

The Role of Gene Transfer in Innovation and Speciation

J. Peter Gogarten

University of Connecticut
Dept. of Molecular and Cell Biol.



Edvard Munch, *The Dance of Life* (1900)

Collaborators:

Greg Fournier (UConn)

Olga Zhaxybayeva (Dalhousie)

Kristen Swithers (UConn)

Thane Papke (UConn)

Jinling Huang (ECU)

Funded through the NASA Exobiology and AISR Programs, and NSF Microbial Genetics

Natural Genetic Engineering and Natural Genome Editing, Salzburg 2008



Outline:

Introduction

Gene transfer and the Tree/Coral of Life

Useful aspects of horizontal gene Transfer (HGT)

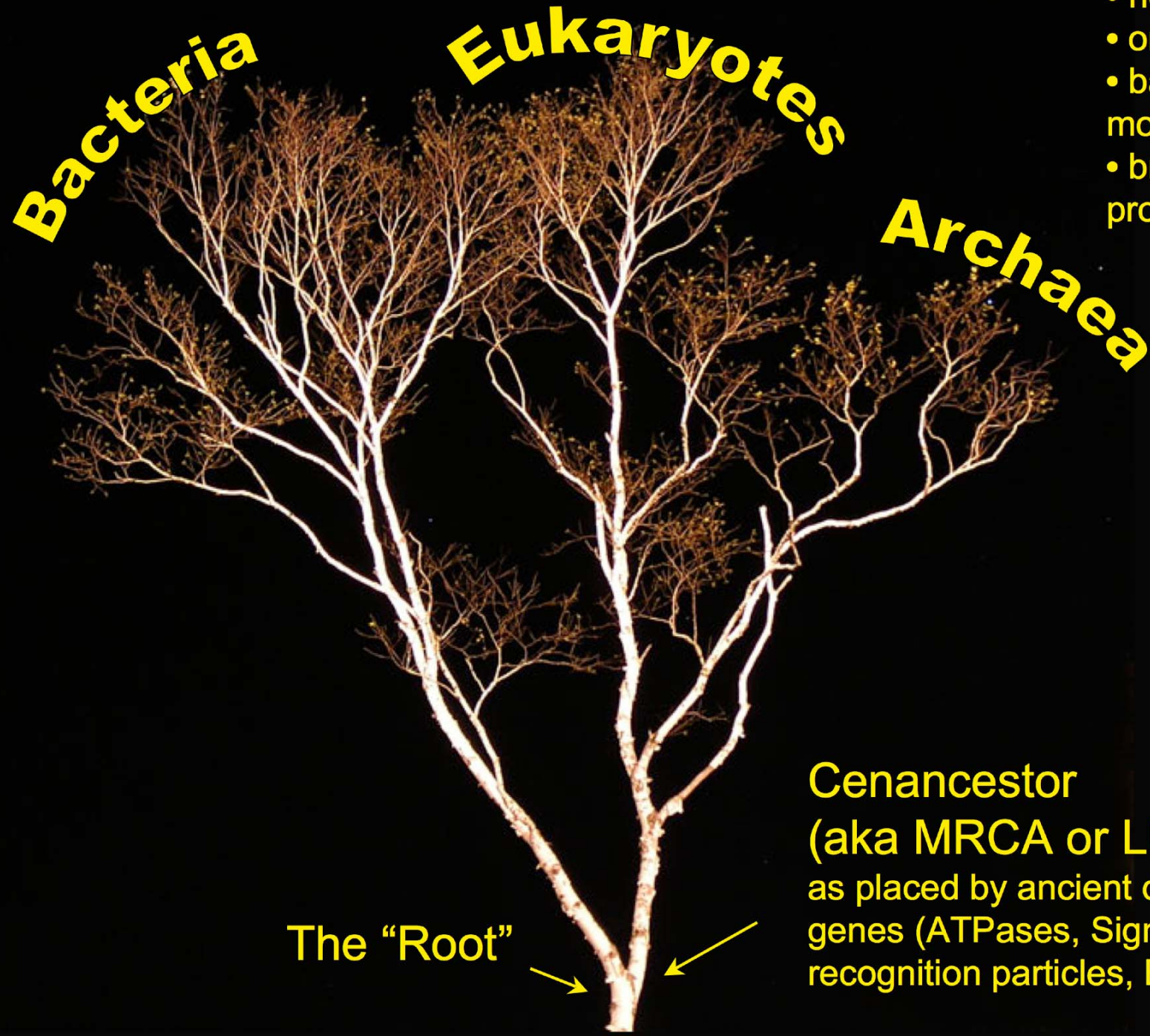
Transferred genes as taxonomic characters

Assembly of new metabolic pathways

Bacterial species, phylo- and ecotypes

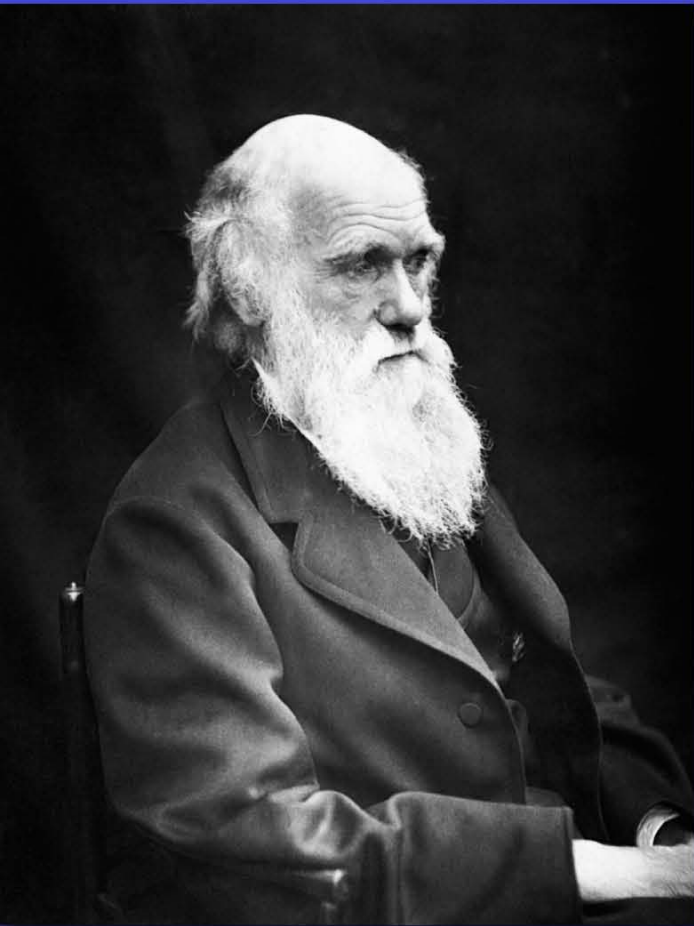
Proposal on the role of bacteriophages in speciation

The Tree of Life according to SSU ribosomal RNA (+)



- strictly bifurcating
- no reticulation
- only extant lineages
- based on a single molecular phylogeny
- branch length is not proportional to time

Tree, Web, or Coral of Life?

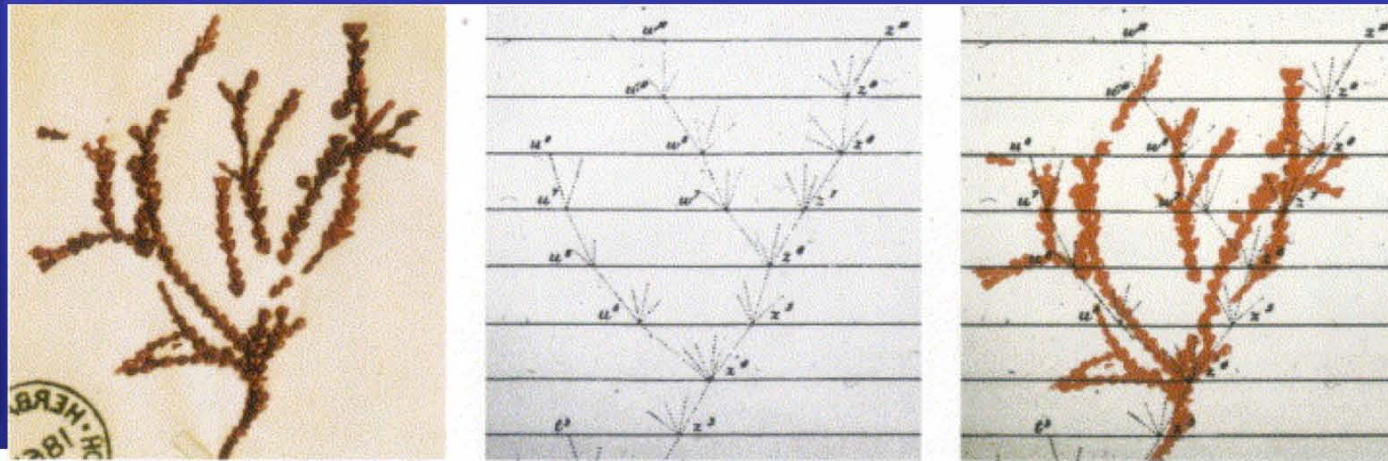


Charles Darwin

Photo by J. Cameron, 1869

“The tree of life should perhaps be called the coral of life, base of branches dead”

Page B26 from Charles Darwin's (1809-1882) notebook (1837)



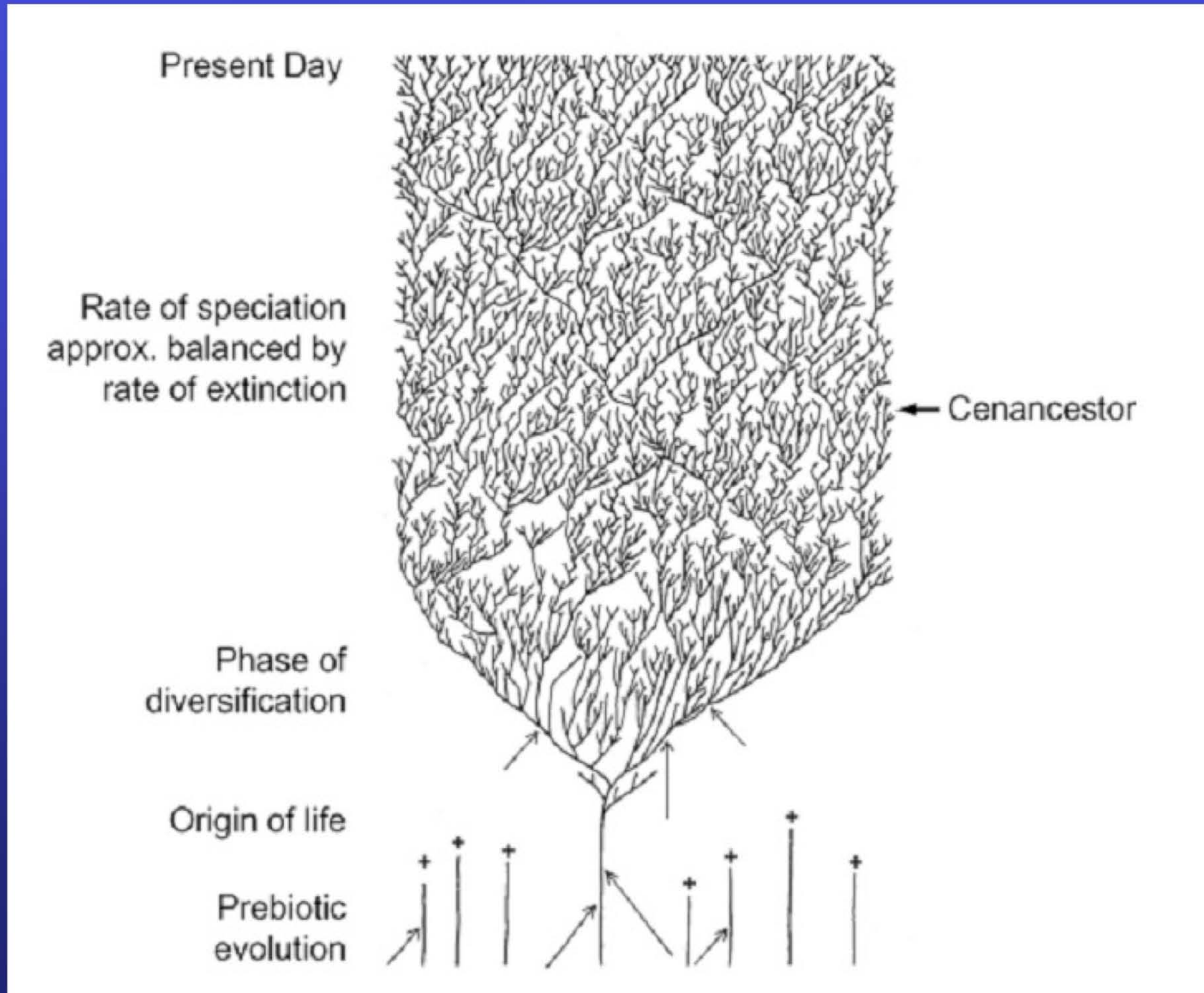
The captivating coral. According to the ideas of Horst Bredekamp, parts of the diagram in Darwin's origin of species (centre) more or less directly reflect the branching properties of a specimen Darwin collected himself.

From Florian Maderspacher:

“The captivating coral--the origins of early evolutionary imagery.”

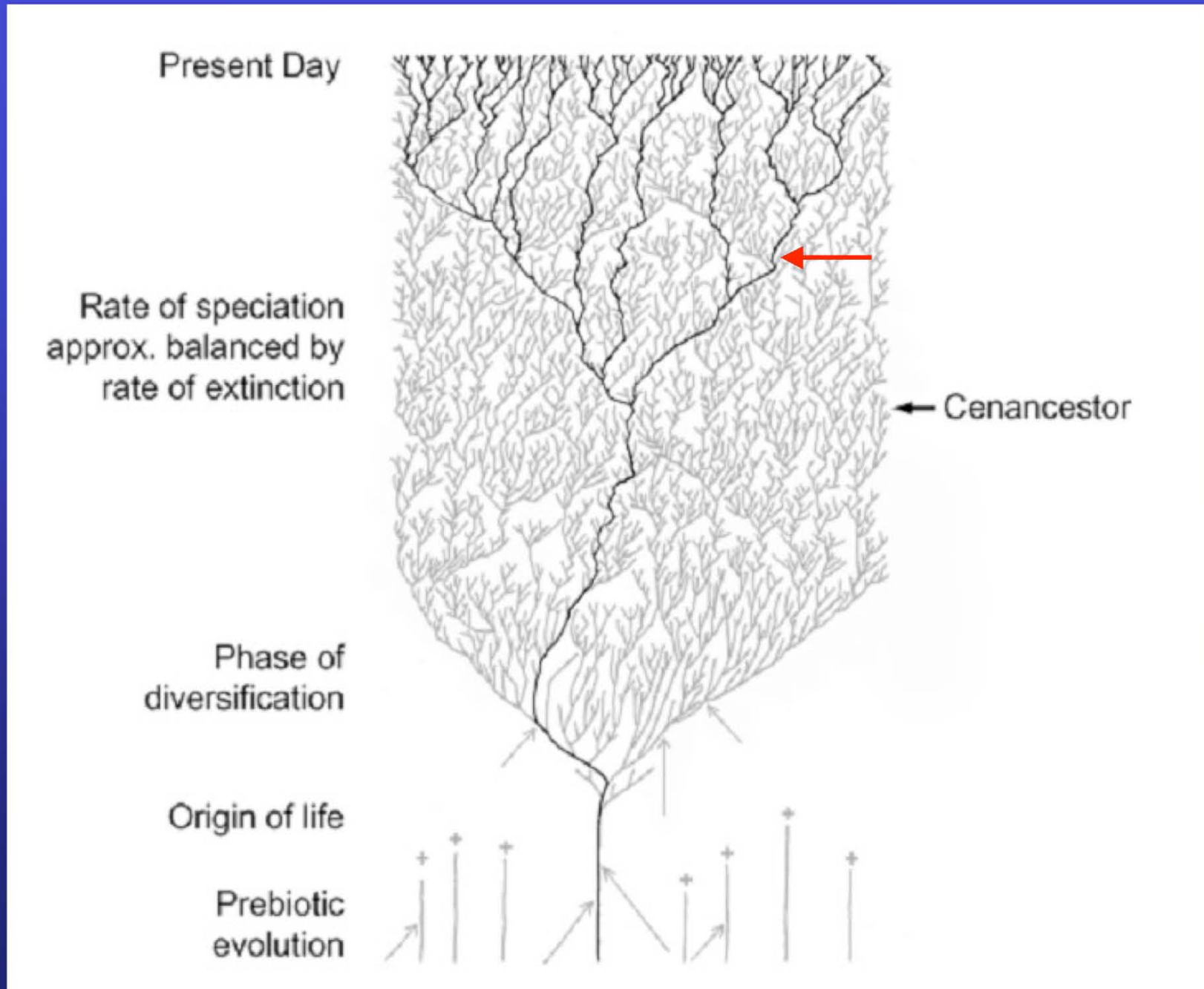
Curr Biol 16: R476-8 2006

The Coral of Life (Darwin)



ZHAXYBAYEVA and GOGARTEN (2004):
Cladogenesis, Coalescence and the Evolution of the Three Domains of Life.
Trends in Genetics 20 (4): 182- 187

The Coral of Life (Darwin)



ZHAXYBAYEVA and GOGARTEN (2004):
Cladogenesis, Coalescence and the Evolution of the Three Domains of Life.
Trends in Genetics 20 (4): 182- 187

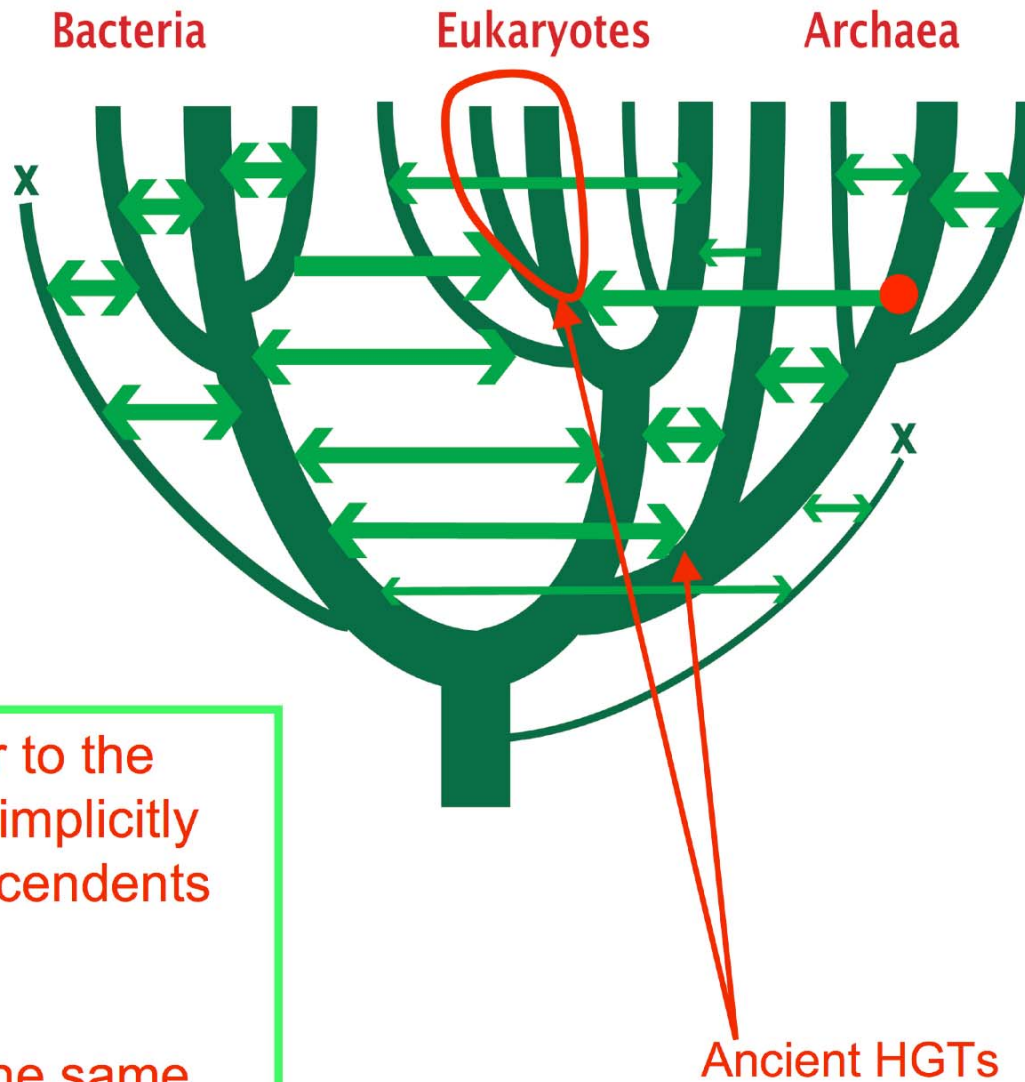
Gene Transfer and Phylogenetic Reconstruction: Friends or Foes?

