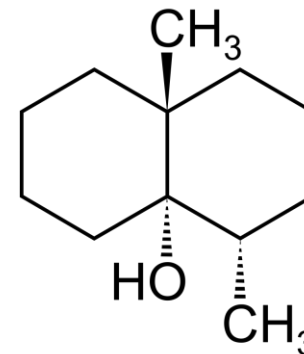
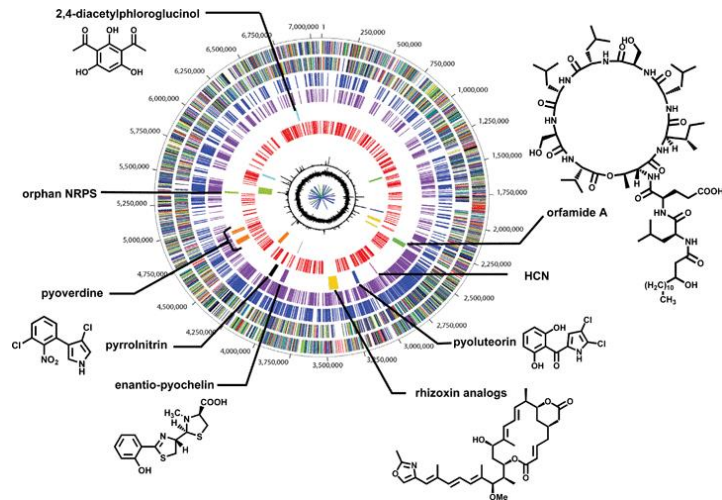


Cyanobacterial genomes

Dr. David P. Fewer

Department of Microbiology, University of Helsinki



Microbial genomic diversity

Bacteria

Uncultured diversity

15% of all bacterial diversity = candidate phyla radiation (CPR)

