

Research article

Open Access

The reach of the genome signature in prokaryotes

Mark WJ van Passel*^{1,3,4}, Eiko E Kuramae¹, Angela CM Luyf², Aldert Bart³ and Teun Boekhout¹

Address: ¹Centraalbureau voor Schimmelcultures (CBS), Uppsalalaan 8, Utrecht, The Netherlands, ²Bioinformatics Laboratory, Academic Medical Center, University of Amsterdam, the Netherlands, ³Center for Infection and Immunity Amsterdam (CINIMA), Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, the Netherlands and ⁴Department of Biochemistry and Molecular Biophysics, University of Arizona, POBox 210088, Tucson, Arizona, USA

Email: Mark WJ van Passel* - mvpassel@email.arizona.edu; Eiko E Kuramae - kuramae@cbs.knaw.nl; Angela CM Luyf - a.c.luyf@amc.uva.nl; Aldert Bart - a.bart@amc.uva.nl; Teun Boekhout - tbo@cbs.knaw.nl

* Corresponding author

Published: 13 October 2006

Received: 19 July 2006

BMC Evolutionary Biology 2006, **6**:84 doi:10.1186/1471-2148-6-84

Accepted: 13 October 2006

This article is available from: <http://www.biomedcentral.com/1471-2148/6/84>

© 2006 van Passel et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: With the increased availability of sequenced genomes there have been several initiatives to infer evolutionary relationships by whole genome characteristics. One of these studies suggested good congruence between genome synteny, shared gene content, 16S ribosomal DNA identity, codon usage and the genome signature in prokaryotes. Here we rigorously test the phylogenetic signal of the genome signature, which consists of the genome-specific relative frequencies of dinucleotides, on 334 sequenced prokaryotic genome sequences.

Results: Intrageneric comparisons show that in general the genomic dissimilarity scores are higher than in intraspecific comparisons, in accordance with the suggested phylogenetic signal of the genome signature. Exceptions to this trend, (*Bartonella* spp., *Bordetella* spp., *Salmonella* spp. and *Yersinia* spp.), which have low average intrageneric genomic dissimilarity scores, suggest that members of these genera might be considered the same species. On the other hand, high genomic dissimilarity values for intraspecific analyses suggest that in some cases (e.g. *Prochlorococcus marinus*, *Pseudomonas fluorescens*, *Buchnera aphidicola* and *Rhodospseudomonas palustris*) different strains from the same species may actually represent different species. Comparing 16S rDNA identity with genomic dissimilarity values corroborates the previously suggested trend in phylogenetic signal, albeit that the dissimilarity values only provide low resolution.

Conclusion: The genome signature has a distinct phylogenetic signal, independent of individual genetic marker genes. A reliable phylogenetic clustering cannot be based on dissimilarity values alone, as bootstrapping is not possible for this parameter. It can however be used to support or refute a given phylogeny and resulting taxonomy.

Background

With the availability of an ever-increasing number of whole genome sequences, evolutionary history can be analysed genetically on a more comprehensive level.

Among microorganisms, phylogenetic relationships have traditionally been defined by phenotypic characters. With current comparative genomics, evolutionary distances may be inferred more thoroughly, independent of varia-

ble expression profiles or morphological characteristics. Recently, a number of studies suggested taxonomy should be based on whole genomic DNA content in order to more properly reflect evolutionary relationships [1-4]. These studies concur with previous observations by Coenye and Vandamme, who tested phylogenetic consistency of a number of parameters in the lactic acid bacteria, including 16S rDNA identity, genome synteny, identity of several house-keeping genes, codon usage and the genome signature similarity. They concluded that these genomic parameters largely agree with each other [5].

The latter parameter, the genome signature, is a compositional parameter reflecting the dinucleotide relative abundance values, which are similar between closely related species, and dissimilar between non-related species [6-8]. The difference in genome signature between two sequences is expressed by the genomic dissimilarity (δ^*), which is the average absolute dinucleotide relative abundance difference between two sequences. This parameter is suitable for intragenomic detection of putative horizontally acquired sequences [7,9], and recently an online tool was made available to detect compositional dissimilarity between input sequences and representative whole genome sequences [10], which has shown for example that plasmids are more dissimilar in composition compared to their host genomes than previously anticipated [11], and that genomic islands can be clustered by similarities in dinucleotide composition [12]. Although genome signature analyses have been used for both intragenomic comparisons [7,10-12] and for intergenomic analyses [5,8,13,14], these analyses were predominantly performed on relatively few, or only partial, prokaryotic genomes. While genome signatures had been analysed prior to the genomic era [6], intra- and intergenomic comparisons based on the genome signature between eukaryotes have been scarce [15]. However, a recent study on several eukaryotic genome sequences (i.e. from human, mouse, rat, fruit fly and nematode) based on short range correlations in DNA sequences, showed that in most cases, chromosomes originating from the same species cluster together [16].

