Molecular taxonomy: the bacterial species concept and polyphasic taxonomy revisited

Peter Vandamme

UNIVERSITEIT GENT

Symposium: **”Fra grundforskning til klinisk anvendelse”**
September 12, Lyngby , Denmark

Definition of a bacterial species

- "The unit of classification is a **coherent group of like individuals**, called a species. The term is difficult to define with precision because a species is not a definite entity, but a taxonomic concept" (Breed et al., 1957)

- "A collection of strains that share many features in common and differ considerably from other strains" (Staley and Krieg, 1984)

Ad Hoc Committees on Reconciliation of Approaches to Bacterial Systematics

(Wayne et al. 1987)

- **The complete genome should be the reference standard to determine phylogeny and taxonomy**
- The phylogenetic definition of a species generally would include strains with at least 60 - 70% DNA-DNA hybridisation
- Phenotypic characteristics should agree with this definition

Polyphasic species definition

- The bacterial species appears to be an assemblage of isolates originating from a common ancestor population in which a steady generation of genetic diversity resulted in clones with different degrees of recombination and characterized by:
  - a certain degree of phenotypic consistency
  - a significant degree of DNA-DNA hybridization
  - over 97% of 16S rRNA gene sequence similarity

Vandamme et al., 1996. Microb. Rev. **60:**407-438

American Academy of Microbiology, March 2007

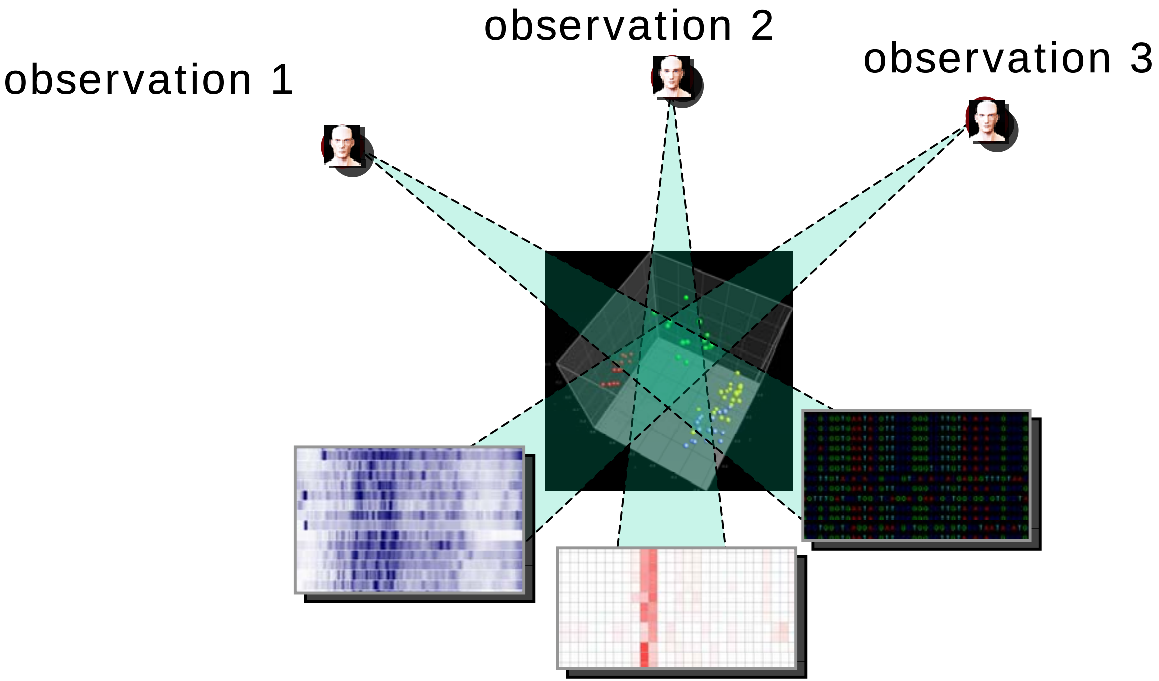
- "Reconciling Microbial Systematics and Genomics“

- http://www.asm.org/Academy/index.asp?bid=49252

Polyphasic taxonomy- Reconciling Microbial Systematics and Genomics

- “ Species are defined by pragmatic, arbitrary, and sometimes artificial methods based on 16S rRNA gene sequences, DNA-DNA hybridisation, morphology, physiology and chemotaxonomy (...)”

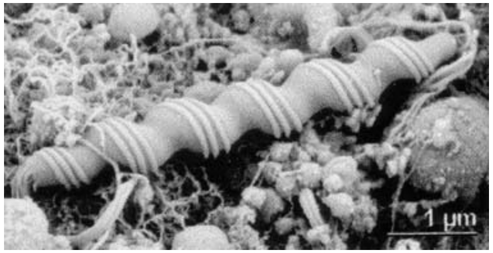
- The system is functional in many ways

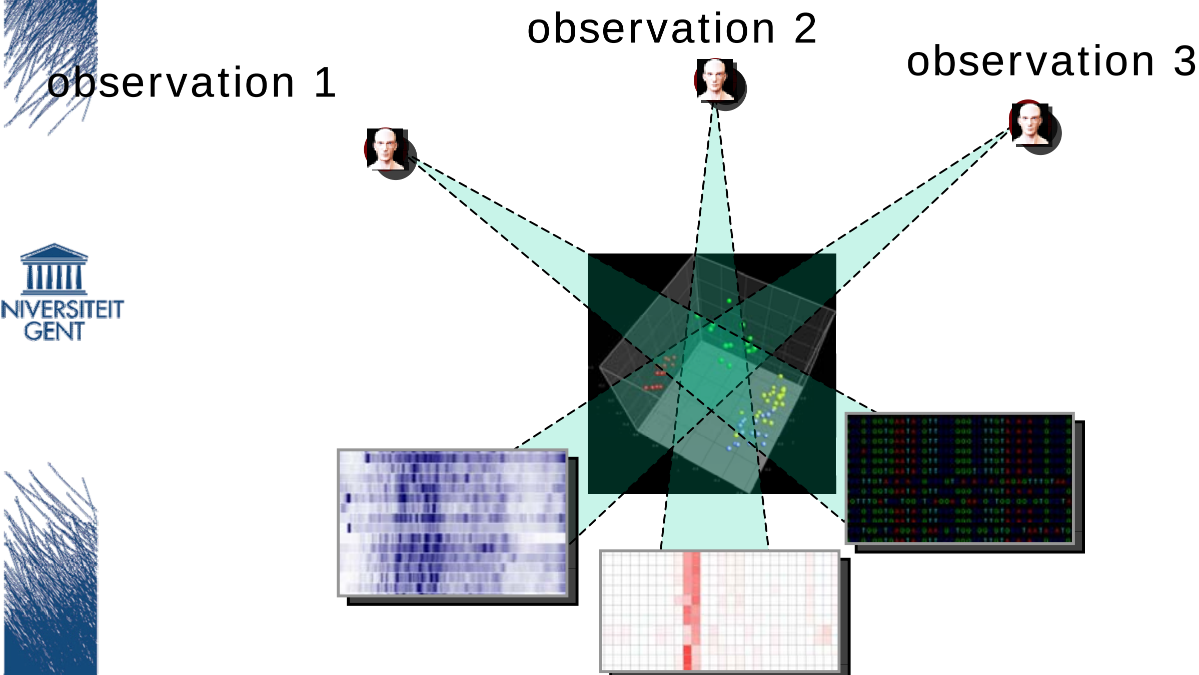


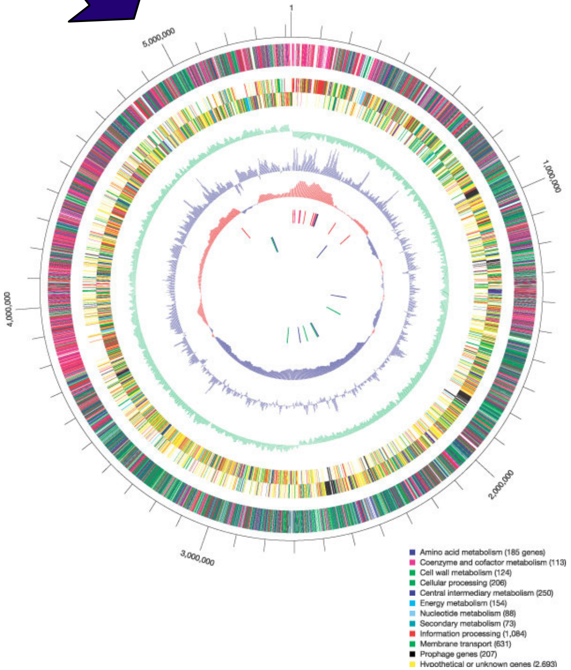
The polyphasic approach (AAM report)

- “ The system is somewhat functional but inadequate:
  - Conflicts between phenotypic and phylogenetic classifications
  - Limited means for classifying uncultured microbes
  - Current species often lack cohesiveness (...) “
- Lack of throughput capacity

?
Polyphasic → Genomic taxonomy



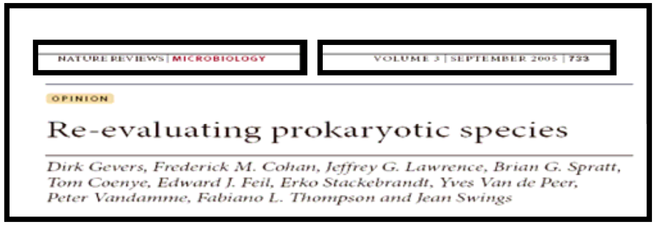




- 🟦 Amino acid metabolism (185 genes)
- 🟪 Coenzyme and cofactor metabolism (113)
- 🟩 Cell wall metabolism (124)
- 🟩 Cellular processing (206)
- 🟦 Central intermediary metabolism (250)
- 🟦 Energy metabolism (154)
- 🟦 Nucleotide metabolism (88)
- 🟦 Secondary metabolism (73)
- 🟥 Information processing (1,084)
- 🟩 Membrane transport (631)
- ⬛ Prophage genes (207)
- 🟨 Hypothetical or unknown genes (2,693)

UNIVERSITEIT GENT

Now that we have access to whole-genome sequences: what do they tell us?



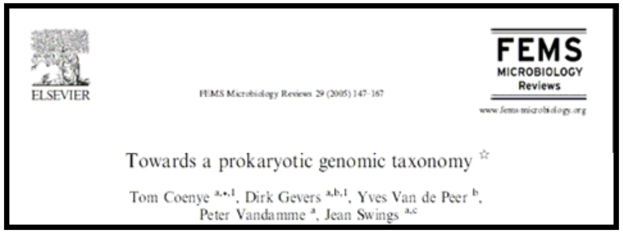
NATURE REVIEWS | MICROBIOLOGY

VOLUME 3 | SEPTEMBER 2005 | 733

OPINION

Re-evaluating prokaryotic species

Dirk Gevers, Frederick M. Cohan, Jeffrey G. Lawrence, Brian G. Spratt, Tom Coenye, Edward J. Feil, Erko Stackebrandt, Yves Van de Peer, Peter Vandamme, Fabiano L. Thompson and Jean Swings



ELSEVIER

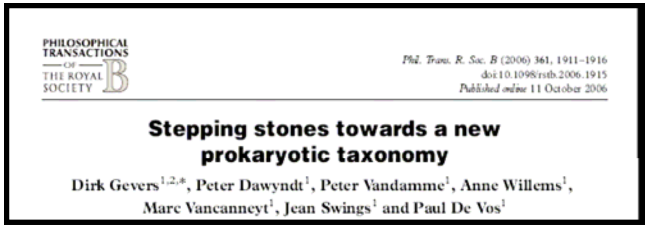
FEMS Microbiology Reviews 29 (2005) 147–167

FEMS MICROBIOLOGY Reviews

www.fems-microbiology.org

Towards a prokaryotic genomic taxonomy ☆

Tom Coenye [a,*,1], Dirk Gevers [a,b,1], Yves Van de Peer [b], Peter Vandamme [a], Jean Swings [a,c]

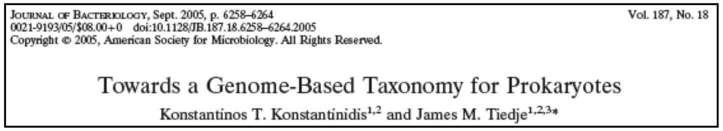


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) 361, 1911–1916
doi:10.1098/rstb.2006.1915
Published online 11 October 2006

Stepping stones towards a new prokaryotic taxonomy

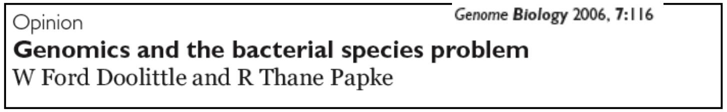
Dirk Gevers[1,2,*], Peter Dawyndt[1], Peter Vandamme[1], Anne Willems[1], Marc Vancanneyt[1], Jean Swings[1] and Paul De Vos[1]



JOURNAL OF BACTERIOLOGY, Sept. 2005, p. 6258–6264 Vol. 187, No. 18
0021-9193/05/$08.00+0 doi:10.1128/JB.187.18.6258–6264.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.

Towards a Genome-Based Taxonomy for Prokaryotes

Konstantinos T. Konstantinidis[1,2] and James M. Tiedje[1,2,3*]

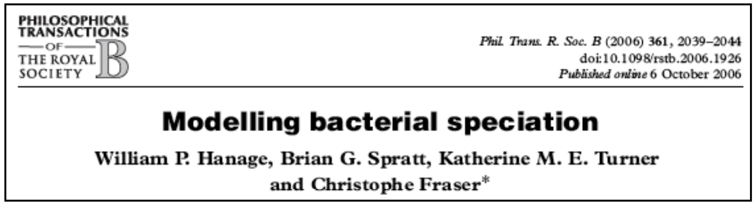


Opinion

Genome ***Biology*** 2006, **7:**116

Genomics and the bacterial species problem

W Ford Doolittle and R Thane Papke

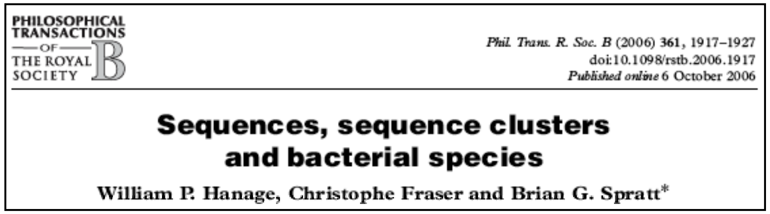


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) 361, 2039–2044
doi:10.1098/rstb.2006.1926
Published online 6 October 2006

Modelling bacterial speciation

William P. Hanage, Brian G. Spratt, Katherine M. E. Turner and Christophe Fraser*

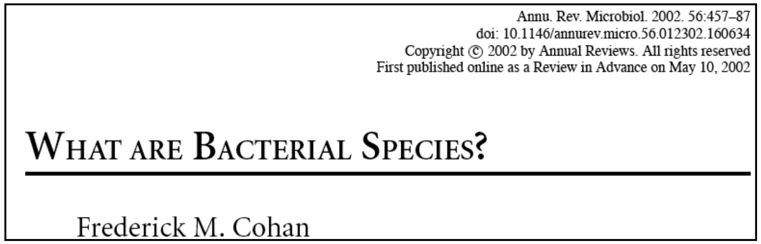


PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) 361, 1917–1927
doi:10.1098/rstb.2006.1917
Published online 6 October 2006

Sequences, sequence clusters and bacterial species

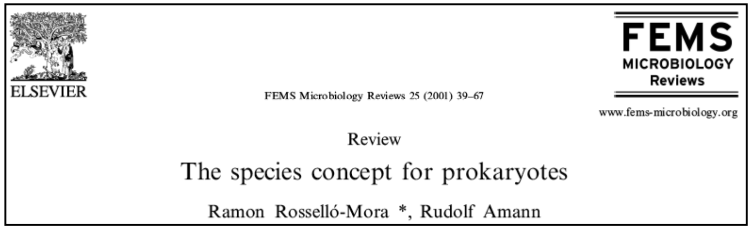
William P. Hanage, Christophe Fraser and Brian G. Spratt*



Annu. Rev. Microbiol. 2002. 56:457–87
doi: 10.1146/annurev.micro.56.012302.160634
Copyright © 2002 by Annual Reviews. All rights reserved
First published online as a Review in Advance on May 10, 2002

WHAT ARE BACTERIAL SPECIES?

