Collapsing the 10,271 16S rRNA Based All-Species Living Tree

Guanghong Zuo[1], Xiaoyang Zhi[2], **Bailin Hao**[1,3]

[1]T-Life Research Center, Fudan University, Shanghai 200433, China

[2]Institute of Microbiology, Yunnan Univrsity, Kunming 650011, China

[3]Institute of Theoretical Physics, Academia Sinica, Beijing 100190, China

Two Web Servers:

- CVTree: Composition Vector Tree

  http://tlife.fudan.edu.cn/cvtree3/

- LVTree: All-Specis Living Tree

  http://tlife.fudan.edu.cn

Prokaryotes: the Most Successful Creatures on Earth

- 10^{30} living cells estimated (Whitman 1998)
- 10^{10} species estimated (Whitman 2011)
- Less than 10^{4} species described so far
- No less than 10^{5} cultures in labs worldwide awaiting description and VP
- Yet culturable species are less than 1% of the total number of species
- New techniques facilitate discovery of novel species, but “Trade Association” regulations hinder the progress

Molecular Phylogeny/Taxonomy

- Inspiring idea of Zukerkandl and Pauling (1965): protein sequences as documents of evolution
- Molecular clocks
- Homologous proteins
- Difficulties for prokaryotes:
- Prokaryotes = Archaea (古菌) + Bacteria (细菌)

Molecular Phylogeny/Taxonomy of Prokaryotes

- Carl R. Woese and coworkers suggested to use small subunit (SSU) RNAs in the late 1970s
- Fox GE, … CR Woese (1977)
- Carl R. Woese, 1928 – 2012

Carl R. Woese (1928 – 30 Dec 2012)

- “Microbiology the science without a past” (Woese, a subtitle in vol. 3 of *The Prokaryotes*, 2000)
- “Genome sequencing has come of age, and genomics will become central to microbiology’s future. It may appear at the moment that the human genome is the main focus and primary goal of genome sequencing, but do not be deceived. The real justification in the long run is microbial genomics.” (Woese, 1998)

16S rRNA (rDNA) Sequence

- Calculated to 1542 nts as of *E. coli*

- Bacteria:
  568 conserved positions
  974 variable positions
- Archaea:
  571 conserved positions
  971 variable positions

- See Bergey’s Manual vol.1 p.52

SSU RNA Databases

- RDP initiated by Woese
- Rel. 11.2 (March 2014)
- 2,929,439 sequences

- SILVA
- Rel. 115 (Aug 2013)
- SSU Parc: 3,808,884
  SSU Ref: 1,426,414
  SSU Ref NR: 479,726

The largest database of a single gene.

The All-Species Living Tree Project

- P. Yarza et al. SAM 31(2008): 241 – 250
- P. Yarza et al. SAM 33(2010): 291– 299
- R. Munoz et al. SAM 34(2011): 169 – 170 (LTPs104)
- P. Yarza et al. SAM 36(2013): 69 – 73

The All-Species Living Tree Project

- LTPs95 (2008)
- LTPs111 (Feb 2013) 361+9340=9701

  Based on 361 Archaea and 9340 Bacteria seqs
- LTPs115 (Mar 2014) 366+9905=10271

  Based on 366 Archaea and 9905 Bacteria seqs

“Demonstration” of LTPs115

- Scrolling up and down the 105 pages of PDF LTPs115_SSU_tree.pdf
- A huge tree with 100271 leaves
- Enlarging the fonts
- … … …

Phylogenetic Definition of The Major Eubacterial Phyla

- Carl Woese and Co., SAM, 1985
- 400 16S rRNA sequences in total
- 10 phyla described

Taxonomic Coverage of 10271 Seqs

The original lineage

Domains:	2
Phyla:	31
Classes:	67
Orders:	142
Families:	341
Genera:	2053
Species:	10067

After taxonomic revisions

Domains:	2
Phyla:	32
Classes:	68
Orders:	144
Families:	345
Genera:	2049
Species:	10068

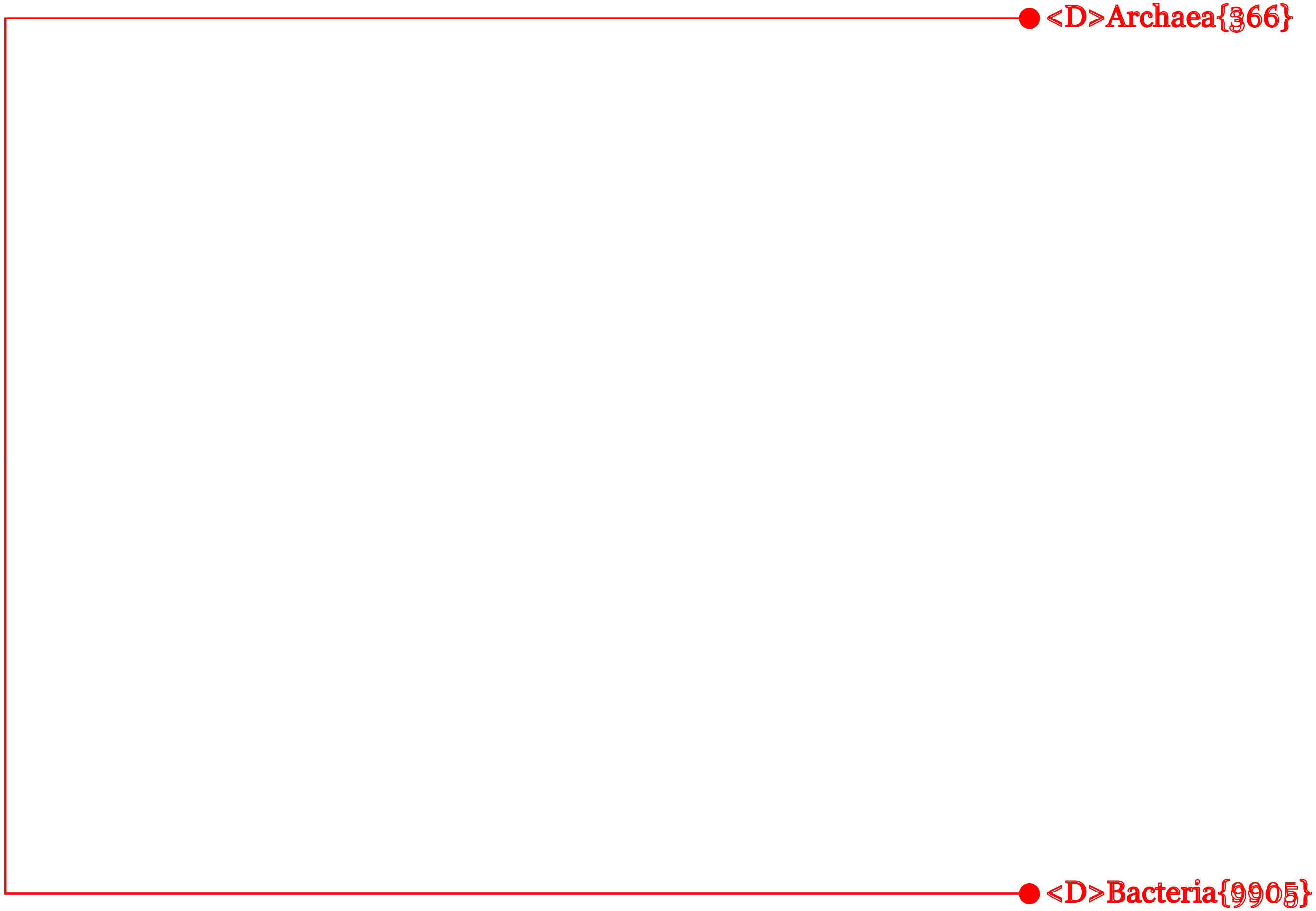
How to Comprehend a Big Tree?