In order to assess the applicability of the genome signature in phylogenetic analyses in general, we performed a comprehensive interchromosomal comparison of prokaryotic genome signatures. Using 334 prokaryotic genome sequences, we tested the congruence of the genome signature with the current genus and species nomenclature, and compared the signal of the genome signature to the phylogenetic signal of 16S rDNA identity scores.

Results

Prokaryotic whole genome signature comparisons

We calculated the genomic dissimilarity values (δ^*) between all 334 available prokaryotic genome sequences available at the time of analysis (May, 2006). We partitioned this large dataset to three smaller subsets to investigate specific topics (see below, see Additional Files 1, 2, 3, 4): the δ^* values between chromosomes from the same genus (containing 36 bacterial and 4 archaeal genera: 130 different species in total), the δ^* values between chromosomes from the same species (containing 33 bacterial species: 111 different strains in total), and δ^* values of prokaryotes with multiple chromosomes (7 different genera, 16 different species). Genera are defined as organisms with the same genus name, and species are defined as organisms with the same genus name and the same specific designation, with two exceptions: *Shigella* spp., which are considered the same species as *Escherichia coli* [17], and the *Bacillus cereus* cluster, consisting of *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringensis*, which are also considered the same species [18,19]. Indeed, for both these exceptions, intraclade δ^* values are very low as mentioned previously [11,13].

Intragenomic compositional comparison of whole genomes

Average intragenomic δ^* values for 40 different genera consisting of both Archaea and Bacteria (Additional File 1) are depicted in Figure 1, indicating a large variation in the genome signature, with an average intragenomic δ^* score of 0.046. All chromosomes from different species of the same genus were compared for these analyses, but intraspecific comparisons were carried out separately. Extreme δ^* values of intergeneric comparisons were highlighted in red (high δ^*) and blue (low δ^*).

The four genera with very high intragenomic δ^* values are *Chlorobium*, *Helicobacter*, *Mycoplasma* and *Treponema*. Although the *Chlorobium* genome sequences are available, the sequences are unpublished, and it is therefore unclear why the genome signature between species of the same genus would differ so greatly. From the general genomic attributes at NCBI [20], it is obvious however that the genomes of *Chlorobium chlorochromatii* and *Chlorobium tepidum* differ substantially in GC% content (44.3% and 56%, respectively), and to a lesser extent in genome size (2.57 Mbp and 2.15 Mbp, respectively), suggesting a substantial phylogenetic distance. This however conflicts with the 16S rDNA sequence identity of 92%.

The reason for the high genomic dissimilarities between the genomes of *Helicobacter hepaticus* and the two *Helicobacter pylori* strains are possibly due to substantial differences in gene complement (possibly due to horizontal gene transfer) between members of this genus [21]. A large phylogenetic distance between these species has