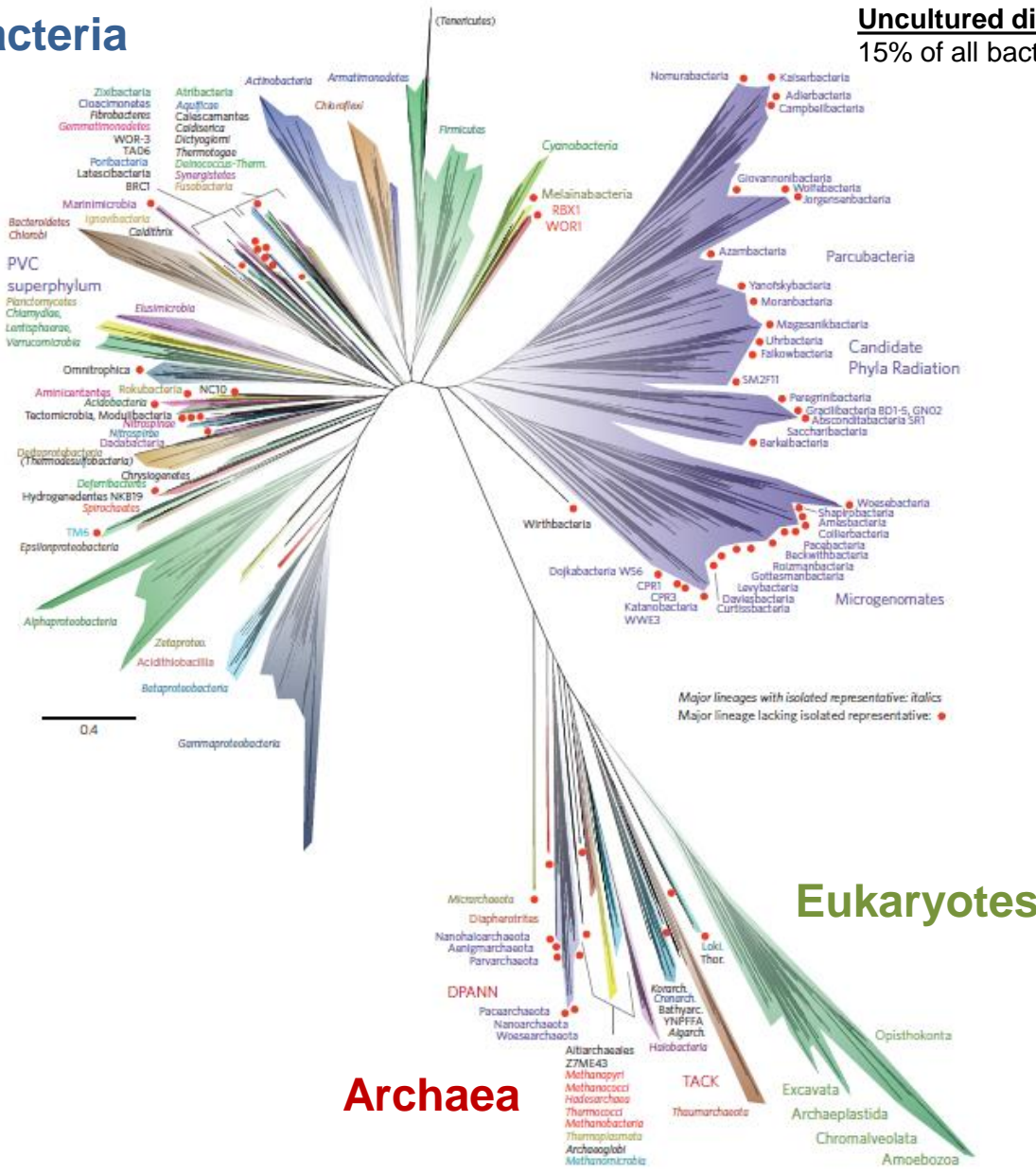
Genome projects

393904 draft/complete genome projects in total
28292 complete genomes

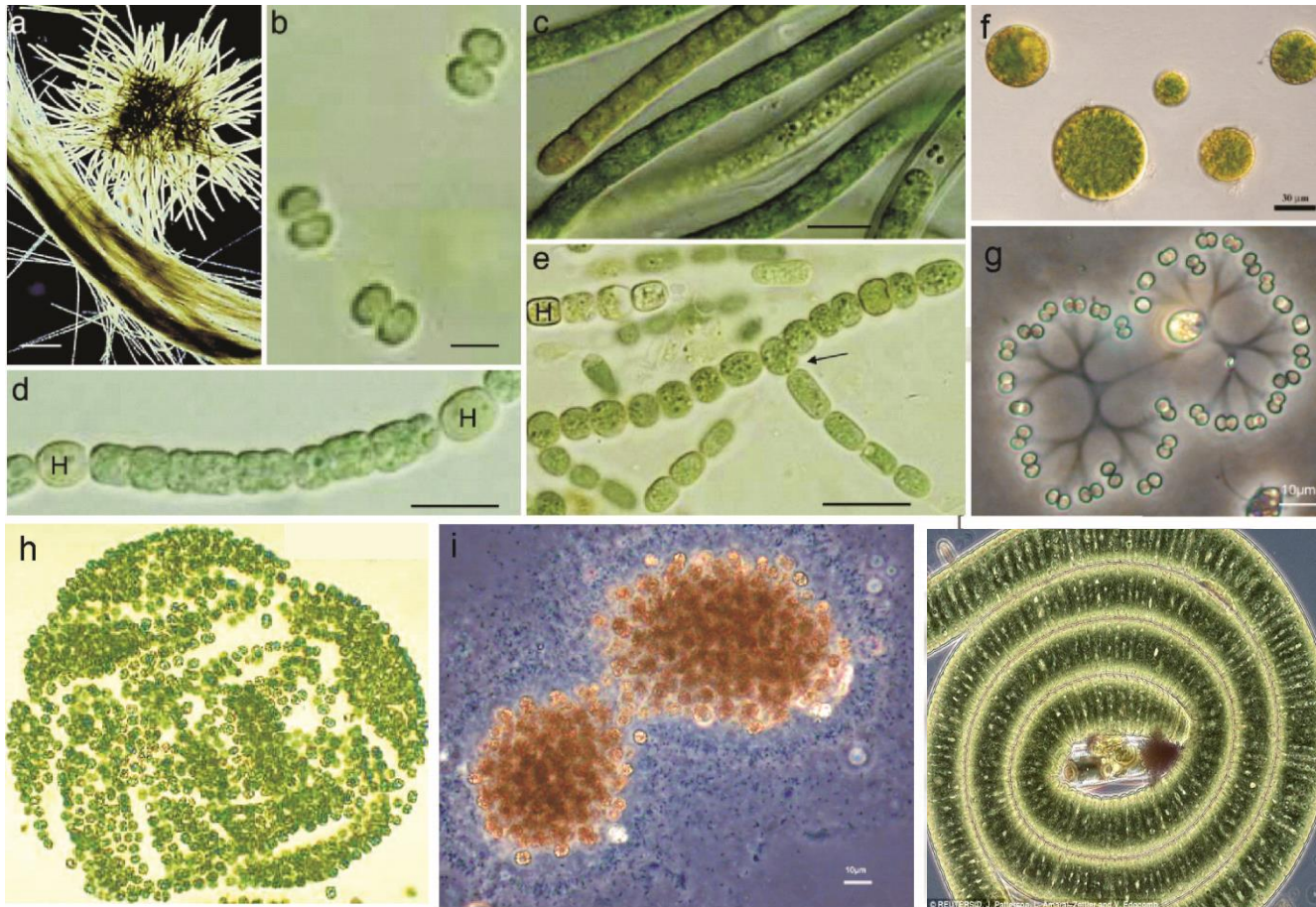
Bacteria	Projects
----------	----------

<i>Escherichia coli</i>	29260
<i>Staphylococcus aureus</i>	14097
<i>Mycobacterium tuberculosis</i>	7068
<i>Neisseria meningitidis</i>	2271
<i>Neisseria gonorrhoeae</i>	934
<i>Yersinia pestis</i>	653
<i>Borrelia burgdorferi</i>	162
<i>Treponema pallidum</i>	90