Popular view

- ⊕ Gene transfer is a disruptive force in phylogenetic reconstruction.

New view

- ⊕ Events of ancient gene transfer are valuable tools for reconstructing organismal phylogeny.



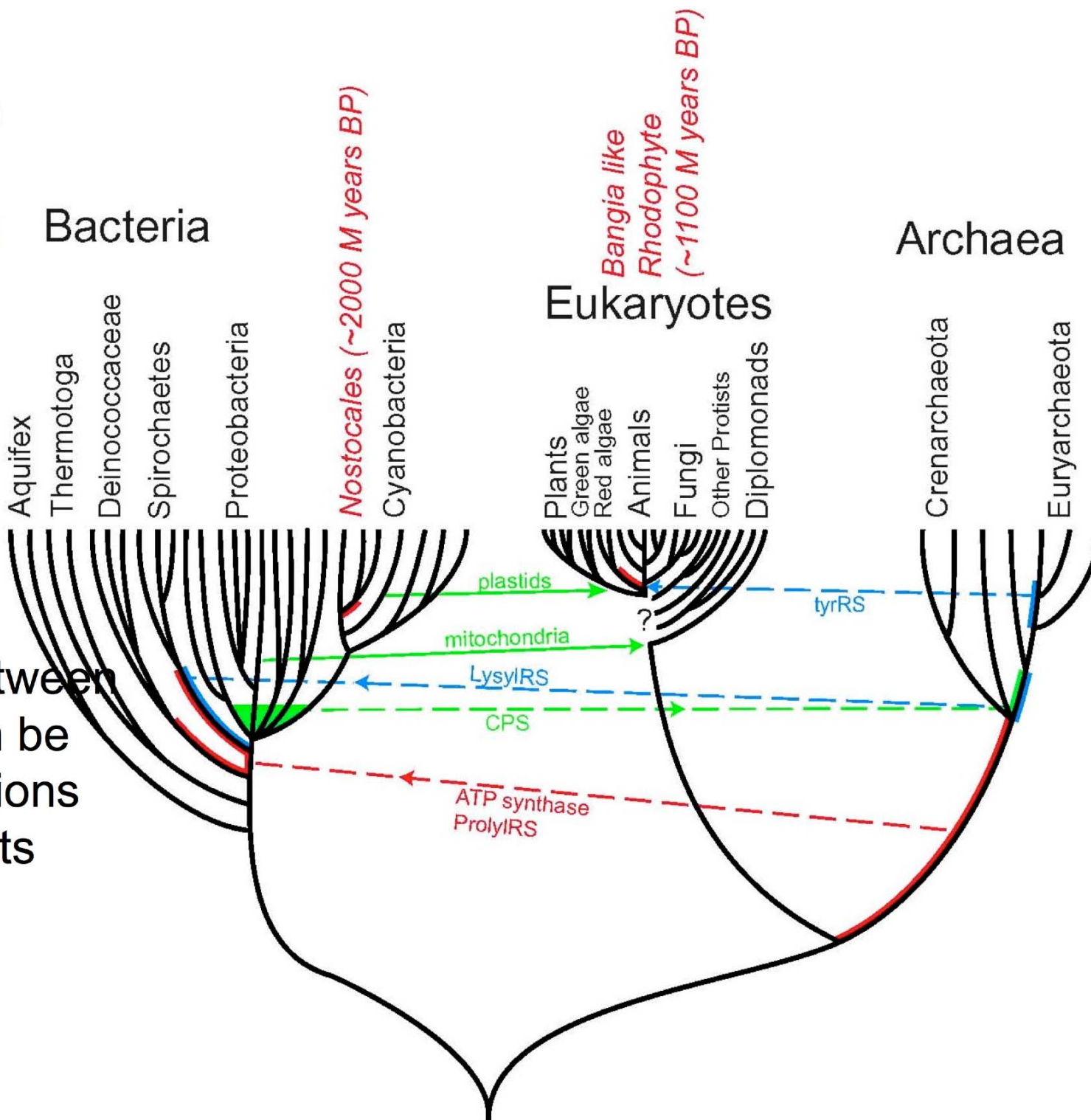
1. Any ancient gene transfer to the ancestor of a major lineage implicitly marks the recipient and descendants as a natural group.

2. The donor must exist at the same time or earlier than the recipient.

Ancient HGTs

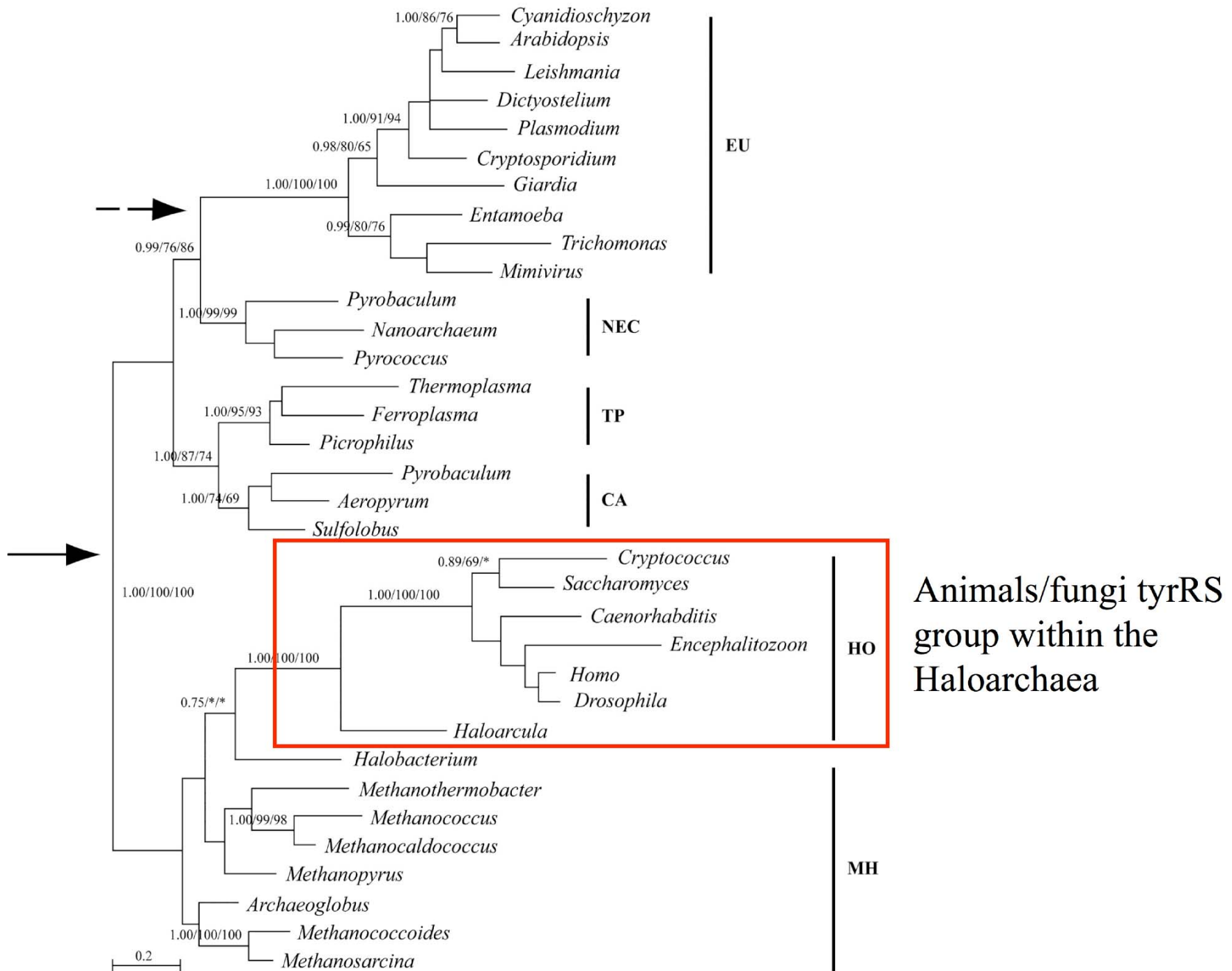
Presence of a transferred gene is a shared derived character that can be useful in systematics.

Gene “ping-pong” between different lineages can be used to build correlations between different parts of the tree/net of life.



Phylogenetic analyses of tyrRS protein sequences

Place where bacterial homologs
join the tree in our analyses



Multiple protein sequence alignment for tyrRS.

Arabidopsis	KYWPLVMDIARKNKLPRILRCVQIMGRSETDE-LSAAQILYPCMQCADIFFLE---ADICQLGMDQRKVNVLAREYCD	234
Cyanidioschyzon	EYWTLVMDIARRNSLARIIRCCQIMGRNESDE-LSAAQILYPCMQCADIFFLR---ADICQLGMDQRKVNVLAREYCD	269
Cryptosporidium	EYWLRFVDISRKFNITRIKRCQIMGRQENDE-QPCASVFYPCMQCADIFQLK---ADICQLGMDQRKVNMLAREYCD	225
Dictyostelium	DYWLRVMDIARKFNVPRIKKCSQIMGRNEEDD-LMCAQLLYPCMQCADIFYLK---ADICQLGMDQRKVNMLAREYCD	234
Leishmania	TYWRMVLDIGRQNTIARIKKCTIMGKTEGT--LTAAQVLYPLMQCCDIFFLK---ADICQLGLDQRKVNMLAREYCD	197
Giardia	SYWRMVNIARSFTVPRMQRCSQIMGRTEGDD-QPISQILYPSMQCADIFELK---LDLVEMGLDQRKVNALARDYCD	221
Mimivirus	SYIERMLDIAEFSTISRVRKCCQIMGRNESDC-LKASQIFYPMQAADVFEVPEGIDICQLGIDQRKVNMLAIEYAN	206
Entamoeba	TYWPLVLDIATKNTVNRIKCSQIMGRDEKDA-LSTSOLFYPMQCADIFELG---ADVCOLGLDQRKVNMLAREYAP	208
Trichomonas	EYFELVHDISSFATLSRIKRIVTAMGRTEGDK-LSLSQLLYPCMQATDVFMLD---VDICQLGVDQRKVNMLAIEYAN	201
Sulfolobus	DYWALVIKVAKNASLARIKRALTIMGRRAEEAEIDASKLIYPAMQVSDIFYLD---LDIALGGTDQRKAHMLARDVAE	201
Pyrobaculum	KYWELVVKIAKETSRLARVKRATPVMGRRAEEVELDFSKLMPYPLMQVADIFYLG---VDVAVGGMDQRRAHMLARDVAE	180
Ferroplasma	DYWELLQATAKNSSLLRIIRSLPIMGRTEEDADKDFSKYIYPLMQVTDVIFYLN---SDIAIGGTDQRHAYMLARDIAE	190
Thermoplasma	EYWRMFIDTAKRSTLKRVIKRSIPIMGRNETDAEKDFSMYLYPIMQVTDIFYLD---VDMAFGGMDQRHAHMLARDIAE	190
Pyrococcus	DYMQTVIDISKNVTLSRVMRSITIMGRQMGES-IDFAKLIYPMQVADIFYQG---VTIAHAGMDQRKAHVIAIEVAQ	202
Nanoarchaeum	EYLENIKISMNTTLNRIKKGITIMGRKQGES-ISFAQLLYVPMQVADIYSLN---VNLAHGGIDQRKAHVIAIEVSD	204
Methanosarcina	EYMLNVLKLSRSVTLNRARRSMDEVGRAMDDP--TVSQMVYPLMQAIDIAMLG---VDIAVGGIDQRKIHMLARENLP	189
Methanococcoides	EYMLNVLKLTQATSLNRAKRSMDEVGRKMEDP--KVSQMVYPLMQAVDIALLG---VDVAVGGIDQRKIHMLAREGLP	189
Archaeoglobus	DYVLDVLKMARITTLNRARRSMDEVSRKEDP--MVSQMIYPLMQALDIAHLG---VDLAVGGIDQRKIHMLARENLP	192
Methanopyrus	DYALDVYRMARHTMRRARRSMDMIARSEENP--PVSQVVYPLMQALDIVHLN---VDLAVGGLEQRKIHMLARDVLP	192
Methanococcus	EYTLIDVYKLALSTTLNRARRSMEVIAREDENP--KVASVVYPLMQVNDIKHLN---ADVAVGGMEQRKIHMLSREILP	193
Methanothermobacter	EYIDLVYRLALITLLRAKRSMQAQITRESRDH--KVAEVIYPLMQVIDMVYLG---VDVALGGMEQRKIHMLARENLP	189
Halobacterium	DYELDLHAMQVETSLKRAQRAMAIEIQ-SGETP--KVSHVVYPLMQALDIEYLD---LDLAIGGMDQRKVHMLAREELP	188
Haloarcula	EYTLVEMRMAAETTISRQRAASEVVRRESPE--NLGGLIYPLMQTLQVLDKALD---ADIAYGGVDQRGIYMLSREILP	196
Drosophila	EYTLVYKLVSSVVTQHDACKKAGAENVKQVEYV--LLSGLLYPGLQALDEEYLYK---VDAQFGGVDQRKIFTFSEKYLP	199
Homo	EYTLVYKLVSSVVTQHDACKKAGAENVKQVEHP--LLSGLLYPGLQALDEEYLYK---VDAQFGGIDQRKIFTFAEKYLP	200
Saccharomyces	DYIMDIFRLSNIVSQNDACKRAGADVVKQVANP--LLSGLIYPLMQALDEQFLD---VDCQFGGVDQRKIFVLAENLP	204
Encephalitozoon	RYIMDLYRILSITSKHDACKKAGAENVRQVENP--MVSLLVYPSMQALDEVHLS---VDAQFGGVDQRKIFTYARKYLP	196
Caenorhabditis	EYTDVLRLLSAQVSQRDALKAGAENVKQVESPE--LLSGLLYPLQALDEQYLYK---VDGQFGGVDQRKIFILAEEQLP	211

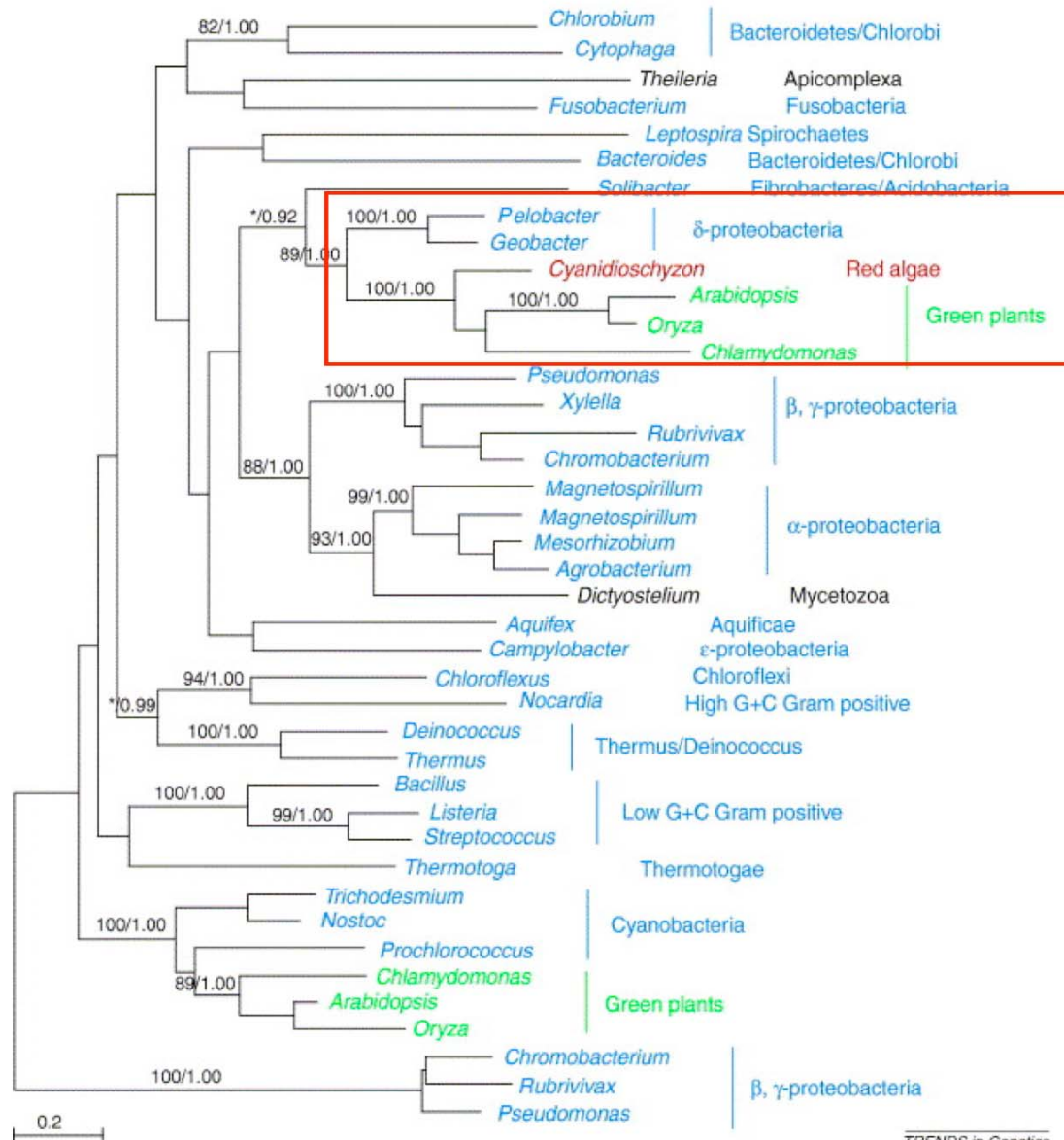
Signature residues for association of metazoan/fungal/haloarchaeal homologs

Transferred tyrRS supports monophyletic opisthokonts

The same conclusion is reached, if haloarchaeal type tyrRS in opisthokonts is explained by ancient paralogy and differential gene loss.

Monophyly of primary photosynthetic eukaryotes is supported by more than 50 ancient gene transfers from different bacterial phyla to the ancestor of the red algae and green plant lineage.