- Collapsing monophyletic branches at all taxonomic ranks from <D>Domain <P>Phylum <C>Class <O>Order <F>Family <G>Genus <S>Species
- Expanding a collapsed branch
- Report taxon monophyly with statistics
- Trial taxnomic revisions
- Re-collapsing and reporting the result

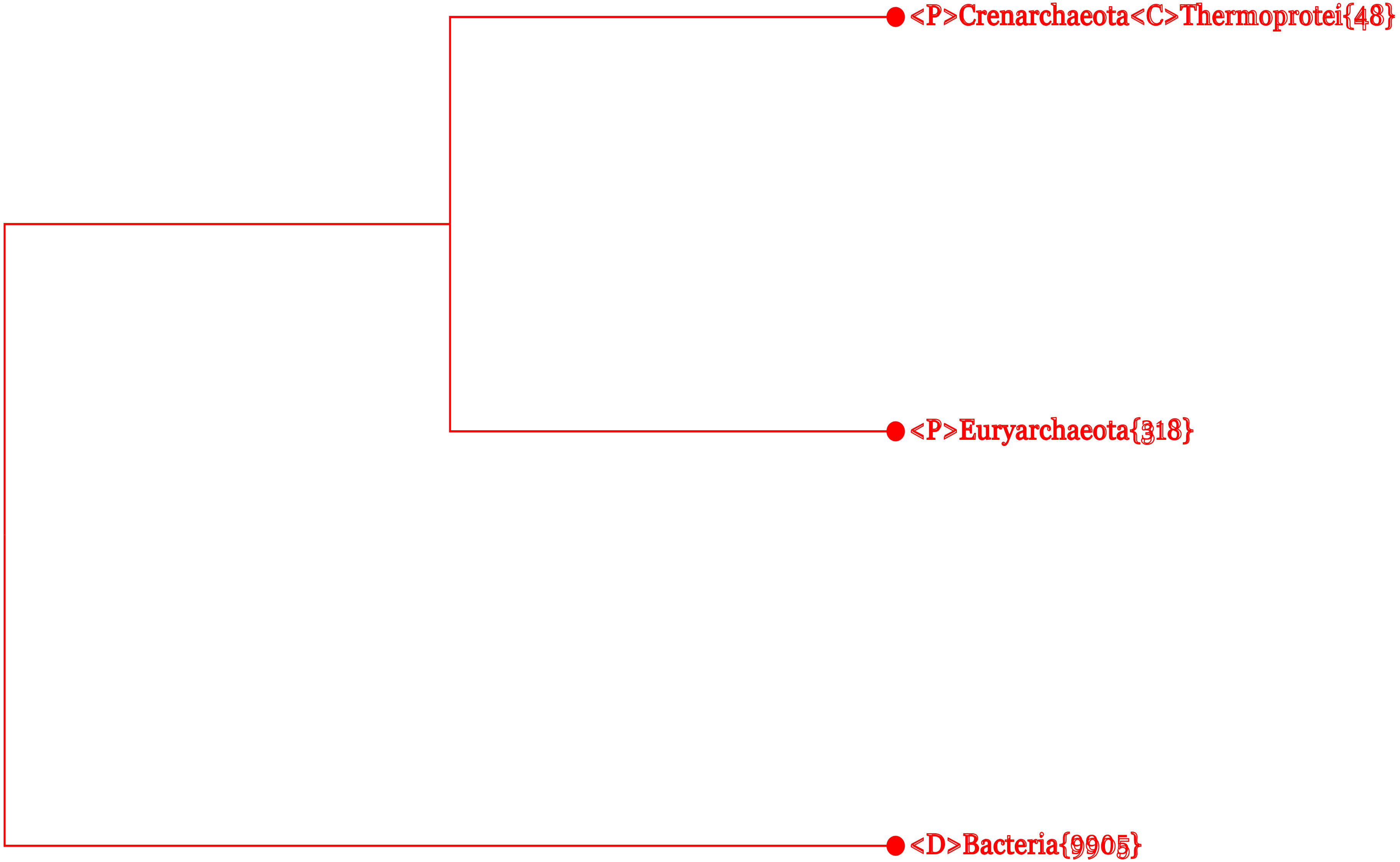
An Interactive LVTree Viewer

- http://tlife.fudan.edu.cn/lvtree/
- Options to pick up LTPs111 or LTPs115
- Let us look at LTPs115 of March 2014 (actually released on 11 April 2014)
- To be published soon