Frederick M. Cohan



ELSEVIER

FEMS Microbiology Reviews 25 (2001) 39–67

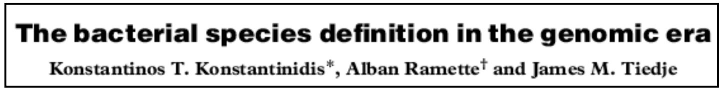
FEMS MICROBIOLOGY Reviews

www.fems-microbiology.org

Review

The species concept for prokaryotes

Ramon Rosselló-Mora *, Rudolf Amann



The bacterial species definition in the genomic era

Konstantinos T. Konstantinidis*, Alban Ramette† and James M. Tiedje

We’re beginning to understand genome content, evolution and diversity of bacterial species

W. Arber / FEMS Microbiology Reviews 24 (2000) 1–7

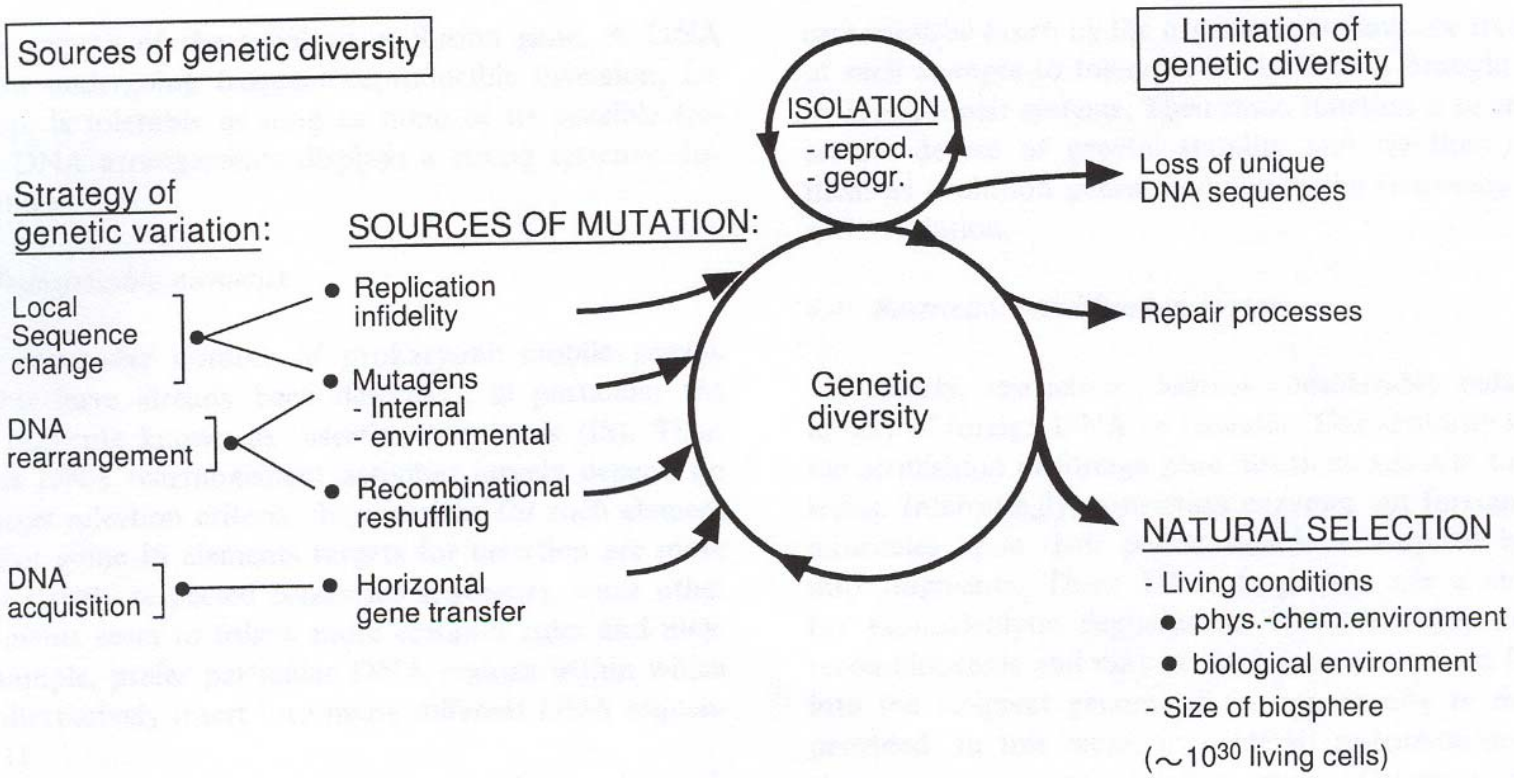


Fig. 2. Synoptic view of the elements of molecular evolution of prokaryotic microorganisms.

Now that we have access to whole-genome sequences: what can they tell us?

- Genomes seems to be composed of a core set of genes that is conserved among strains of the same species and accessory genes that are strain specific
- Phylogenetic signal present in core genes (ANI values*) does not necessarily correlate with gene content
  - ANI values reflect phylogeny
  - Gene content reflects ecology
- The basic taxonomy parameters are being confirmed: there is a core set of genes which, together, reflect 16S rRNA gene sequence similarity and whole genome DNA-DNA hybrid stability (‘relatedness’)

* Konstantinidis and Tiedje. 2005. Genomic insights that advance the species definition for prokaryotes PNAS 102:2567-2572

Real life... the lactic acid bacteria as test case

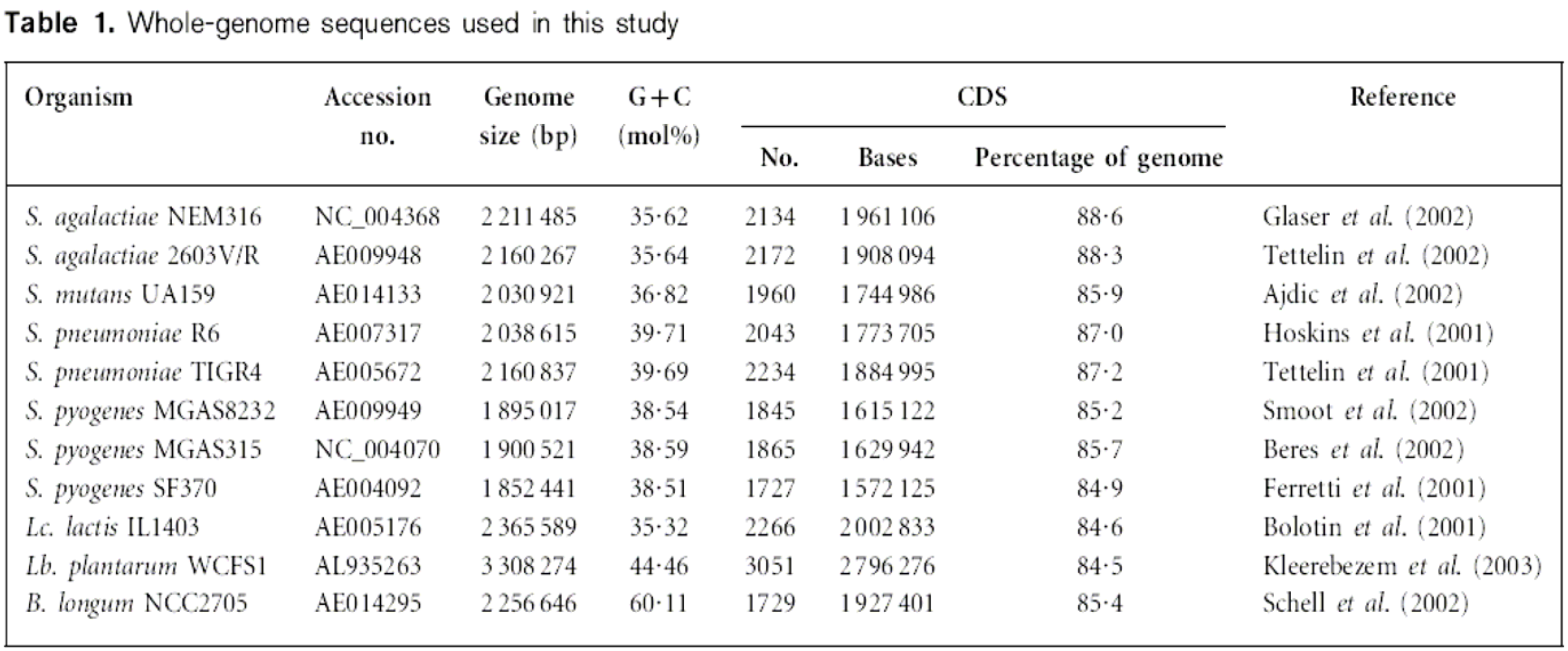


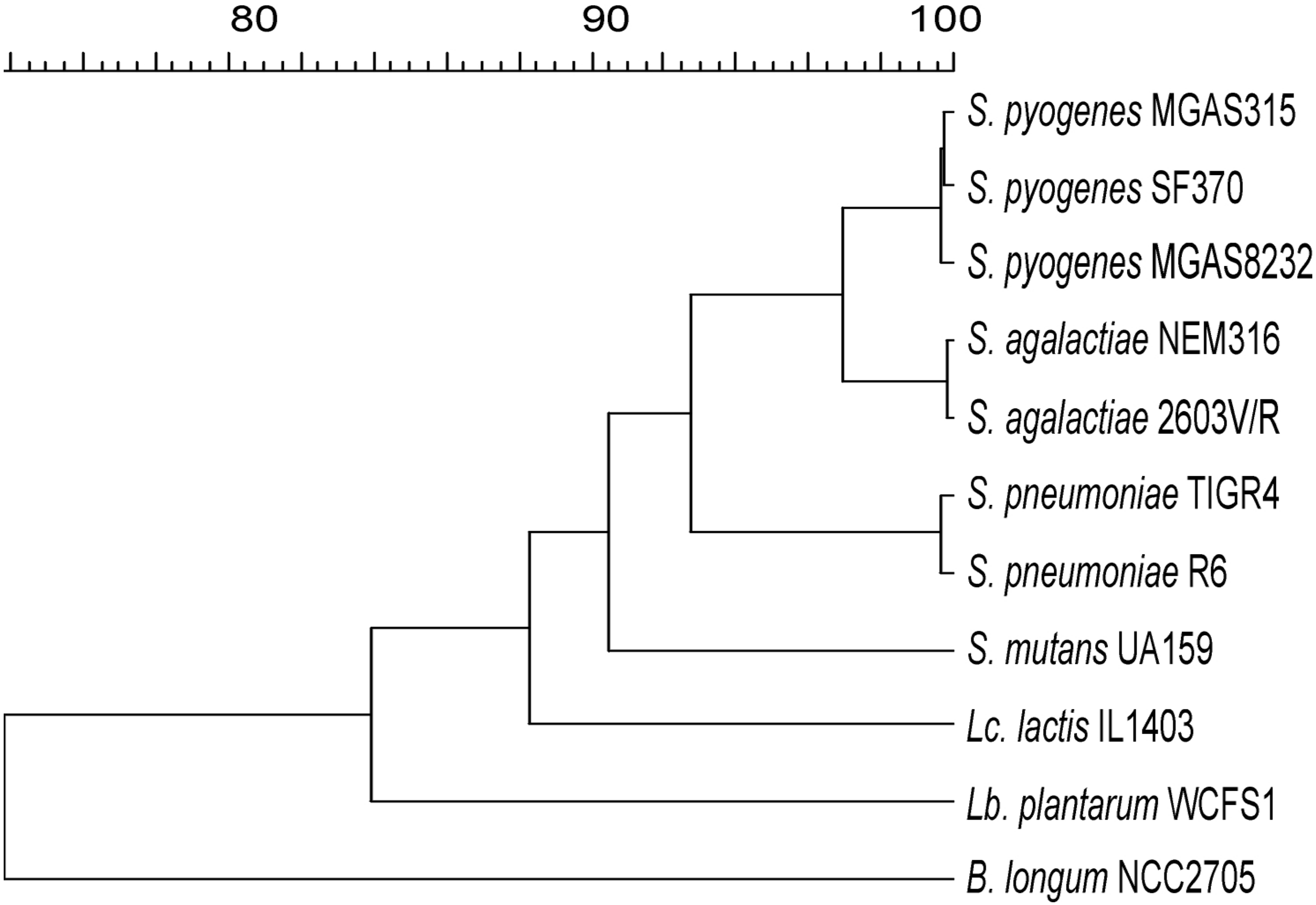
Table 1. Whole-genome sequences used in this study

Organism	Accession no.	Genome size (bp)	G+C (mol%)	CDS			Reference
				No.	Bases	Percentage of genome	
S. agalactiae NEM316	NC_004368	2 211 485	35·62	2134	1 961 106	88·6	Glaser *et al.* (2002)
S. agalactiae 2603V/R	AE009948	2 160 267	35·64	2172	1 908 094	88·3	Tettelin *et al.* (2002)
S. mutans UA159	AE014133	2 030 921	36·82	1960	1 744 986	85·9	Ajdic *et al.* (2002)
S. pneumoniae R6	AE007317	2 038 615	39·71	2043	1 773 705	87·0	Hoskins *et al.* (2001)
S. pneumoniae TIGR4	AE005672	2 160 837	39·69	2234	1 884 995	87·2	Tettelin *et al.* (2001)
S. pyogenes MGAS8232	AE009949	1 895 017	38·54	1845	1 615 122	85·2	Smoot *et al.* (2002)
S. pyogenes MGAS315	NC_004070	1 900 521	38·59	1865	1 629 942	85·7	Beres *et al.* (2002)
S. pyogenes SF370	AE004092	1 852 441	38·51	1727	1 572 125	84·9	Ferretti *et al.* (2001)
Lc. lactis IL1403	AE005176	2 365 589	35·32	2266	2 002 833	84·6	Bolotin *et al.* (2001)
Lb. plantarum WCFS1	AL935263	3 308 274	44·46	3051	2 796 276	84·5	Kleerebezem *et al.* (2003)
B. longum NCC2705	AE014295	2 256 646	60·11	1729	1 927 401	85·4	Schell *et al.* (2002)

Coenye and Vandamme. 2003. Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case. Microbiology **149:**3507-3517

Comparing sequences

16S rRNA gene sequence similarity



Wole-genome analyses

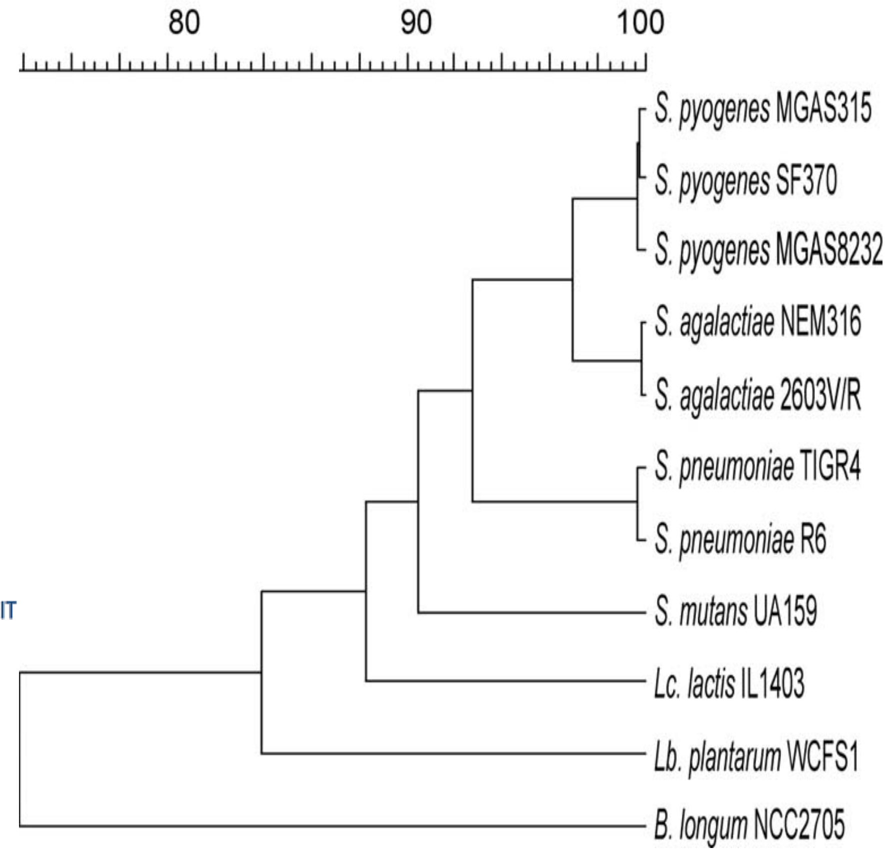
- Comparison of the sequence of 16S rRNA genes (nucleotides) and nine house-keeping proteins (gyrB, rpoD, sodA, dnaK, recA, gki, ddl, alaS and ileS) (amino acids) + construction of a supertree
- Detection of orthologous genes by bidirectional genome-to-genome BLASTP analysis
- Determination of dinucleotide relative abundance values
- Determination of codon usage statistics
- Determination of conservation of gene order