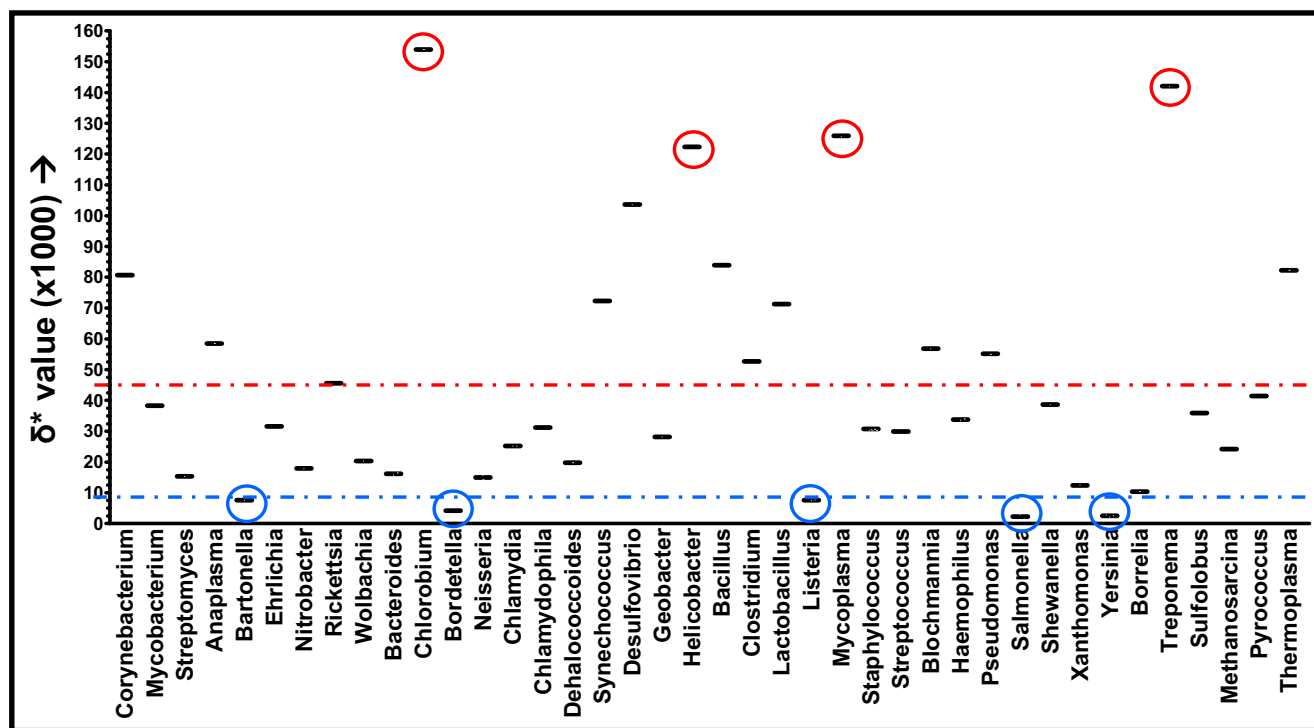


Figure 1

Intrageneric comparison of the genome signature within 40 prokaryotic genera. Species with high genomic dissimilarity scores are in red circles (11: *Chlorobium*, 20: *Helicobacter*, 25: *Mycoplasma* and 36: *Treponema*) species with low genomic dissimilarity scores are in blue circles (5: *Bartonella*, 12: *Bordetella*, 31: *Salmonella* and 34: *Yersinia*). The average genomic dissimilarity between different species of the same genus is 0.046 (red line). The blue line represents the average intraspecific genomic dissimilarity from figure 2 ($\delta^* = 0.009$).

been observed previously [22,23], and whole genome analyses by Dewhirst and co-workers support a phylogeny that places *H. hepaticus* more closely related to *Wolinella succinogenes* than to *Helicobacter pylori* [24]. The latter study actually supports HGT of 16S sequences, thus obscuring single locus-based phylogenies for this genus.

The *Mycoplasma* spp. (including *Ureaplasma parvum*) and the *Treponema* spp. show low intrageneric 16S rDNA sequence identity scores (between 70–98.3% sequence identity and 89.1% identity, respectively), which had already been noted for a subset of the *Mycoplasma* spp. genome sequences [8,13]. These two genera also show large intrageneric phylogenetic distances in a Tree of Life study based on universally conserved genes [22].

The four genera with very low intrageneric δ^* values are *Bartonella* (0.0077), *Bordetella* (0.0042), *Salmonella* (0.0023) and *Yersinia* (0.0025). These values are lower than the average intraspecific δ^* values (see below). All species that constitute these four genera have nearly identical 16S rDNA sequences (i.e. show 98.9%, 99.8%, 98.2% and 99.9% intrageneric identity, respectively), thus

corroborating the established close, possibly intraspecific, evolutionary relationship.

Intraspecific compositional comparisons of whole genomes

Although no archaeal genome sequences of the same species are currently available, for 33 bacterial species there are two or more genome sequences available (Additional File 2), allowing the detection of variation of δ^* of genome sequences from the same species. The average intraspecific genomic dissimilarity values are depicted in Figure 2, with an average intraspecific δ^* of 0.009. Four clear outliers are detected with substantially higher δ^* values (red circles in Figure 2), originating from intraspecific comparisons between the genomes from *Rhodopseudomonas palustris*, *Prochlorococcus marinus*, *Buchnera aphidicola* and *Pseudomonas fluorescens* (intraspecific δ^* values of 0.0295, 0.0639, 0.0396 and 0.0822, respectively).