<i>Prochlorococcus</i> spp.	1097
<i>Oscillatoria</i> spp.	9



Cyanobacteria



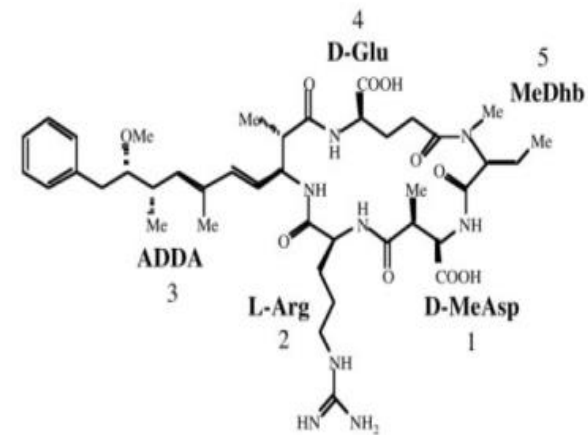
Cyanobacteria are morphological diverse.
Unicellular, filamentous, colonial, branching forms

Cyanobacterial blooms

BALTIC SEA



Nodularia spumigena



First toxic cyanobacterial blooms reported in 1878.

Produce a range of toxins.

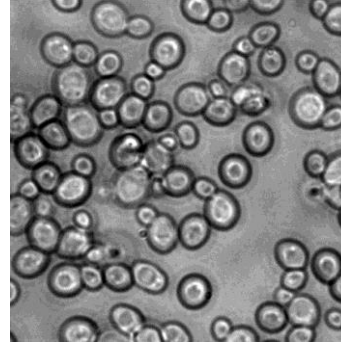
Place	Year	Animals affected	Toxin	Producer
Australia (L. Alexandrina)	1878	Sheep, horses, dogs, pigs	Hepatotoxin	<i>Nodularia</i>
Finland (L. Vesijärvi)	1928	40 cows	Neurotoxin	<i>Anabaena</i>
Baltic Sea	1975 1982-84	Dogs, 16 young cattle	Hepatotoxin	<i>Nodularia</i>
Norway (Roagland)	1978	4 heifers	Hepatotoxins	<i>Microcystis</i>
Finland (L. Sääskj. & L. Säyhteenj.)	1985-86	5 cows	Anatoxin-a	<i>Anabaena</i>
Finland (Åland Island)	1985	Fish, birds, muskrats	Hepatotoxins	<i>Planktothrix</i>
Denmark (L. Knud sø)	1993 1994	Over 20 birds, 1 dog	Anatoxin-a(S)	<i>Anabaena</i>
Australia (Darling river)	1991	2000 cattle and sheep	saxitoxins	<i>Anabaena</i>

Animal poisonings

Genomic size



Phascolarctos cinereus
3420 Mb genome



Saccharomyces cerevisiae
12 Mb Genome



Prochlorococcus marinus
1.4 Mb genome

One of these organisms is the most abundant photosynthetic organism on the planet and responsible for 5 % of global photosynthesis.

Genomic size

Cyanobacteria: 1.44 Mb to

Smallest: Candidatus *Atelocyanobacterium thalassa* isolate ALOHA 1.44 Mb, 1156 proteins)

Largest: *Calothrix* sp. PCC 7716 (12.37 Mb, 10667 proteins)

Bacteria: ~0.11 Mb to 16.0 Mb

Smallest: Candidatus *Hodgkinia cicadicola* (0.11 Mb, 135),

insect endosymbiont

Largest: *Minicystis rosea* DSM 24000 (16 Mb, 14018),

saprophytic soil bacterium

Archaea: ~0.5 Mb to 5.75 Mb

Smallest: *Nanoarchaeum equitans* Kin4-M (0.49 Mb, 536 proteins)

nanoarchaeum from hydrothermal vent

Largest: *Haladaptatus* sp. PSR5 (7.28 Mb, 5032 proteins)

halophilic archaeal species

Eukaryotes: ~7.7 Mb to 40000 Mb

Smallest: *Pneumocystis wakefieldiae* (7.3 Mb, 3182 proteins)

Parasitic fungus isolated from rats

Largest: *Protopterus annectens* (40054 Mb, 37069 proteins)*

West African lungfish

*N.B not based on completely sequenced genomes

Human X chromosome is 155 Mb (5% of genetic material)