Differences in gene content

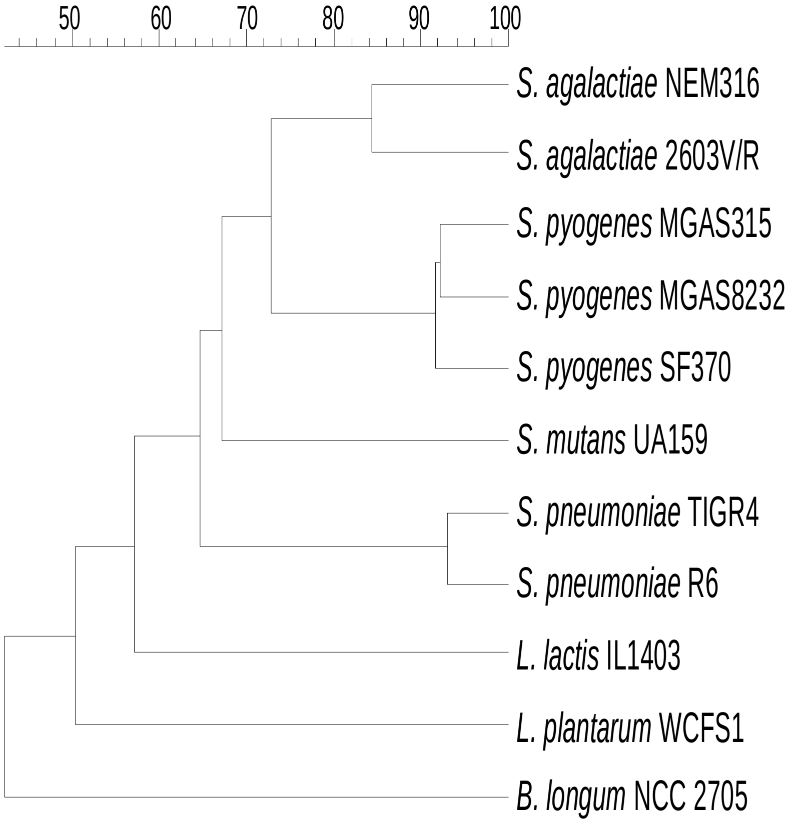
Species and strain designation	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. *S. agalactiae* NEM316	-										
2. *S. agalactiae* 2603V/R	84.39	-									
3. *S. mutans* UA159	68.73	68.83	-								
4. *S. pneumoniae* R6	66.72	67.76	67.32	-							
5. *S. pneumoniae* TIGR4	61.19	61.94	59.78	93.05	-						
6. *S. pyogenes* MGAS8232	70.79	74.04	64.77	64.93	65.20	-					
7. *S. pyogenes* MGAS315	69.60	72.92	64.34	63.24	63.97	92.28	-				
8. *S. pyogenes* SF370	74.23	75.90	69.20	67.85	68.08	92.01	91.31	-			
9. *Lc. lactis* IL1403	59.44	60.24	58.47	55.96	56.49	55.69	56.13	54.63	-		
10. *Lb. plantarum* WCFS1	53.03	53.10	51.98	48.90	49.79	47.10	46.90	46.28	56.54	-	
11. *B. longum* NCC2705	43.96	43.96	43.78	42.97	43.20	38.98	39.33	39.56	39.98	48.41	-

Differences in gene content

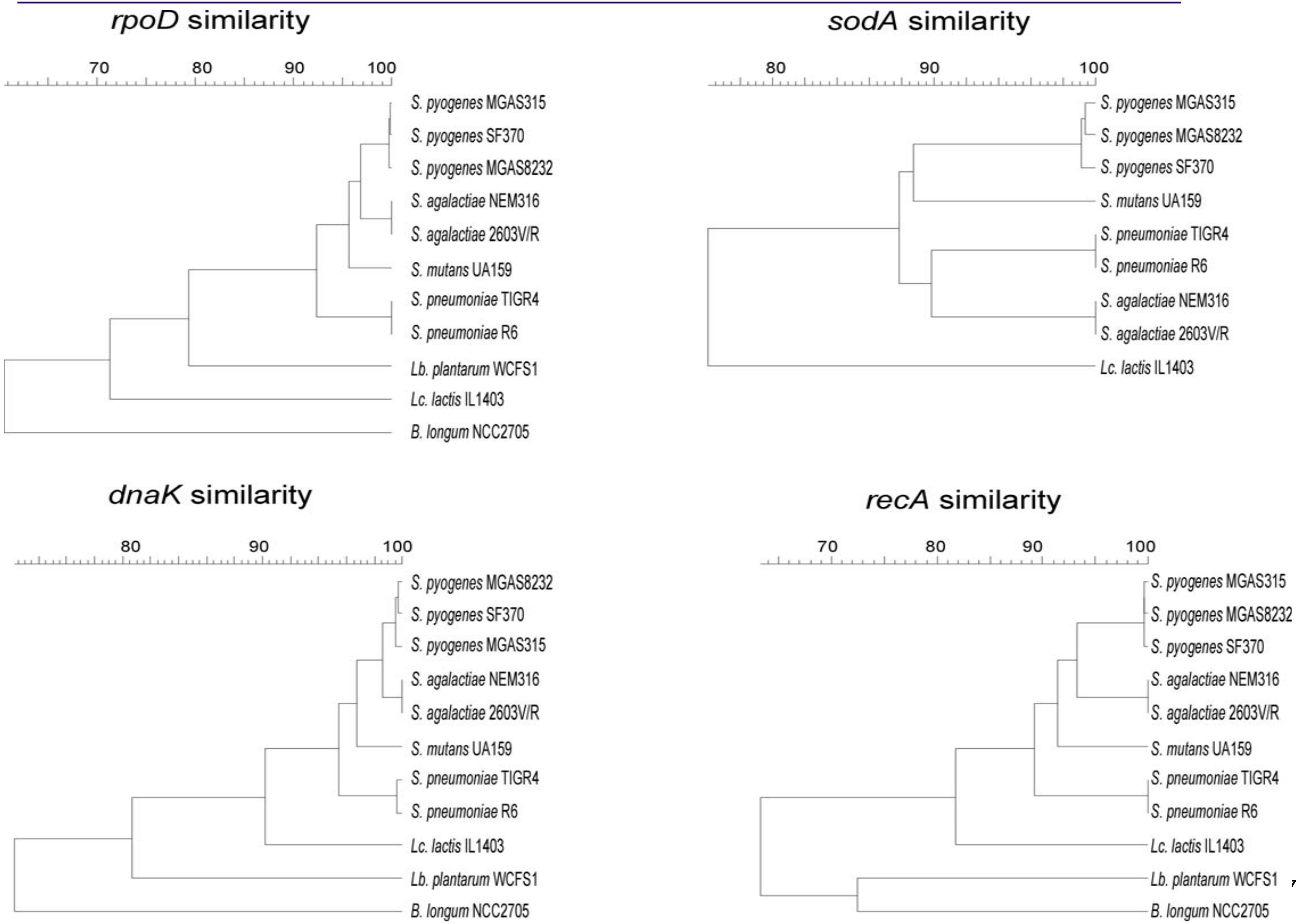
16S rDNA similarity



fraction of shared orthologous genes



Comparing sequences

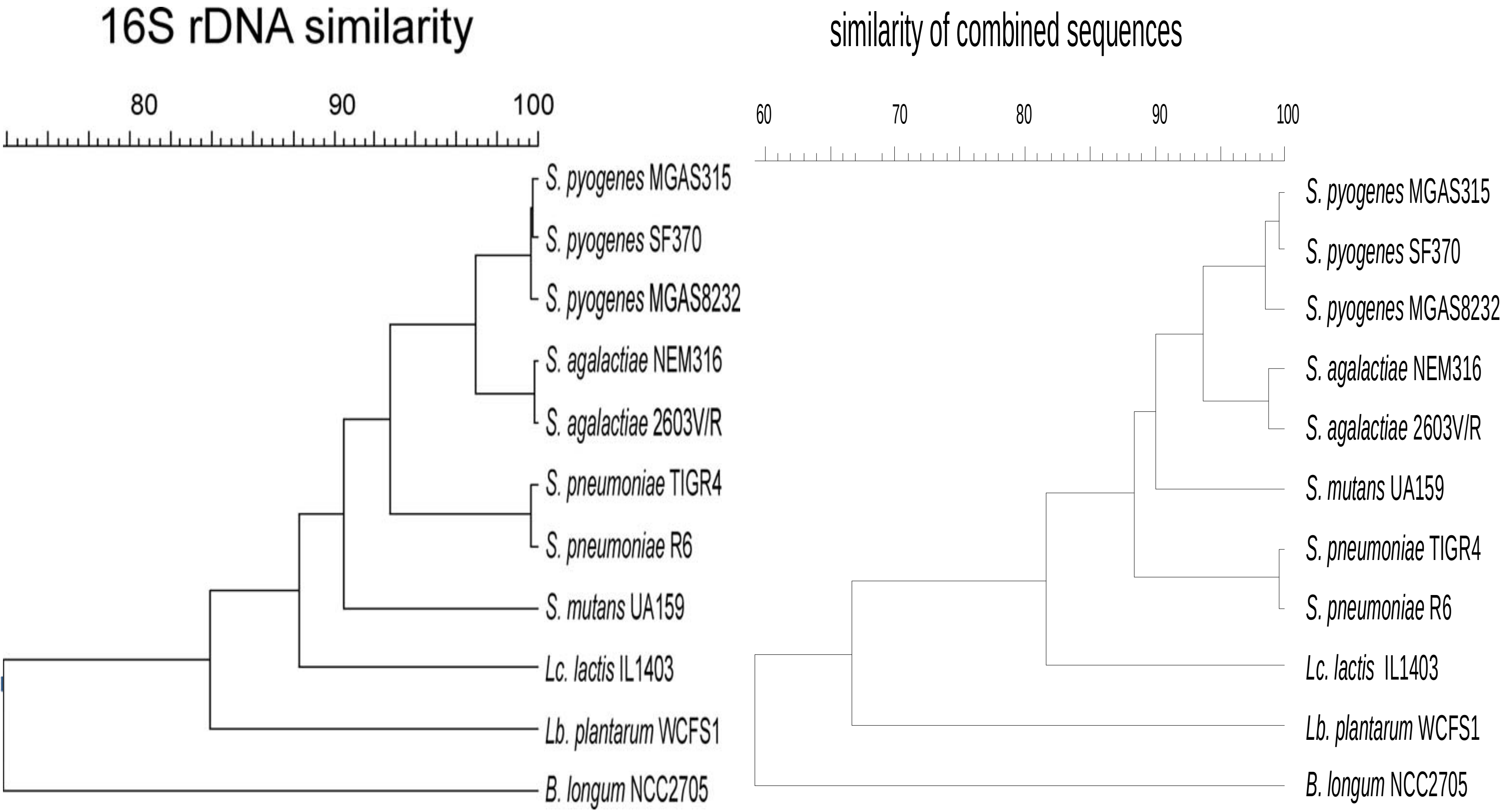

rpoD similarity

sodA similarity

dnaK similarity

recA similarity

The supertree



16S rDNA similarity

similarity of combined sequences

- Konstantinidis et al., 2006a. Towards a more robust assessment of intraspecies diversity using fewer genetic markers. AEM 72:7286-93
- Konstantinidis et al., 2006b. The bacterial species definition in the genomic era. Phil. Trans. Royal Soc. B 361:1929-40.

Compositional bias (Karlin signatures)

- Relative abundance values of di/tri/tetranucleotides constitute a genomic signature; hence dissimilarity in relative abundance

- Most used : set of dinucleotide values (easiest to compute!)

- Mathemical :

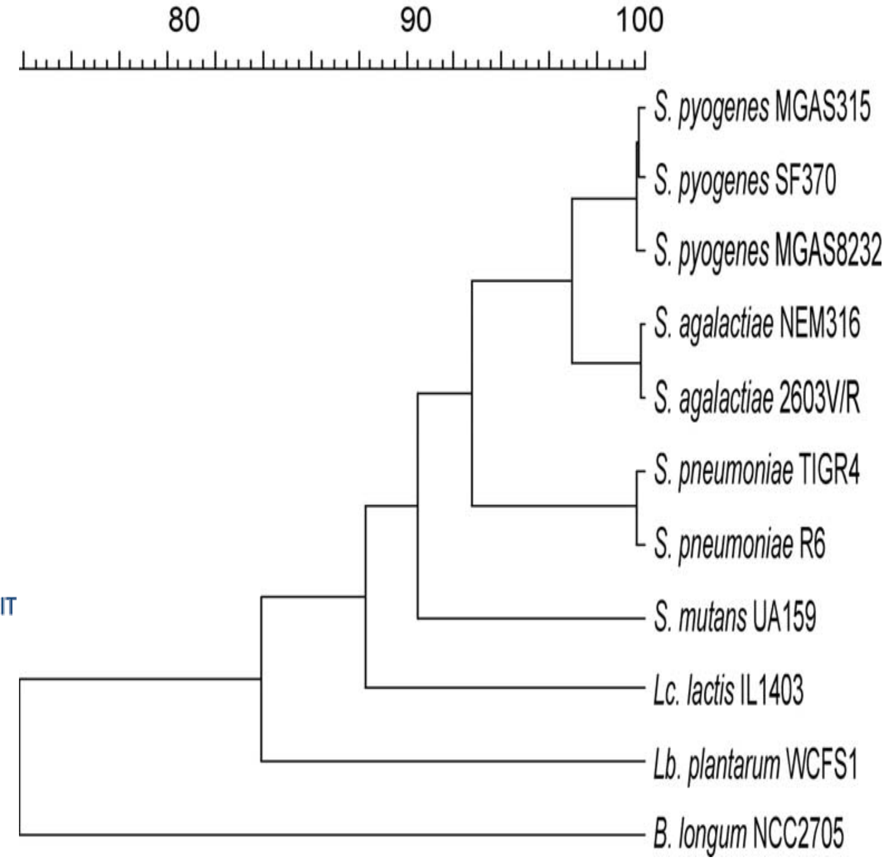
  $\rho^*_{XY} = f_{XY}/f_X f_Y$ (normal range 0.78 – 1.23)
  (X, Y = A, C, G, T ; XY = AA, AC, AG, AT, ..., TT)

  $\delta^*(f,g) = 1/16\ \Sigma\ |\ \rho^*_{XY}(f) - \rho^*_{XY}(g)|$ (within species < 20)
  (measure of dissimilarity between genomes)

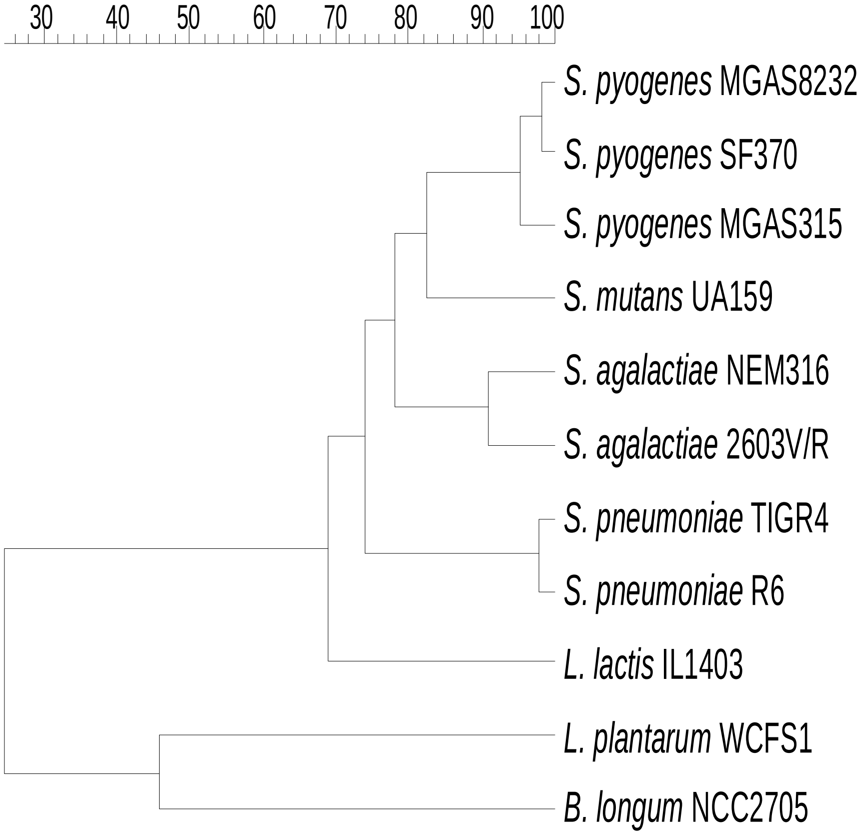
Karlin et al., 1997. Compositional biases of bacterial genomes and evolutionary implications. JB 179:3899-913