Except for *B. aphidicola*, for which the intraspecific 16S rDNA sequence identity is low for members of the same species (91%), the other 16S rDNA sequence identities are >96%. However, for the *P. marinus* genomes, the range of

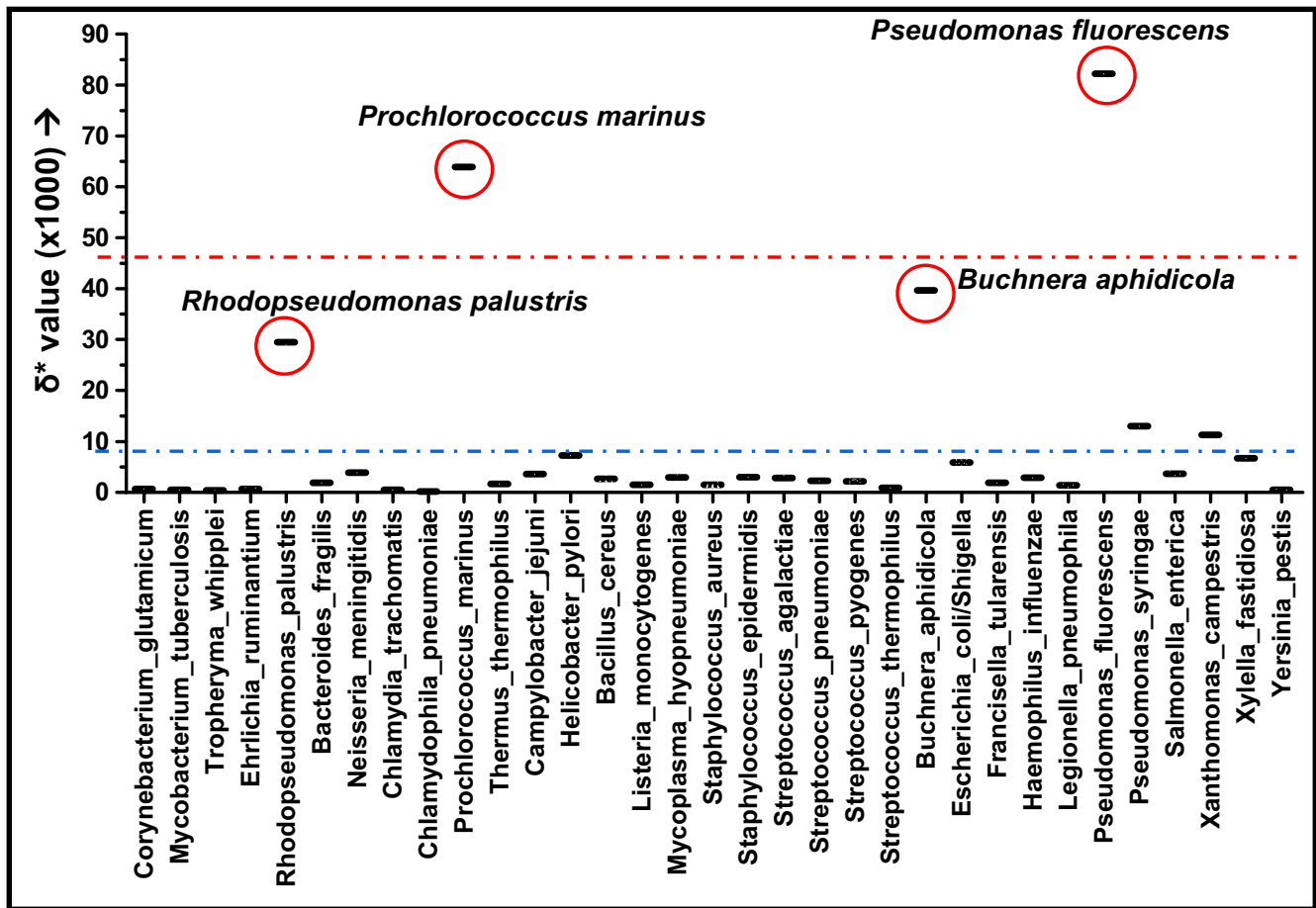


Figure 2
 Intraspecific comparisons of genome signature of 33 bacterial species. Species with high genomic dissimilarity scores are in red circles. The average genomic dissimilarity between genomes of the same species is 0.009 (blue line). The red line depicts the average intragenomic genomic dissimilarity from Figure 1 ($\delta^* = 0.046$).