- E.g., ancient gene transfer of frp-gene (florfenicol resistance protein)



Gene from Green plants and Red algae groups with delta proteobacteria

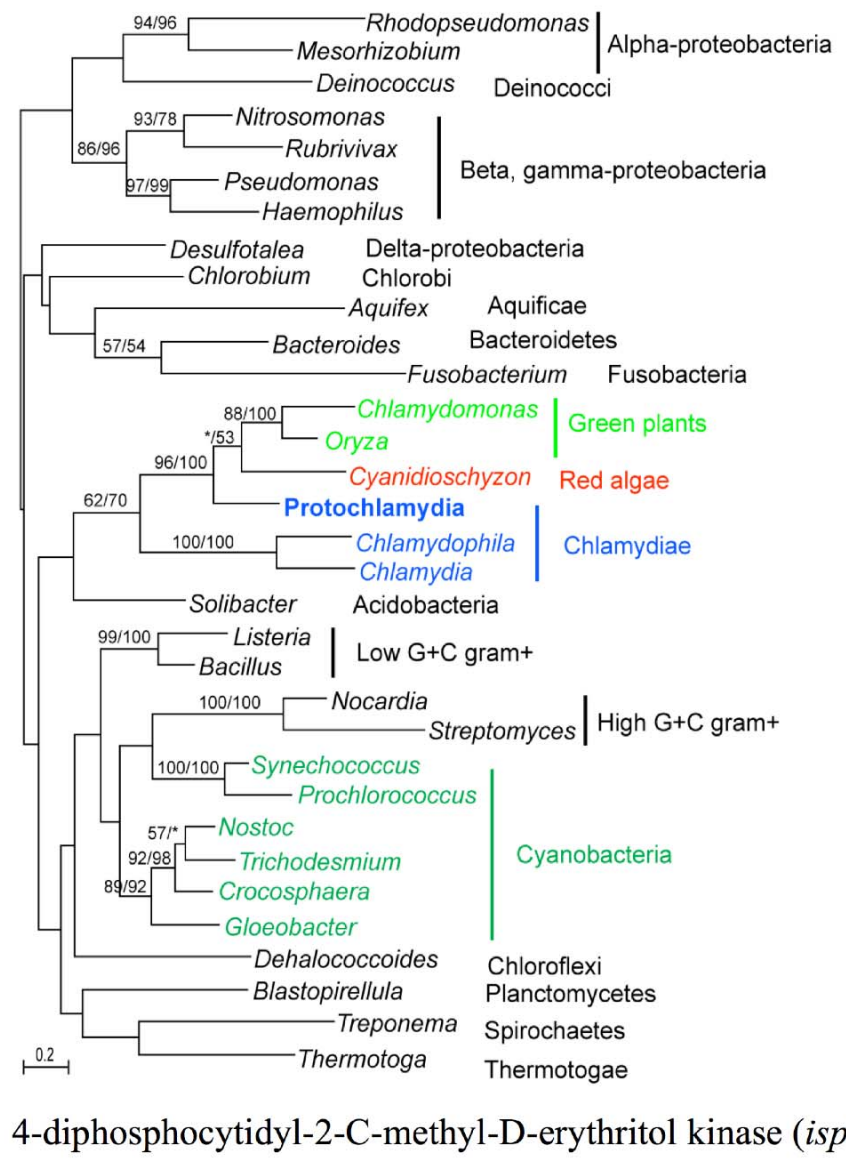
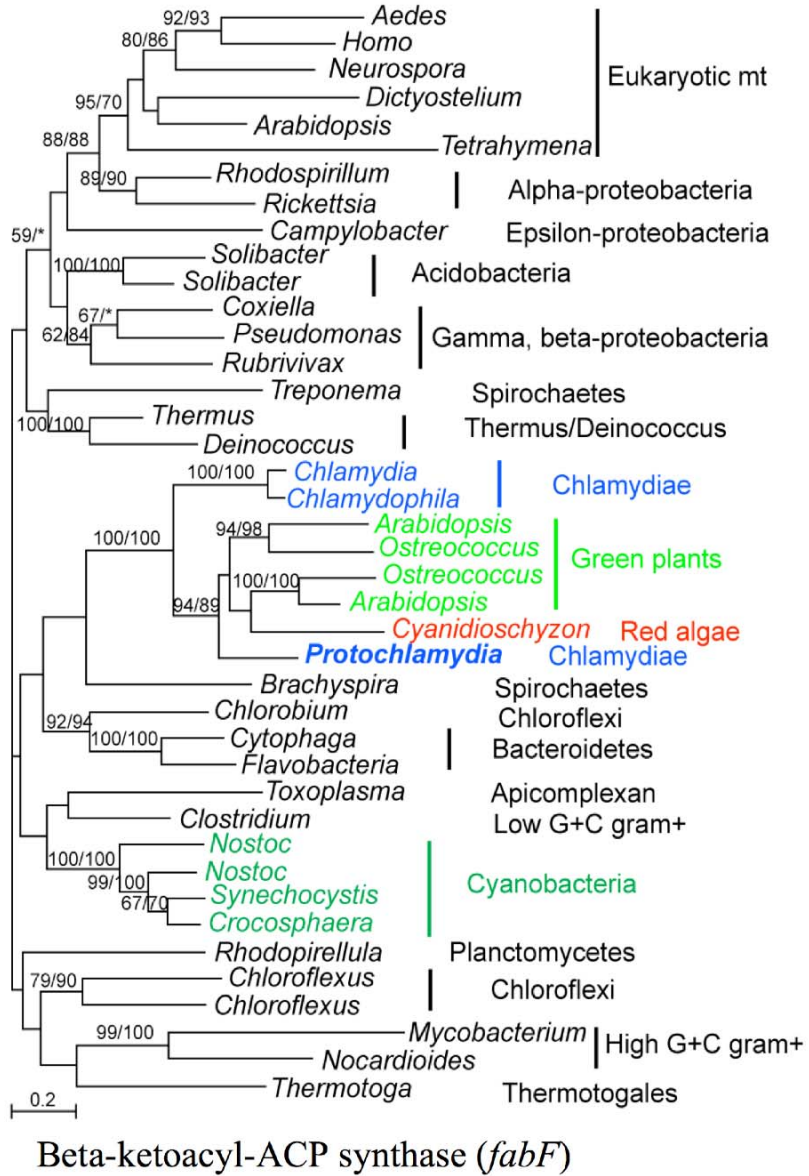
modified from Huang and Gogarten (2006), Trends in Genetics 22, 361- 366

Red Algal and Green Plant Genes of Chlamydial Origin

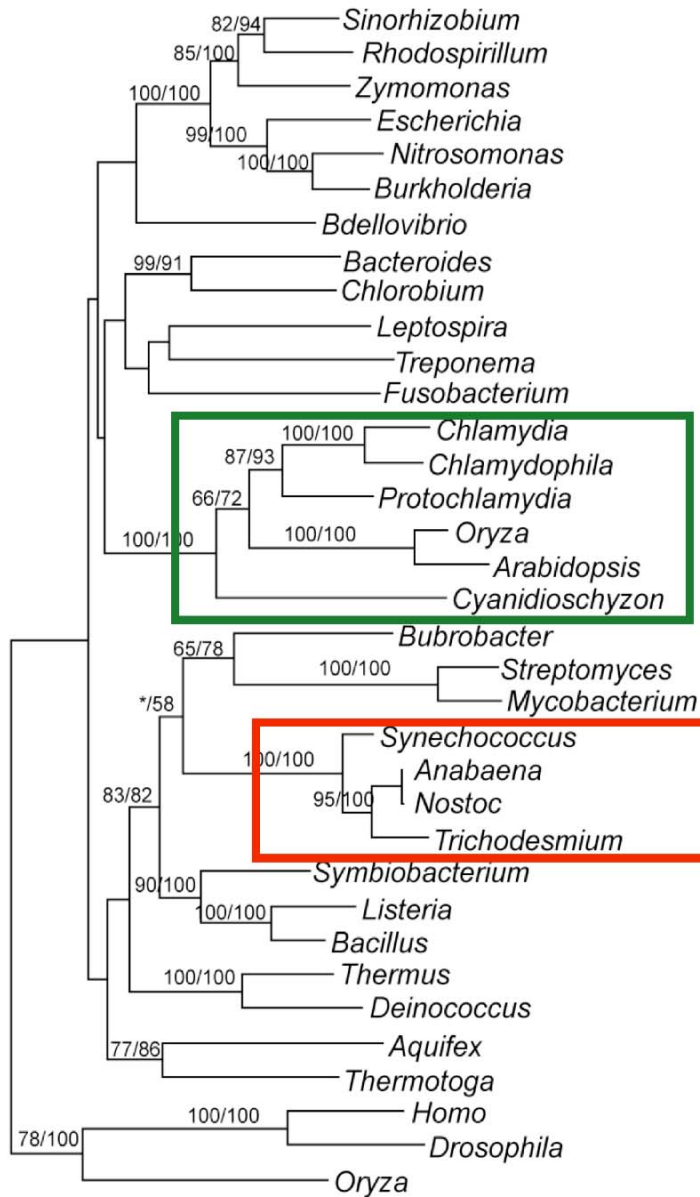
Gene Name or Gene Product	Presence	Putative Function
ADT/ATP translocase	R and G	ATP/ADP transport
Phosphate transporter	G	Phosphate transport
Sodium:hydrogen antiporter	R and G	Ion transport
Cu-ATPase	R and G	Ion transport
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (<i>gcpE</i>)	R and G	Isoprenoid biosynthesis
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (<i>ispE</i>)	R and G	Isoprenoid biosynthesis
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (<i>ispD</i>)	R and G	Isoprenoid biosynthesis
Enoyl-ACP reductase (<i>fabI</i>)	R and G	Fatty acid biosynthesis
Beta-ketoacyl-ACP synthase (<i>fabF</i>)	R and G	Fatty acid biosynthesis
Glycerol-3-phosphate acyltransferase	R and G	Phospholipid
Polynucleotide phosphorylase	R and G	RNA degradation
Phosphoglycerate mutase	G	Glycolysis
Oligoendopeptidase F	R	Amino acid biosynthesis
Aspartate transaminase	R and G	Amino acid metabolism
Malate dehydrogenase	G	Energy conversion
Tyrosyl-tRNA synthetase	R and G	Translation
23S rRNA (Uracil-5-)-methyltransferase	R and G	RNA modification
Isoamylase	R and G	Starch biosynthesis
Hypothetical protein	R	Unknown
Sugar phosphate isomerase	G	Sugar interconversion
CMP-KDO synthetase	G	Cell envelope formation

Consistent phylogenetic signal links
Chlamydiae, red algae and green plants.

Chlamydial-type genes in red algae and plants are often specifically associated with *Protochlamydia* (*Parachlamydia*)



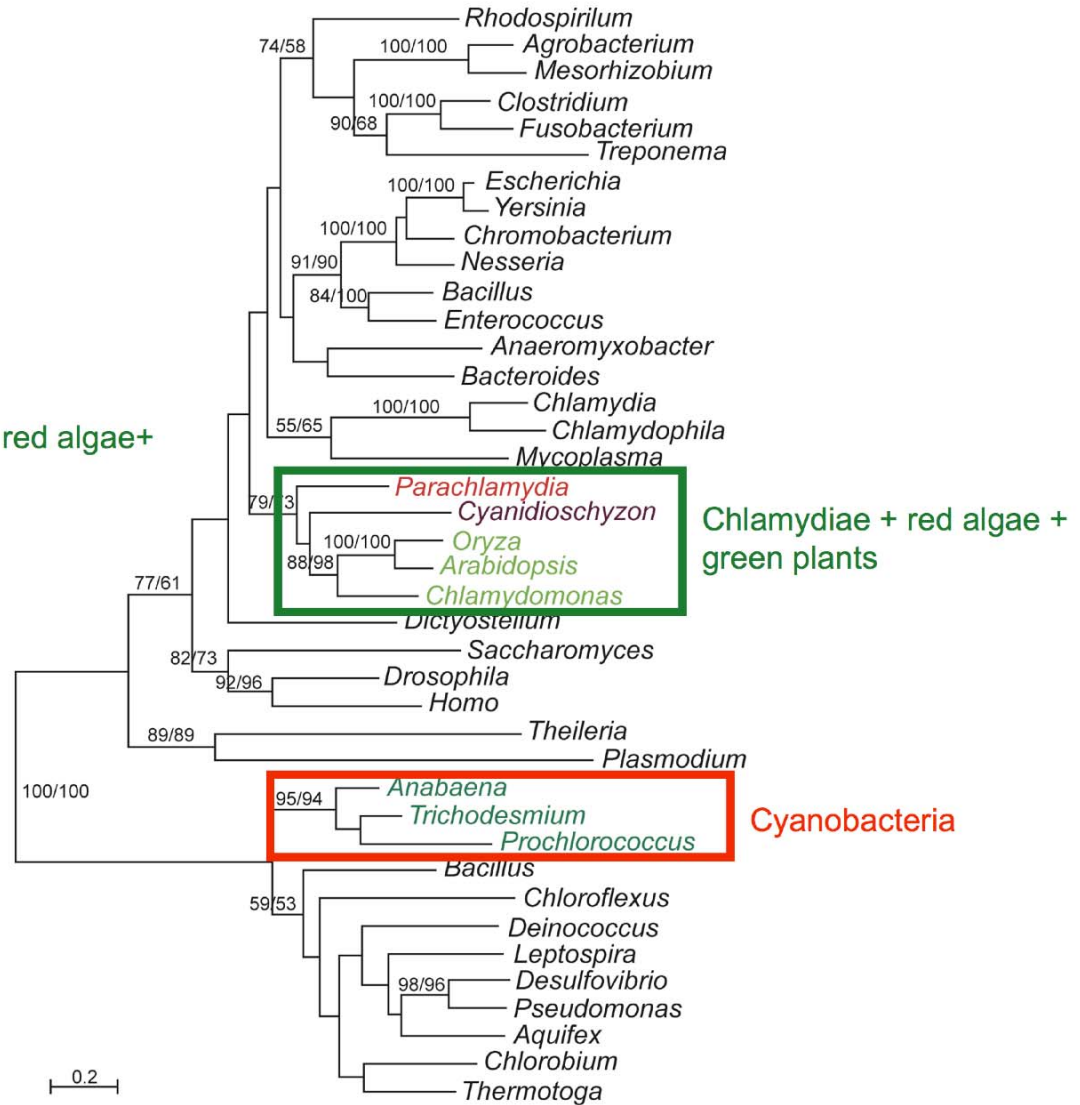
The chlamydial genes (plant & red algae) group separate from the cyanobacterial homologs



Polynucleotide Phosphorylase

Chlamydiae, red algae + green plants

Cyano-bacteria



Tyrosyl-tRNA synthetase

Chlamydiae + red algae + green plants

Cyanobacteria

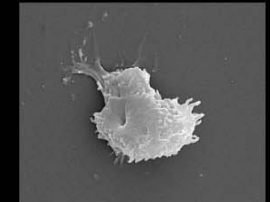
Examining possible Hypotheses

1. Plants acquired chlamydial genes via insect feeding activities (Everett et al. 2005).



No. The ancestor of red algae and green plants is much older than insects.

2. Chlamydiae acquired plant-like genes via Acanthamoeba hosts (Stephens et al. 1999; Wolf et al. 1999; Ortutay et al. 2003).



No. All these genes are of bacterial origin. The direction of gene transfer is from bacteria to eukaryotes.

3. Chlamydial and plant sequence similarities reflect an ancestral relationship between chlamydiae and cyanobacteria (Brinkmann et al. 2002; Horn et al. 2004).



No. Genes of chloroplast ancestry should still be more similar to cyanobacterial than to chlamydial sequences. In many instances the cyanobacterial homologs form a clearly distinct, and separate clade.

What do the data suggest?

(Chlamydial genes in red algal and plant genomes)

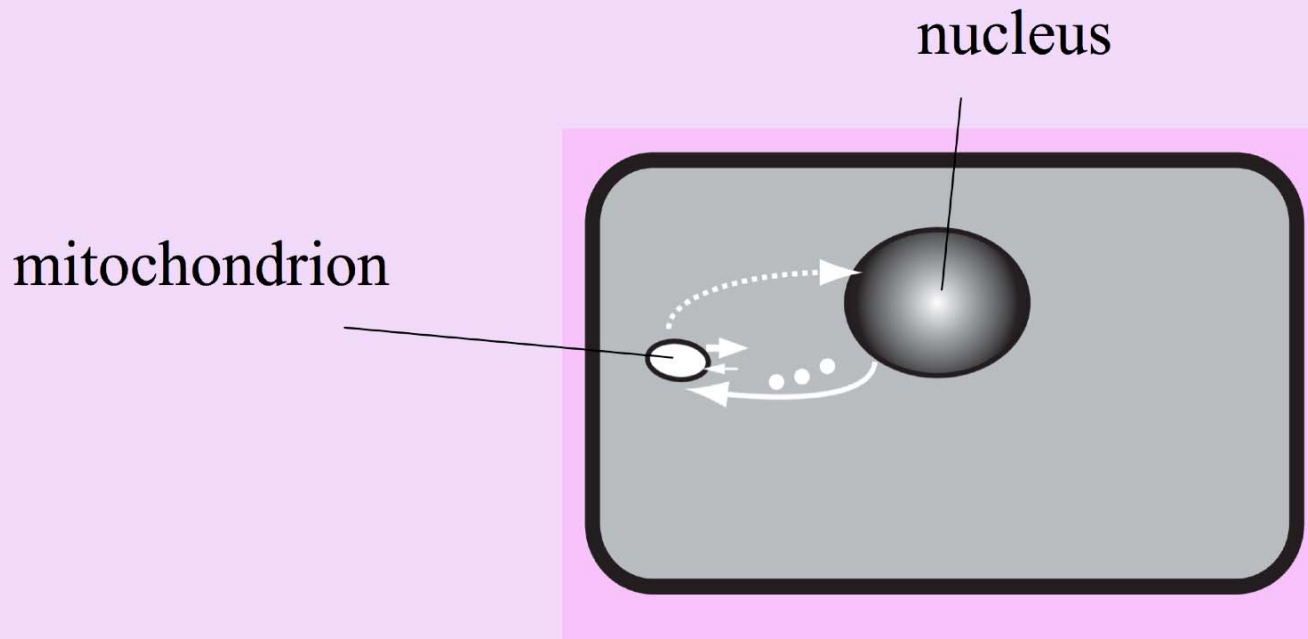
Unless a stable physical association existed, it is highly unlikely for any single donor to transfer such a number of genes to a single recipient.

Our Hypothesis: An ancient, unappreciated symbiotic association existed between chlamydiae and the ancestor of red algae and green plants.

Genes from this chlamydial symbiont might have been crucial to establish communication between host and the cyanobacterial cytoplasm.

Hypothesis: Chlamydiae and the primary plastids

A) The Host



White: α -proteobacterial (mitochondrial) symbiont

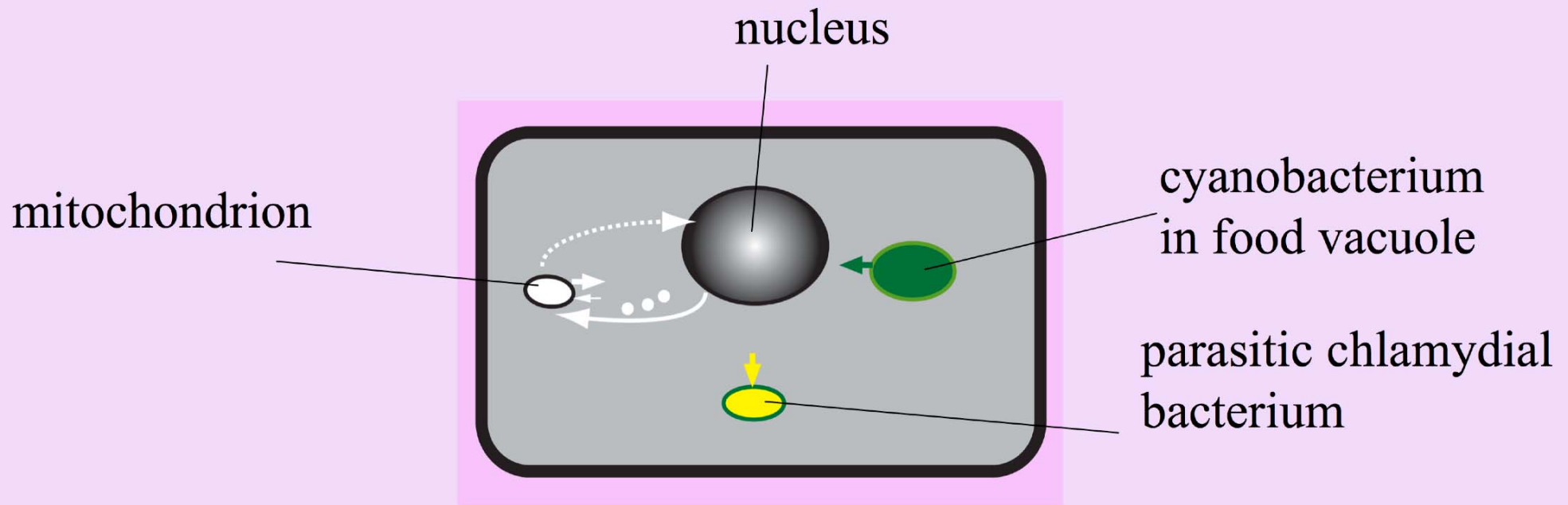
.....▶ Gene transfer to the nucleus

└▶ Transport of nuclear encoded proteins to symbiont

↔ Direction of symbiotic benefit



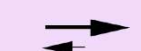
Hypothesis: Chlamydiae and the primary plastids

B) The Host invaded by a parasite



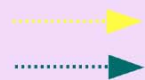
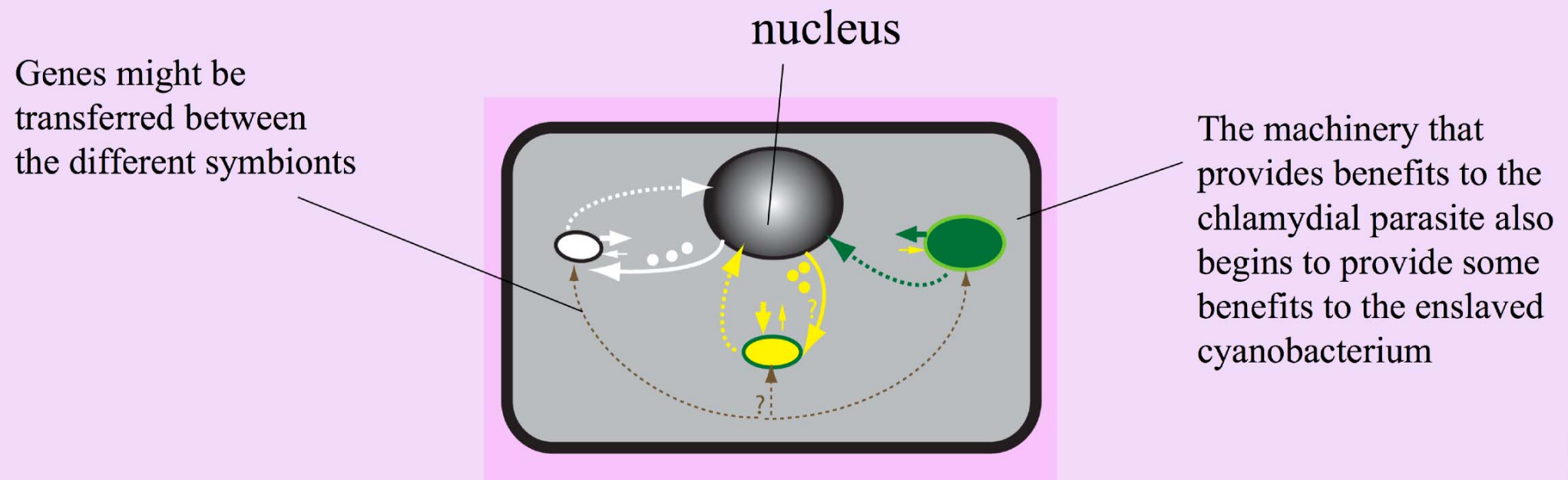
Yellow: parasitic chlamydial bacterium