Dinucleotide relative abundance

16S rDNA similarity



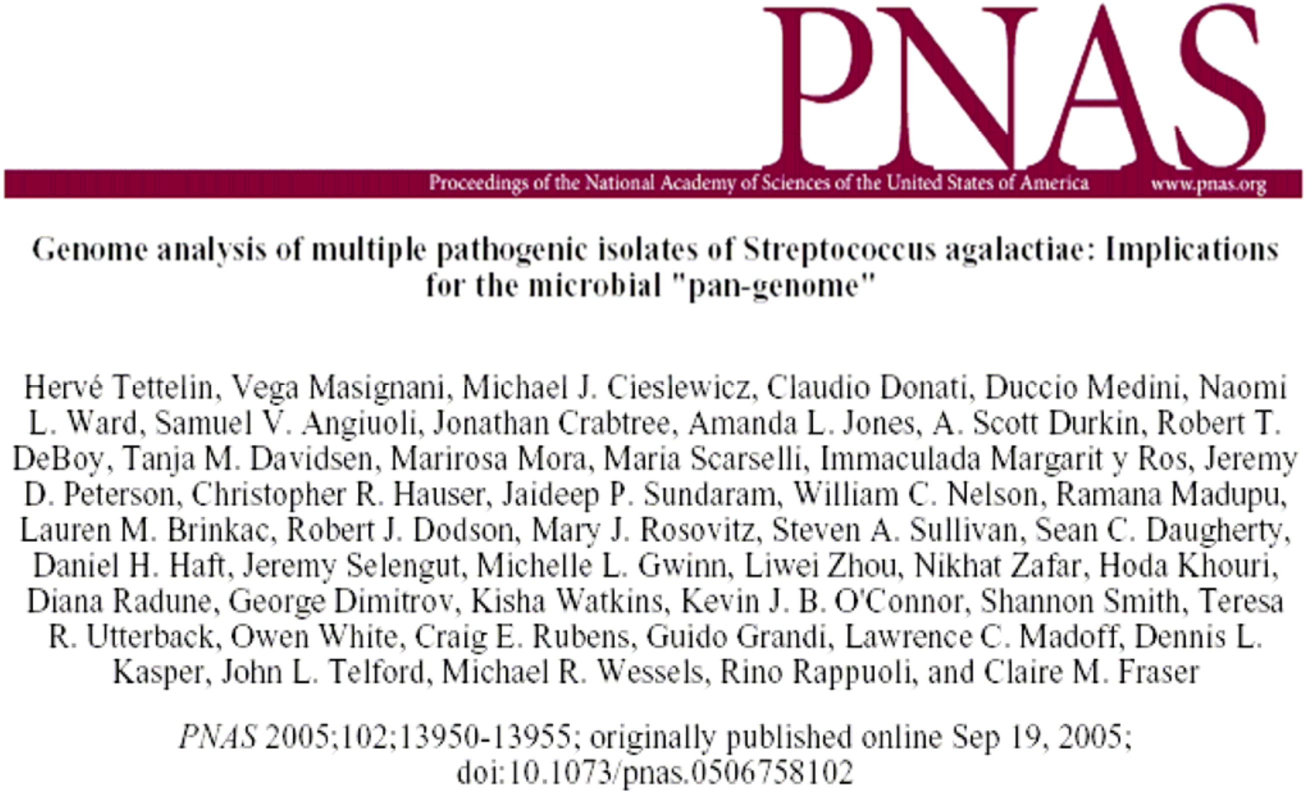
Normalised δ^*(f,g) values



Now that we have access to whole-genome sequences: what do they tell us?

- Some basic taxonomic parameters are being confirmed: high DNA-DNA hybridisation levels and highly similar 16S rRNA gene sequences are reflected in the whole genome content
- Core & accessory genomes, open & closed pan-genomes

The species “pan-genome”



PNAS

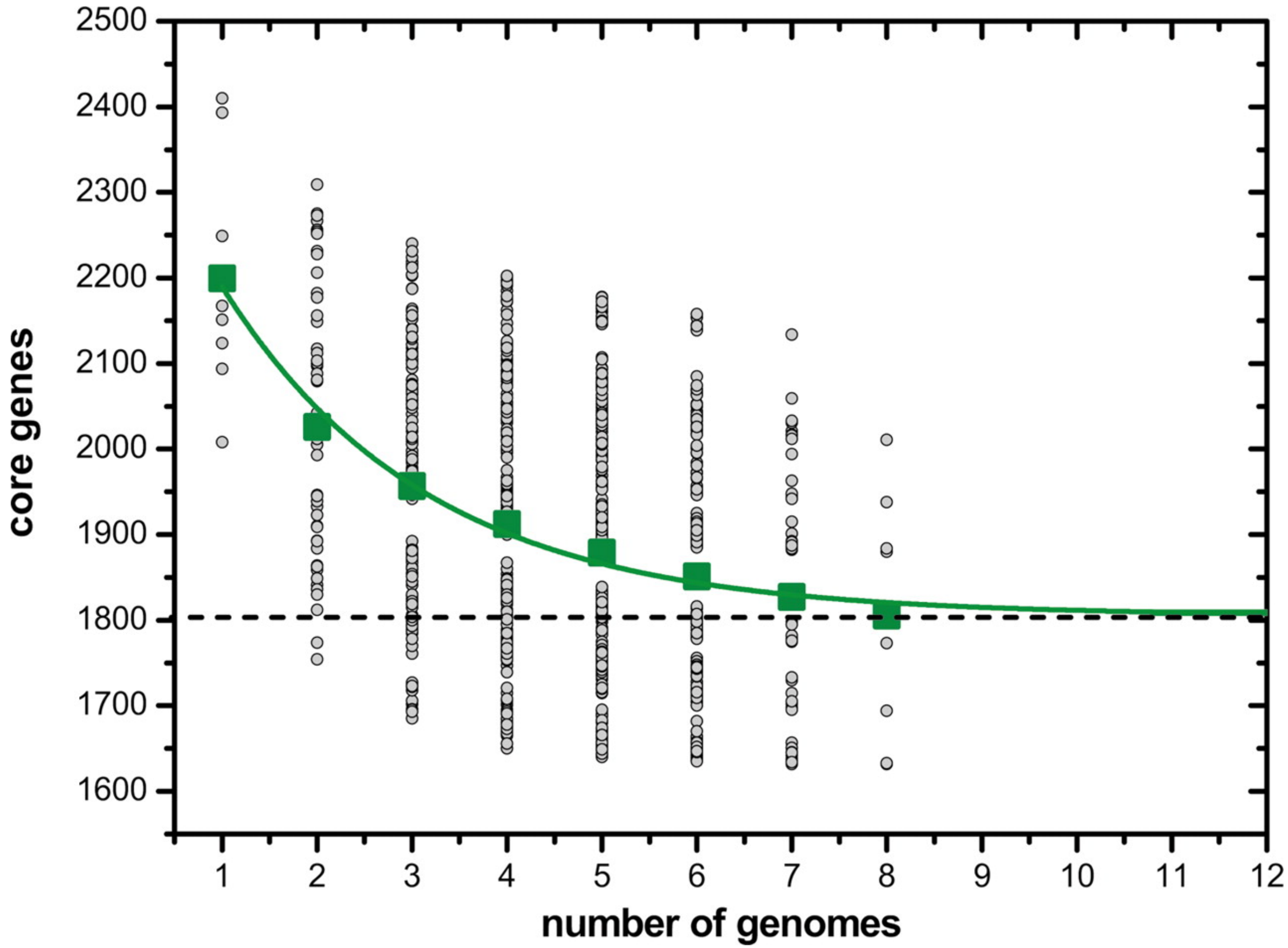
Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome"

Hervé Tettelin, Vega Masignani, Michael J. Cieslewicz, Claudio Donati, Duccio Medini, Naomi L. Ward, Samuel V. Angiuoli, Jonathan Crabtree, Amanda L. Jones, A. Scott Durkin, Robert T. DeBoy, Tanja M. Davidsen, Marirosa Mora, Maria Scarselli, Immaculada Margarit y Ros, Jeremy D. Peterson, Christopher R. Hauser, Jaideep P. Sundaram, William C. Nelson, Ramana Madupu, Lauren M. Brinkac, Robert J. Dodson, Mary J. Rosovitz, Steven A. Sullivan, Sean C. Daugherty, Daniel H. Haft, Jeremy Selengut, Michelle L. Gwinn, Liwei Zhou, Nikhat Zafar, Hoda Khouri, Diana Radune, George Dimitrov, Kisha Watkins, Kevin J. B. O'Connor, Shannon Smith, Teresa R. Utterback, Owen White, Craig E. Rubens, Guido Grandi, Lawrence C. Madoff, Dennis L. Kasper, John L. Telford, Michael R. Wessels, Rino Rappuoli, and Claire M. Fraser

PNAS 2005;102;13950-13955; originally published online Sep 19, 2005;
doi:10.1073/pnas.0506758102

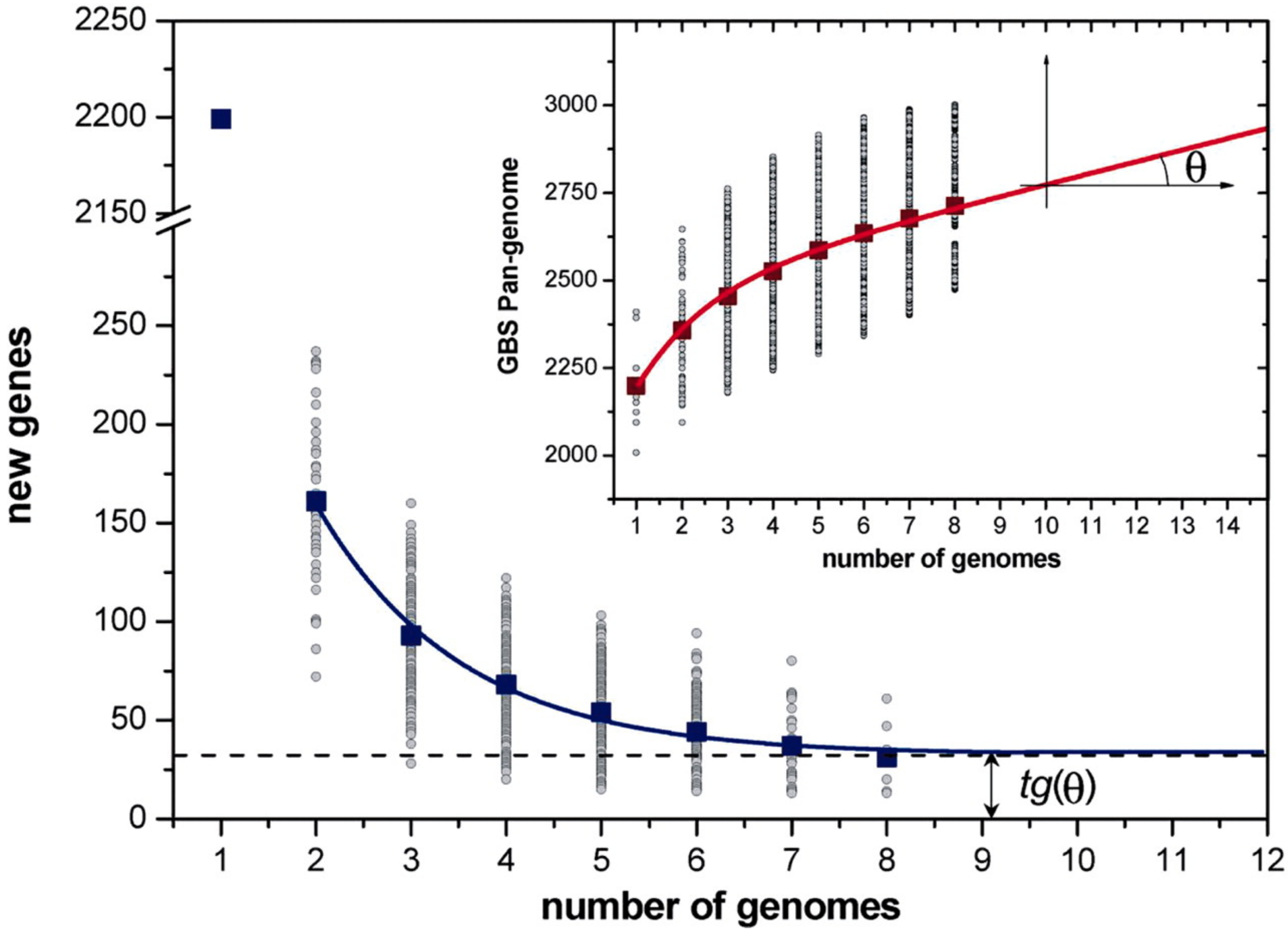
Fig. 2. GBS core genome



Tettelin et al. (2005) Proc. Natl. Acad. Sci. USA 102, 13950-13955

Fig. 3. GBS pan-genome

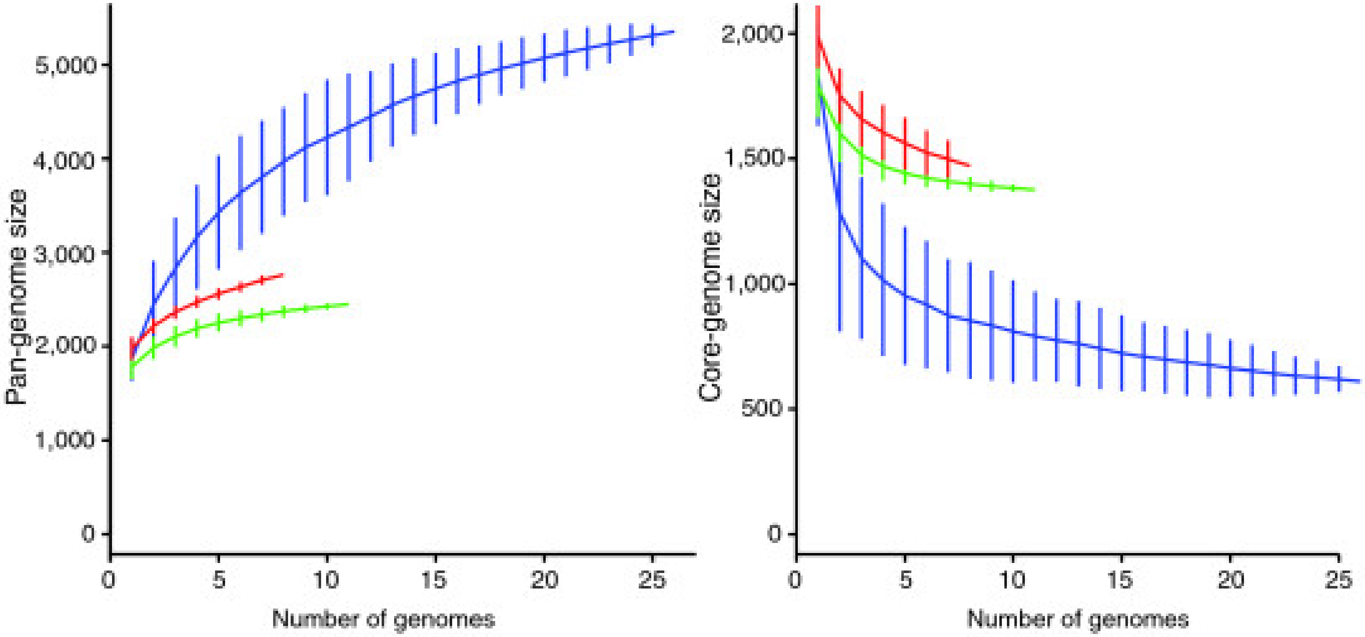
Open pan-genome



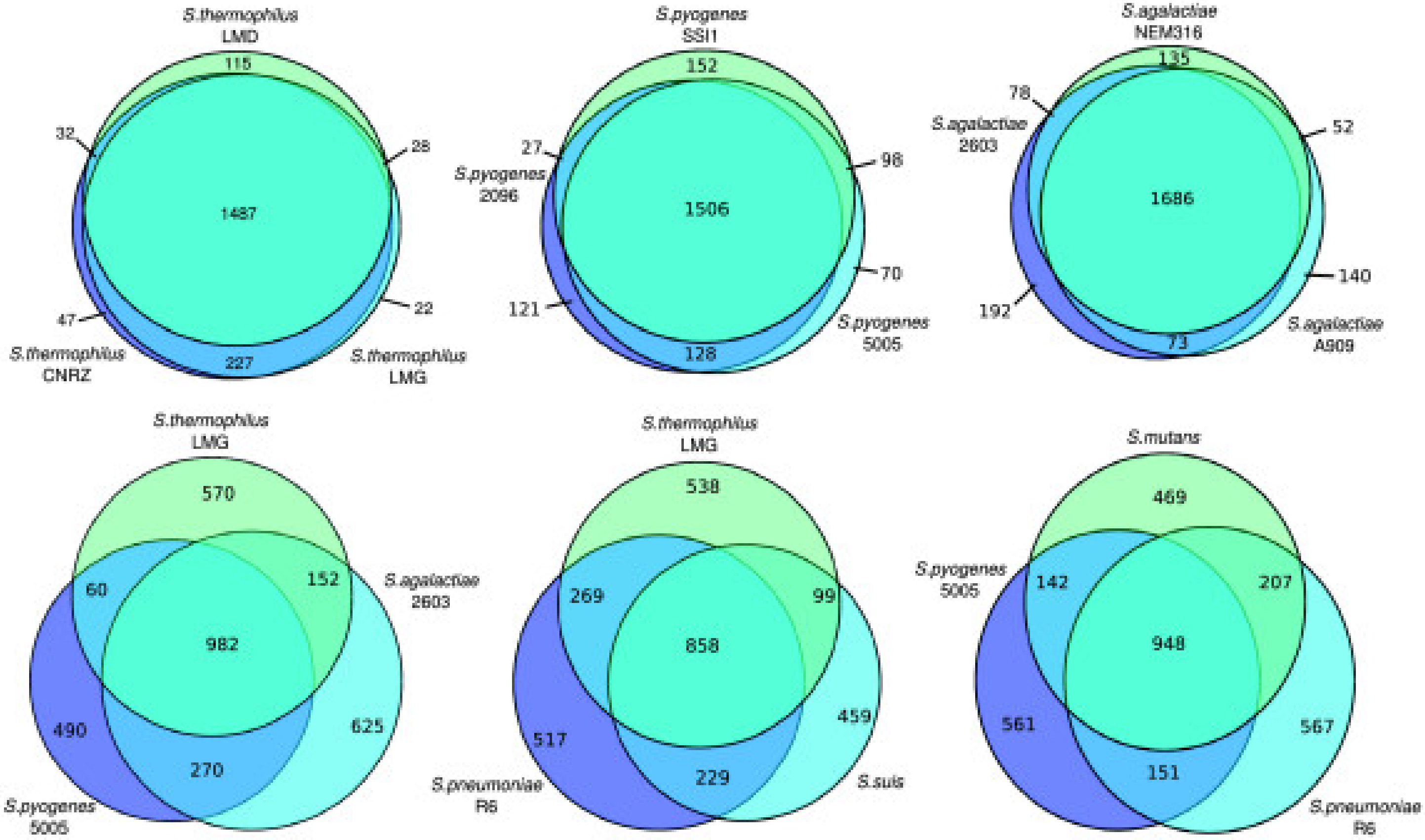
Tettelin et al. (2005) Proc. Natl. Acad. Sci. USA 102, 13950-13955