genomic GC percentages is considerable (30.8–50.7% GC), and the genome sizes differ substantially (1.66–2.41 Mbp), suggesting that these organisms may not be the same species, they contain similar 16S rDNA sequences due to HGT of these loci or that acquisition of large amounts of anomalous DNA plays an important role in the genome organization and nucleotide composition.

For both *P. fluorescens* genomes, a difference in genome size is observed (>600 Mbp), but the difference in GC percentage is relatively small.

Congruence between genomic dissimilarity values and 16S rDNA identity values of different bacterial species

It had been suggested previously that various genomic parameters are congruent in their phylogenetic signal [5]. We compared eight sets of genomic dissimilarity values with 16S rDNA sequence identity scores between eight different groups of species; one group containing the *B.*

cereus cluster, and seven groups of several related species (the *Chlamydia/Chlamydophila* clade, the genus *Mycoplasma*, the genus *Staphylococcus*, the *E. coli/Shigella/Salmonella* clade, the genus *Mycobacterium*, the genus *Rickettsia*, the genus *Lactobacillus* and the genus *Streptococcus*; seven groups in total. see also Additional File 3). An inverse correlation between these two parameters is observed (Fig. 3, note the differences in scale of the 16S rDNA identity and δ^*), although incongruities are visible (e.g. within the *E. coli/Shigella/Salmonella* clade).

Whole genome comparisons of prokaryotic chromosomes of species with multiple chromosomes

Sixteen prokaryotic species contain more than one chromosome and the δ^* values between the two chromosomes of the same strains are illustrated in Figure 4 (for the species list, see Additional File 4). On average, the δ^* value between two chromosomes from the same species is 0.0156. Four outliers are circled; the two higher outliers