Green: cyanobacterium (as food)

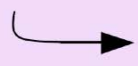
-  Gene transfer to the nucleus
-  Transport of nuclear encoded proteins to symbiont
-  Direction of symbiotic benefit

Hypothesis: Chlamydiae and the primary plastids

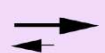
C) The chlamydial symbiont becomes mutualistic



Gene transfer from the symbionts to the nucleus



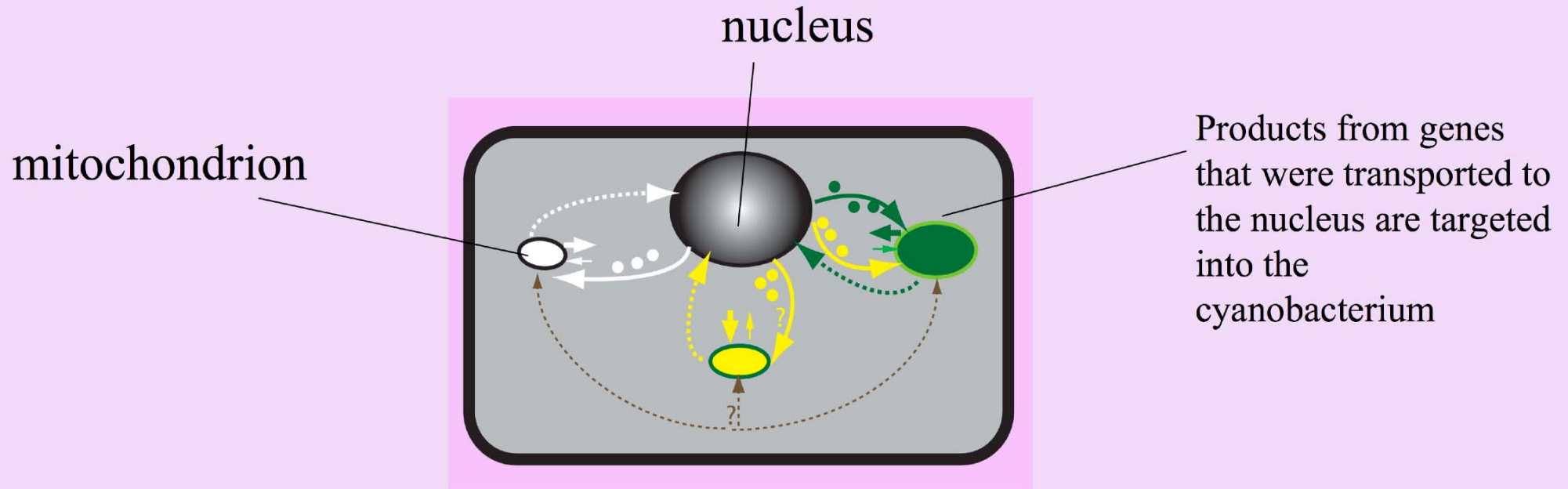
Transport of nuclear encoded proteins to symbiont



Direction of symbiotic benefit

Hypothesis: Chlamydiae and the primary plastids

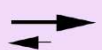
D) The cyanobacterium evolves into an organelle



Chlamydial and cyanobacterial type gene products are targeted to the plastid



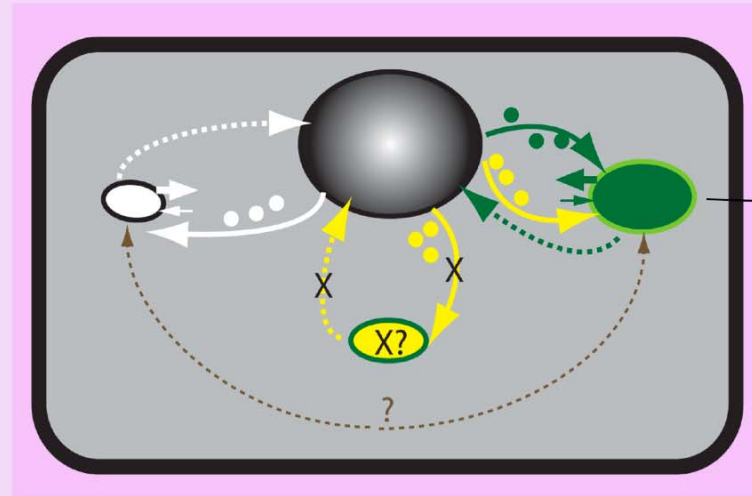
Gene transfer from the symbionts to the nucleus



Direction of symbiotic benefit

Hypothesis: Chlamydiae and the primary plastids

D) Loss of chlamydial symbiont



With the help of genes contributed by the chlamydial parasite, the cyanobacterium has evolved into a photosynthetic organelle

The chlamydial symbiont is lost - possibly without trace, except for the genes that facilitated the integration of the metabolism of the host with that of a photoautotroph.

This hypothesis also explains why the evolution of an organelle from a primary endosymbiont is rare. A photoautotroph with a single compartment has few transporters available that would allow integration with the host metabolism.

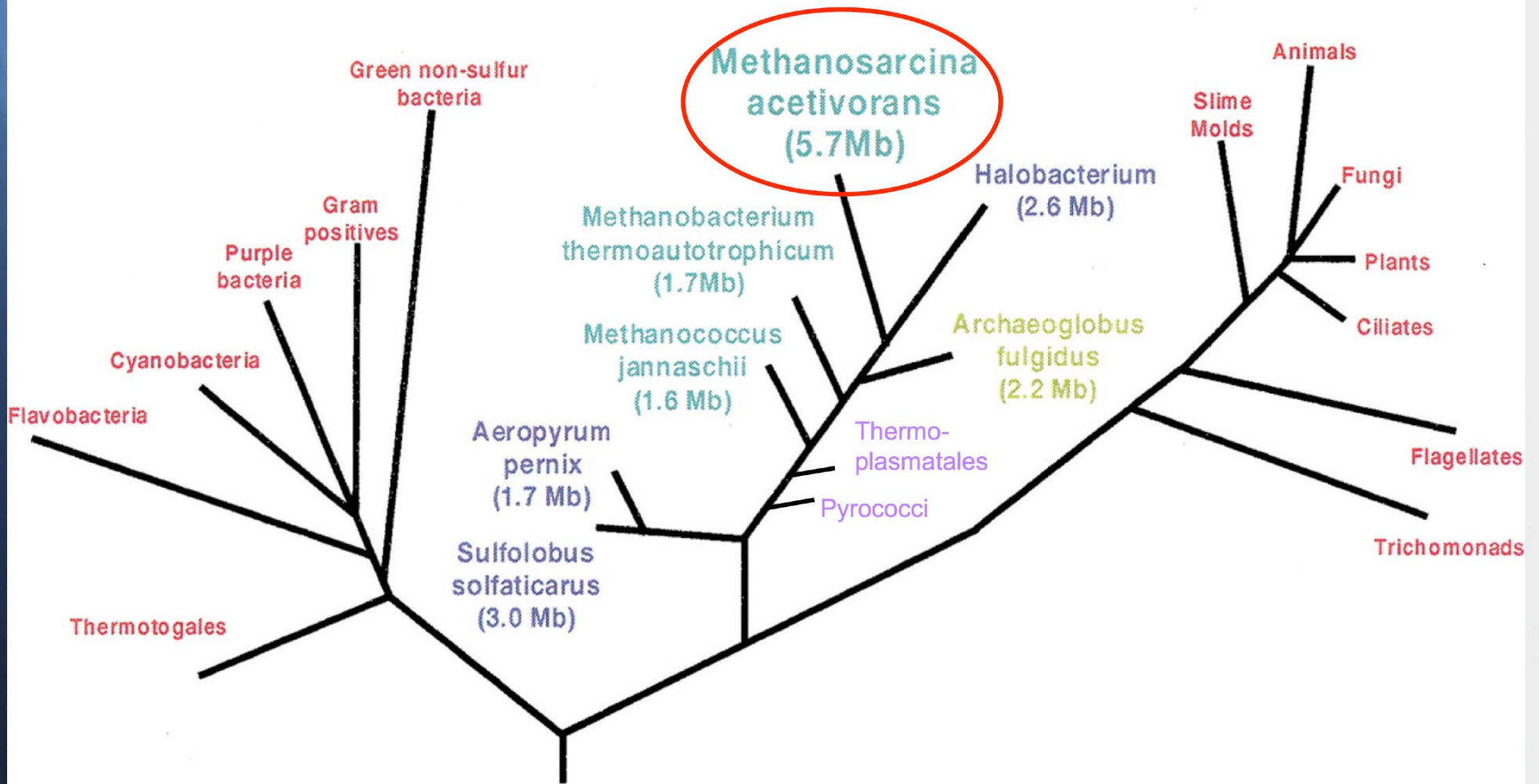
The simultaneous presence of an intracellular parasite allows for the integration of the two cytoplasms.

Evolution of Aceticlastic Methanogenesis in
Methanosarcinales via Horizontal Gene
Transfer from *Clostridia*

Bacteria

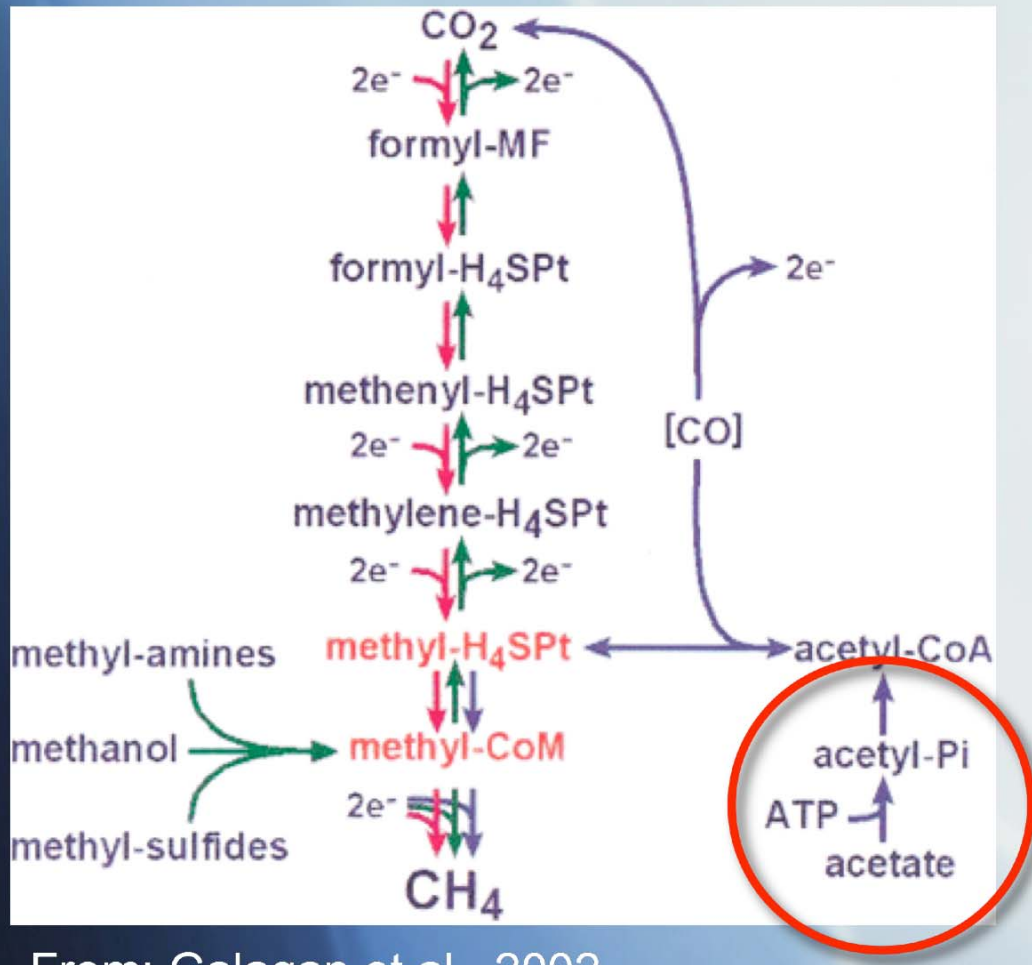
Archaea

Eucarya



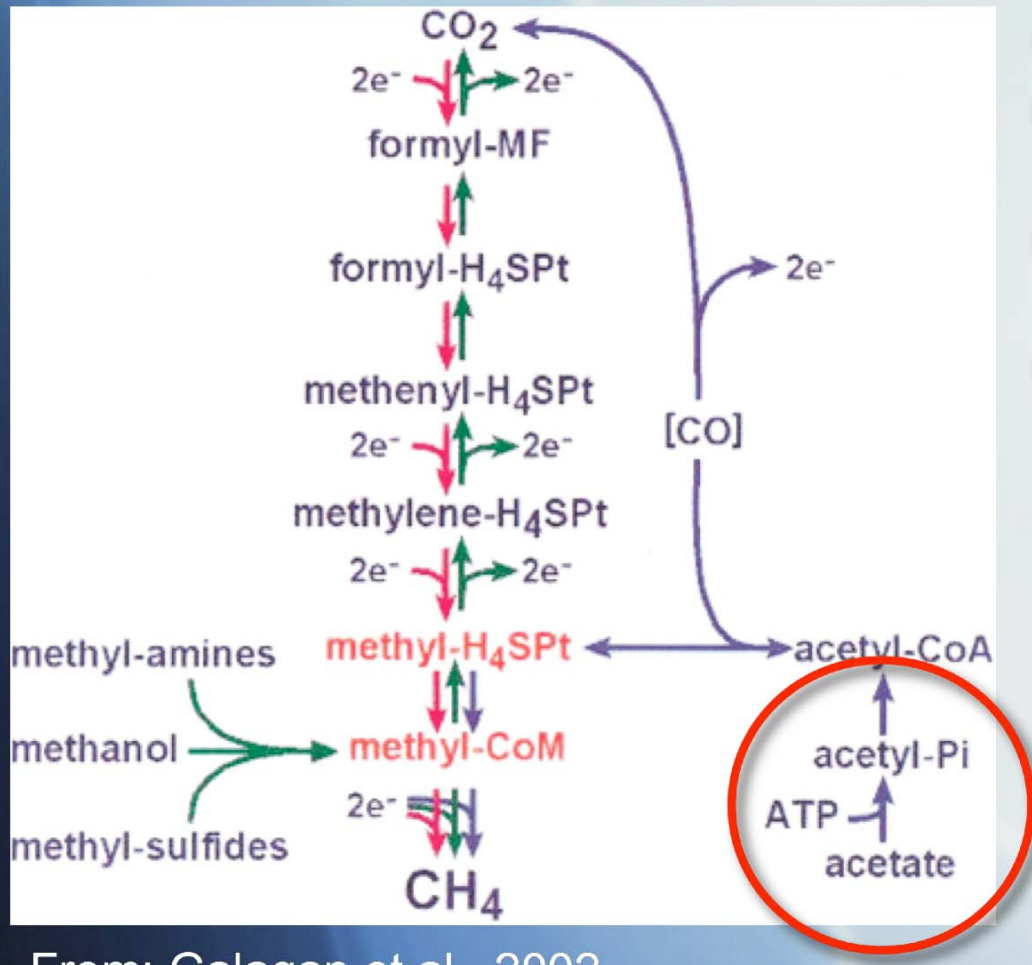
Modified from: Galagan et al., 2002

Methanogenesis



From: Galagan et al., 2002

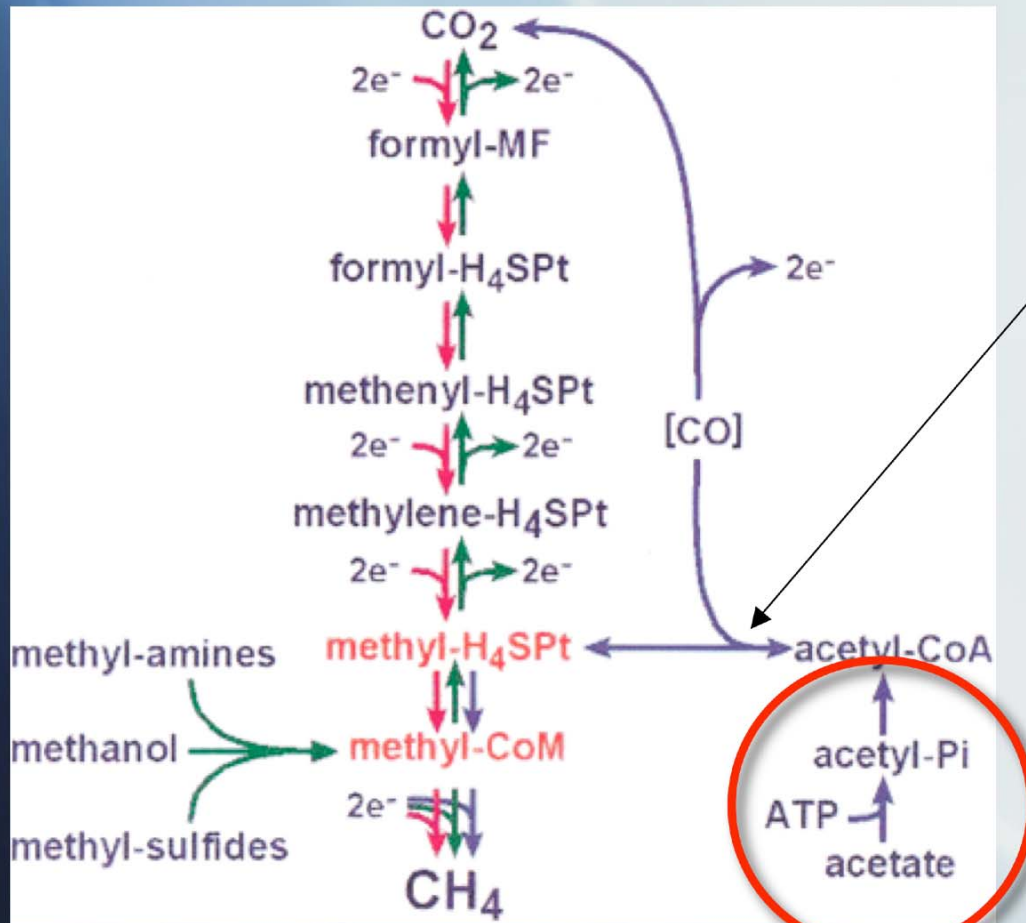
Methanogenesis



- Unique to subset of Archaea
- Energy production via reduction of multiple carbon substrates to CH₄
- 900 Million metric tons of biogenic methane produced annually.
- Over 66% of biogenic methane is produced from acetate, mostly by *Methanosarcina* genera.