Lefébure and Stanhope 2007 Genome Biol. 8: R71

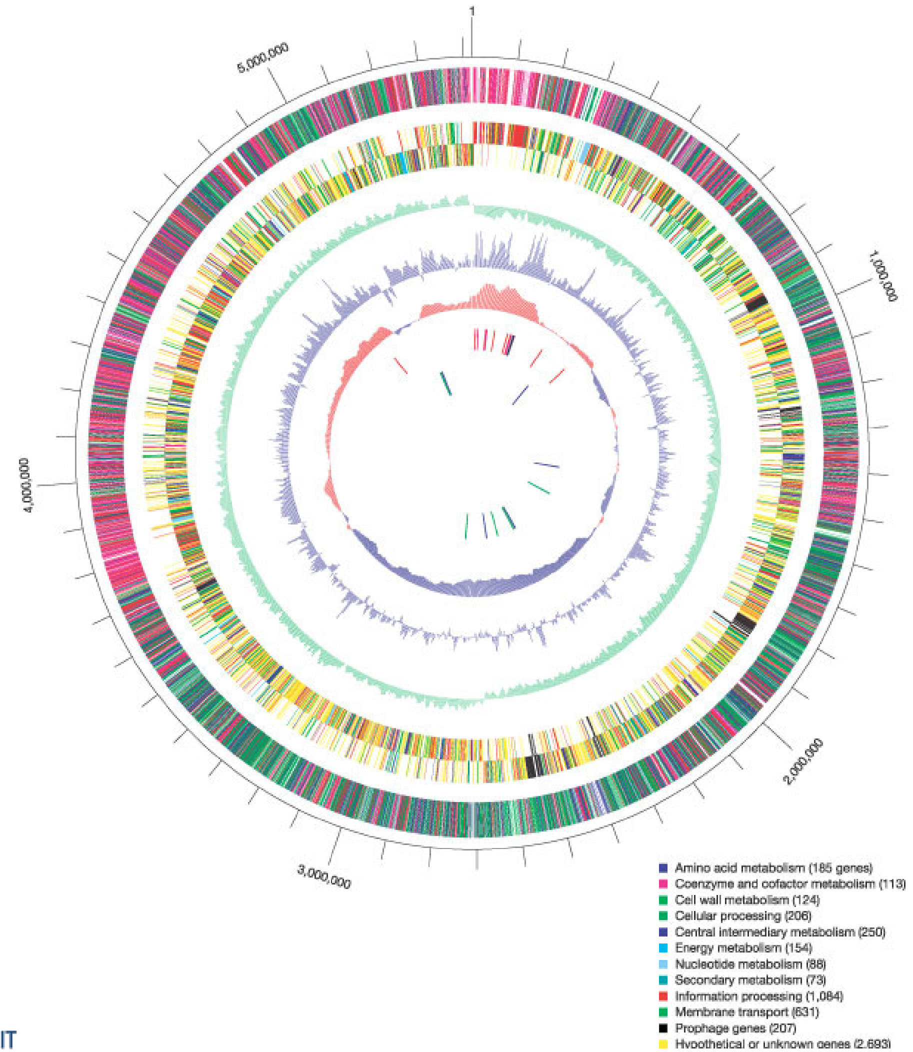
- 26 *Streptococcus* genomes:
  - 11 *S. pyogenes*
  - 8 *S. agalactiae*
  - 2 *S. pneumoniae*
  - 1 *S. mutans*
  - 3 *S. thermophilus*
  - 1 *S. suis*



- Accumulation curves for the total number of genes (left) or the number of genes in common (right) given a number of genomes analyzed for the different species of Streptococcus (in blue), the different strains of *S. agalactiae* (in red) and *S. pyogenes* (in green). The vertical bars correspond to standard deviations after repeating one hundred random input orders of the genomes (Lefébure and Stanhope 2007 Genome Biol. 8: R71)



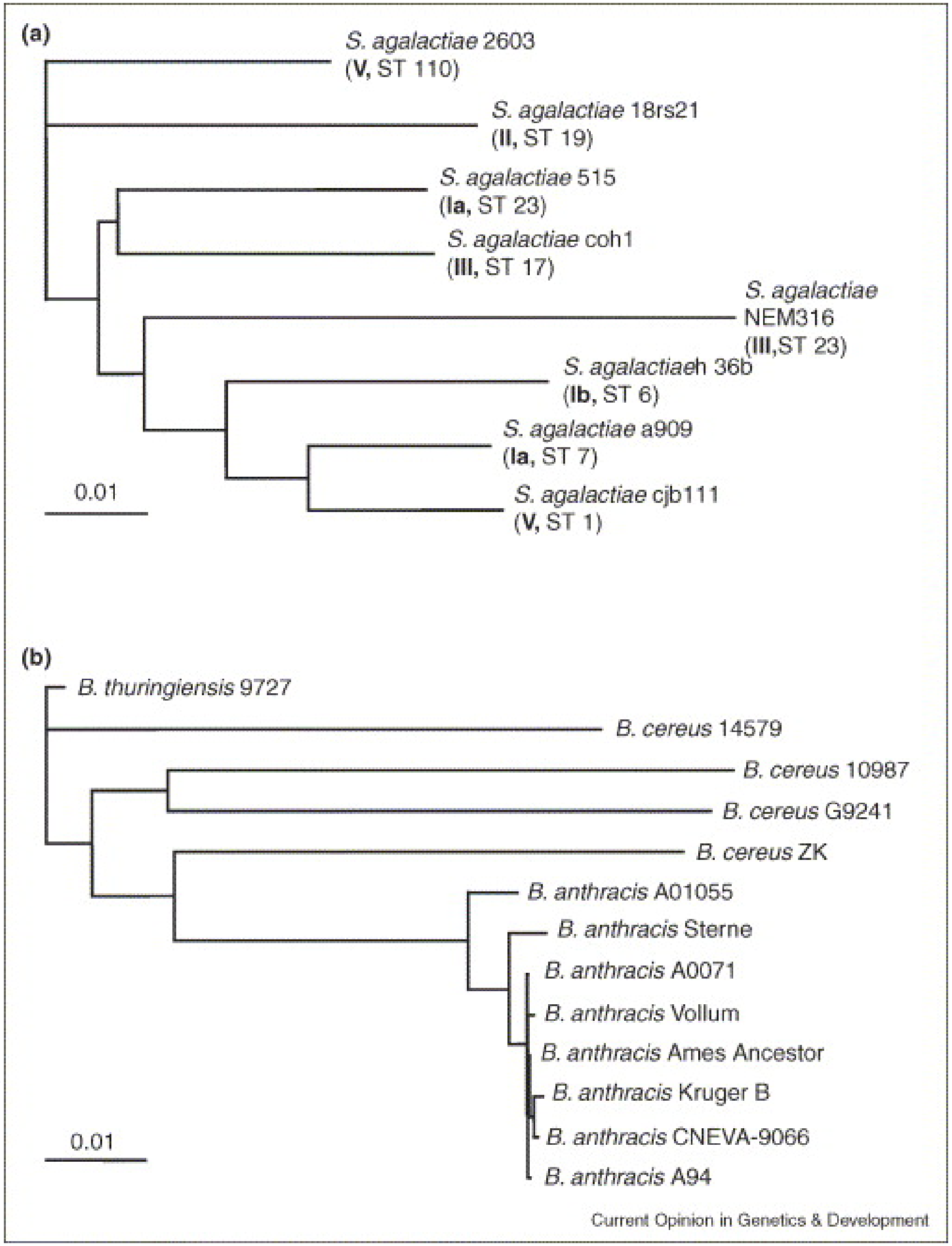
- Venn diagram for six sets of three taxa. Above are taxa of the same species and below are taxa of different species. The surfaces are approximately proportional to the number of genes (Lefébure and Stanhope 2007 Genome Biol. 8: R71)


1
1,000,000
2,000,000
3,000,000
4,000,000
5,000,000
Amino acid metabolism (185 genes)
Coenzyme and cofactor metabolism (113)
Cell wall metabolism (124)
Cellular processing (206)
Central intermediary metabolism (250)
Energy metabolism (154)
Nucleotide metabolism (88)
Secondary metabolism (73)
Information processing (1,084)
Membrane transport (631)
Prophage genes (207)
Hypothetical or unknown genes (2,693)

- The large core set of genes (75–80%) conserved between *B. cereus* ATCC 14579 and *B. anthracis* A2012 could have been inherited from a common ancestor (Ivanova et al. 2003 Nature **423**, 87-91)

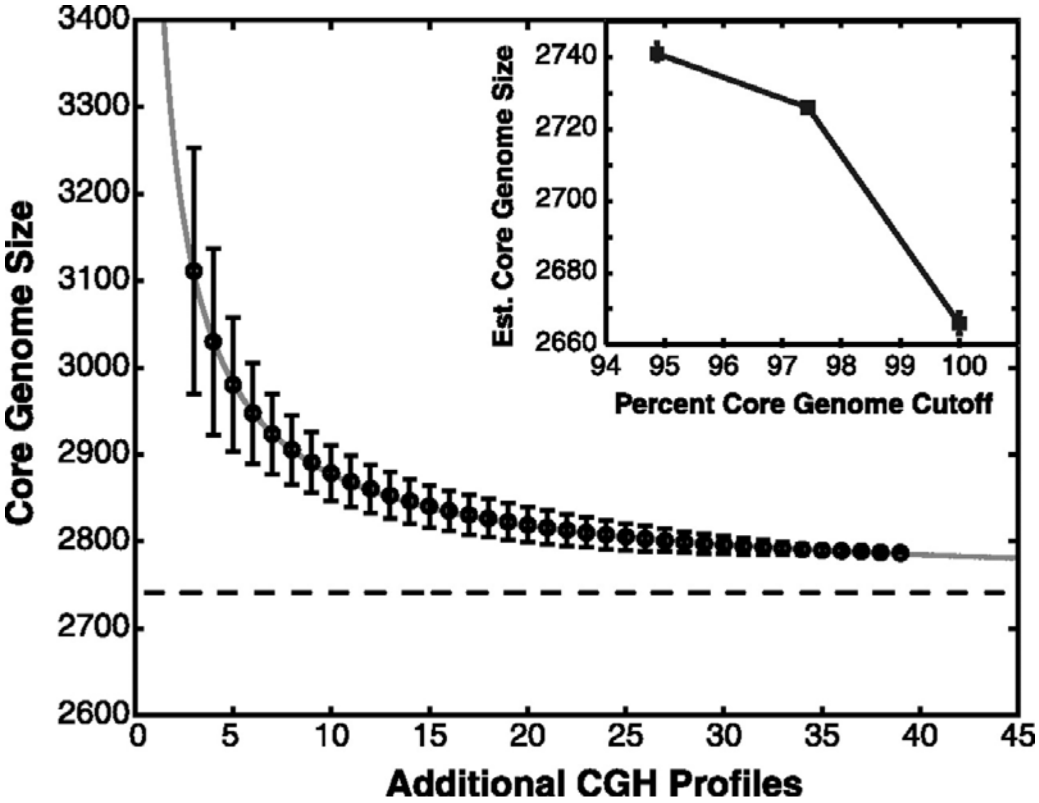
Figure 3. Dendrograms of the eight *Streptococcus agalactiae* **(a)** and thirteen *B*. *cereus* group **(b)** genomes. The fraction of genes of one strain that is not shared with other strains was used to define a distance matrix.

From the figure, it is evident that the distance between two *S*. *agalactiae* strains is comparable to the distance between *B*. *anthracis* stains and other *B*. *cereus* group species, making the definition of *B*. *anthracis* as an autonomous species questionable.

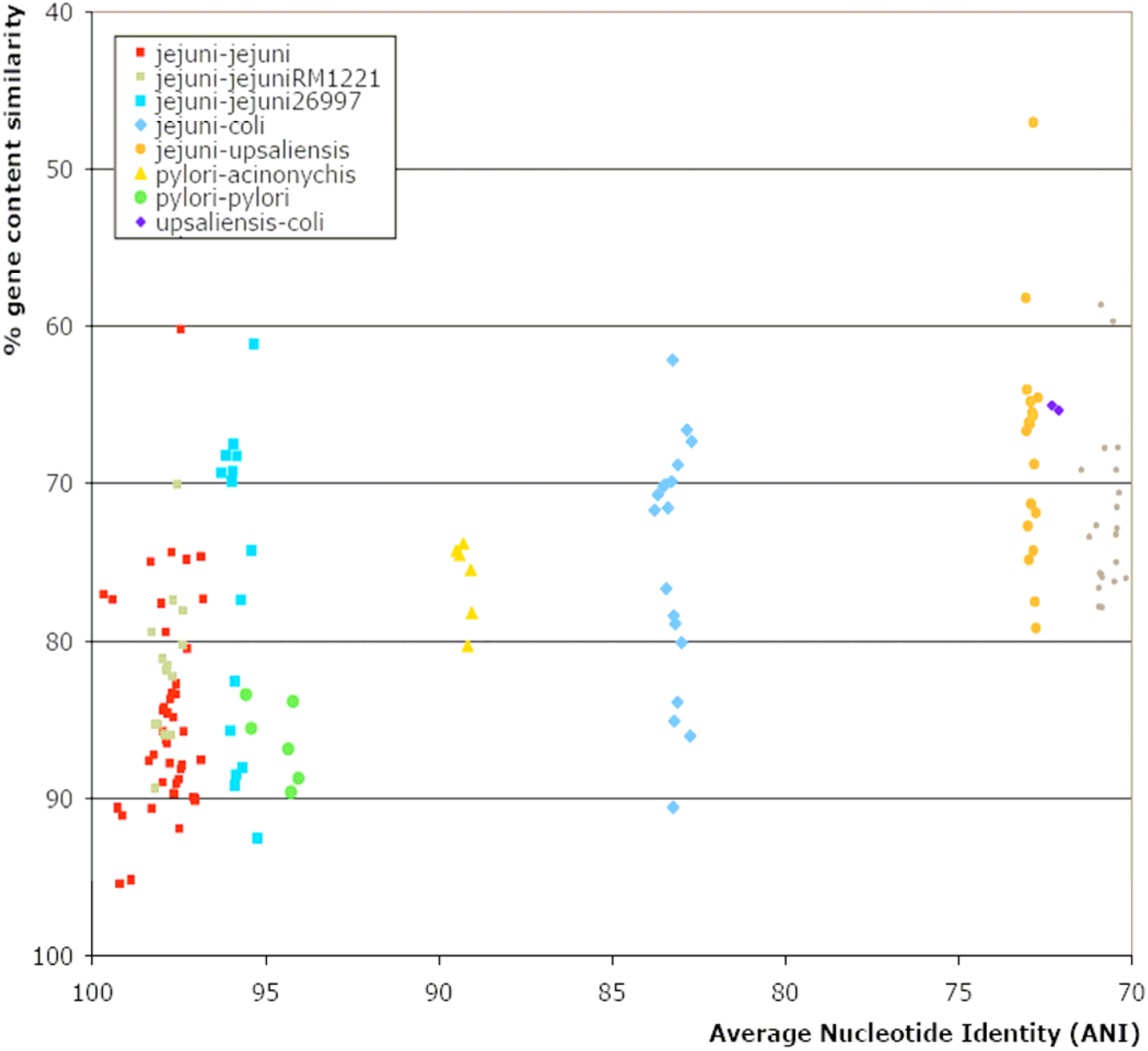

(a)
S. agalactiae 2603 (V, ST 110)
S. agalactiae 18rs21 (II, ST 19)
S. agalactiae 515 (Ia, ST 23)
S. agalactiae coh1 (III, ST 17)
S. agalactiae NEM316 (III,ST 23)
S. agalactiaeh 36b (Ib, ST 6)
S. agalactiae a909 (Ia, ST 7)
S. agalactiae cjb111 (V, ST 1)
0.01
(b)
B. thuringiensis 9727
B. cereus 14579
B. cereus 10987
B. cereus G9241
B. cereus ZK
B. anthracis A01055
B. anthracis Sterne
B. anthracis A0071
B. anthracis Vollum
B. anthracis Ames Ancestor
B. anthracis Kruger B
B. anthracis CNEVA-9066
B. anthracis A94
0.01