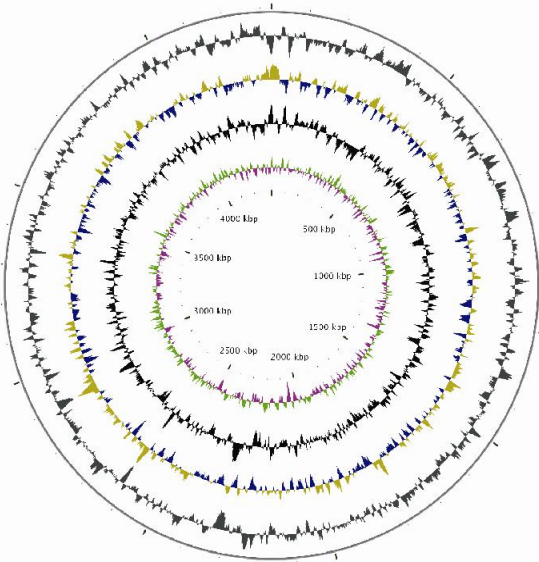
Browse prokaryotic genomes

<https://www.ncbi.nlm.nih.gov/genome/microbes/>

Genomic partitions in prokaryotes

CHROMOSOME A

4,329,685 bp



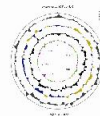
PLASMID C

20,025 bp



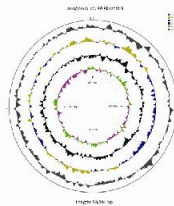
PLASMID B

56,038 bp



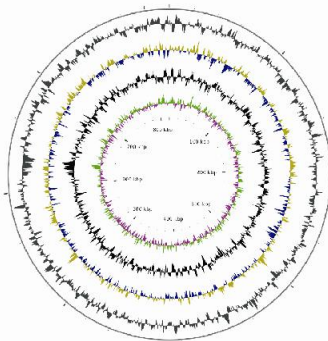
PLASMID A

80,384 bp



CHROMOSOME B

819,965 bp



Anabaena sp. UHCC-0090
Hepatotoxic cyanobacterium
Lake Vesijärvi



- Genome can be comprised of one or more chromosomes and many species harbour plasmids (extrachromosomal genetic elements)
- Plasmids and chromosomes can also be linear (e.g. *Borrelia*, *Streptomyces*, and *Agrobacteria*)

Genomic partitions in prokaryotes

Agrobacterium tumefaciens (5.75 Mb genome)

- One linear chromosome (2.1 Mb)
- One circular chromosome (3.0 Mb)
- Two circular plasmids (0.2 & 0.45 Mb)

***Borrelia burgdorferi* B31** (1.5 Mb genome)

- Linear chromosome (0.91 Mb)
- More than 20 circular & linear plasmids (5-54 kb)

***Minicystis rosea* DSM 24000** (16 Mb genome)

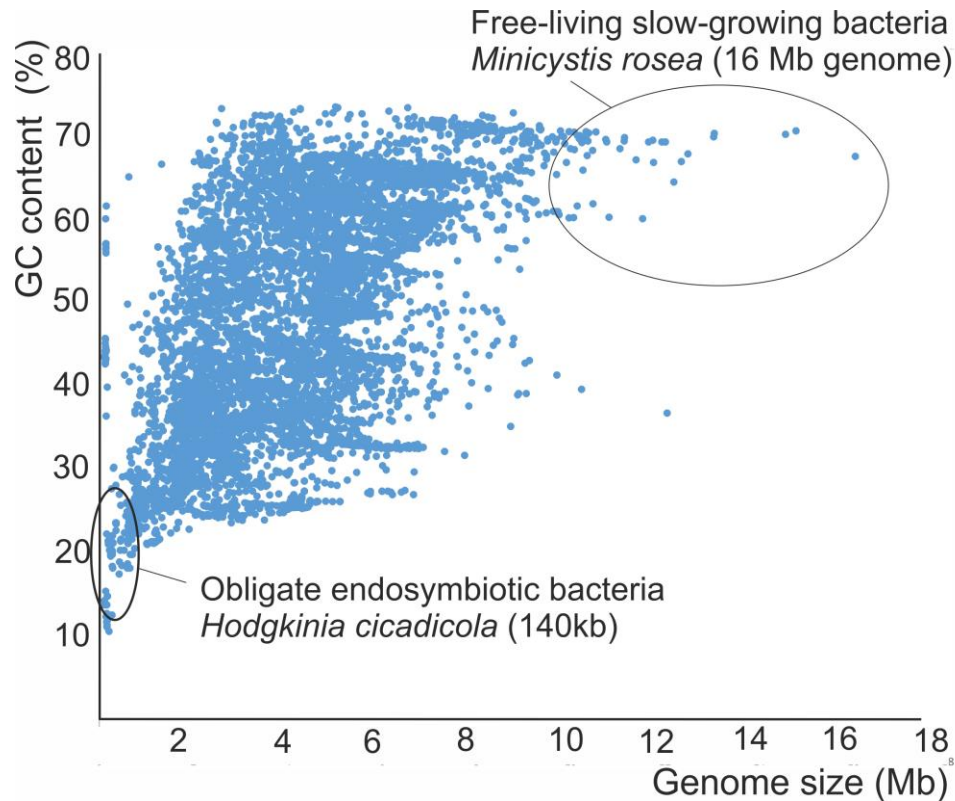
- Single circular chromosome (16 Mb)