From: Galagan et al., 2002

Methanogenesis

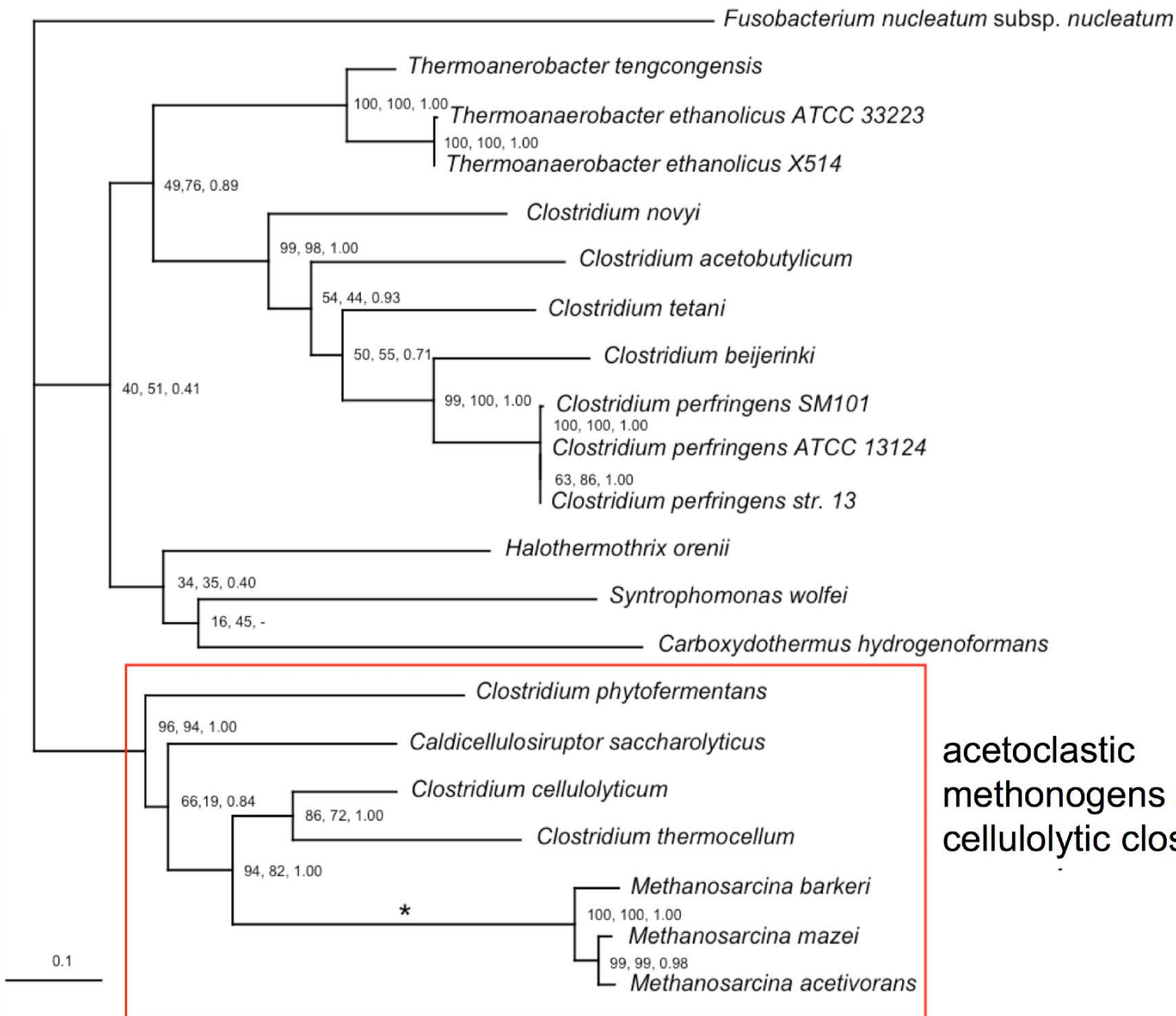


CODH complex:
reversible Acetyl-CoA
synthesis

Aceticlastic methanogenesis-
specific pathway:
acetate kinase (AckA) and
phosphotransacetylase (Pta)

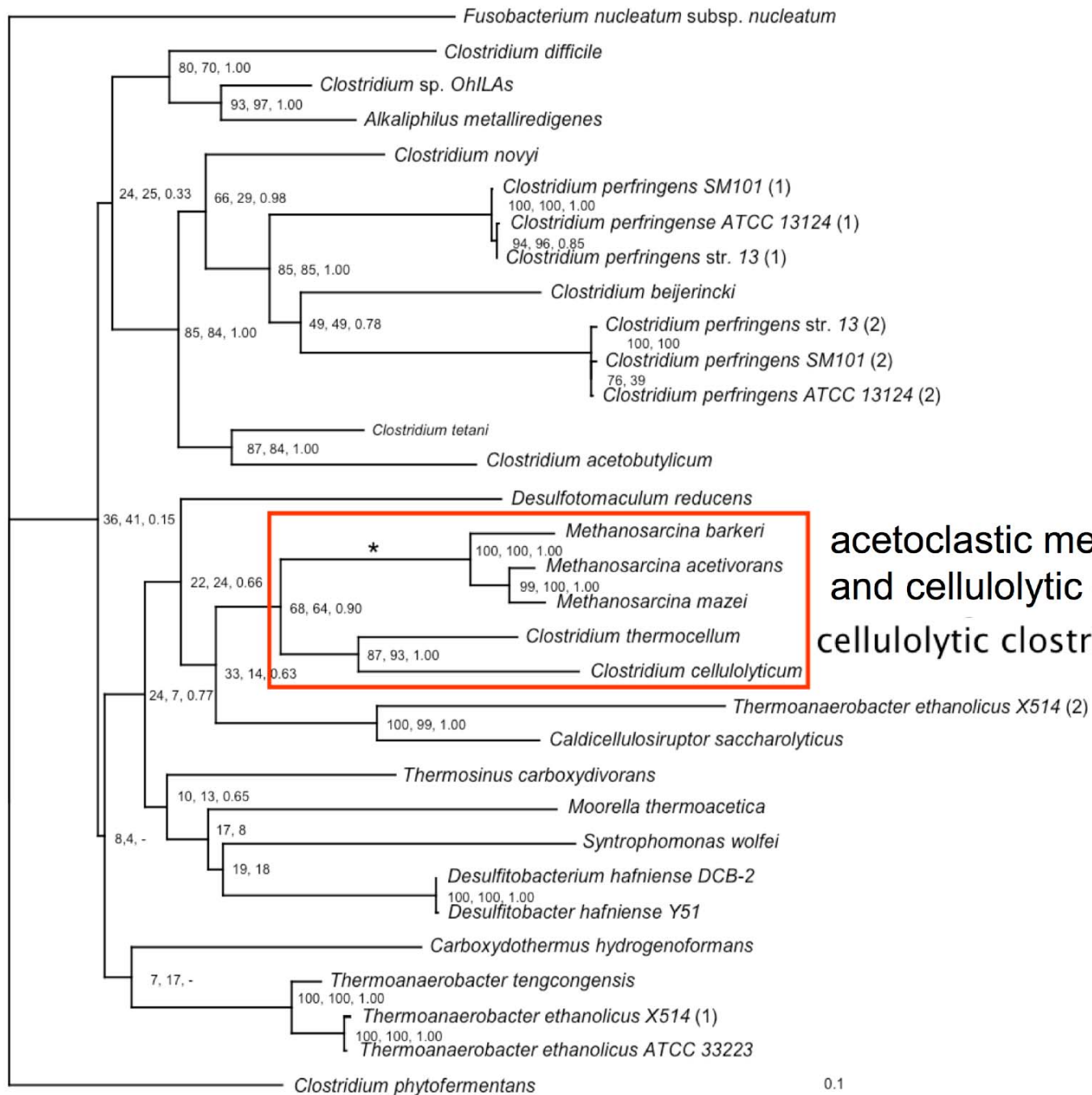
From: Galagan et al., 2002

Phosphotransacetylase (Pta)



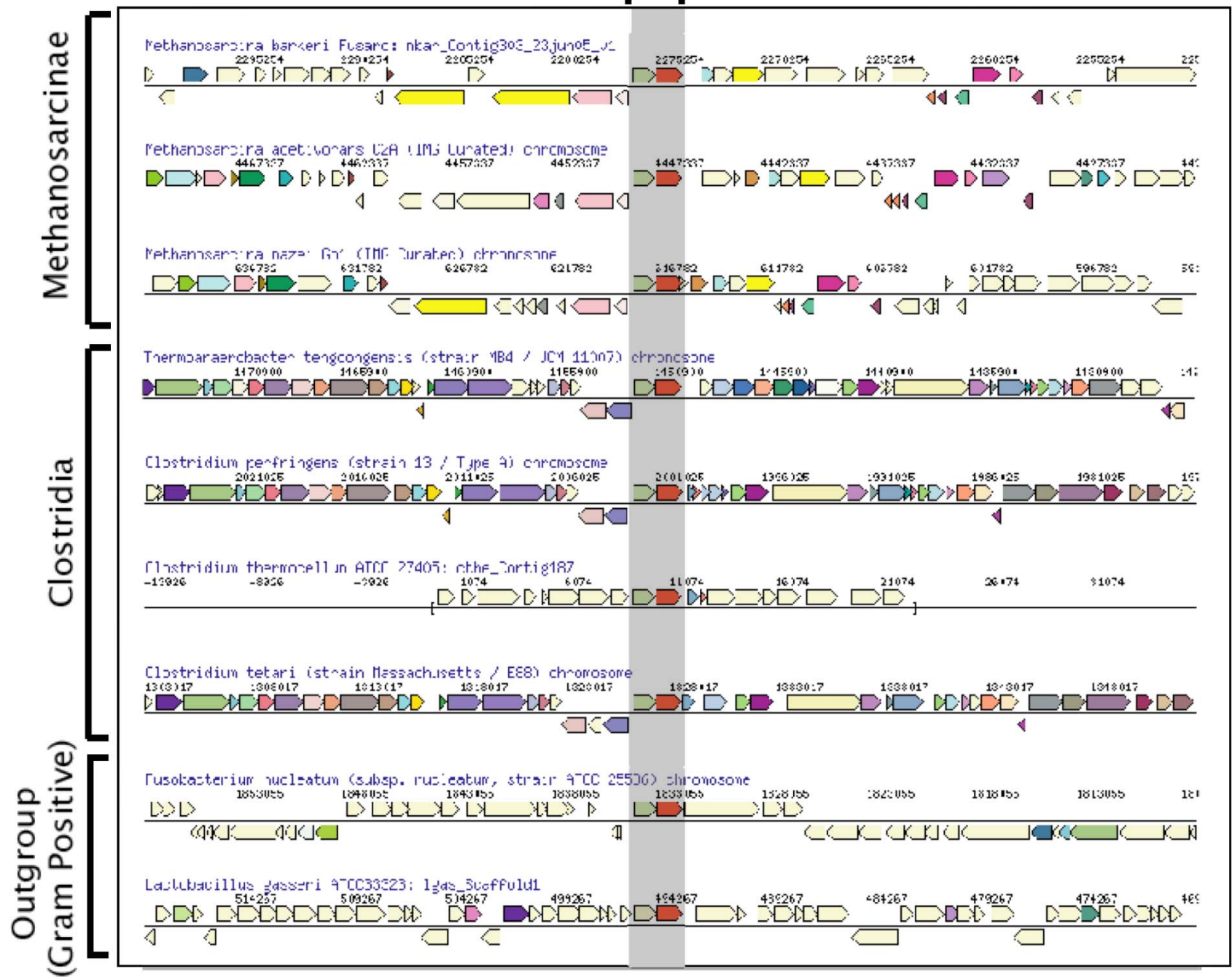
acetoclastic
methonogens and
cellulolytic clostridia

acetate kinase (AckA)

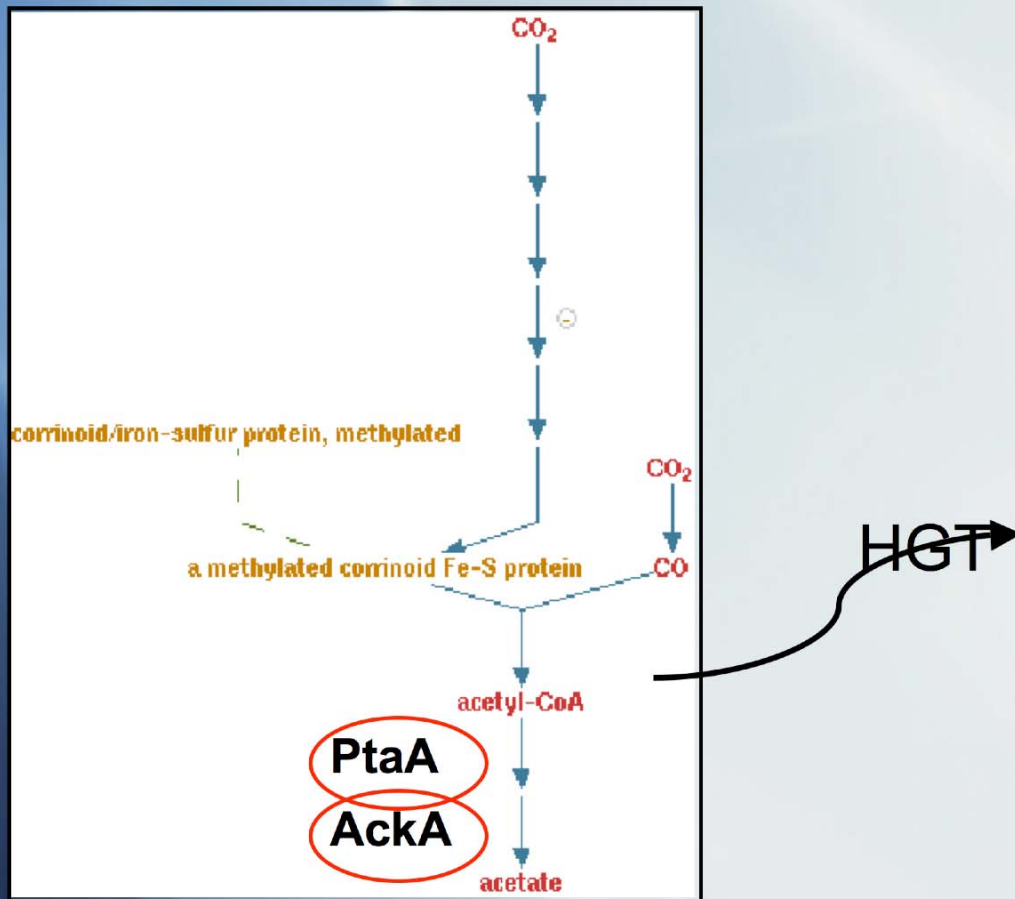


acetoclastic methonogens
and cellulolytic clostridia
cellulolytic clostridia

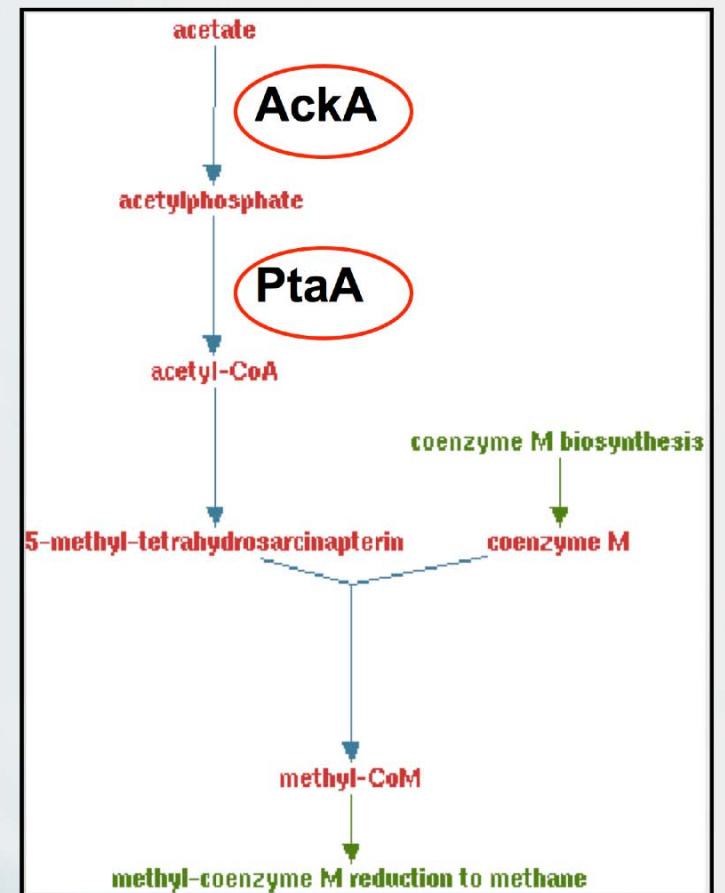
AckA/PtaA



Clostridia acetigenic pathway



Methanosarcina aceticlastic pathway



Figures drawn with *Metacyc* (www.metacyc.org)

Can one reconstruct microbial phylogeny?

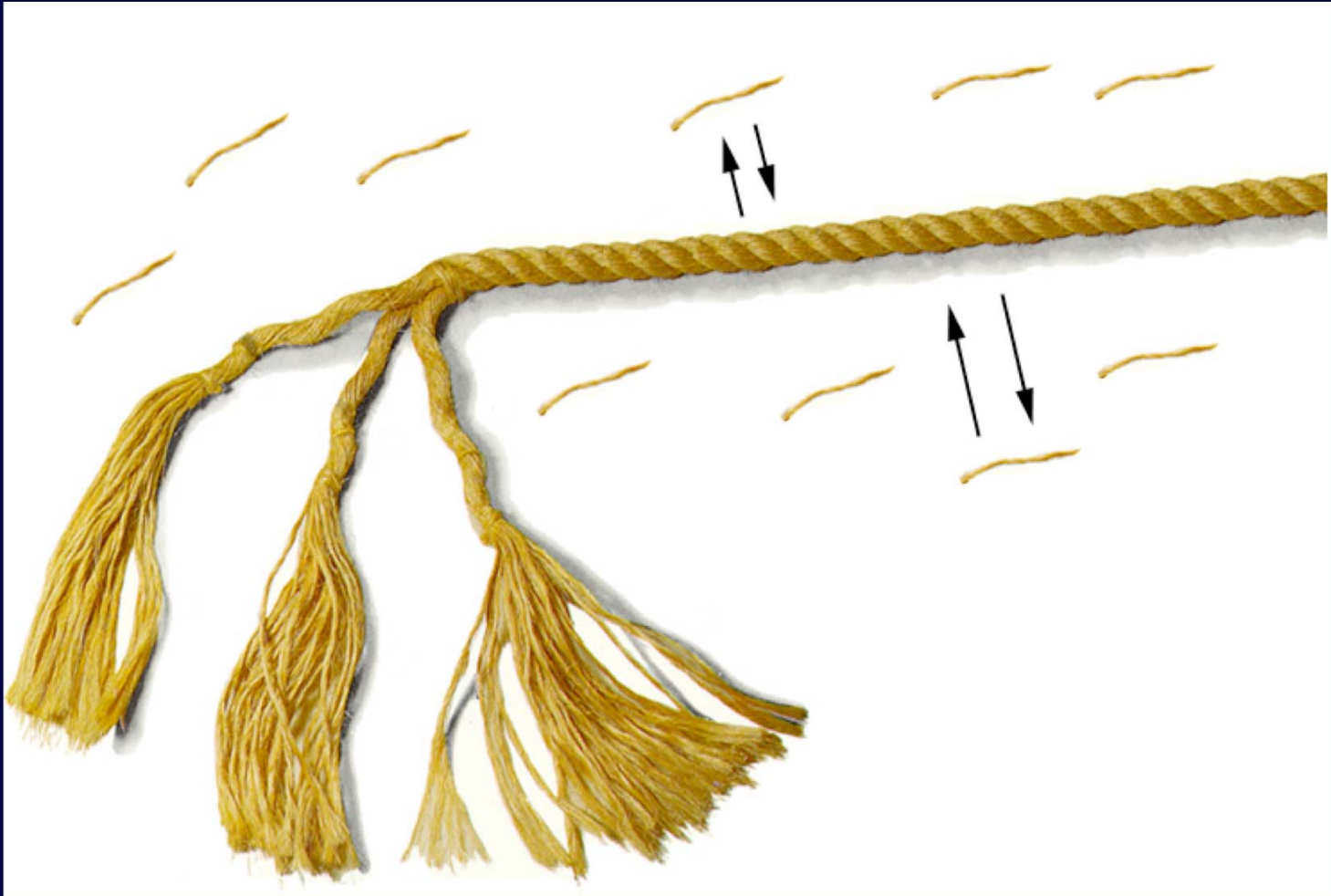
PHYLOGENY: from Greek phylon, race or class, and -geneia, born.

“the origin and evolution of a set of organisms, usually of a species” (Wikipedia);

The phylogeny of a species does not necessarily occur in a tree-like process!