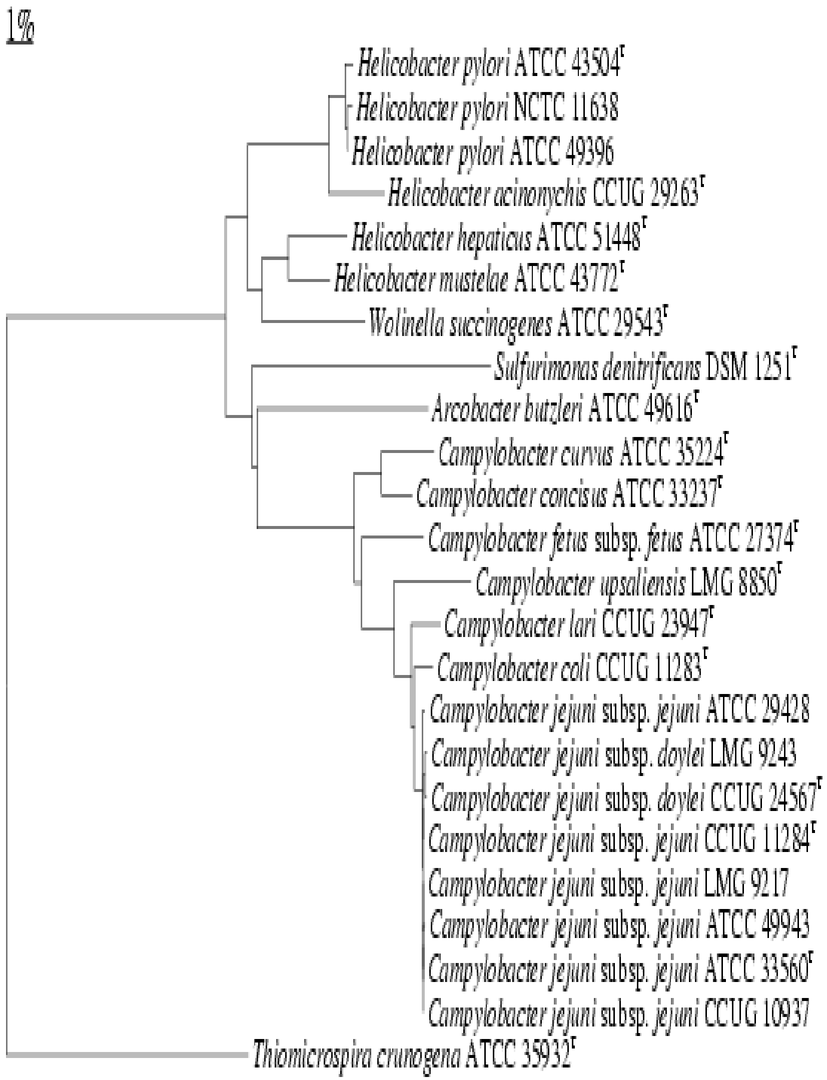
Current Opinion in Genetics & Development

Closed pan-genome

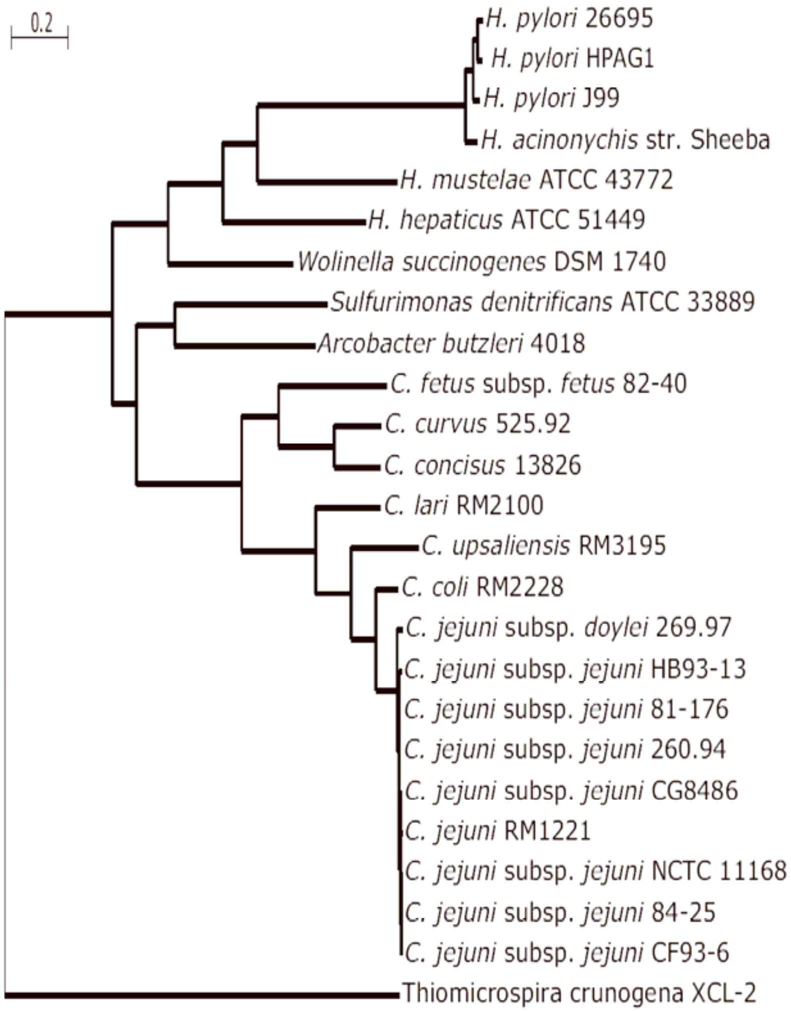


- FIG. 2. Estimation of *Vibrio cholerae* core genome size by regression analysis. Open circles with 95% confidence limits represent the mean number of core genes with increasing numbers of genomes sampled for 10,000 random permutations of sampling order. A power law regression fit [y = a x (b) + c] with an *R*-squared value of 0.9998 is included. Regression coefficients with 95% confidence limits (CL) are as follows: a, 906.1 (CL, 894.1, 918.0); b, –0.8215 (CL, –0.8348, –0.8083); and c, 2,741 (CL, 2,739, 2,744). The horizontal dashed line represents the extrapolated core genome size for *Vibrio cholerae*, which is equal to 2,741 genes for a threshold of genes shared among 95% of sampled genomes. (Inset) Closed squares show the reduction in projected core genome size with increased stringency for gene ubiquity from 95% to 100% of strains Keymer et al. 2007. AEM 73, 3705-3714

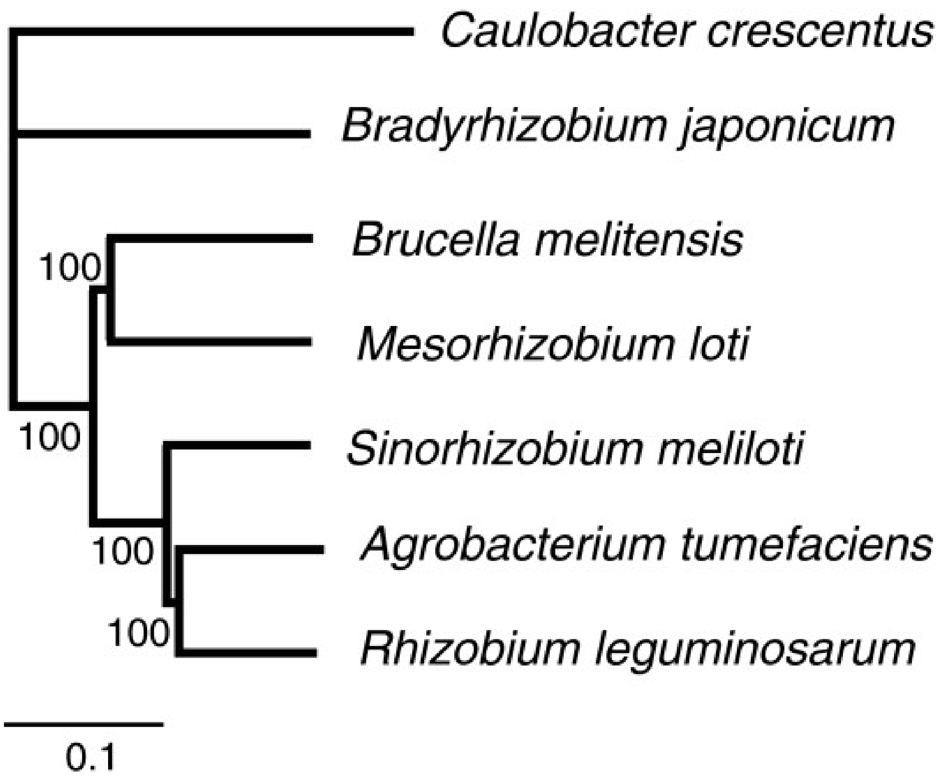

% gene content similarity
40
50
60
70
80
90
100
100
95
90
85
80
75
70
Average Nucleotide Identity (ANI)
■ jejuni-jejuni
■ jejuni-jejuniRM1221
■ jejuni-jejuni26997
◆ jejuni-coli
● jejuni-upsaliensis
▲ pylori-acinonychis
● pylori-pylori
◆ upsaliensis-coli


1%
Helicobacter pylori ATCC 43504T
Helicobacter pylori NCTC 11638
Helicobacter pylori ATCC 49396
Helicobacter acinonychis CCUG 29263T
Helicobacter hepaticus ATCC 51448T
Helicobacter mustelae ATCC 43772T
Wolinella succinogenes ATCC 29543T
Sulfurimonas denitrificans DSM 1251T
Arcobacter butzleri ATCC 49616T
Campylobacter curvus ATCC 35224T
Campylobacter concisus ATCC 33237T
Campylobacter fetus subsp. fetus ATCC 27374T
Campylobacter upsaliensis LMG 8850T
Campylobacter lari CCUG 23947T
Campylobacter coli CCUG 11283T
Campylobacter jejuni subsp. jejuni ATCC 29428
Campylobacter jejuni subsp. doylei LMG 9243
Campylobacter jejuni subsp. doylei CCUG 24567T
Campylobacter jejuni subsp. jejuni CCUG 11284T
Campylobacter jejuni subsp. jejuni LMG 9217
Campylobacter jejuni subsp. jejuni ATCC 49943
Campylobacter jejuni subsp. jejuni ATCC 33560T
Campylobacter jejuni subsp. jejuni CCUG 10937
Thiomicrospira crunogena ATCC 35932T

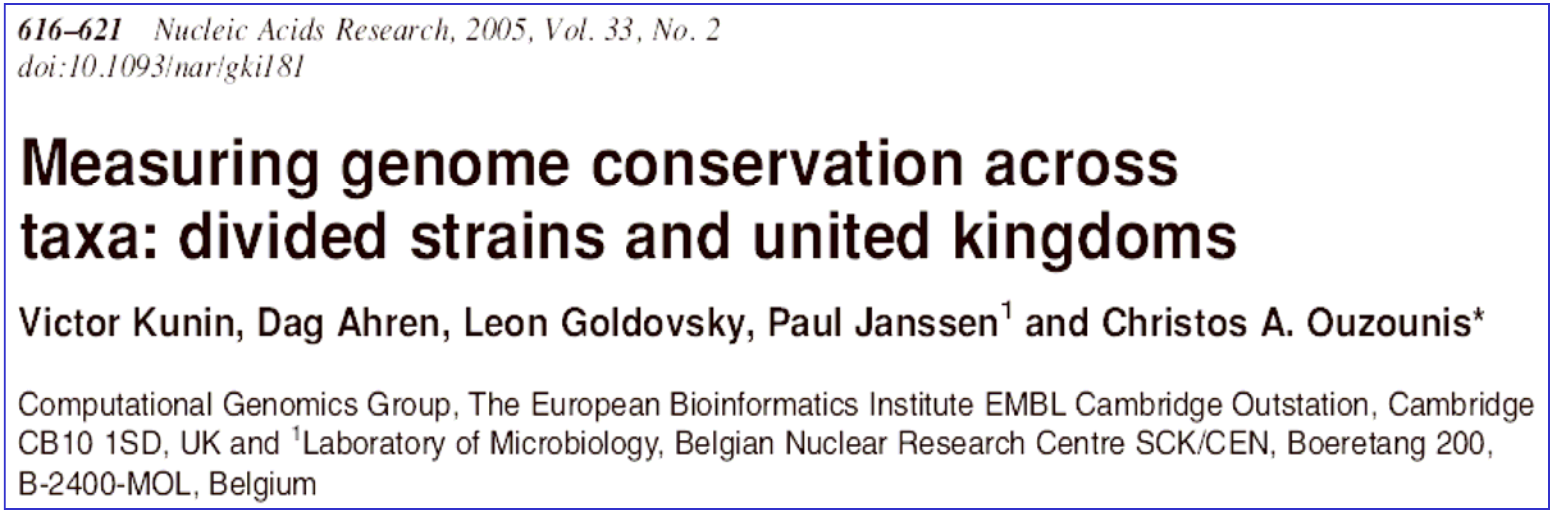
16S rRNA tree


0.2
H. pylori 26695
H. pylori HPAG1
H. pylori J99
H. acinonychis str. Sheeba
H. mustelae ATCC 43772
H. hepaticus ATCC 51449
Wolinella succinogenes DSM 1740
Sulfurimonas denitrificans ATCC 33889
Arcobacter butzleri 4018
C. fetus subsp. fetus 82-40
C. curvus 525.92
C. concisus 13826
C. lari RM2100
C. upsaliensis RM3195
C. coli RM2228
C. jejuni subsp. doylei 269.97
C. jejuni subsp. jejuni HB93-13
C. jejuni subsp. jejuni 81-176
C. jejuni subsp. jejuni 260.94
C. jejuni subsp. jejuni CG8486
C. jejuni RM1221
C. jejuni subsp. jejuni NCTC 11168
C. jejuni subsp. jejuni 84-25
C. jejuni subsp. jejuni CF93-6
Thiomicrospira crunogena XCL-2

Supertree based on 60 protein sequences



- Phylogeny of completely sequenced genomes of selected α-proteobacteria. The phylogeny is based on the concatenated sequences of 648 orthologous proteins. Neighbor-Joining method with % bootstrap support indicated. Scale indicates substitutions (Young et al. 2006. Genome Biol. 2006; 7(4): R34)
- Overall, a phylogeny based on all of these 648 proteins (Figure 7) is consistent with the species relationships inferred from 16S ribosomal RNA, in which the closest relative of *R. leguminosarum* is *A. tumefaciens*, followed by *S. meliloti*, and then *M. loti*. However, many individual proteins actually support different phylogenetic relationships.