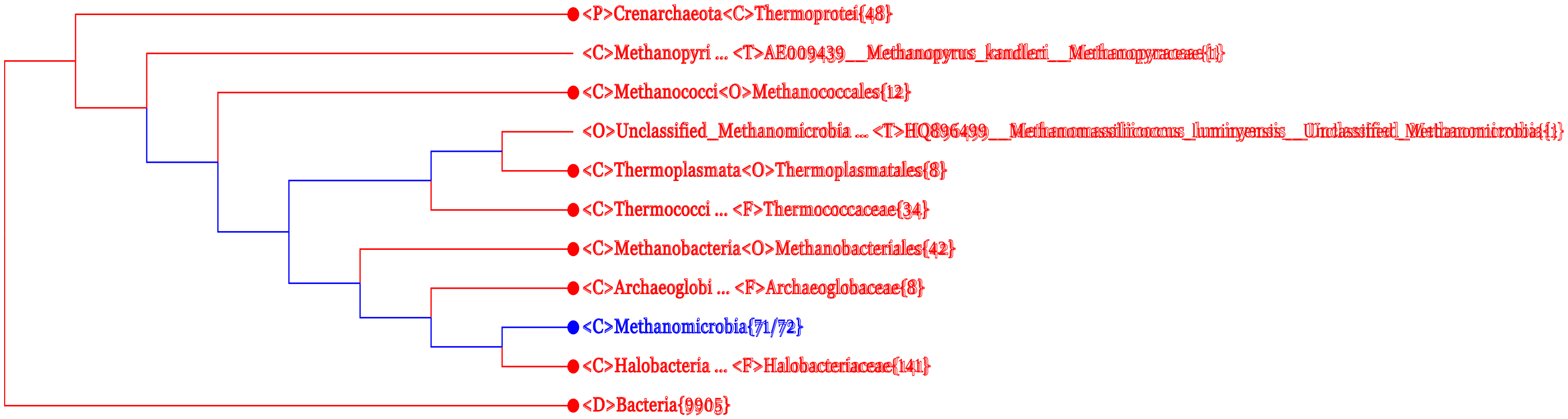
The 1st Screen with 2 Leaves

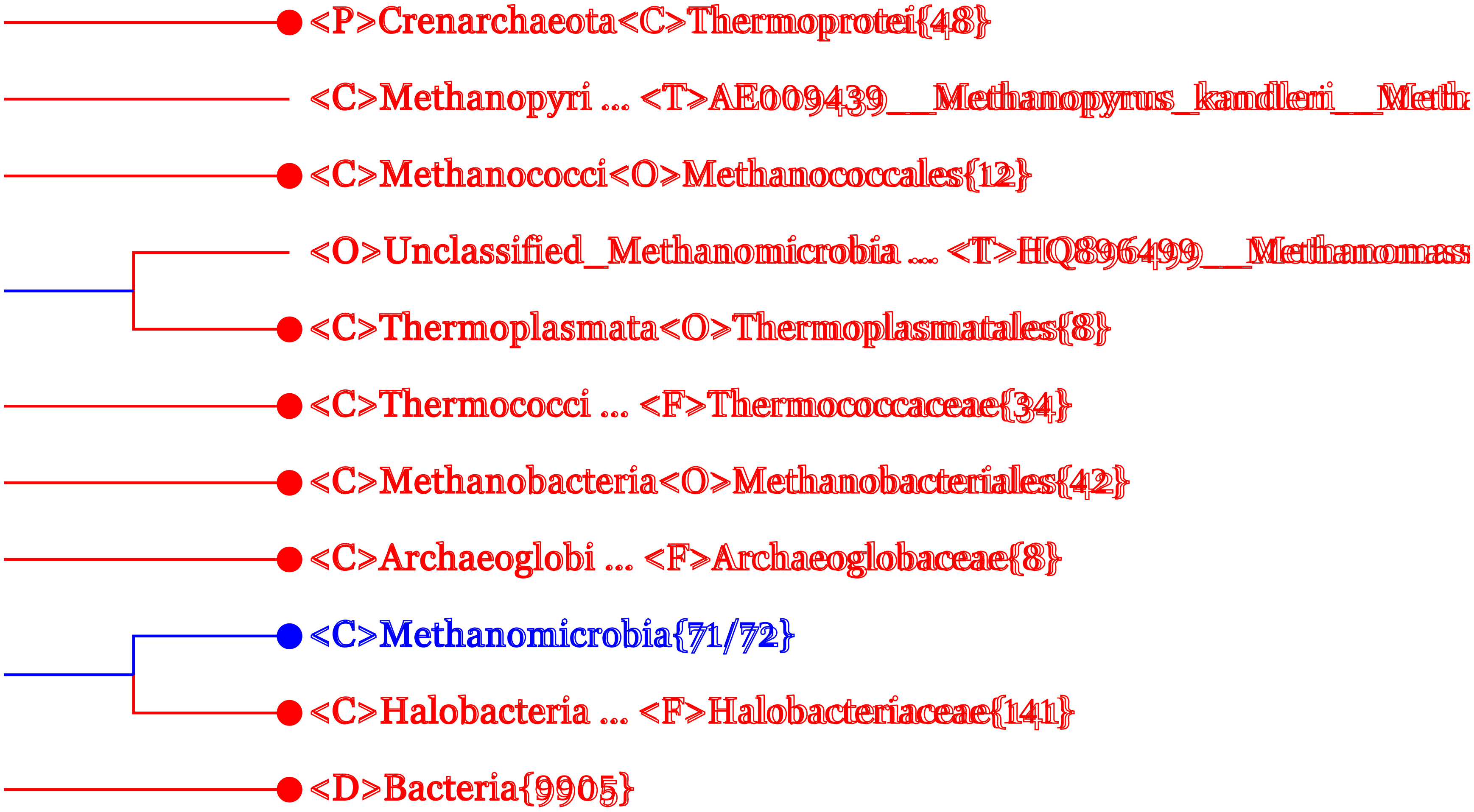


Expand the Archaea Branch

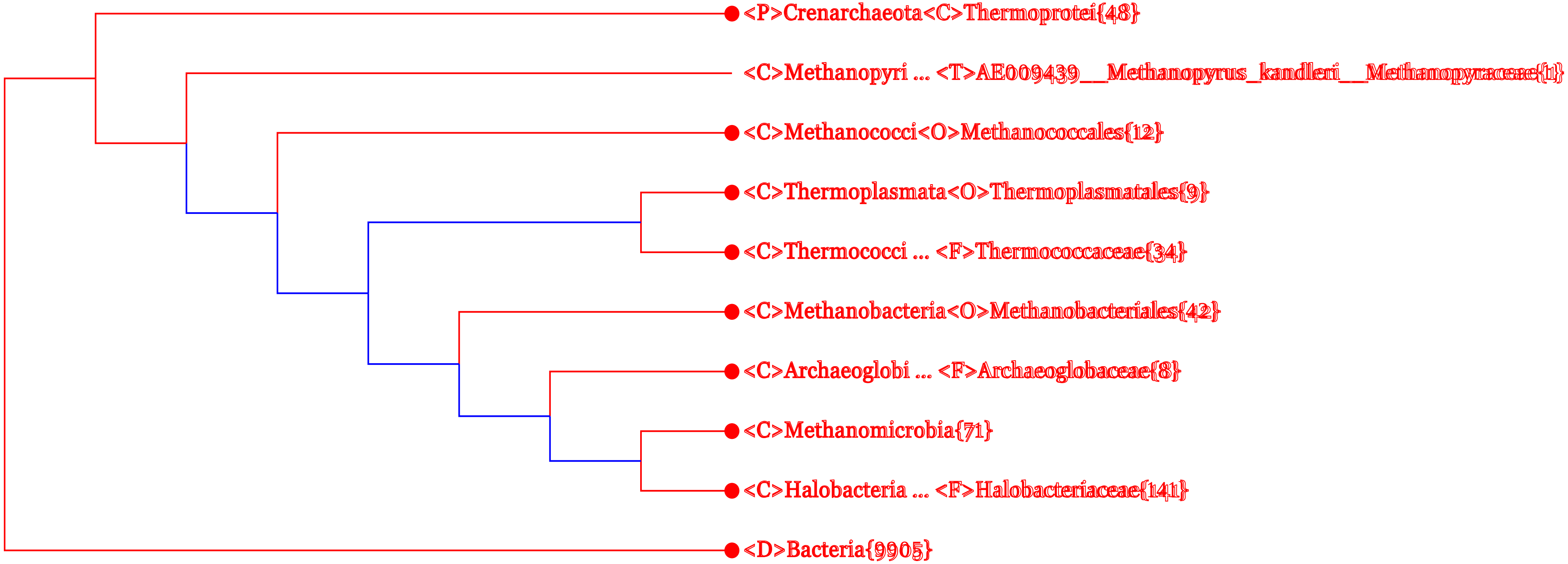


Phylum Euryarchaeota: <C> + <O>




<P>Crenarchaeota<C>Thermoprotei{48}
<C>Methanopyri ... <T>AE009439__Methanopyrus_kandleri__Meth
<C>Methanococci<O>Methanococcales{12}
<O>Unclassified_Methanomicrobia .... <T>HQ896499__Methanomass
<C>Thermoplasmata<O>Thermoplasmatales{8}
