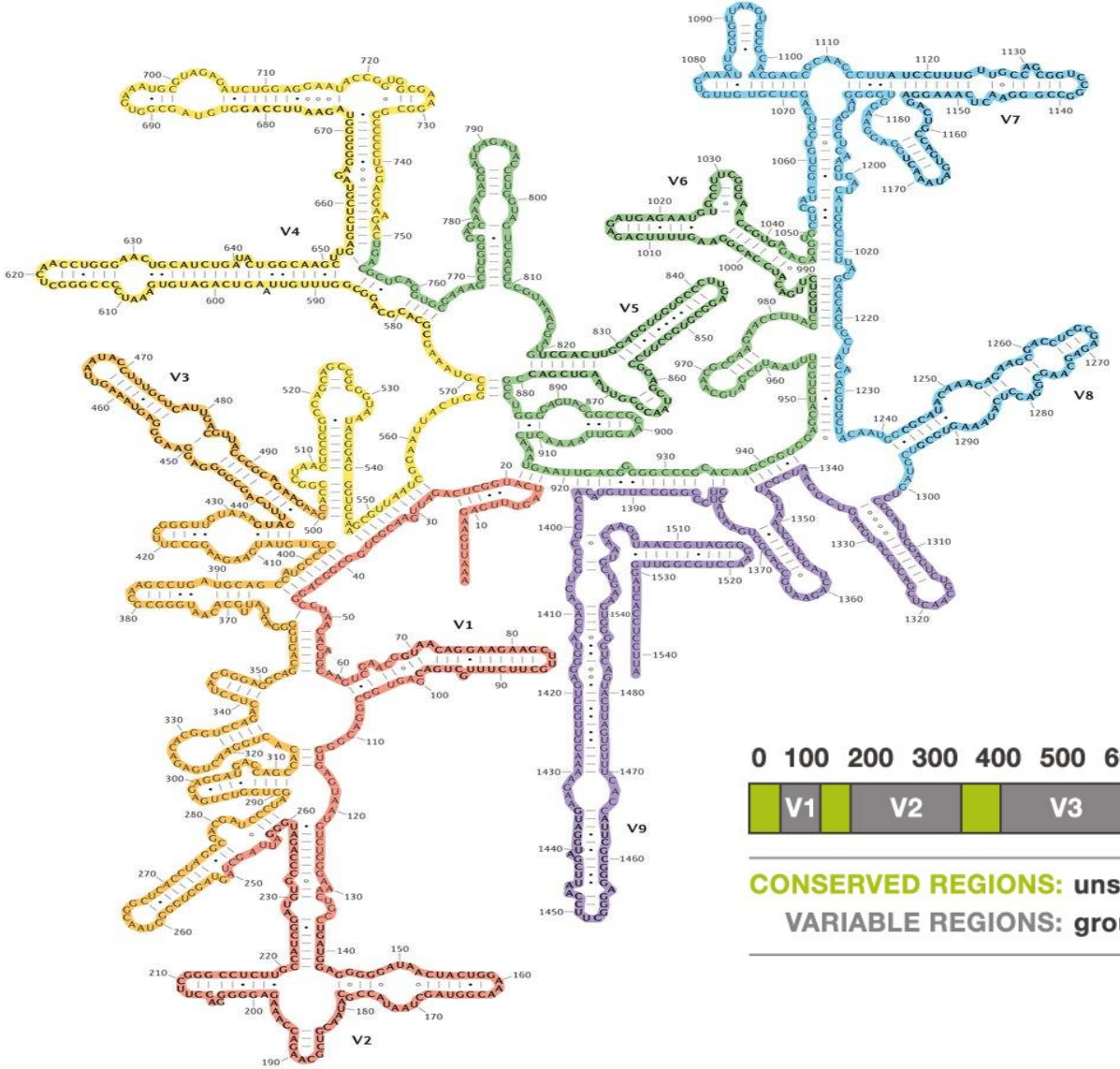


Figure 2.22 The impact of lateral gene transfer on the content of prokaryotic genomes

The chart shows the DNA that is unique to a particular species in blue and the DNA that has been acquired by lateral gene transfer in red. The number at the end of each bar indicates the percentage of the genome that derives from lateral transfer. Note that intergenic regions are omitted from this analysis. Redrawn from Ochman *et al.* (2000).

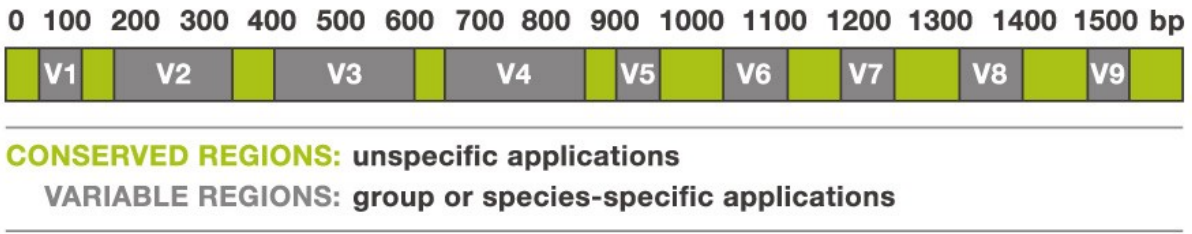

Methionine metabolism
RLP alpha
Ancestral gene
Duplicate gene
RLP beta and gamma ancestor
RLP beta
RLP gamma
Unknown function
RubisCO (large subunit) ancestor
RubisCO Form II
Purple bacteria
RubisCO duplicate
RubisCO Form I
Cyanobacteria
RubisCO Form III
Methanogens
(b)

Marker genes



The **16S rRNA** gene is the most widely used marker gene to investigate prokaryotic diversity in a sample and to compare it across samples

Different regions are used depending on sequencing technology, scientific question, group investigated



Nature Reviews | Microbiology

This week read

Welch, R.A., Burland, V., Plunkett, G.I.I.I., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J. and Stroud, D., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences*, 99(26), pp. 17020-17024

doi: 10.1073/pnas.252529799