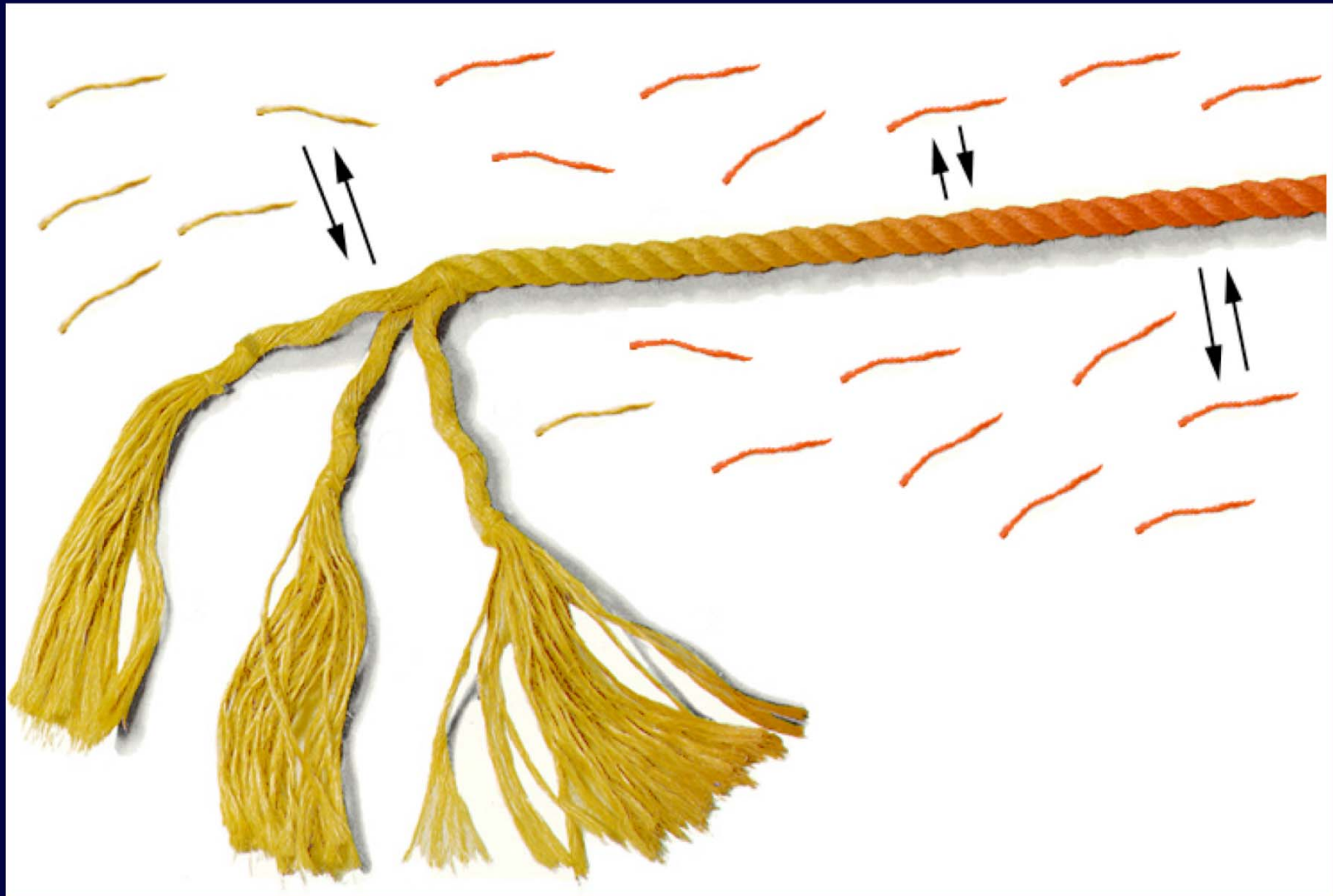
Rope as a metaphor to describe an organismal lineage (*Gary Olsen*)

Individual fibers = genes that travel for some time in a lineage.

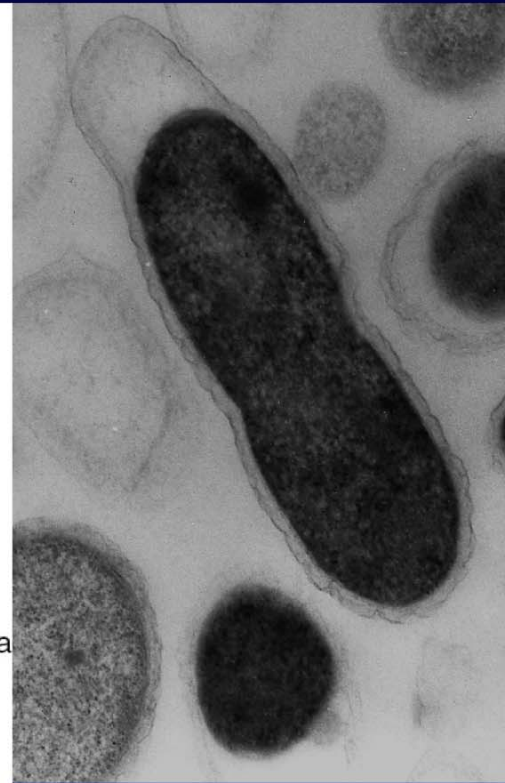
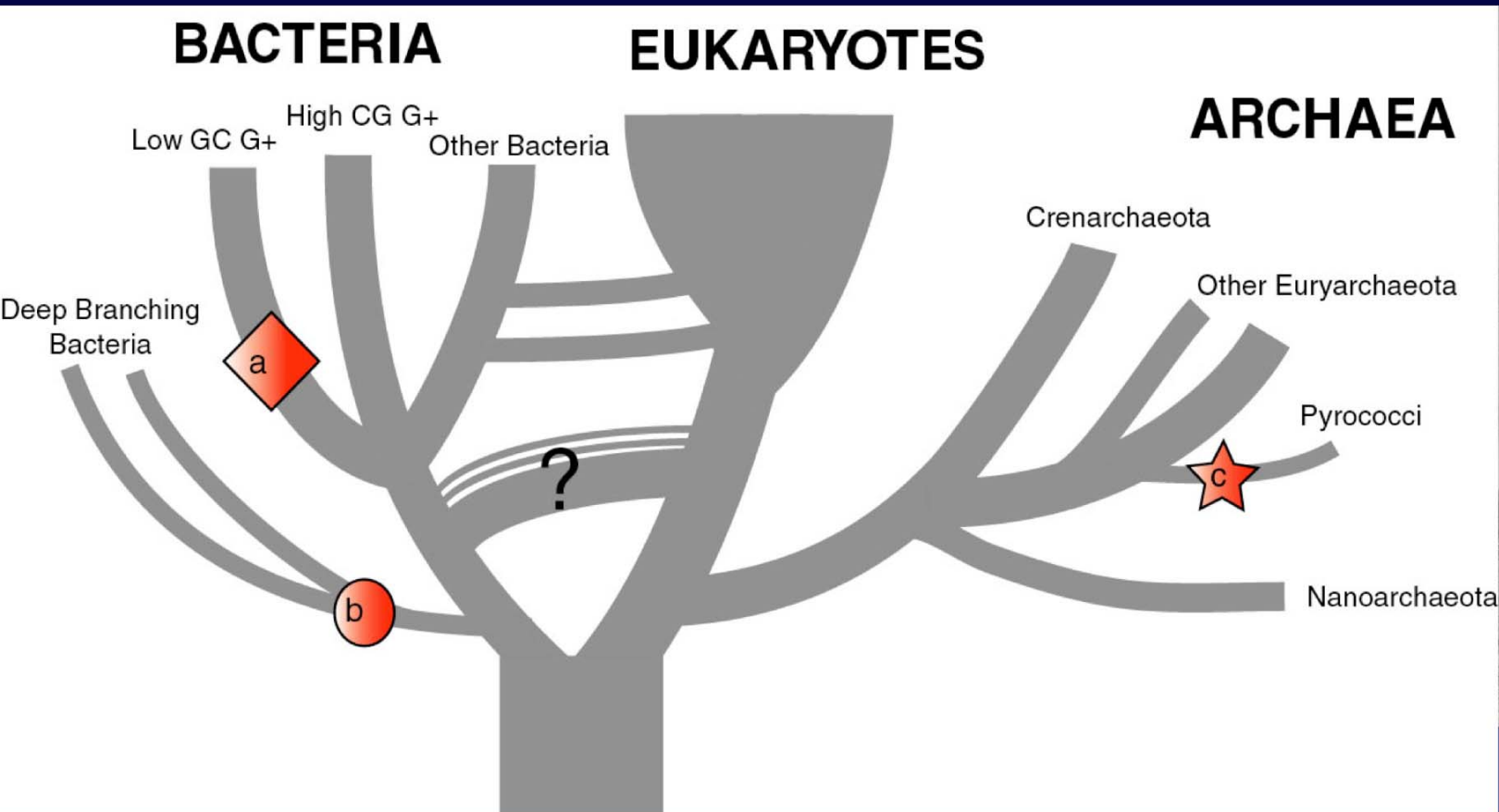


While no individual fiber present at the beginning might be present at the end, the rope (or the organismal lineage) nevertheless has continuity.

However, the genome as a whole will acquire the character of the incoming genes (the rope turns solidly red over time).



The Phylogenetic Position of *Thermotoga*



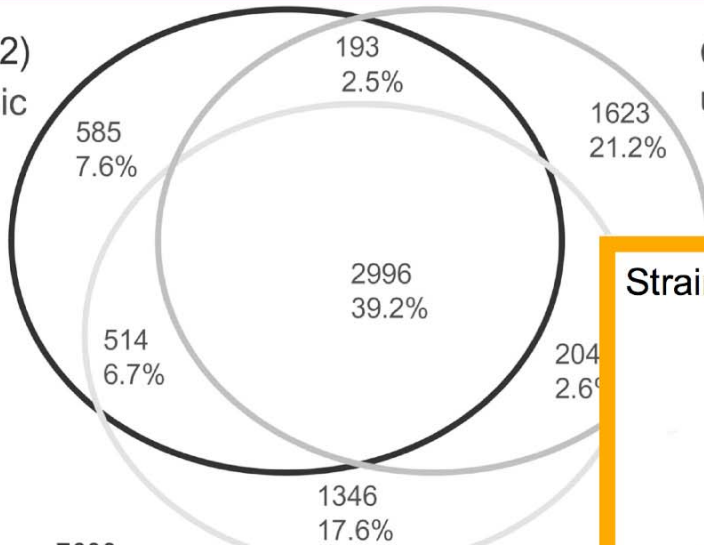
(a) Some genome based approaches (favoring concordant genes) (b) according to 16S & other genome based approaches (c) according to phylogenetically discordant genes

Gophna, U., Doolittle, W.F. & Charlebois, R.L.:

Weighted genome trees: refinements and applications. *J. Bacteriol.* (2005)

MG1655 (K-12)
non-pathogenic

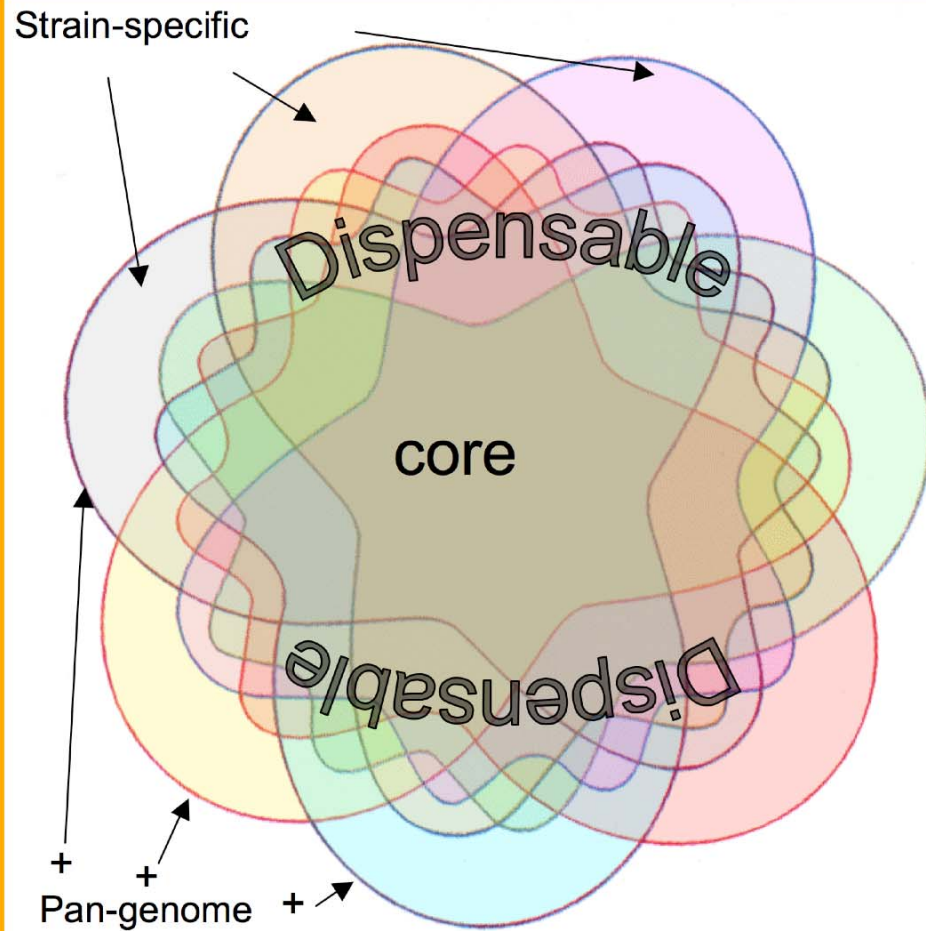
CFT073
uropathogenic



Total proteins = 7638
2996 (39.2%) in all 3
911 (11.9%) in 2 out of 3
3554 (46.5%) in 1 out of 3

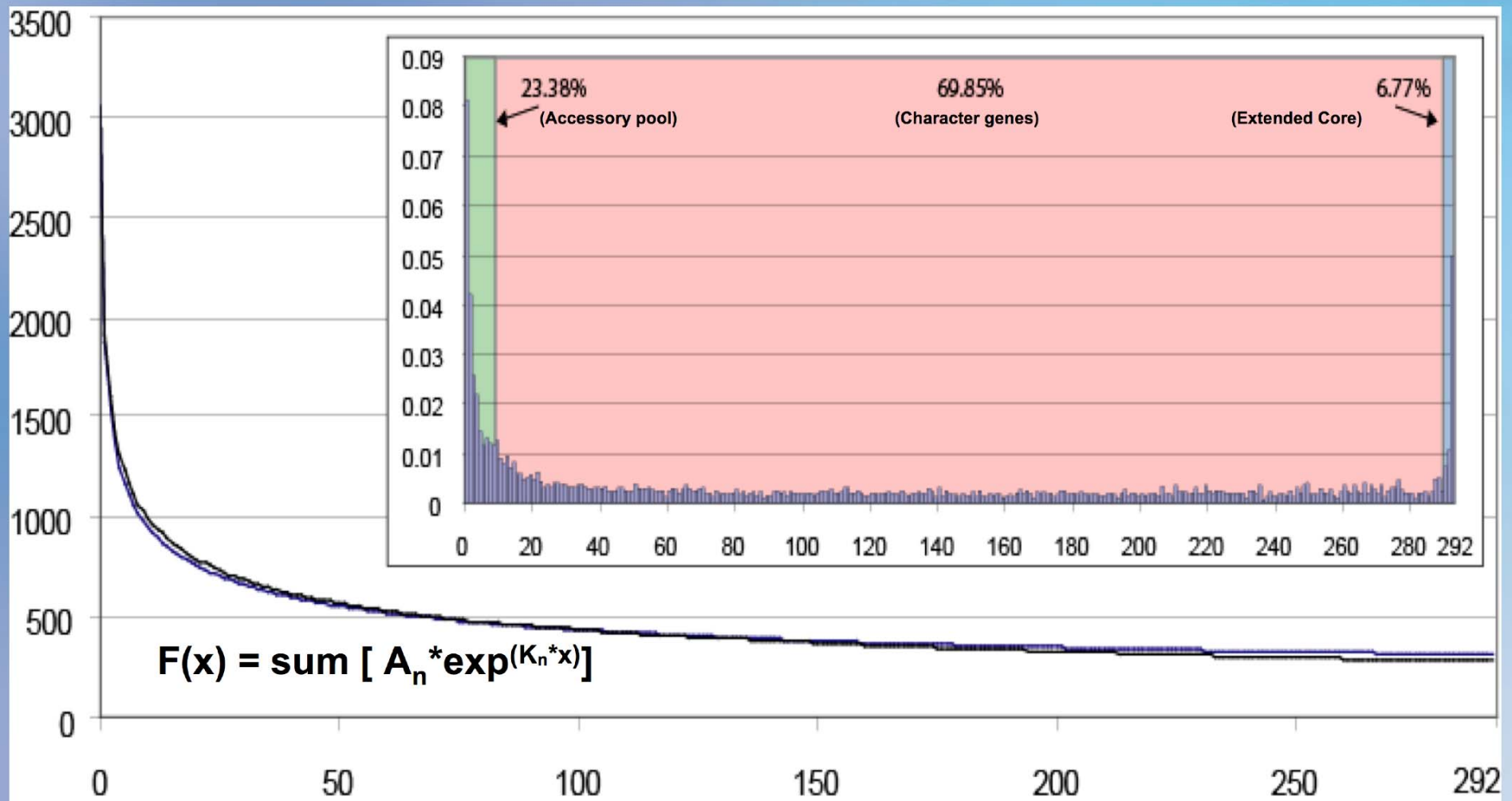
EDL933 (O157:H7)
enterohaemorrhagic

Strain-specific

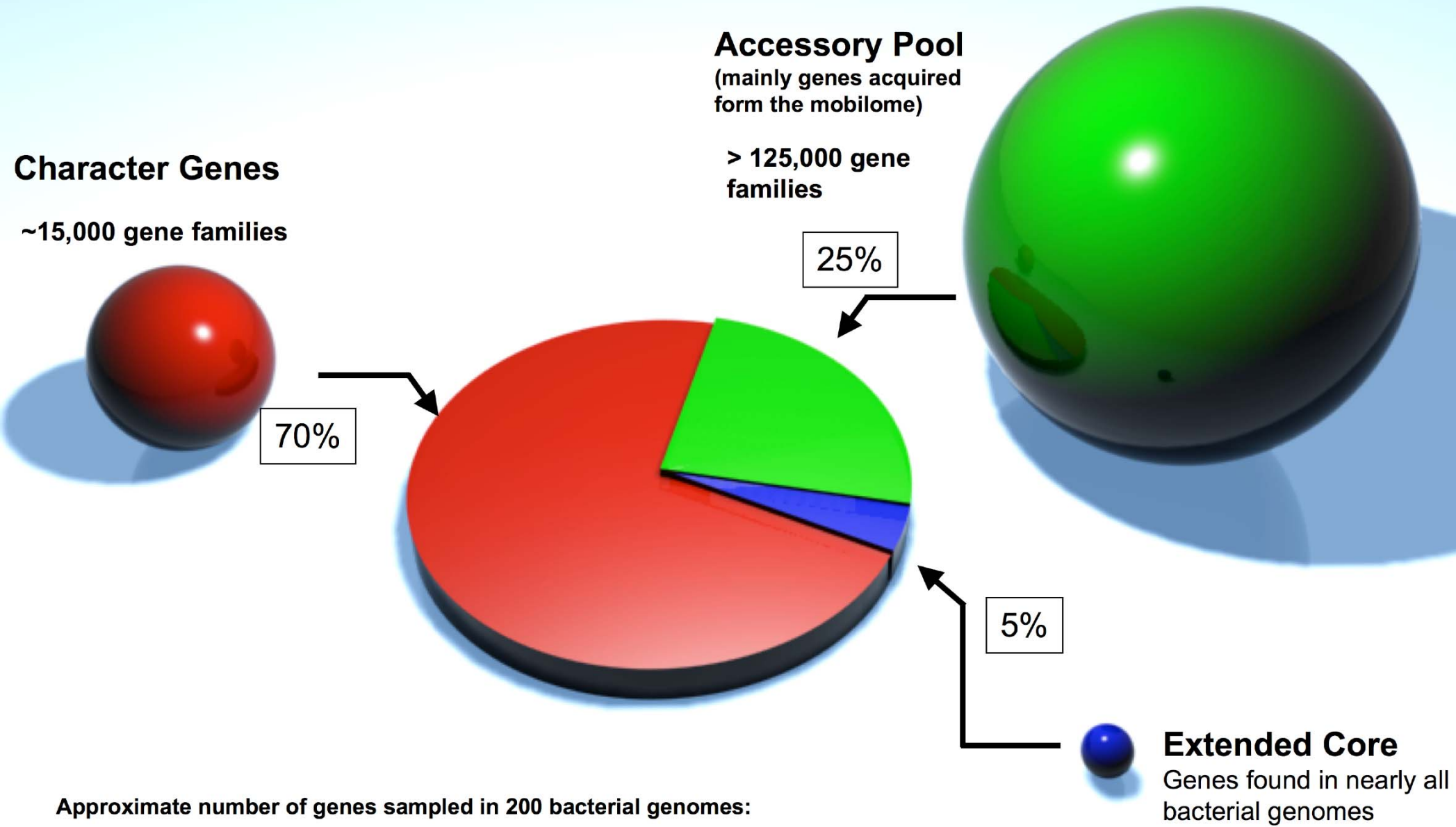


Gene frequency in a typical genome

- Pick a random gene from any of the 293 genomes
- Search in how many genomes this gene is present
- Sampling of 15,000 genes