<C>Thermococci ... <F>Thermococcaceae{34}
<C>Methanobacteria<O>Methanobacteriales{42}
<C>Archaeoglobi ... <F>Archaeoglobaceae{8}
<C>Methanomicrobia{71/72}
<C>Halobacteria ... <F>Halobacteriaceae{141}
<D>Bacteria{9905}

Euryarchaeota: all <C>



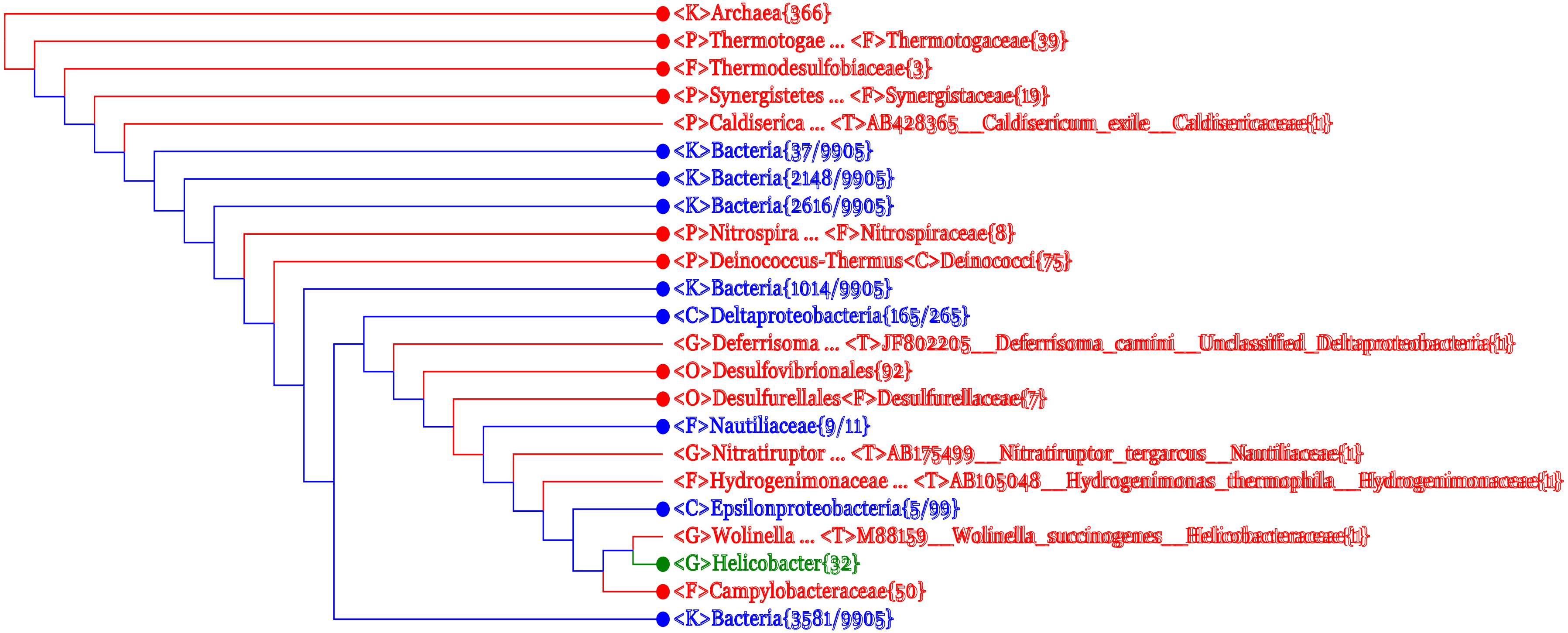
- Note: all trees shown in this lecture PPT are full trees based on 10271 16S rRNA sequences, collapsed or expanded appropriately in order to highlight a certain aspect of the phylogeny