Distinction between plasmid, mega-plasmid & chromosome (100 kb cutoff)

Browse prokaryotic genomes

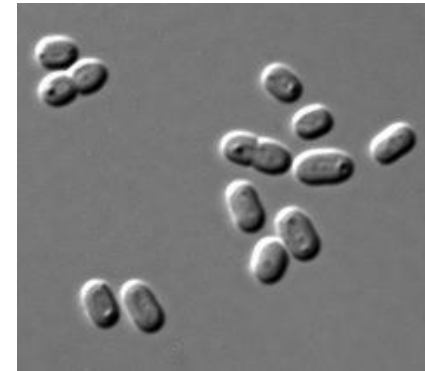
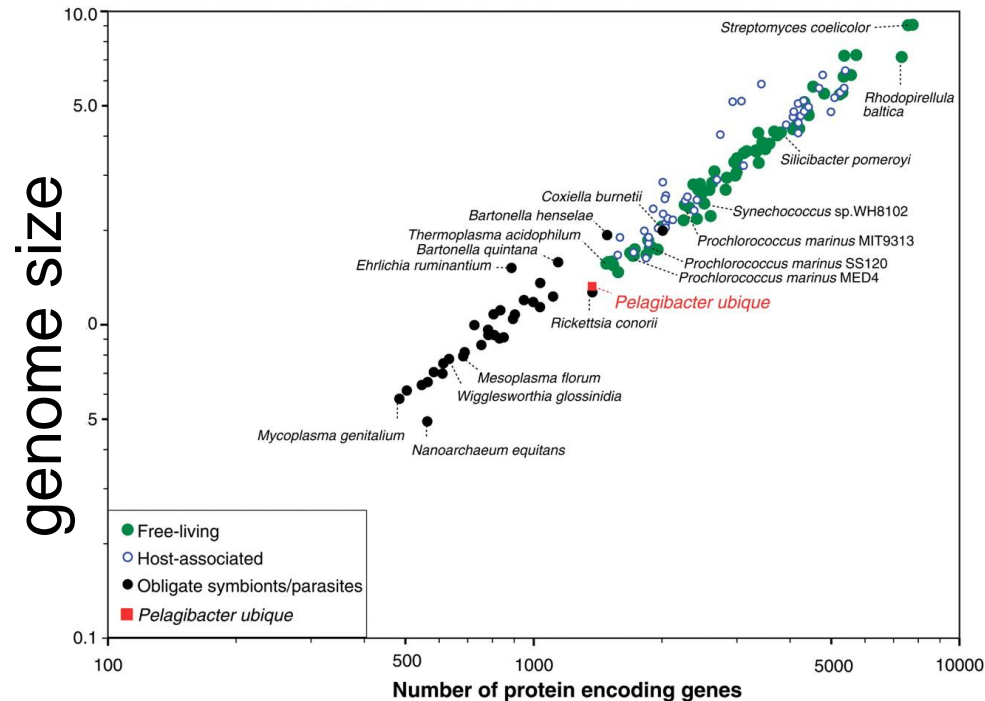
<https://www.ncbi.nlm.nih.gov/genome/microbes/>

Genome size and nucleotide content



- There is correlation between genome size and GC content of prokaryotic genomes (12563 complete genomes; $R^2=0.3697$)
- Bacterial species with the lowest effective population sizes, such as endosymbiotic bacteria, typically have the smallest and most compact genomes, whereas those with the largest populations exhibit more expansive genomes