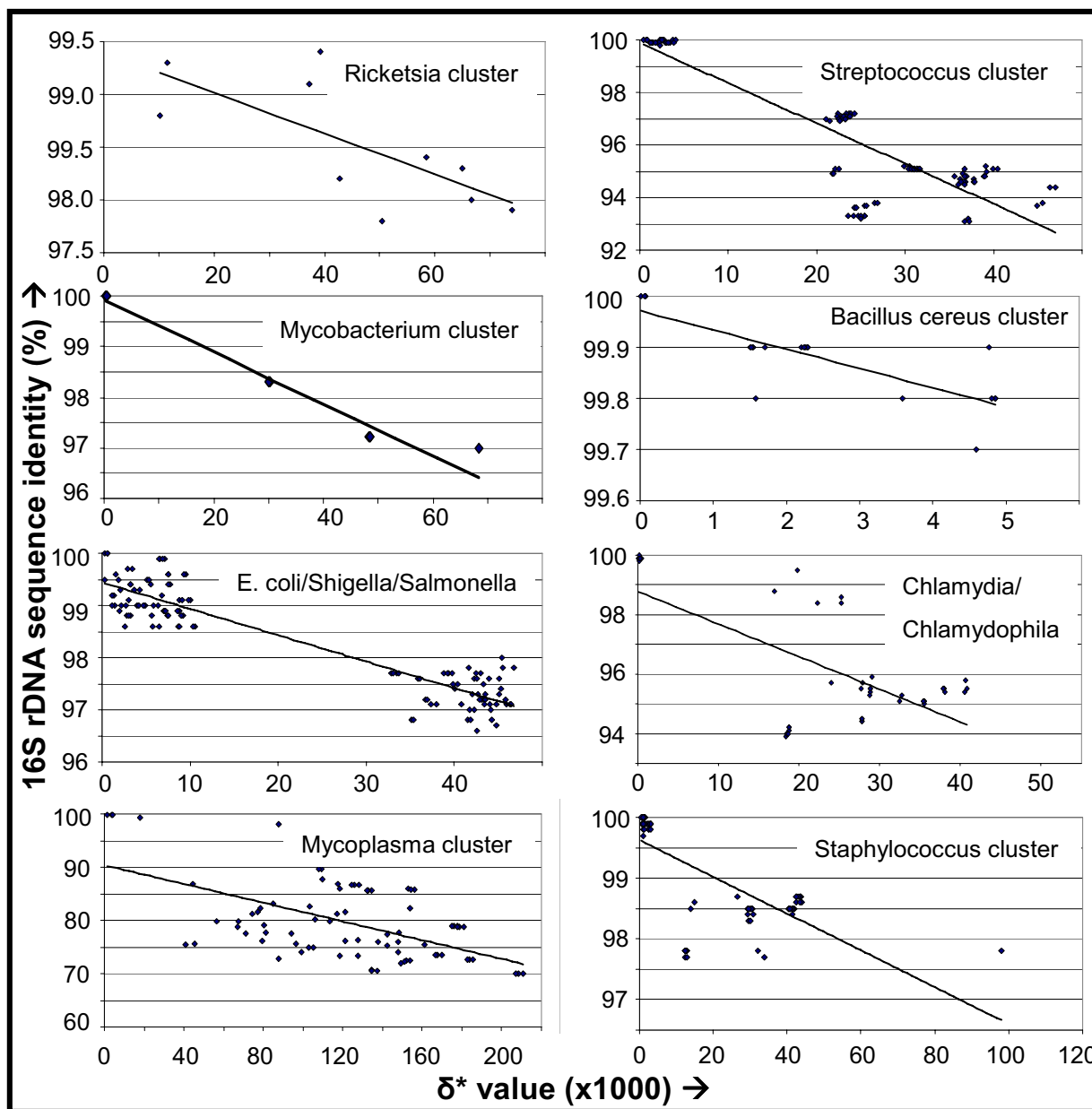


Figure 3
 Congruence between 16S rDNA sequence identity (y-axis) and δ^* (x-axis) within different groups (note the different scales on the axes). See Additional File 3 for more information.

are comparisons between chromosomes 1 and 3 from *Burkholderia* spp. 383 and chromosomes 1 and 3 from *Burkholderia xenovorans*, indicating a very high genomic dissimilarity between these chromosomes and thereby suggesting an exogenous origin for chromosome 3 in these species (δ^* is 0.0289 and 0.0296 for *B.* spp. 383 and *B. xenovorans*, respectively). The two lower outliers are the two different chromosome I and chromosome II comparisons from the two *Leptospira interrogans* species. The

intragenomic δ^* value is 0.0059 for *L. interrogans* serovar Copenhageni and 0.0047 for *L. interrogans* serovar Lai suggesting a longer shared evolutionary history.

Conclusion

Traditionally, the assignment of taxonomical appropriate names to microbes was based on phenotypic characters, such as Gram stains or the possession of cell walls. Currently, with over 350 whole genomes sequenced, there is