- The easiest way to get to an entry consists in using the Search Query function

Query: Helicobacter

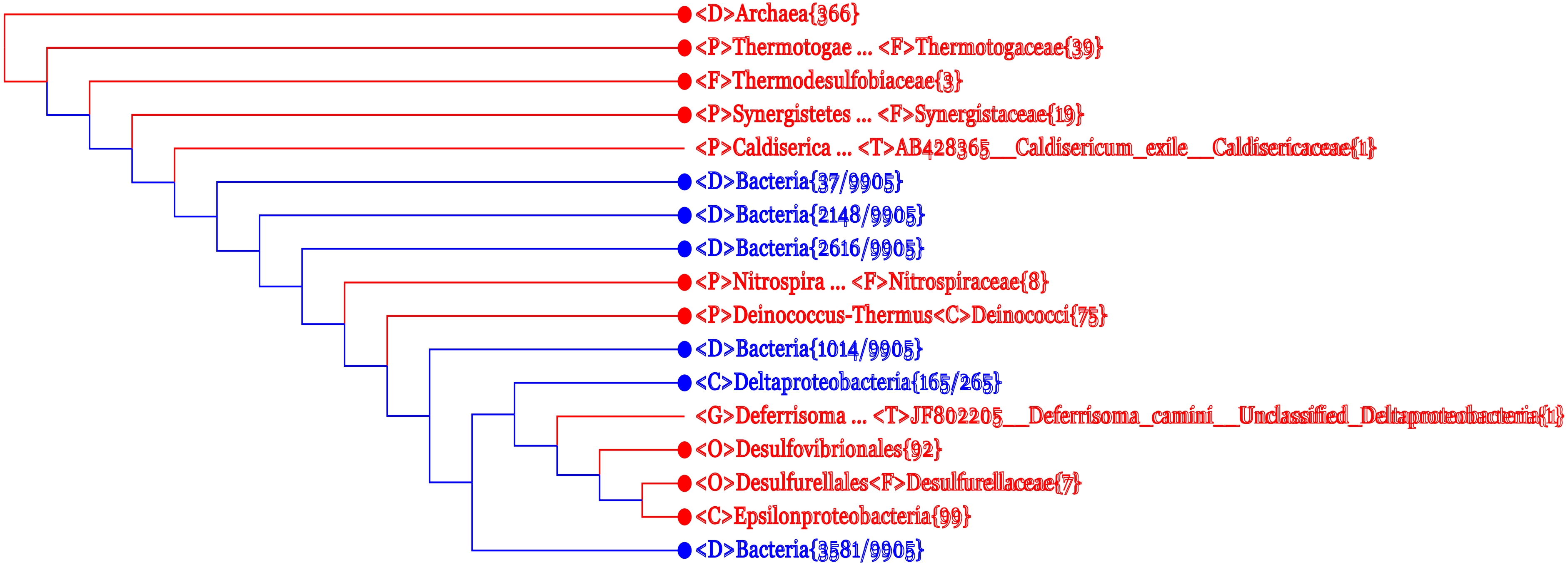
- <P>Proteobacteria
- <C>Epsilonproteobacteria
- <O>Campylobacterales
- <F>Helicobacteraceae
- <G>Helicobacter
- <S>Helicobacter pylori
- <T>Helicobacter pylori SouthAfrica7

Search Query: Helicobacter



<K>Archaea{366}
<P>Thermotogae ... <F>Thermotogaceae{39}
<F>Thermodesulfobiaceae{3}
<P>Synergistetes ... <F>Synergistaceae{19}
<P>Caldiserica ... <T>AB428365__Caldisericum_exile__Caldisericaceae{1}
<K>Bacteria{37/9905}
<K>Bacteria{2148/9905}
<K>Bacteria{2616/9905}
<P>Nitrospira ... <F>Nitrospiraceae{8}
<P>Deinococcus-Thermus<C>Deinococci{75}
<K>Bacteria{1014/9905}
<C>Deltaproteobacteria{165/265}
<G>Deferrisoma ... <T>JF802205__Deferrisoma_camini__Unclassified_Deltaproteobacteria{1}
<O>Desulfovibrionales{92}
<O>Desulfurellales<F>Desulfurellaceae{7}
<F>Nautiliaceae{9/11}
<G>Nitratiruptor ... <T>AB175499__Nitratiruptor_tergarcus__Nautiliaceae{1}
<F>Hydrogenimonaceae ... <T>AB105048__Hydrogenimonas_thermophila__Hydrogenimonaceae{1}
<C>Epsilonproteobacteria{5/99}
<G>Wolinella ... <T>M88159__Wolinella_succinogenes__Helicobacteraceae{1}
<G>Helicobacter{32}
<F>Campylobacteraceae{50}
<K>Bacteria{3581/9905}

<C>Epsilonproteobacteria{99}



<D>Archaea{366}
<P>Thermotogae ... <F>Thermotogaceae{39}
<F>Thermodesulfobiaceae{3}
<P>Synergistetes ... <F>Synergistaceae{19}
<P>Caldiserica ... <T>AB428365__Caldisericum_exile__Caldisericaceae{1}
<D>Bacteria{37/9905}
<D>Bacteria{2148/9905}
<D>Bacteria{2616/9905}
<P>Nitrospira ... <F>Nitrospiraceae{8}
<P>Deinococcus-Thermus<C>Deinococci{75}
<D>Bacteria{1014/9905}
<C>Deltaproteobacteria{165/265}
<G>Deferrisoma ... <T>JF802205__Deferrisoma_camini__Unclassified_Deltaproteobacteria{1}
<O>Desulfovibrionales{92}
<O>Desulfurellales<F>Desulfurellaceae{7}
<C>Epsilonproteobacteria{99}
<D>Bacteria{3581/9905}

Monophyly of Phyla

- Of the 31 phyla represented by the 10271 16S rRNA sequences 22 are monophyletic, including 3 phyla represented by a single sequence.

- There are 9 phyla that are non-monophyletic.

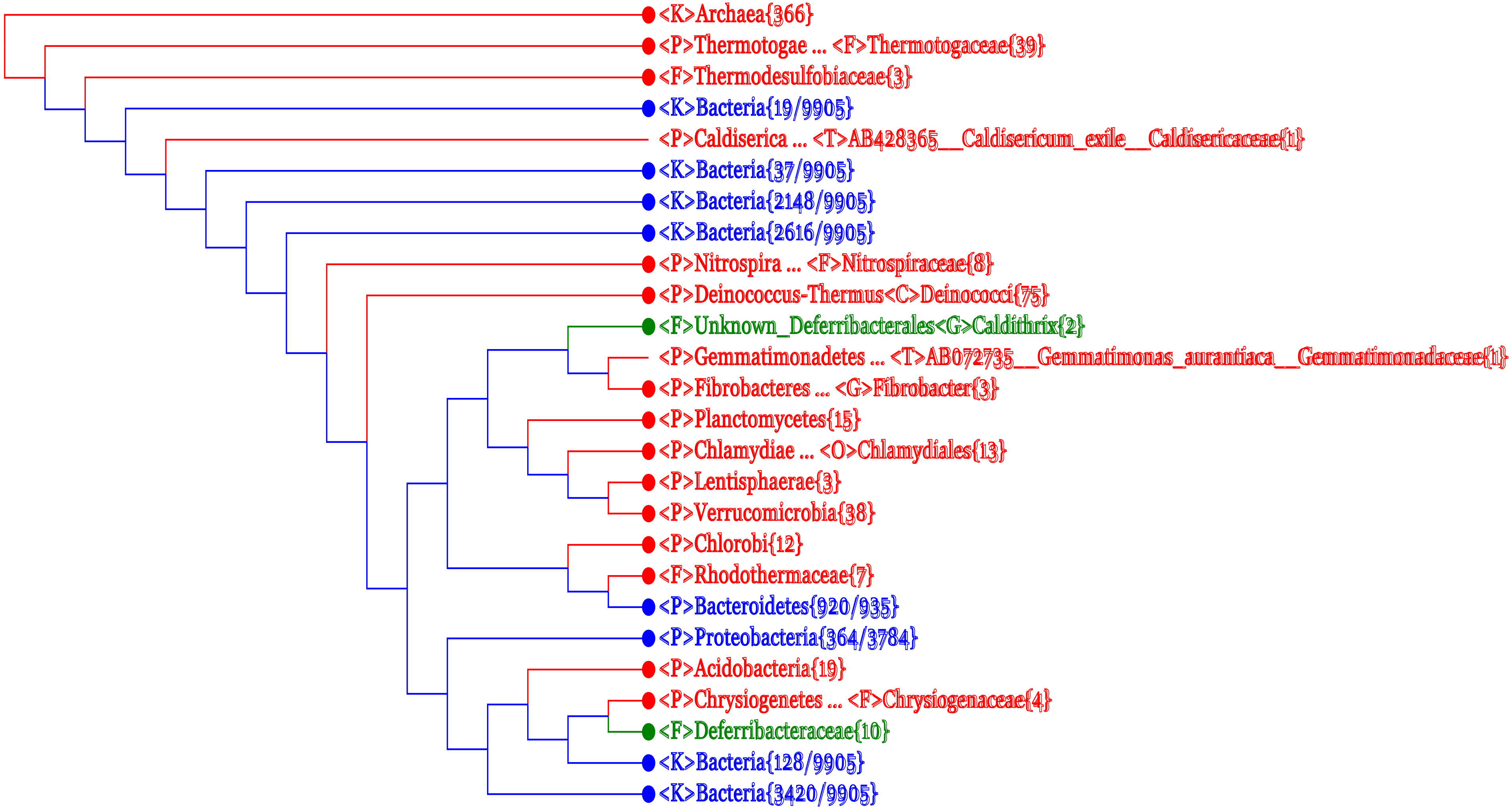
- 24 possible revisions would bring 7 out of these 9 phyla to monophyletic.