Gene content and genome streamlining

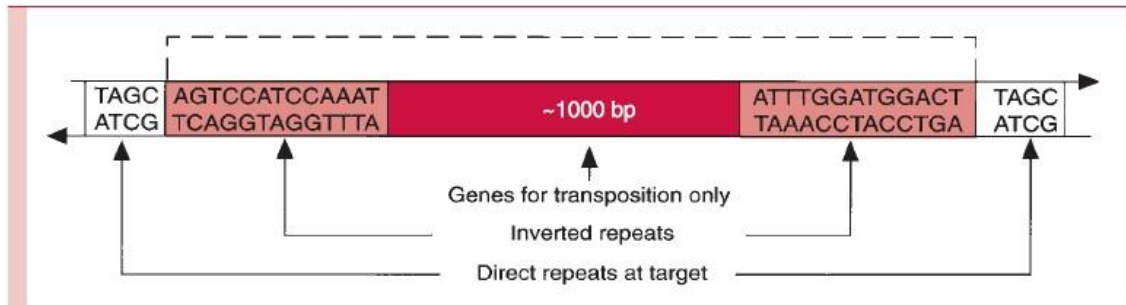


Pelagibacter ubique

Length 0.37–0.89 μm , diameter 0.12–0.20 μm

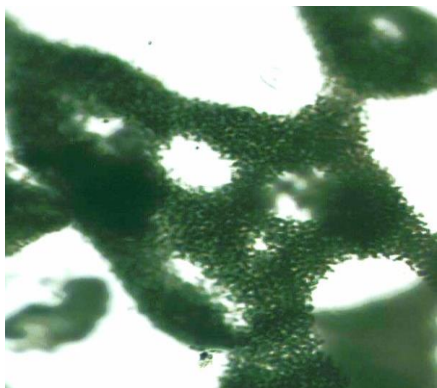
- Bacterial genomes consisted mostly of functional protein-coding regions and little non-coding or intervening sequences
- As a rule of thumb 1 kb per gene in bacteria (not true for eukaryotes)
- Prokaryotic genomes encode 170–14,000 protein coding genes
- Linear relationship between genome size and number of genes
- Prokaryotic genes do not encode introns
- Coding density is 87 % with a typical range of 85–90 % (versus 1–3% for humans)
- *Pelagibacter ubique* 1.3 Mb genome encodes 1354 genes, one of the smallest for any free-living organism (Giovannoni 2005 *Science* 309:1242)

Repetitive DNA and transposable elements



- An **insertion sequence** (IS) is a common type of bacterial transposable element and encodes only the gene required for its own transposition, flanked by insertion sequence repeats (9-40 bp)
 - **Miniature Inverted-repeat Transposable Elements** (MITEs) are non-autonomous transposable elements
 - MITEs have 4-30 bp inverted repeats, no transposase and range from 100-400 bp
 - Some prokaryotic genomes can encode hundreds of such IS and MITE transposable elements
- > promote genome rearrangement, gene deletions & regulate gene expression
-> promote genome streamlining