616–621 Nucleic Acids Research, 2005, Vol. 33, No. 2
doi:10.1093/nar/gki181

Measuring genome conservation across taxa: divided strains and united kingdoms

Victor Kunin, Dag Ahren, Leon Goldovsky, Paul Janssen[1] and Christos A. Ouzounis*

Computational Genomics Group, The European Bioinformatics Institute EMBL Cambridge Outstation, Cambridge CB10 1SD, UK and [1]Laboratory of Microbiology, Belgian Nuclear Research Centre SCK/CEN, Boeretang 200, B-2400-MOL, Belgium

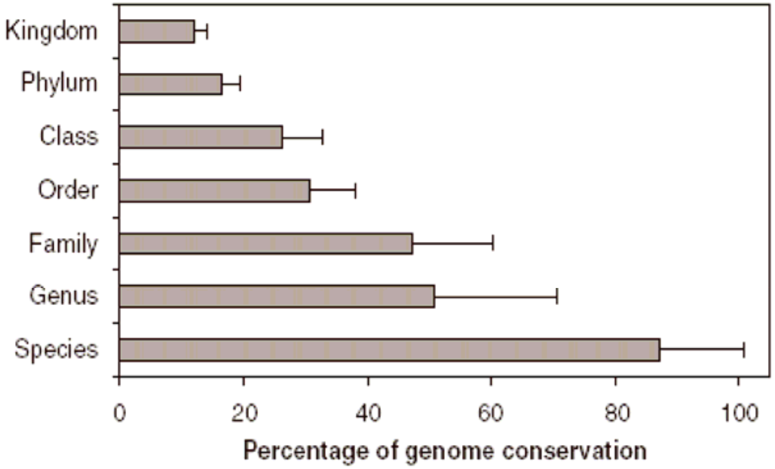


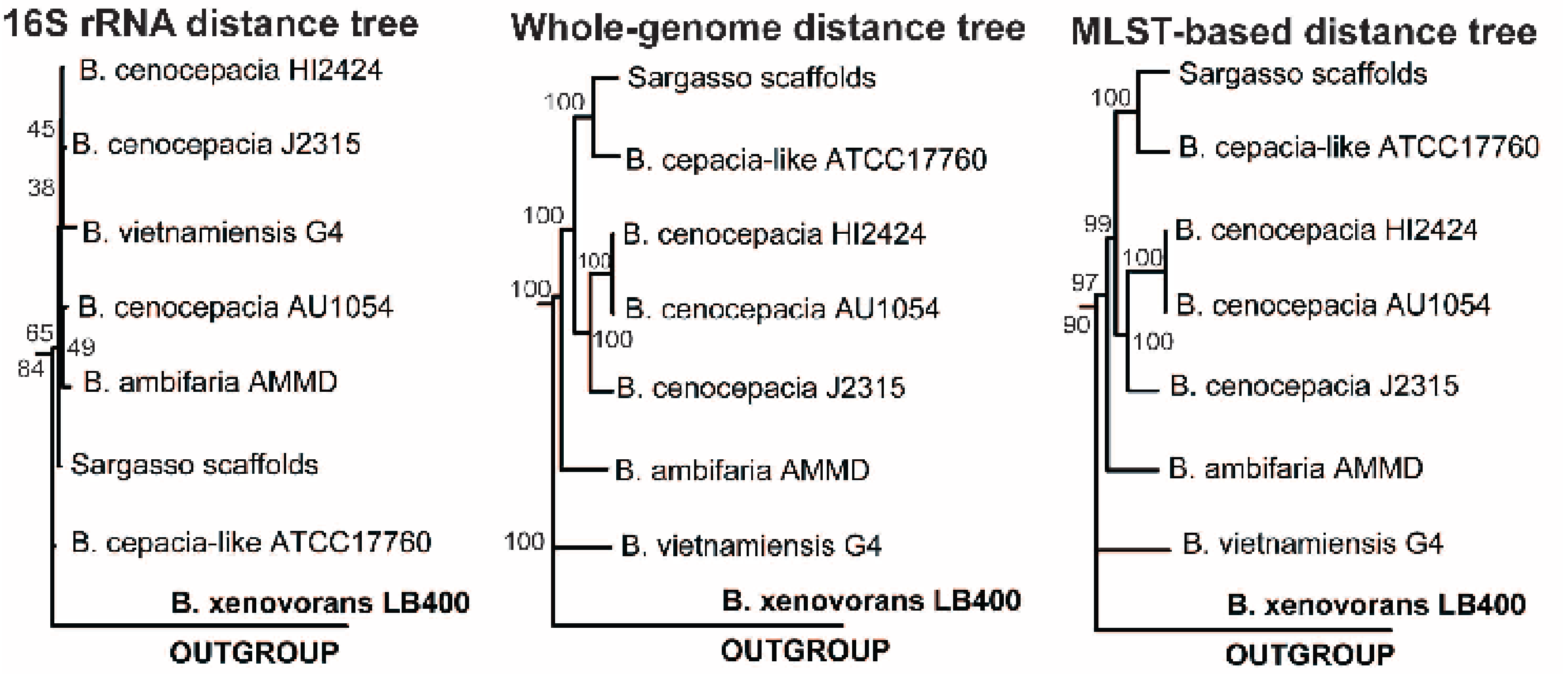
Figure 3. Genome conservation within bacterial taxonomic ranks. Error bars mark standard deviations. See text for discussion, genome conservation computed using D1 normalization (see Materials and Methods).

Now that we have access to whole-genome sequences: what do they tell us?

- Some basic taxonomic parameters are being confirmed: high DNA-DNA hybridisation levels and highly similar 16S rRNA gene sequences are reflected in the core genome content

Lack of throughput capacity???

(K. Konstantinidis, unpublished)



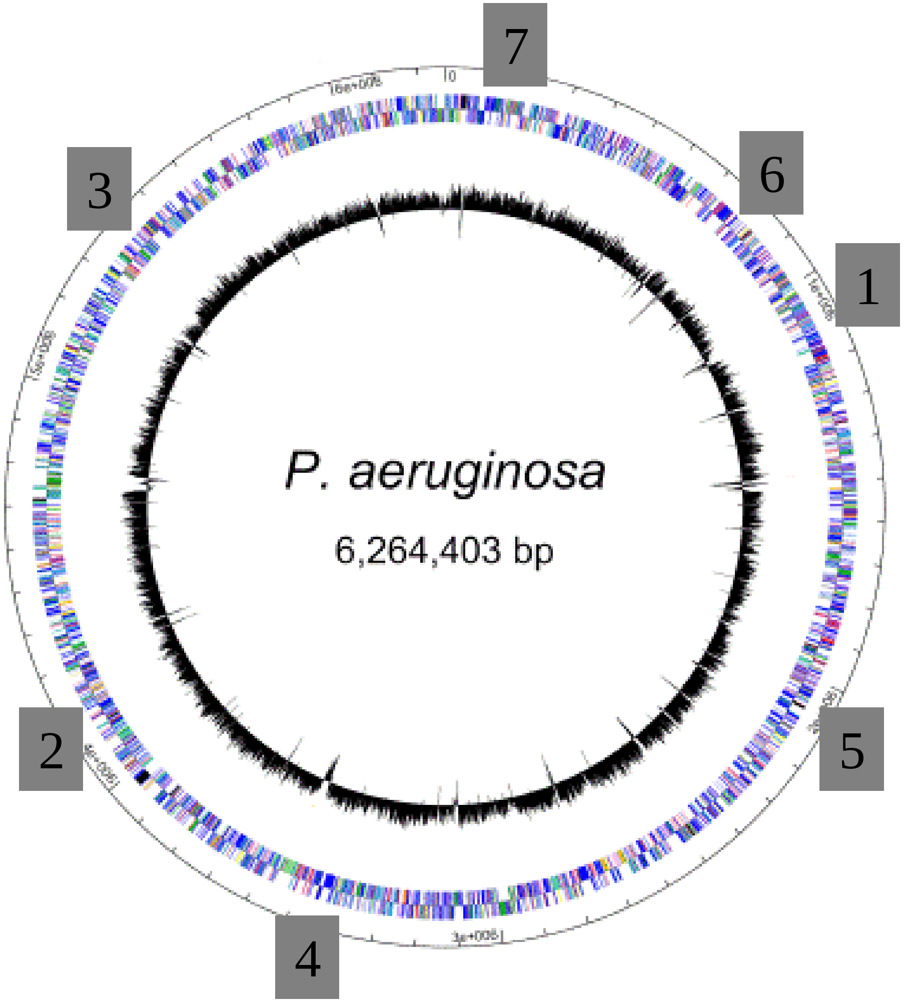
16S rRNA tree: DnaDist (Phylip package) of full-length 16S rRNA genes aligned with Clustalw. Nodes on the nodes denote statistical support by 500 bootstrap replicates.

Whole-genome-tree: Dnadist (phylip Package) of concatenated alignments (with Clustalw) of the 2,183 core genes. Nodes on the nodes denote statistical support by 100 bootstrap replicates.

MLST tree: DnaDist (Phylip package) of concatenated alignments (with Clustalw) of full-length RecA, GyrB, LepA, PhaB, TrpB, GtlB, GyrB. Nodes on the nodes denote statistical support by 500 bootstrap replicates.

- Konstantinidis et al., 2006. Towards a more robust assessment of intraspecies diversity using fewer genetic markers. AEM 72:7286-93

Selection of Seven Loci



- Acetyl-coenzyme A synthetase (acsA)
- GMP synthase (guaA)
- DNA mismatch repair protein (mutL)
- NADH dehydrogenase I chain C,D (nuoD)
- Phosphoenolpyruvate synthase (ppsA)
- Anthralite synthetase component I (trpE)
- Shikimate dehydrogenase (aroE)

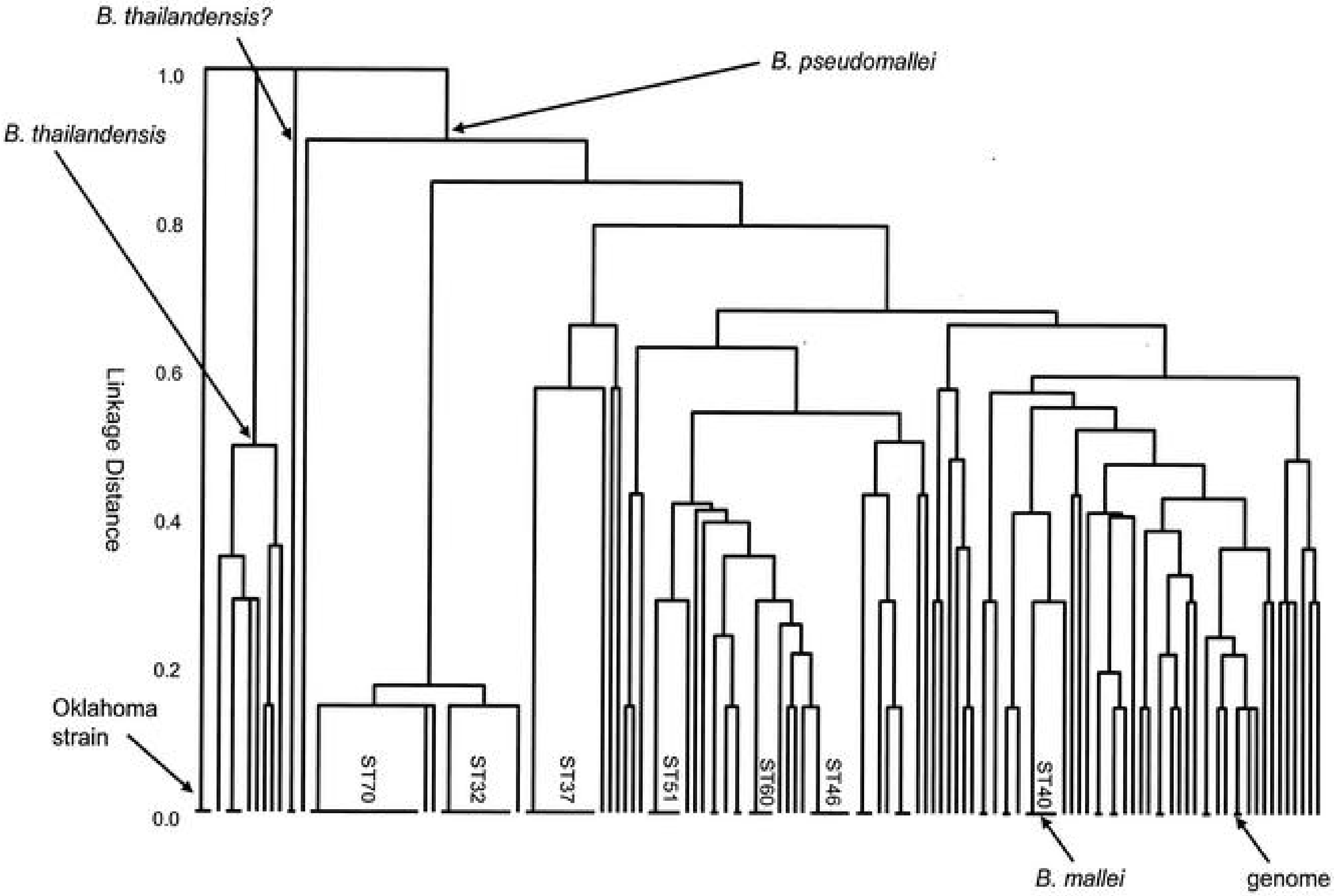
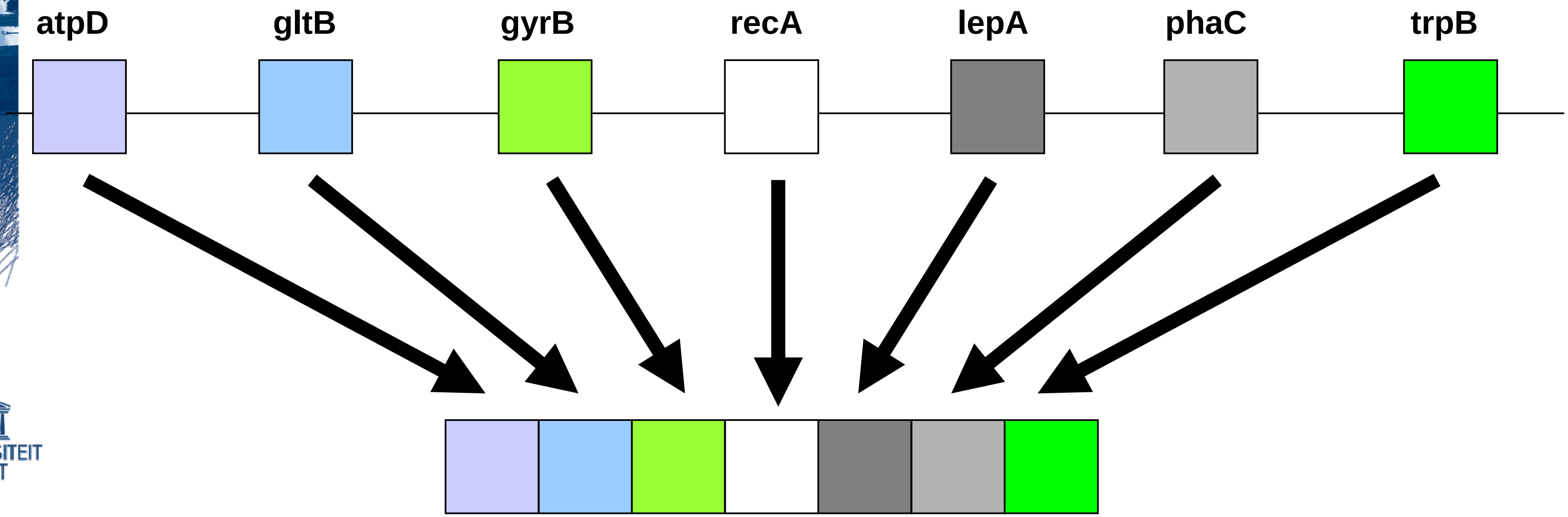


FIG 2.
Relationships among Burkholderia isolates. A UPGMA tree was constructed from the matrix of pairwise differences in the allelic profiles of the 147 Burkholderia isolates. The nodes from which all *B. pseudomallei* and *B. thailandensis* isolates descend are marked. The five *B. mallei* isolates (ST40) have identical allelic profiles and cluster among the *B. pseudomallei* isolates. Two isolates that were assigned to the species *B. pseudomallei* but which in this study were found to be closely allied with *B. thailandensis* (shown as *B. thailandensis*?) and three isolates from Oklahoma that originally were tentatively assigned to the species *B. pseudomallei* had divergent allelic profiles and differed from all *B. pseudomallei* and *B. thailandensis* isolates at all seven loci. The STs that include at least four isolates and the strain used to obtain the genome sequence (K96243) are shown (Godoy et al. J Clin Microbiol. 2003 May;41(5):2068-79)

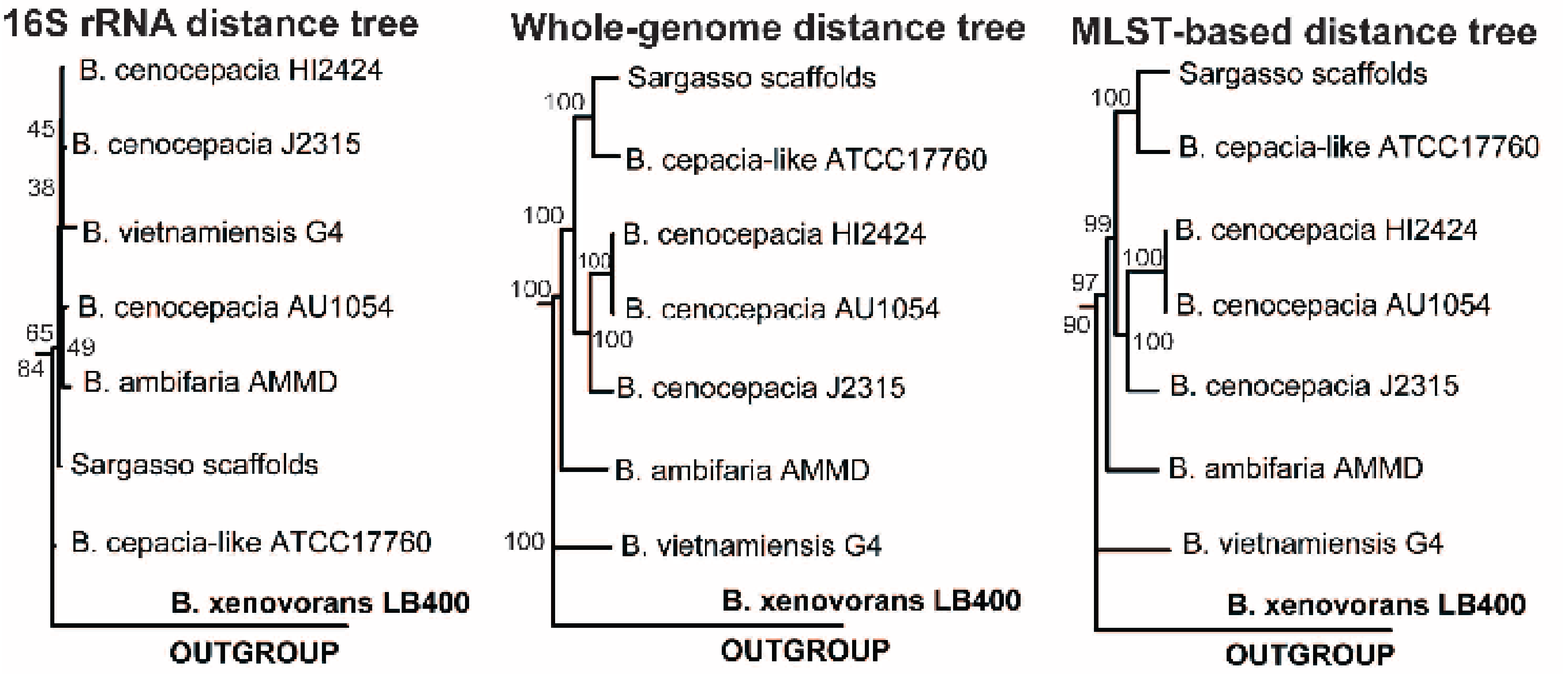
MLST loci - concatenated sequence analysis

Concatenation

The sequences of the seven loci are put end to end to form one large sequence which can be used in base pair comparisons.