The 9 Non-Monophyletic Phyla

Phylum	#(Sequences)	Monophyletic	Non-Monophyletic
Actinobacteria	2597	2593	
Aquificae	28	29	
Bacteroidetes	935	927	
Deferribacteres	12	10	
Firmicutes	1940		1953
Fusobacteria	38	38	
Proteobacteria	3785		3784
Synergistetes	18	19	
Tenericutes	188	186	

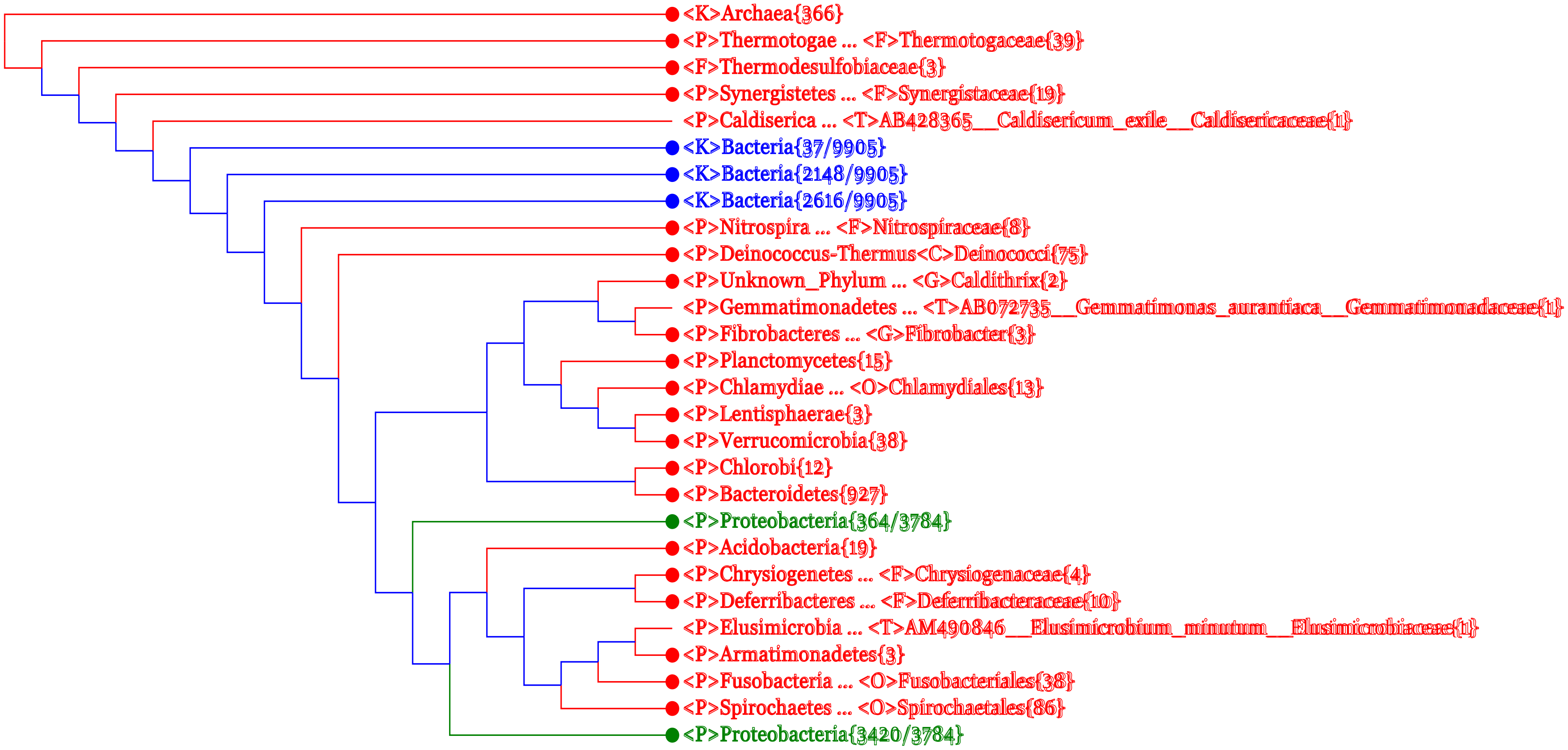
Unclassified phylum 2

Caldithrix not in <P>Deferribacteres



<K>Archaea{366}
<P>Thermotogae ... <F>Thermotogaceae{39}
<F>Thermodesulfobiaceae{3}
<K>Bacteria{19/9905}
<P>Caldiserica ... <T>AB428365__Caldisericum_exile__Caldisericaceae{1}
<K>Bacteria{37/9905}
<K>Bacteria{2148/9905}
<K>Bacteria{2616/9905}
<P>Nitrospira ... <F>Nitrospiraceae{8}
<P>Deinococcus-Thermus<C>Deinococci{75}
<F>Unknown_Deferribacterales<G>Caldithrix{2}
<P>Gemmatimonadetes ... <T>AB072735__Gemmatimonas_aurantiaca__Gemmatimonadaceae{1}
<P>Fibrobacteres ... <G>Fibrobacter{3}
<P>Planctomycetes{15}
<P>Chlamydiae ... <O>Chlamydiales{13}
<P>Lentisphaerae{3}
<P>Verrucomicrobia{38}
<P>Chlorobi{12}
<F>Rhodothermaceae{7}
<P>Bacteroidetes{920/935}
<P>Proteobacteria{364/3784}
<P>Acidobacteria{19}
<P>Chrysiogenetes ... <F>Chrysiogenaceae{4}
<F>Deferribacteraceae{10}
<K>Bacteria{128/9905}
<K>Bacteria{3420/9905}