Repetitive DNA and transposable elements



Microcystis aeruginosa

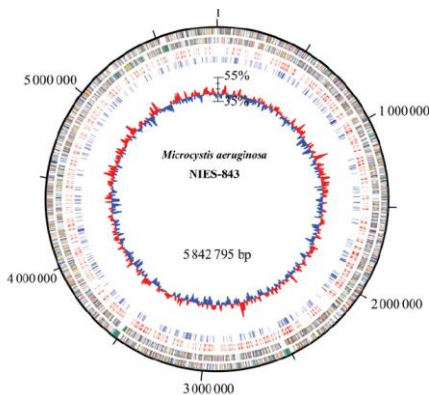


Table 1. Summary of ISs and MITEs in the *M. aeruginosa* genome

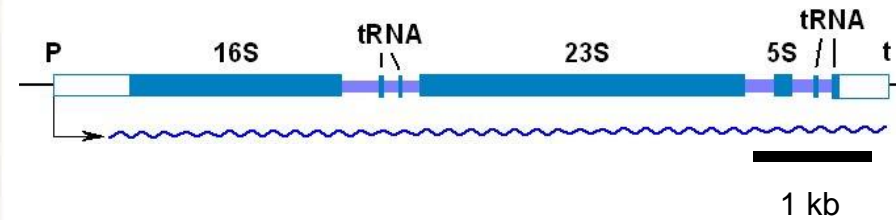
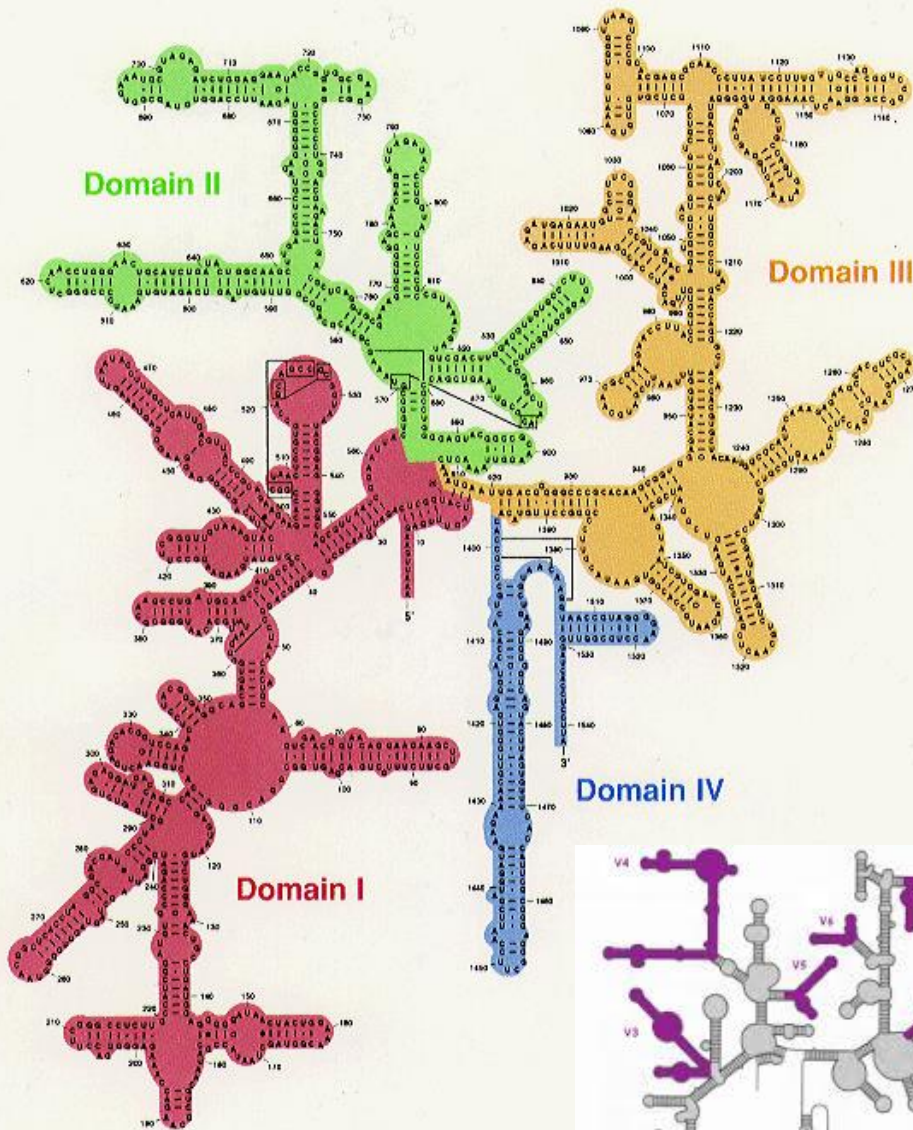
Name	Length (bp)	Family	Copy number
ISMae1	955	IS1	10
ISMae4	1223	IS5	33
ISMae5	1694	IS4	19
ISMae6	1686	IS5	8
ISMae7	519	IS5	3
ISMae8	1093	IS30	8
ISMae9	1370	IS110	7
ISMae10	1330	IS605	49
ISMae11	1473	IS605	9
ISMae12	1835	IS605	6
ISMae13	1295	IS605	5
ISMae14	1366	IS605	4
ISMae15	1431	IS605	3
ISMae16	1379	IS605	2
ISMae17	1510	IS607	9
ISMae18	1240	IS607	2
ISMae19	1410	IS607	2
ISMae20	1897	IS607	4
ISMae21	956	IS630	22
ISMae22	1144	IS630	21
ISMae23	1237	IS630	15
ISMae24	1089	IS630	14
ISMae25	1262	IS630	13
ISMae26	1041	IS630	12
ISMae27	1148	IS630	9
ISMae28	1599	IS630	8
ISMae29	1224	IS630	5
ISMae30	1091	IS1200	16

- *Microcystis aeruginosa* NIES-843 has 704 IS/MITE elements belonging to 18 families
- 11.8% of genome is IS and MITEs

Discuss with partner:
What implications do repetitive elements have for genome sequencing?

288 draft genomes, 12 complete

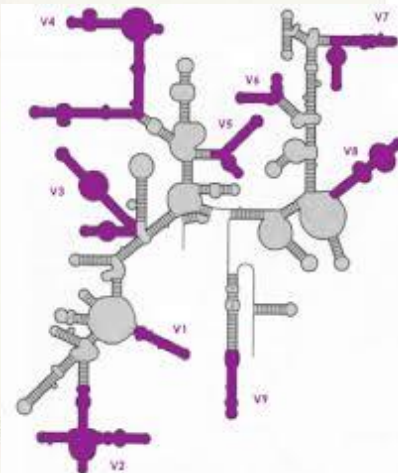
Repetitive DNA and ribosomal operons



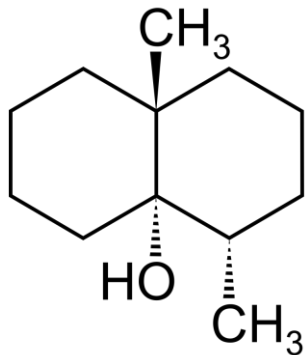
- The ribosomal operon gene has a conserved secondary structure (4-7 kb)
- Primary sequence of 16S rRNA gene is highly conserved and widely used in phylogenetics
- The ribosomal operon can be present in multiple copies in bacteria
- In cyanobacteria 1-4 copies is normal.

Question

What implications does this have for genome sequencing?