Gene frequency in individual genomes

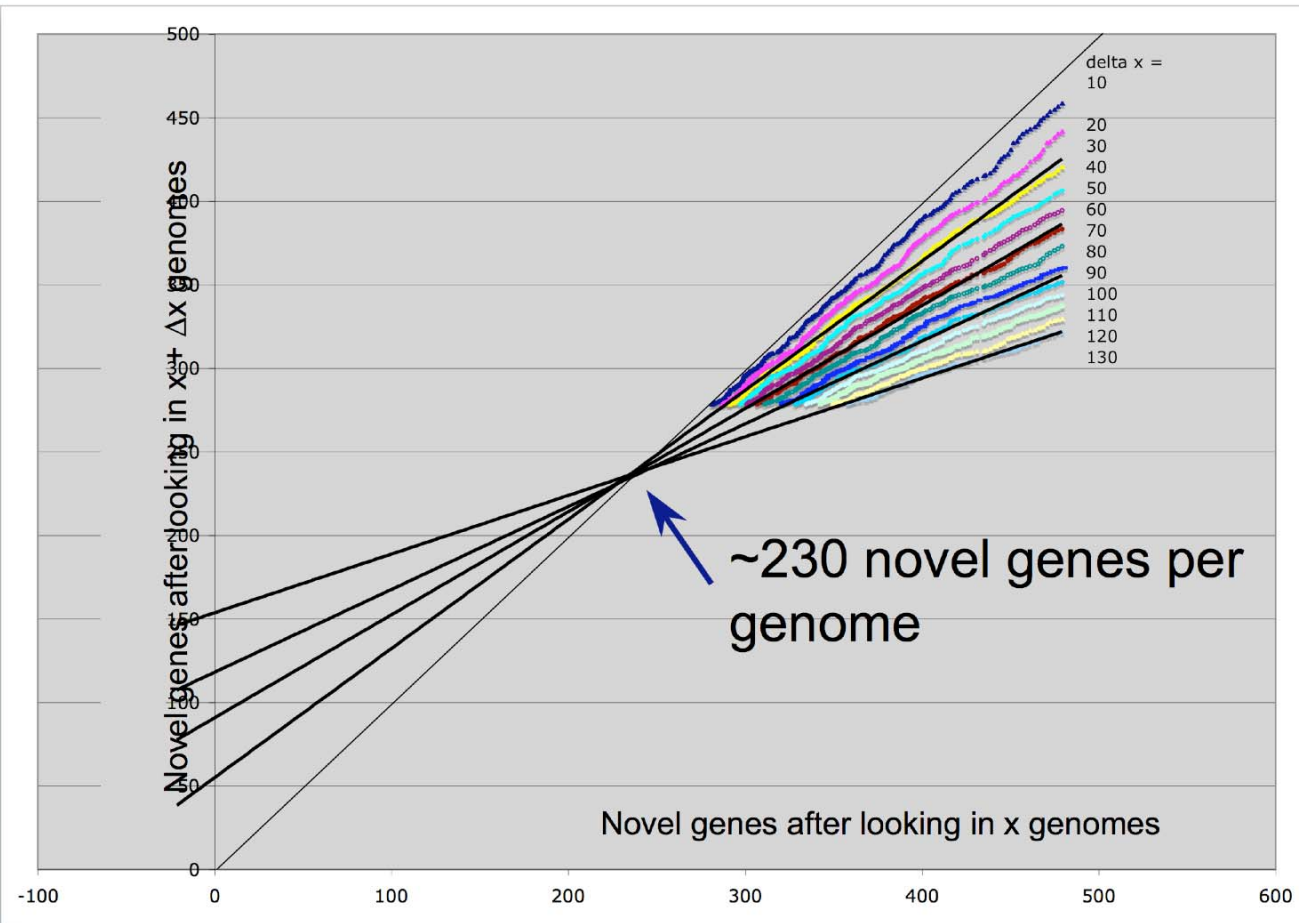


Approximate number of genes sampled in 200 bacterial genomes:

- 25,160 core genes
- 453,781 extended core genes
- 156,259 accessory genes

~150 gene families

Kézdy-Swinbourne Plot



If $f(x) = K + A \cdot \exp(-k \cdot x)$, then
 $f(x + \Delta x) = K + A \cdot \exp(-k \cdot (x + \Delta x))$.

Through elimination of A:

$$f(x + \Delta x) = \exp(-k \cdot \Delta x) \cdot f(x) + K'$$

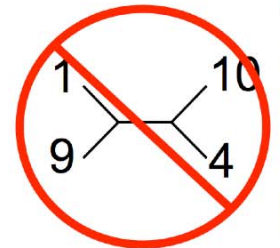
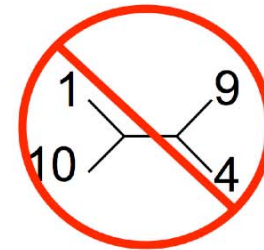
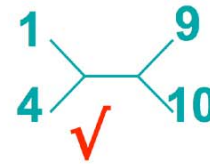
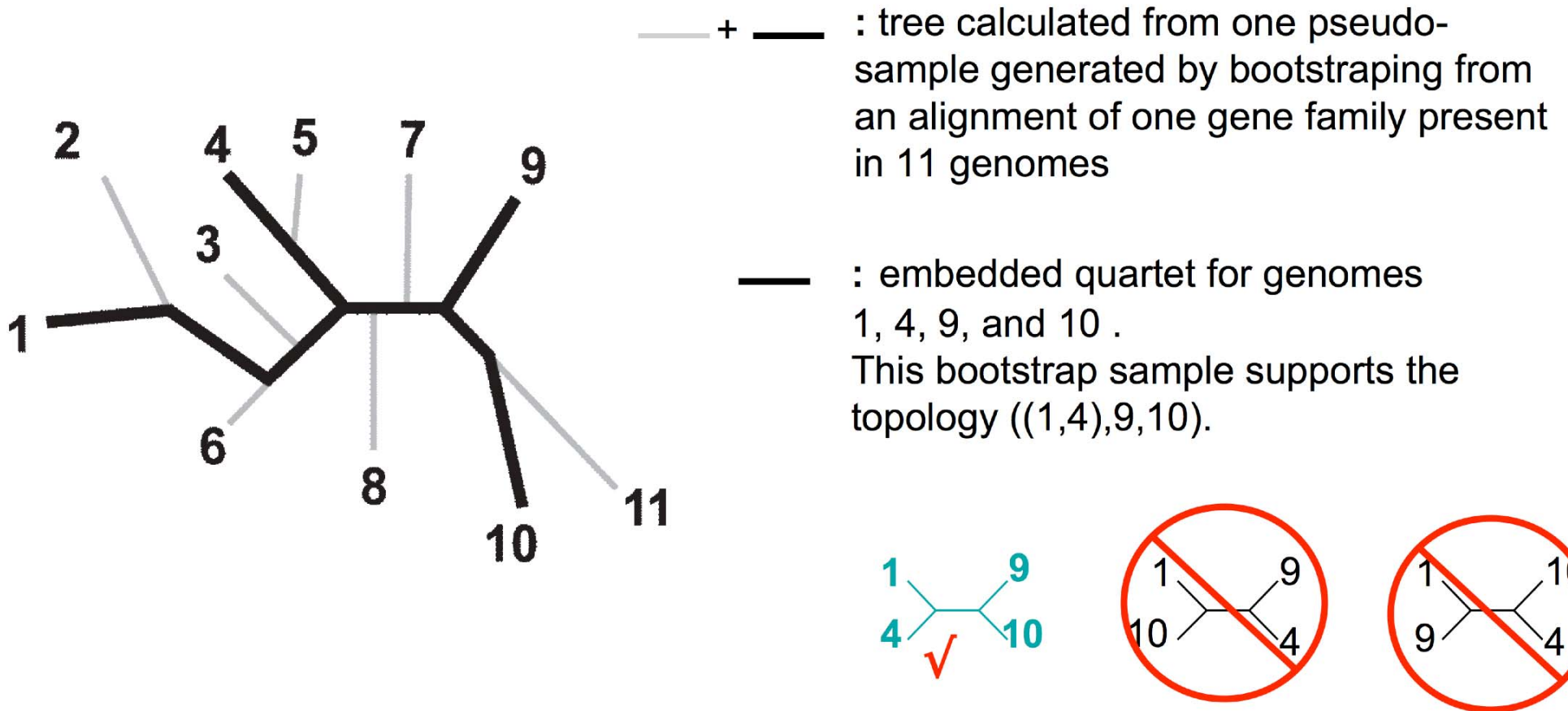
And for $x \rightarrow \infty$,

$$f(x) \rightarrow K, f(x + \Delta x) \rightarrow K$$

only values with $x \geq 80$ genomes were included

Even after comparing to a very large (infinite) number of bacterial genomes, on average, each new genome will contain about **230** genes that do not have a homolog in the other genomes.

Bootstrap support values for embedded quartets



Quartet spectral analyses of genomes iterates over three loops:

- Repeat for all bootstrap samples.
- Repeat for all possible embedded quartets.
- Repeat for all gene families.

Illustration of one component of a quartet spectral analyses

Summary of phylogenetic information for one genome quartet for all gene families

Total number of gene families containing the species quartet

Number of gene families supporting the same topology as the plurality (colored according to bootstrap support level)

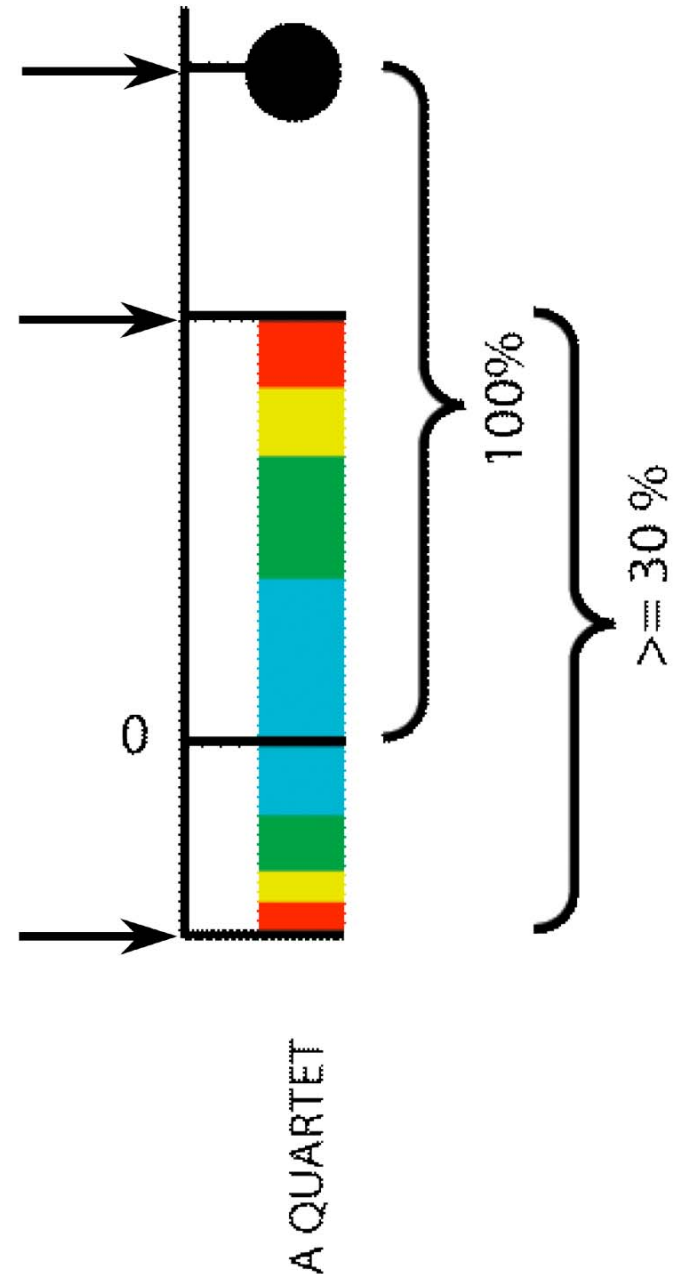
80%

90%

95%

99%

Number of gene families supporting one of the two alternative quartet topologies



Quartet Spectrum of 11 cyanobacterial genomes

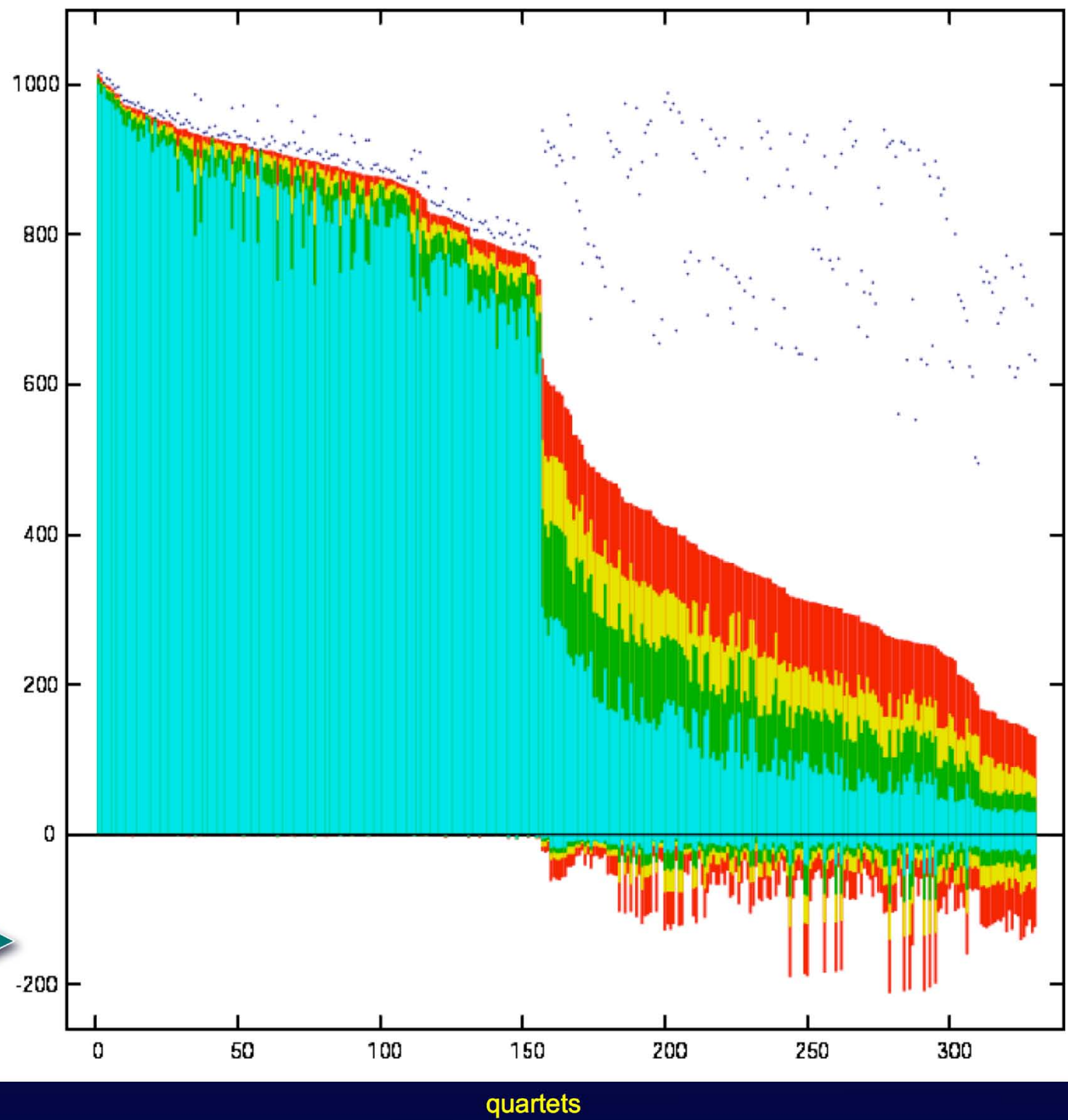
330 possible quartets

1128 datasets (present in all taxa + datasets with one or two taxa missing)

685 datasets show conflicts with plurality

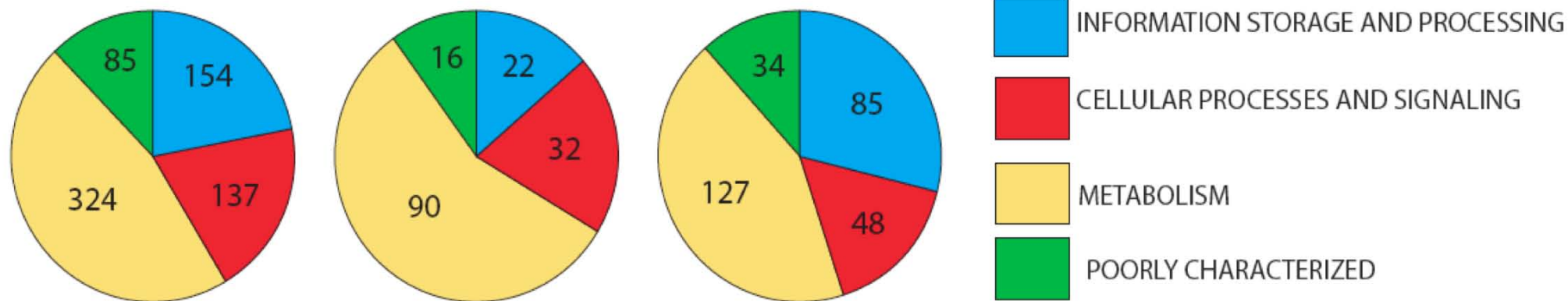


Number of datasets



Genes with orthologs outside the cyanobacterial phylum: Distribution among Functional Categories

(using COG db, release of March 2003)

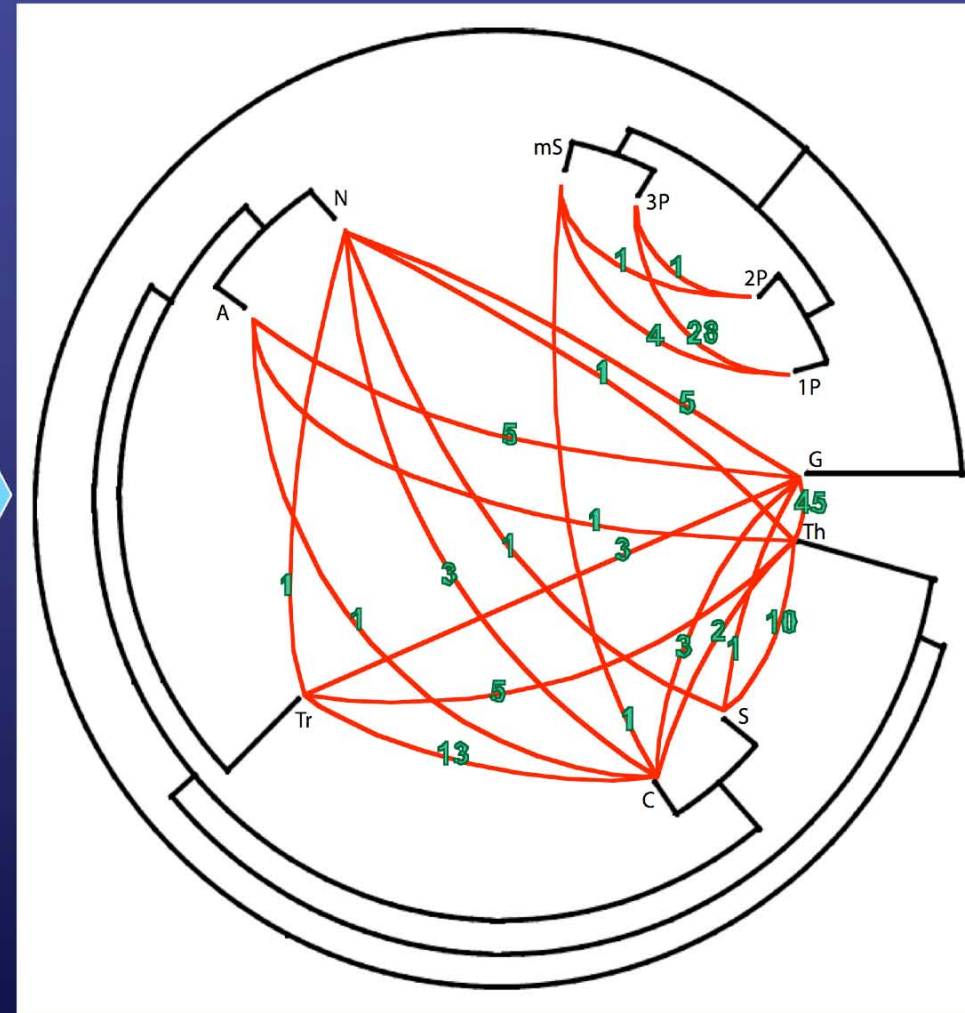
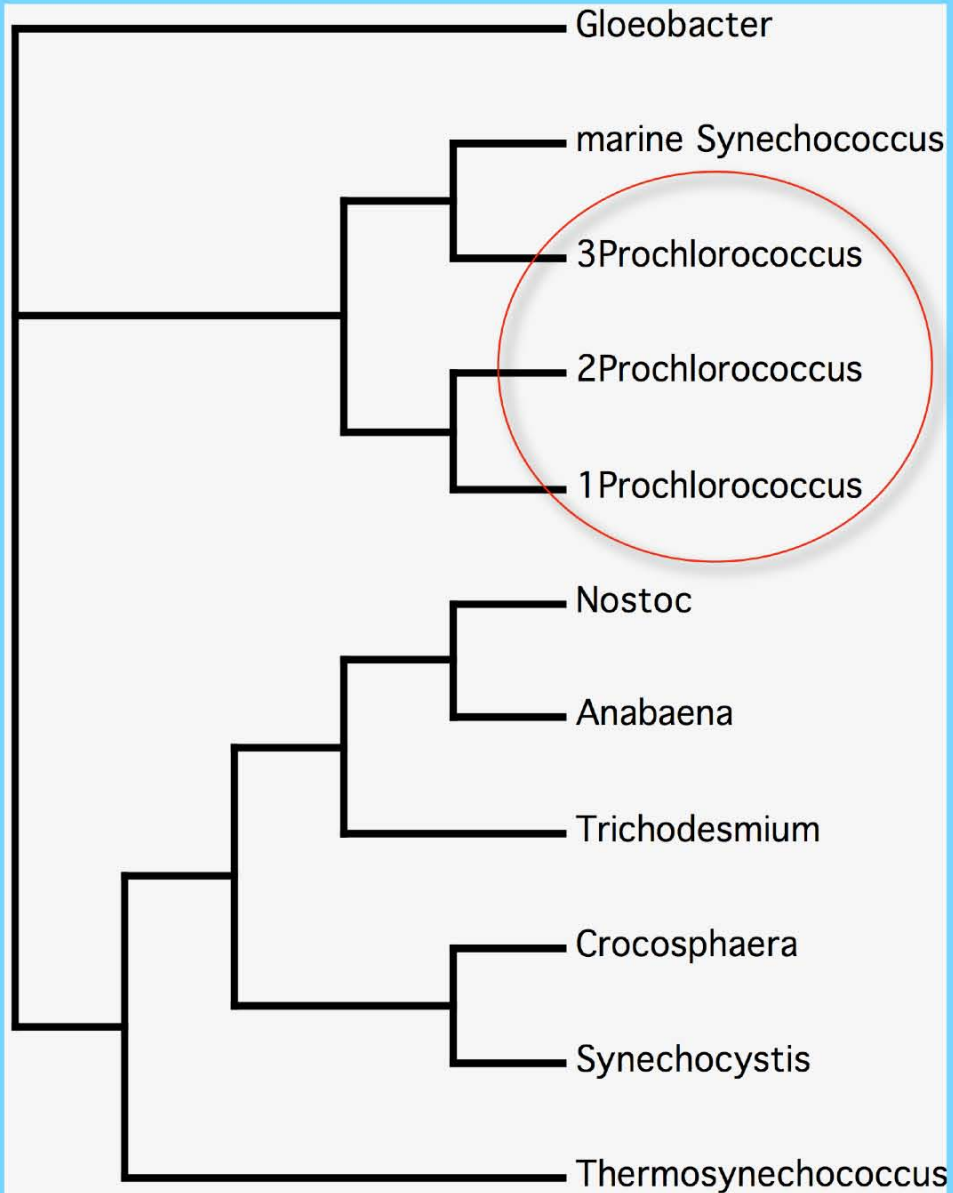