<P>Proteobacteria{3784}



Proteobacteria: 3784 = 364 + 3420

Delta	265
Epsilon	99
Total	364

Alpha	1209
Beta	529
Gamma	1681
Zeta	1
Total	3420

<P>Firmicutes{1943}

- The only bacterial phylum with many taxonomic problems
- Historically, many phyla have been taken out from *Firmicutes*, e.g., the *Actinobacteria*, the *Tenericutes*
- It calls for further study

- Problem with *Cyanobacteria*


<D>Archaea{366}
<P>Thermotogae ... <F>Thermotogaceae{39}
<F>Thermodesulfobiaceae{3}
<P>Synergistetes ... <F>Synergistaceae{19}
<P>Caldiserica ... <T>AB428365__Caldisericum_exile__Caldisericaceae{1}
<D>Bacteria{37/9905}
<G>Thermaerobacter{5}
<G>Caldicellulosiruptor{8}
<G>Thermanaeromonas ... <T>AB062280__Thermanaeromonas_toyohensis__Thermoanaerobacteraceae{1}
<P>Dictyoglomi ... <G>Dictyoglomus{2}
<C>Clostridia{8/618}
<F>Thermoanaerobacteraceae{25/44}
<O>Thermoanaerobacterales{12/70}
<C>Clostridia{7/618}
<C>Clostridia{4/618}
<P>Firmicutes{5/1953}
<C>Clostridia{35/618}
<F>Peptococcaceae{2/58}
<O>Clostridiales{25/518}
<G>Thermincola{2}
<C>Unclassified_Firmicutes ... <T>HM038000__Vampirovibrio_chlorellavorus__Bdellovibrionaceae{1}
<P>Cyanobacteria{10}
<O>Halanaerobiales{27}
<C>Negativicutes<O>Selenomonadales{75}
<P>Firmicutes{443/1953}
<O>Bacillales{4/819}
<F>Thermoactinomycetaceae{20}
<G>Caldalkalibacillus{2}
<G>Thermicanus ... <T>AJ242495__Thermicanus_aegyptius__Unclassified_Bacillales{1}
<G>Microaerobacter ... <T>FN552009__Microaerobacter_geothermalis__Bacillaceae{1}
<O>Bacillales{187/819}
<G>Vulcanibacillus ... <T>AM050346__Vulcanibacillus_modesticaldus__Bacillaceae{1}
<G>Bacillus{2/186}
<O>Bacillales{29/819}
<F>Sporolactobacillaceae{3/12}
<F>Bacillaceae{136/396}
<S>Bacillus_taeanensis<T>AY603978__Bacillus_taeanensis__Bacillaceae{1}
<O>Bacillales{49/819}
<O>Bacillales{183/819}
<F>Bacillaceae{54/396}
<O>Bacillales{13/819}
<O>Bacillales{39/819}
<F>Staphylococcaceae{87}
<G>Gemella{7}
<O>Haloplasmatales ... <T>EF999972__Haloplasma_contractile__Haloplasmataceae{1}
<G>Turicibacter ... <T>AF349724__Turicibacter_sanguinis__Erysipelotrichaceae{1}
<P>Firmicutes{10/1953}
<P>Firmicutes{14/1953}
<P>Tenericutes<C>Mollicutes{186}
<O>Lactobacillales{420/423}
<D>Bacteria{7658/9905}

Warning: Taxonomy an Emotional Science

- Play with taxonomic revisions for yourselves

- Comply to the International Code for bacterial nomenclature if you want to publish something

- Polyphasic taxonomy/phylogeny

Limitations of 16S rRNA Analysis

- Lateral gene transfer of ribosomal operons: not known in Nature, but realized in laboratory
- Species with 2 or more different (>6%) active 16S rRNA sequences
- No or weak resolution at species level and below