Geosmin, the aroma of rain



Geosmin

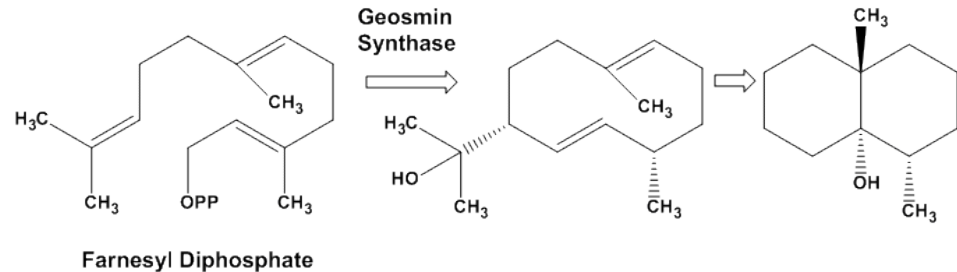
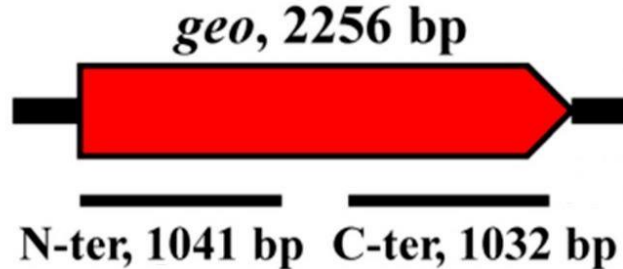
Monoisotopic mass: 182.167

Elemental composition: C₁₂H₂₂O



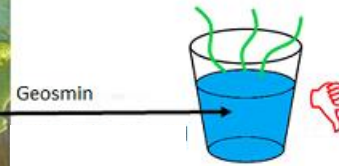
- Geosmin is a volatile organic compound
- Causes petrichor (earthy, muddy smell after rain)
- It gets its name from the Greek for "earth odour"
- Geosmin was first identified in 1891, the structure of geosmin determined in 1965, and total synthesis in 1968
- Human nose can detect less than 10-100 parts per trillion
- Produced by cyanobacteria and actinobacteria
- Camels and mosquitoes

Geosmin, the aroma of rain



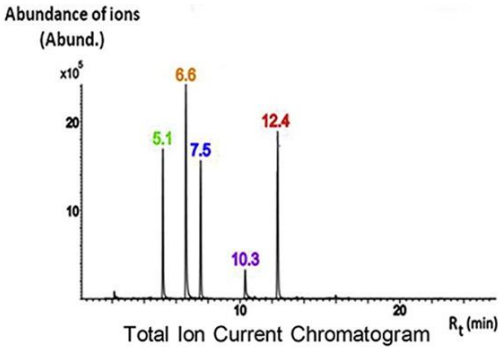
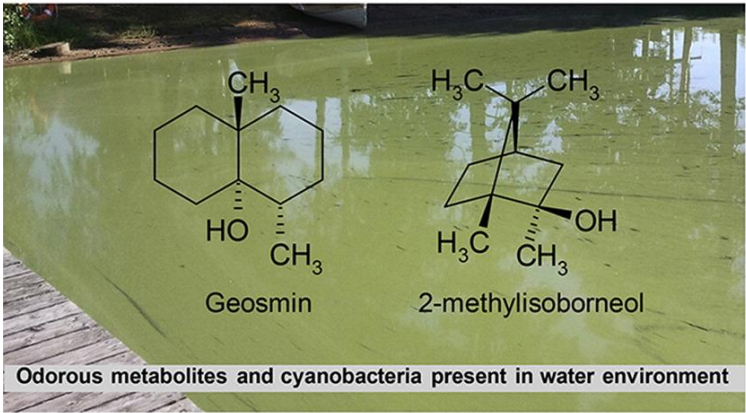
- The conversion of FPP to geosmin was previously thought to involve multiple enzymes in a biosynthetic pathway
- Biosynthesis reported from *Streptomyces coelicor* in 2003-2006 (e.g. *JACs*, 2006 28: 8128-8129)
- Geosmin synthase (GeoA, 726 aa) is a bifunctional enzyme that transforms farnesyl diphosphate into geosmin
- The N-terminal half of the protein catalyzes the conversion of farnesyl diphosphate to germacradienol and germacrene D, the C-terminal-mediated conversion of germacradienol to geosmin
- Geosmin synthases, found in cyanobacteria, actinobacteria, myxobacteria
- 45-78% identity, hypothesized that all geosmin synthases function in the same manner

Geosmin, the aroma of rain



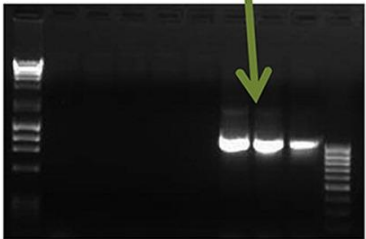
- Cyanobacterial taste and odor compounds, such as geosmin, adversely impact water quality.
- Human nose can detect less than 10-100 parts per trillion
- Accumulates in drinking water and fish
- Earthy and musty off-odors are a common cause of consumer complaints
- Problem in production of drinking water and aquaculture
- Causes economic losses

Geosmin, the aroma of rain



Chemical identification of compounds

Producers



Molecular identification of the producers with PCR targeting biosynthetic genes

Suurnäkki et al. 2015 *Water Research* 1: 68: 56-66



Sequence complete genome of three geosmin producers

- *Oscillatoria* sp. 193
- *Planktothrix* sp. 328
- *Oscillatoria* sp. 327/2

Use a combination of Illumina and Nanopore data