700 phylogenetically
useful extended
datasets

Cyanobacteria do not
form a coherent group
(160)

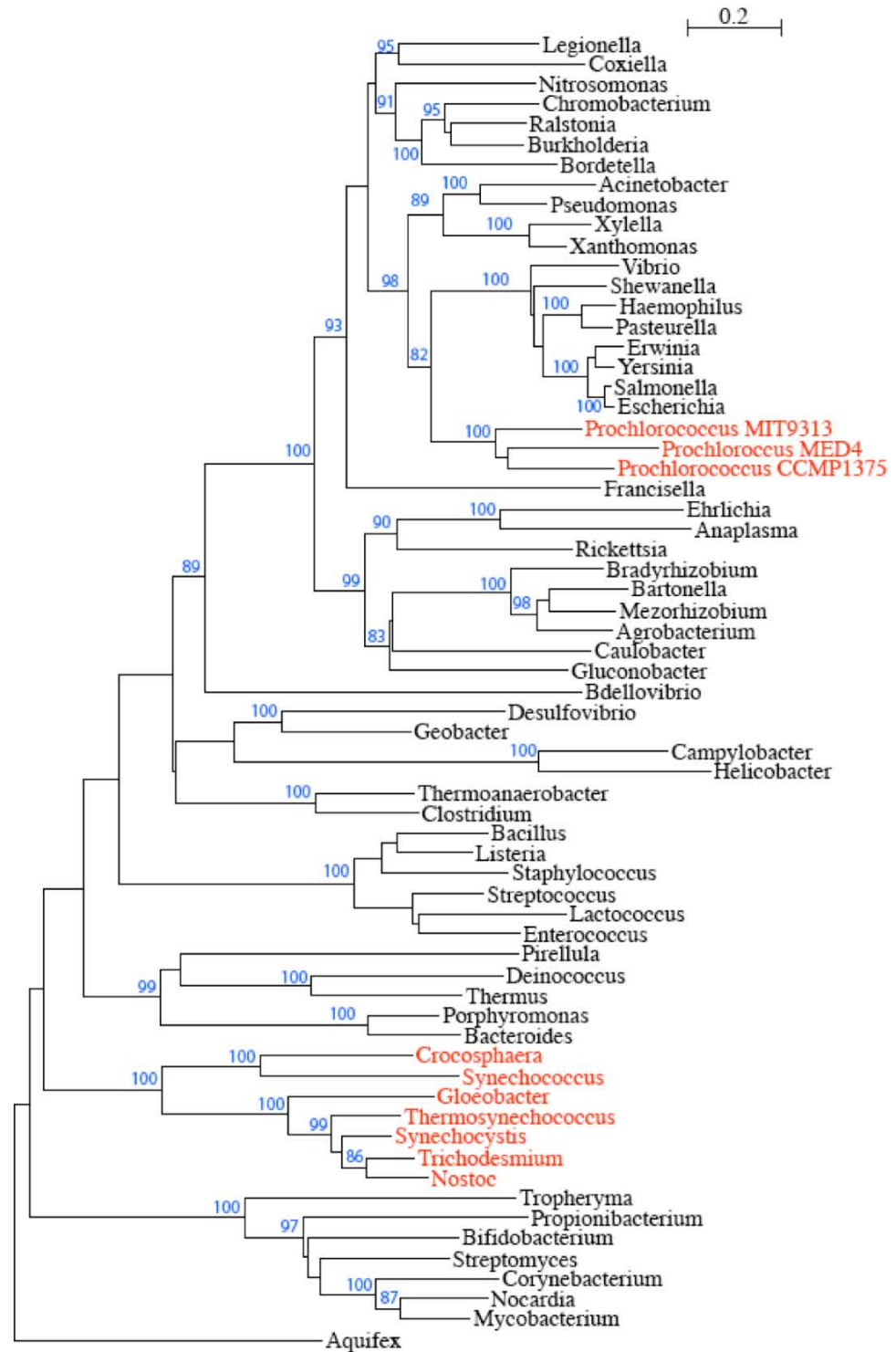
Cyanobacteria do form
a coherent group, but
conflict with plurality
(294)

Plurality Signal form Quartet Decomposition



Modified from Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., Papke, R.T. (2006)
Genome Res. 16(9):1099-108

Example of inter-phylum transfer: threonyl tRNA synthetase



Species evolution *versus* plurality consensus

⊕ In case of the marine *Synechococcus* and *Prochlorococcus* spp. the plurality consensus is unlikely to reflect organismal history.

⊕ This is probably due to frequent gene transfer mediated by phages e.g.:

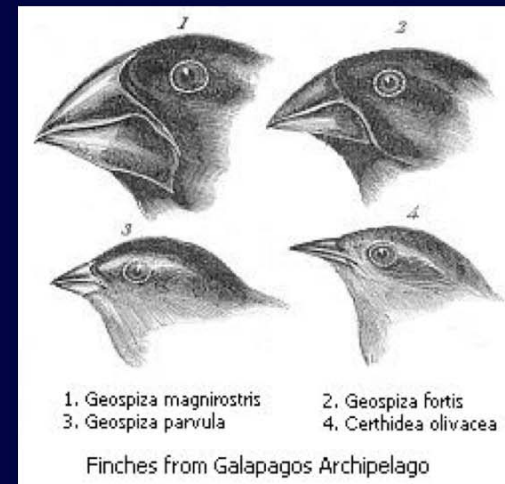
Transfer of photosynthesis genes to and from *Prochlorococcus* viruses

Debbie Lindell^{*†}, Matthew B. Sullivan^{†‡}, Zackary I. Johnson^{*}, Andrew C. Tolonen[‡], Forest Rohwer[§], and Sallie W. Chisholm^{*¶}

www.pnas.org/cgi/dol/10.1073/pnas.0401526101

PNAS | July 27, 2004 | vol. 101 | no. 30 | 11013–11018

⊕ These conflicting observations are not limited to prokaryotes. In incipient species of Darwin's finches frequent introgression can make some individuals characterized by morphology and mating behavior as belonging to the same species genetically more similar to a sister species (Grant et al. 2004 "*Convergent evolution of Darwin's finches caused by introgressive hybridization and selection*" *Evolution Int J Org Evolution* 58, 1588-1599).

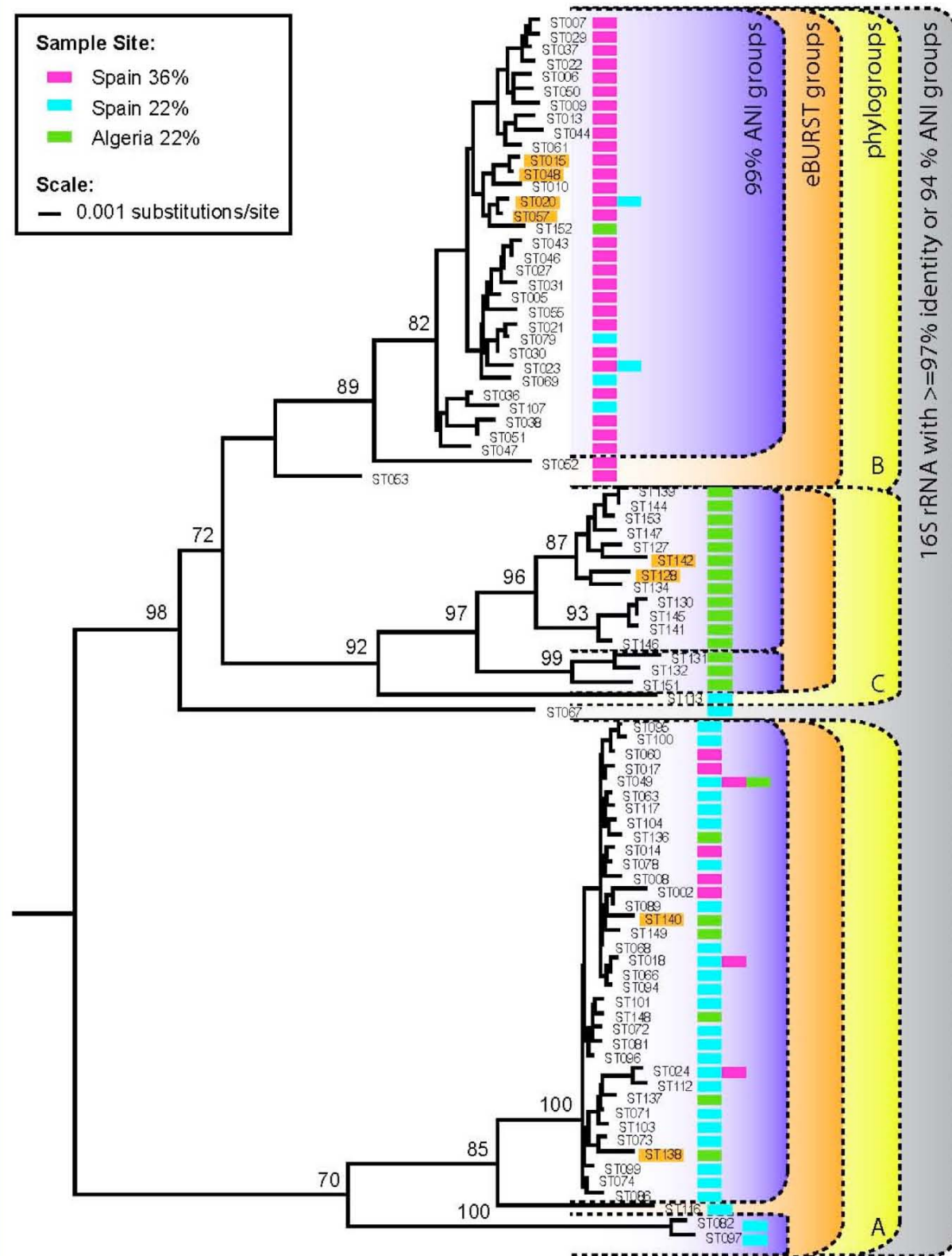


Time, Population Size, and Speciation

- Animals: example Galapagos finches, speciation in progress is accompanied with frequent introgression, only defining characteristics (song, beak shape) group according to species. A few million years post-hybridization barriers block frequent introgression and species and gene trees become congruent (e.g., marine iguanas).
- Bacteria: Retchless and Lawrence estimate that speciation between *E. coli* and *Salmonella* was a process that lasted over 70 million years. (Science. 2007 317:1093-6). (Species specific genes separate first, other genes, not located in the vicinity of specific genes separate later).
- How long would it take to have speciation in case of gigantic population sizes, as in the case of *Prochlorococcus* or *Halorubrum*?

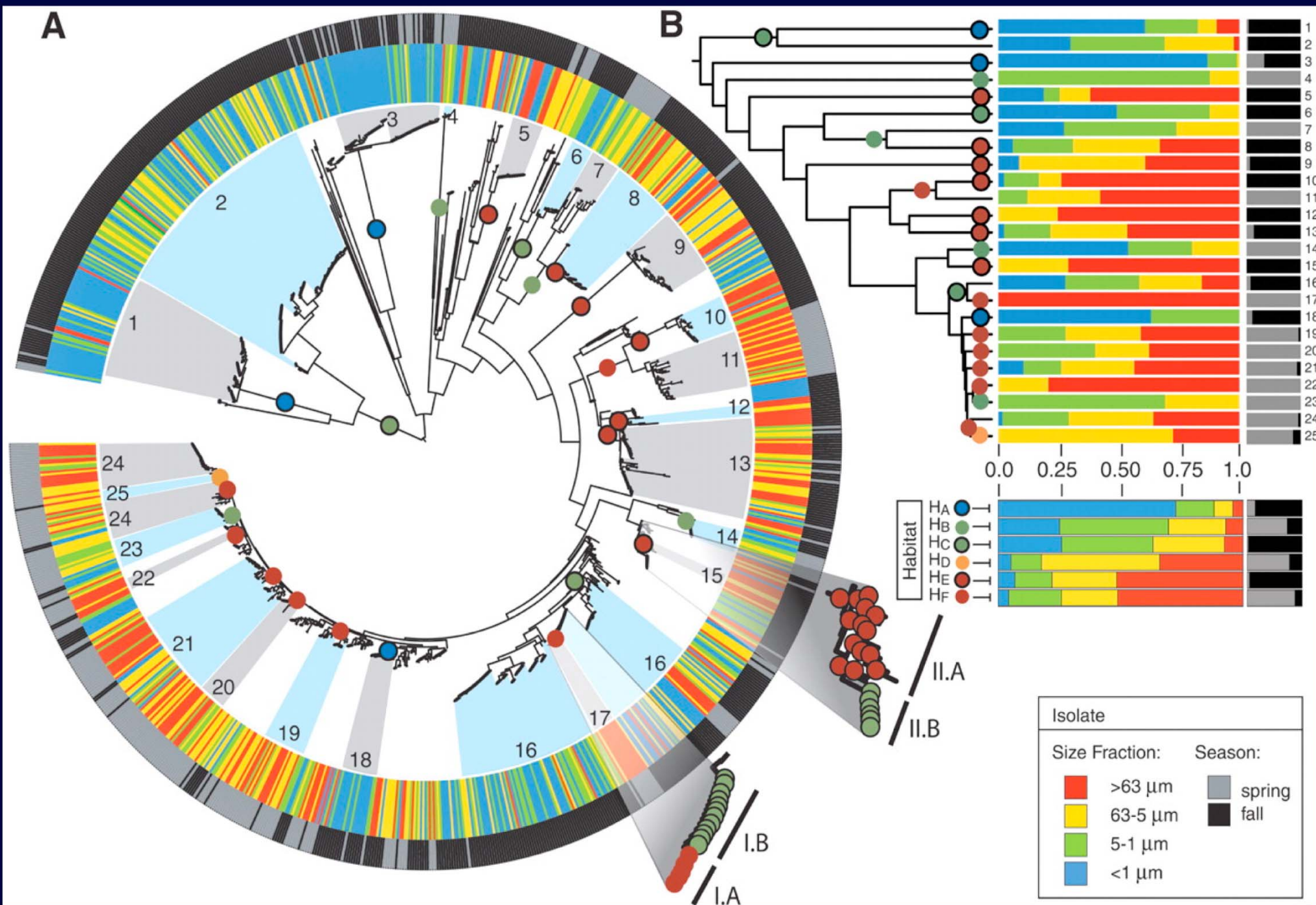
Multi Locus Sequence Analysis of *Halorubrum* strains

⊕ Recombination occurs within (selective sweep of one marker did not carry the other markers along) and between phylotypes. Nevertheless distinct types persist.



Papke et al.

PNAS 104, 14092-14097 2007



Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF.

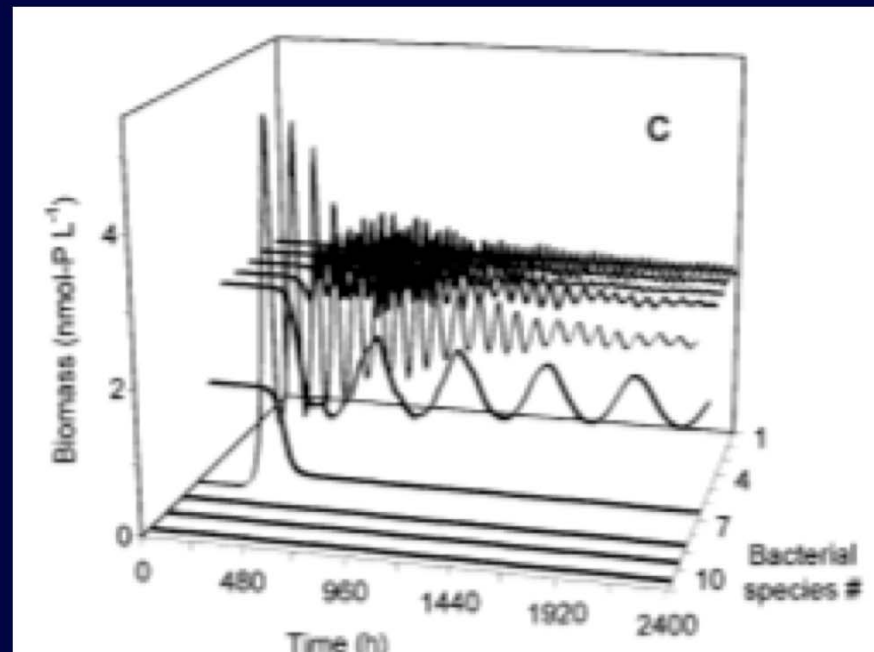
Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320:1081-5 (2008)

Does the partitioning into fuzzy phylotypes reflect ecological adaptation?

“Kill the Winner”

- ⊕ Phages (bacterial viruses) mainly attack the dominant bacterial type.
- ⊕ This type is thereby reduced and another type becomes dominant.
- ⊕ Phages evolve to kill this bacterial type
- ⊕

T. Frede Thingstad (2000):
Elements of a theory for the
mechanisms controlling
abundance, diversity, and
biogeochemical role of lytic
bacterial viruses in aquatic
systems. *Limnol. Oceanogr.*
45, 1320 -1 328



Biomass of individual bacterial species. // The simulation was started in the theoretical steady state discussed in the text. At time 100 h, the affinity of bacterial species No. 7 was changed from 0.065 to 0.11. // bacterial species No. 7 replaced bacterial species No. 6, and virus No. 7 replaced virus No. 5.

Does the partitioning into fuzzy phylotypes reflect ecological adaptation?

- ⊕ Phages are one of the main conduits for gene transfer.
- ⊕ Phage receptors frequently are also involved in conjugation (e.g. F-phage).
- ⊕ In many communities phages partition the bacterial population (temporal and spatial).
- ⊕ This partitioning based on phage susceptibility creates groups with limited between group genetic exchange.
- ⊕ The exchange groups created by phage predation might precede the adaptation to other environmental conditions.
- ⊕ Different ecotypes might evolve within these partitions.

Acknowledgements



Jinling Huang (ECU)
Olga Zhaxybayeva (Dalhousie)
Pascal Lapierre (UConn)
Greg Fournier (UConn)
Kristen Swithers (UConn)
Maria Poptsova (UConn)

Thane Papke (UConn)
Ken Noll (UConn)
Robert Charlesbois (Ottawa)
Jeff Townsend (UConn/Yale)
Jason Raymond (ASU)
Bob Blankenship (ASU)
Ford Doolittle (Dalhousie)
Jeffery Lawrence (Pittsburgh)
Gary Olsen (Urbana)

NSF Microbial Genetics,
NASA Exobiology & AISR Programs