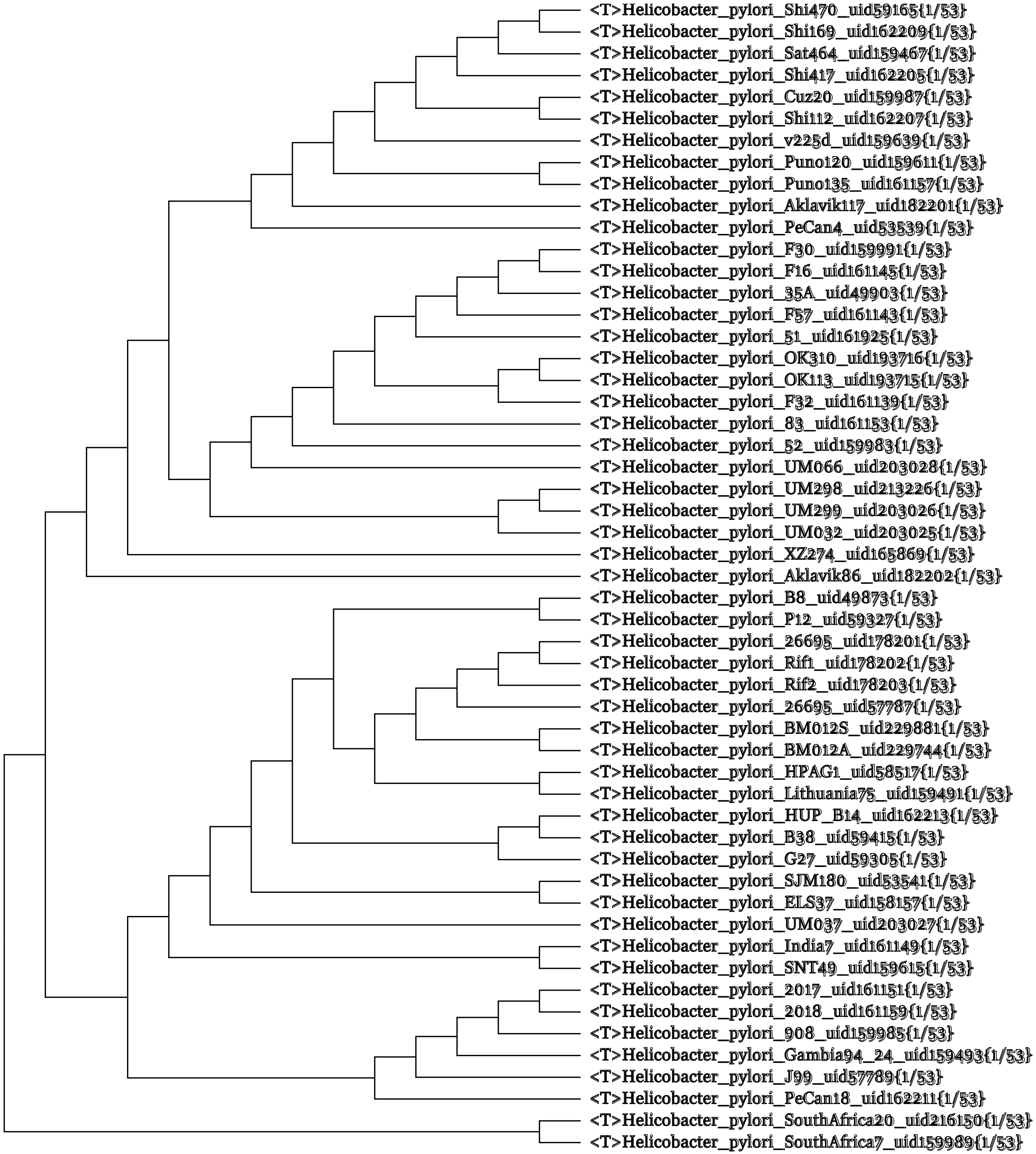
Whole-Genome-Based Phylogeny

- Necessity:
- Feasibility: more than 16,000 sequenced bacterial genomes; the number grows almost everyday
- Alignment-free
- Direct comparison with taxonomy
- The CVTree Web Server at

  http://tlife.fudan.edu.cn/cvtree3/

Bacteria and Archaea Genome Resources

- NCBI: ftp.ncbi.nih.gov/genomes/Bacteria
- GOLD: www.genomesonline.org
- PATRIC: patricbrc.org
- EzBioCloud: www.ezbiocloud.net


<T>Helicobacter_pylori_Shi470_uid59165{1/53}
<T>Helicobacter_pylori_Shi169_uid162209{1/53}
<T>Helicobacter_pylori_Sat464_uid159467{1/53}
<T>Helicobacter_pylori_Shi417_uid162205{1/53}
<T>Helicobacter_pylori_Cuz20_uid159987{1/53}
<T>Helicobacter_pylori_Shi112_uid162207{1/53}
<T>Helicobacter_pylori_v225d_uid159639{1/53}
<T>Helicobacter_pylori_Puno120_uid159611{1/53}
<T>Helicobacter_pylori_Puno135_uid161157{1/53}
<T>Helicobacter_pylori_Aklavik117_uid182201{1/53}
<T>Helicobacter_pylori_PeCan4_uid53539{1/53}
<T>Helicobacter_pylori_F30_uid159991{1/53}
<T>Helicobacter_pylori_F16_uid161145{1/53}
<T>Helicobacter_pylori_35A_uid49903{1/53}
<T>Helicobacter_pylori_F57_uid161143{1/53}
<T>Helicobacter_pylori_51_uid161925{1/53}
<T>Helicobacter_pylori_OK310_uid193716{1/53}
<T>Helicobacter_pylori_OK113_uid193715{1/53}
<T>Helicobacter_pylori_F32_uid161139{1/53}
<T>Helicobacter_pylori_83_uid161153{1/53}
<T>Helicobacter_pylori_52_uid159983{1/53}
<T>Helicobacter_pylori_UM066_uid203028{1/53}
<T>Helicobacter_pylori_UM298_uid213226{1/53}
<T>Helicobacter_pylori_UM299_uid203026{1/53}
<T>Helicobacter_pylori_UM032_uid203025{1/53}
<T>Helicobacter_pylori_XZ274_uid165869{1/53}
<T>Helicobacter_pylori_Aklavik86_uid182202{1/53}
<T>Helicobacter_pylori_B8_uid49873{1/53}
<T>Helicobacter_pylori_P12_uid59327{1/53}
<T>Helicobacter_pylori_26695_uid178201{1/53}
<T>Helicobacter_pylori_Rif1_uid178202{1/53}
<T>Helicobacter_pylori_Rif2_uid178203{1/53}
<T>Helicobacter_pylori_26695_uid57787{1/53}
<T>Helicobacter_pylori_BM012S_uid229881{1/53}
<T>Helicobacter_pylori_BM012A_uid229744{1/53}
<T>Helicobacter_pylori_HPAG1_uid58517{1/53}
<T>Helicobacter_pylori_Lithuania75_uid159491{1/53}
<T>Helicobacter_pylori_HUP_B14_uid162213{1/53}
<T>Helicobacter_pylori_B38_uid59415{1/53}
<T>Helicobacter_pylori_G27_uid59305{1/53}
<T>Helicobacter_pylori_SJM180_uid53541{1/53}
<T>Helicobacter_pylori_ELS37_uid158157{1/53}
<T>Helicobacter_pylori_UM037_uid203027{1/53}
<T>Helicobacter_pylori_India7_uid161149{1/53}
<T>Helicobacter_pylori_SNT49_uid159615{1/53}
<T>Helicobacter_pylori_2017_uid161151{1/53}
<T>Helicobacter_pylori_2018_uid161159{1/53}
<T>Helicobacter_pylori_908_uid159985{1/53}
<T>Helicobacter_pylori_Gambia94_24_uid159493{1/53}
<T>Helicobacter_pylori_J99_uid57789{1/53}
<T>Helicobacter_pylori_PeCan18_uid162211{1/53}
<T>Helicobacter_pylori_SouthAfrica20_uid216150{1/53}
<T>Helicobacter_pylori_SouthAfrica7_uid159989{1/53}

Tripartite Comparison

- The All-Species Living Tree
- The CVtree Web Server
- The 2nd Edition of the Bergey’s Manual of Systematic Bacteriology

Bergey's Manuals (1923-2012)

- Bergey's Manual of Deterministic Bacteriology: 1st Ed. 1923; 9th Ed. 1994
- Bergey's Manual of Systematic Bacteriology: 1st Ed. 1980s, 4 volumes
- Bergey's Manual of Systematic Bacteriology: 2nd Ed. 2001-2012, 5 volumes, 7600 pp.
- The end of classic bacterial taxonomy.
- The 4th Edition will be electronic.

Taxonomy at Large

- Different questions at different scales
- Phylogeny/taxonomy across 30-40 and more phyla
- Topology versus branch lengths

Bacterial Taxonomy

- At the beginning of 20th century: in a chaotic state
- Great achievement of 20th century: the international bacterial code of nomenclatuer
- 16S rRNA sequence analysis in 1970s
- Whole-genome approaches since 21st century

- Good hope to surpass other branches of the Tree of Life (animals, plants, fungi, … …)

The Endless End

THANKS FOR ATTENTION