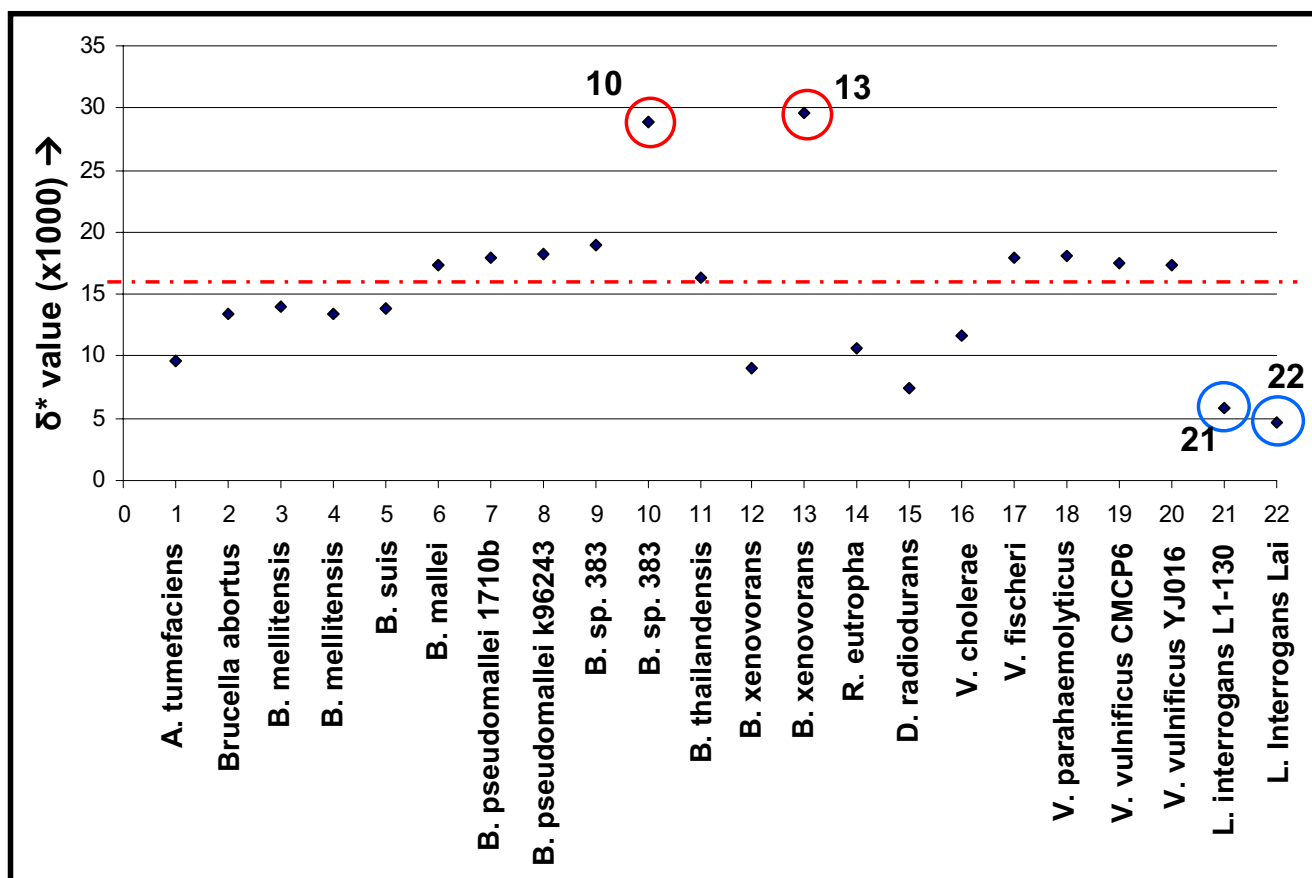


Figure 4
 Intragenomic comparisons of the genome signature of 20 bacterial species (see Additional File 4). Highlighted are 10 (*Burkholderia* sp. 383 chromosome I vs. 3), 13 (*Burkholderia xenovorans* LB400 chromosome I vs. 3), 21 (*Leptospira interrogans* serovar Copenhageni str. Fiocruz LI-130 chromosome I vs. II) and 22 (*Leptospira interrogans* serovar Lai str. 5660I chromosome I vs. II). The average genomic dissimilarity between chromosomes of the same species is 0.016 (red line).

ongoing debate to re-evaluate the prokaryotic species definition [2,25]. One such attempt, considering different genomic parameters, was performed by Coenye and Vandamme [5], who compared different phylogenetic approaches on the lactic acid bacteria. They concluded that the different approaches agreed well, although these do not necessarily provide much additional information about phylogenetic relationships. In our study, we analysed the overall consistency of the phylogenetic signal of the genome signature in 334 prokaryotic. We tested the congruence between the δ^* intragenerically, intraspecifically, and the δ^* values with their corresponding 16S rDNA sequence identity values.

For some species (*E. coli*/*Shigella* group, *Bacillus cereus* cluster) it was known that they are probably the same species, and the low values for δ^* corroborate this [11,13]. Comparison of the intrageneric δ^* (Fig. 1) and intraspecific δ^* (Fig. 2) shows various intrageneric values to be

well in the range of intraspecific values, suggesting that there are more clusters that may actually constitute one species. This is the case for the different *Bartonella* spp., *Bordetella* spp., *Salmonella* spp. and *Yersinia* spp., This may also hold true for *M. bovis* and *M. tuberculosis*, and the *L. innocua* and *L. monocytogenes* species. For these six groups, the 16S rDNA data support the notion of very close phylogenetic relationships.