(K. Konstantinidis, unpublished)



16S rRNA tree: DnaDist (Phylip package) of full-length 16S rRNA genes aligned with Clustalw. Nodes on the nodes denote statistical support by 500 bootstrap replicates.

Whole-genome-tree: Dnadist (phylip Package) of concatenated alignments (with Clustalw) of the 2,183 core genes. Nodes on the nodes denote statistical support by 100 bootstrap replicates.

MLST tree: DnaDist (Phylip package) of concatenated alignments (with Clustalw) of full-length RecA, GyrB, LepA, PhaB, TrpB, GtlB, GyrB. Nodes on the nodes denote statistical support by 500 bootstrap replicates.

- Konstantinidis et al., 2006. Towards a more robust assessment of intraspecies diversity using fewer genetic markers. AEM 72:7286-93

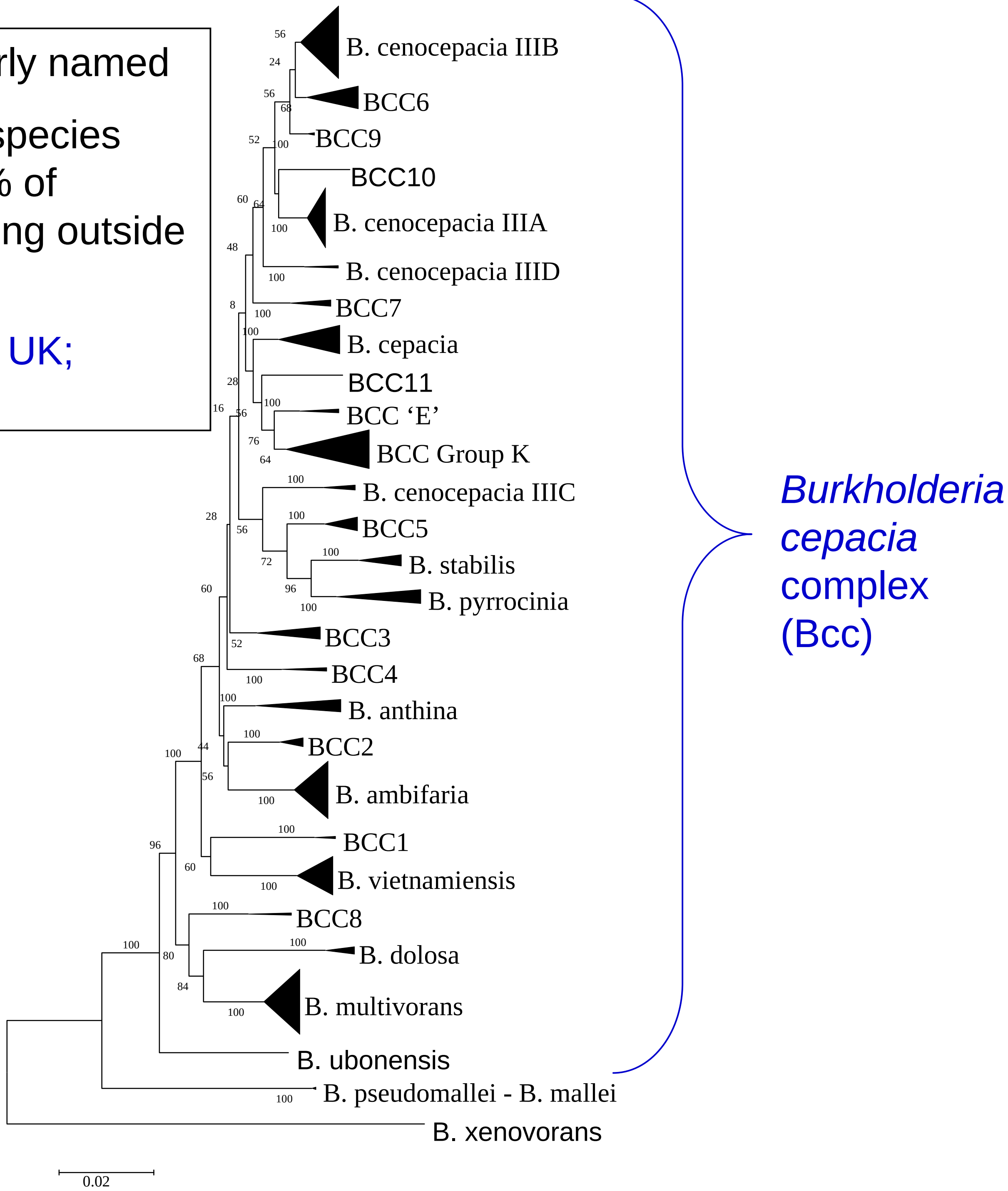
MLST/A is attractive...

- Reflects whole genome content similarity
- Compared to DNA-DNA hybridisation and 16S rRNA gene sequencing: increased resolution
- Fast (large biodiversity to explore): increased throughput capacity
- Portable (”online taxonomy”)

- 9 Bcc species formerly named

- about 15 novel Bcc species pending including 20% of isolates examined falling outside of the named species

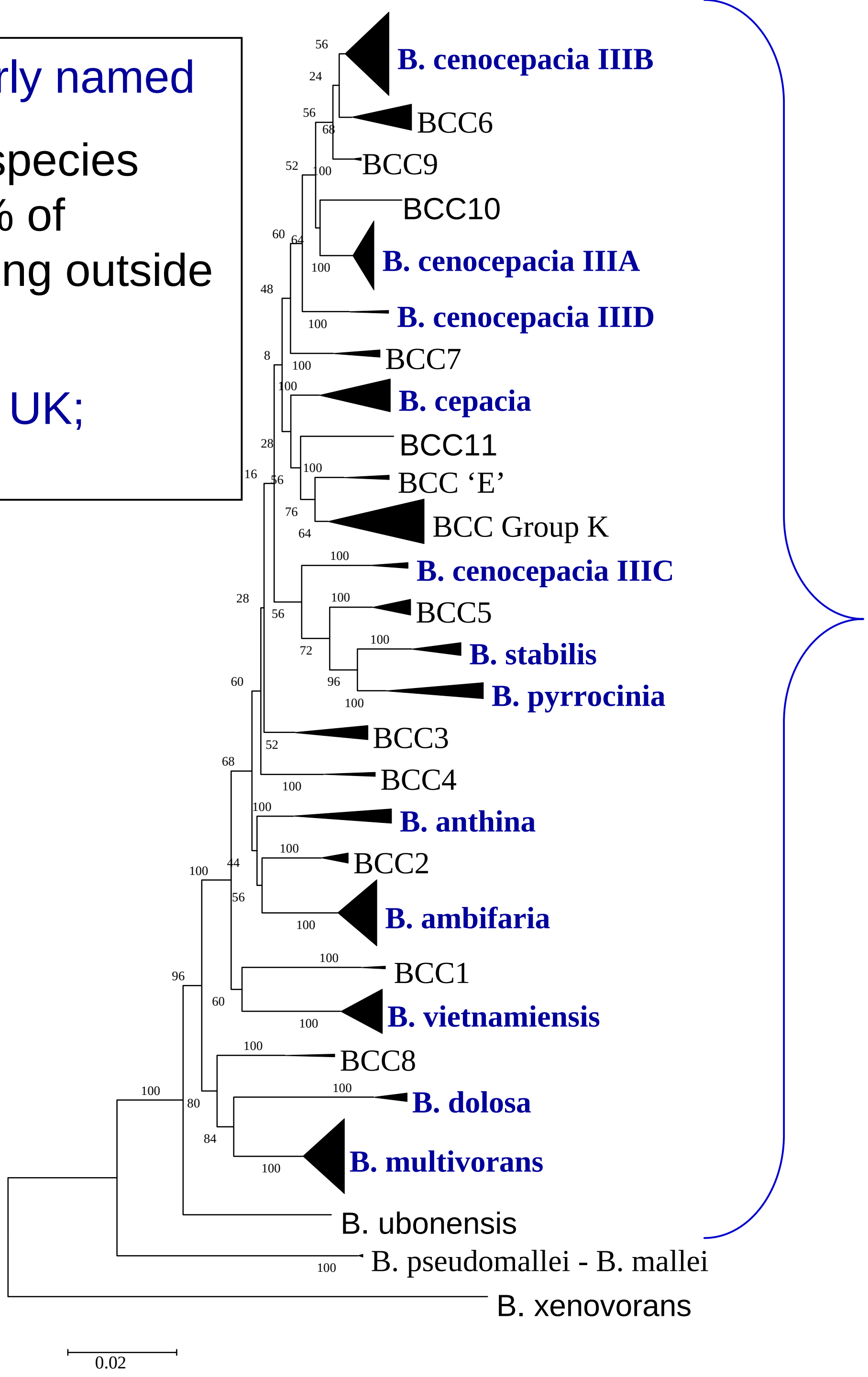
(A. Baldwin, Warwick, UK; unpublished)



- 9 Bcc species formerly named

- about 15 novel Bcc species pending including 20% of isolates examined falling outside of the named species

(A. Baldwin, Warwick, UK; unpublished)

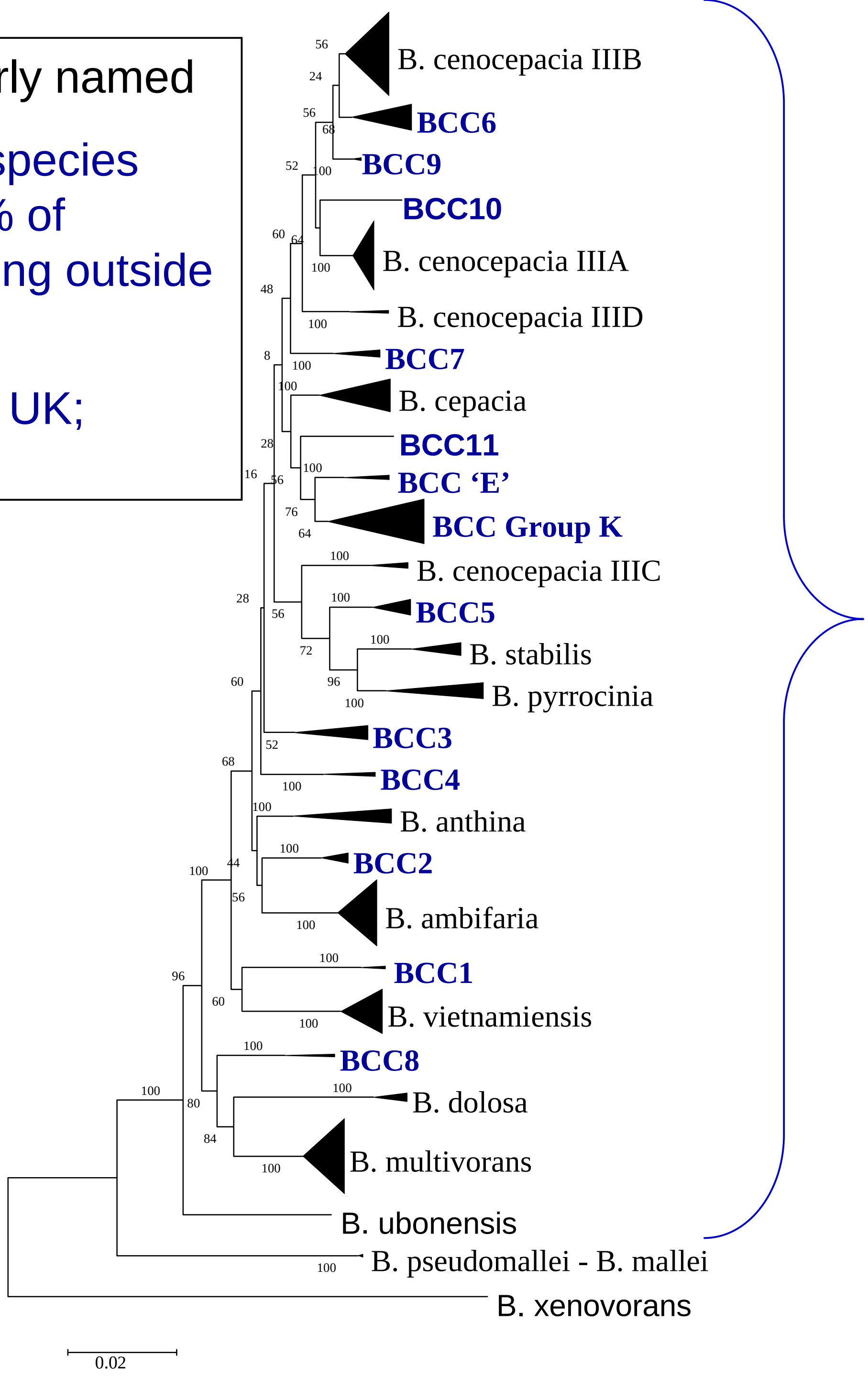


Burkholderia cepacia complex (Bcc)

- 9 Bcc species formerly named

- about 15 novel Bcc species pending including 20% of isolates examined falling outside of the named species

(A. Baldwin, Warwick, UK; unpublished)



Burkholderia cepacia complex (Bcc)

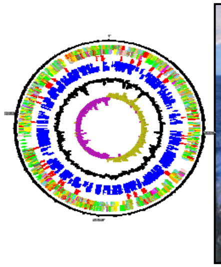
Manufacturer's Recall of Nasal Spray
Contaminated with
Burkholderia cepacia Complex

Industrial contaminant
(cultivated)

Clones of the same strain! (ST102)

Infectious for CF and Non-CF people 1999-2005
(cultivated)

Spanish sheep mastitis
(cultivated)



SAR1 Metagenome
(hypothetical, not cultivated)

Conclusion

- For two decades complete genome sequences have been considered the reference standard to determine phylogeny and taxonomy
- In spite of genome evolution, lateral gene transfer and recombination, genomes contain substantial information that seems mainly inherited vertically: the core genome
- The core genome varies in size between species but its total content is the most likely reference material for future genome based species definitions
- Sequence information as derived from shared core gene or protein sequences can be used to reconstruct organismal phylogeny and reflects 16S rRNA based schemes. It therefore has the potential to be used to construct an ordered scheme (‘taxonomy’) of prokaryotic diversity
- MLSA schemes have the potential to reflect relationships as imprinted in shared genome content and have a superior throughput capacity

Acknowledgements

- Dr. D. Gevers (UGent, Belgium & MIT, Cambridge, USA)
- Prof. Dr. T. Coenye (UGent, Belgium)

- http://www.asm.org/Academy/index.asp?bid=49252

"A classification that is of little use to microbiologists no matter how fine a scheme or who devised it, will soon be ignored or significantly modified"

Staley & Krieg, 1984