In contrast, four extreme intraspecific δ^* values are within the intrageneric range. The different *B. aphidicola* species display high genomic dissimilarity values as well as low 16S rDNA sequence identity scores, suggesting these might actually be different species. This is in agreement with an estimated divergence time of over 150 million years [26]. The reason why the different species of *R. palustris*, *P. marinus* and *P. fluorescens* display high δ^* values, while the ribosomal sequences of the individual species are nearly identical remains unclear, although very long

branches have been observed between members of these species in different phylogenetic studies [22,23]. It is of note that between the different *P. marinus* genomes substantial differences in size and GC percentage are observed.

Generally, an inverse relation between δ^* and 16S is found, but the perceived resolution of this relation seems low and therefore δ^* values alone seem insufficient to infer reliable phylogenetic relationships. Also, it is not possible to infer a reliable phylogenetic clustering based on distance matrices as bootstrapping is not possible.

For some prokaryotic species containing multiple chromosomes it had been suggested that the secondary chromosome may have been acquired via horizontal gene transfer [12,27]. We find that the genomic dissimilarity between the two primary chromosomes in bacteria is generally low, but it is higher than the genomic dissimilarity between chromosomes from the same species, supporting the HGT hypothesis. Intragenomic dispersal of DNA can ameliorate the dissimilarity in genome signature, obscuring compositional dissimilarities over time [12]. A consequence of this is that different chromosomes found in metagenomic analyses can not readily be grouped into genomes for prokaryotes, though this is a minor problem as most prokaryotes have single chromosome genomes. We find that each intragenomic comparison of the two chromosomes of the different *Vibrio* genomes yields a higher δ^* value than the average of $\delta^* = 0.009$ in intraspecific comparisons (data not shown). Secondary chromosomes are present in all sequenced genomes of *Vibrio* spp., and if they have been present in each genome since the split of the different *Vibrio* spp. the different chromosomes in each genome should have had ample time to ameliorate towards more similar dinucleotide frequencies. The fact that the different chromosomes of each *Vibrio* genome are still dissimilar from each other in composition may be caused by an instable chromosome II, which is known to be less well-conserved between the different *Vibrio* species than chromosome I [28].

The precise origin of the genome signature is still unknown. For the GC percentage it has been suggested that certain environmental conditions shape the nucleotide composition [29]. This has also been found to be the case for the genome signature [30], although the exact effect of different conditions on different genome sequences remains unknown. It is likely that mutational pressures direct the shape of the genome signature, but the fact that secondary chromosomes in most cases remain dissimilar from the primary chromosomes underscores our lack of understanding of the factors that shape the nucleotide composition.

In conclusion, the genome signature is more similar between closely related species, and increases with larger phylogenetic distances, but this relation seems inadequate to infer phylogenetic relationships by itself. Unfortunately, distance matrices based on single values, as is the case with δ^* scores, are not amenable to bootstrapping, so robust phylogenetic analysis can not be inferred from δ^* values for prokaryotes. This parameter does however have a strong phylogenetic signal and can therefore be used to support or contradict a given phylogeny and resulting taxonomy. The combination of δ^* and 16S rDNA data given above for *Mycobacteria*, *Listeria*, *Prochlorococcus* and *Buchnera* provide convincing evidence for a re-evaluation of these taxonomic relationships. Also, if there are no additional ways to infer relationships (e.g. in the absence of comparable markers, as with multiple chromosomes in metagenomic analyses) the genome signature may help to cluster chromosomes, although the intragenomic δ^* scores can be relatively high in multichromosomal genomes from prokaryotes.

Methods

Sequences and data analyses

Prokaryotic chromosomal sequences were obtained from NCBI [20] (Additional Files 1 to 4, August 2005). Genome signature comparisons were performed using the online tool Compare_Islands [10,12,22], and all prokaryote genome comparisons are available at the Compare_Islands website [31]. The 16S rDNA sequences were obtained from the NCBI website [20] and trimmed after alignment by using the MEGA3 software package [32]. Identity percentages were calculated using MATGAT with the BLOSUM50 scoring matrix [33].

Authors' contributions

MWJvP, EEK and TB conceived the study and carried out the analyses. ACML carried out the bioinformatical analyses. MWJvP, EEK, AB and TB analysed the data and wrote the manuscript. All authors have read and approved the final manuscript.

Additional material

Additional File 1

Prokaryotic intragenomic genome signature comparisons.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-84-S1.pdf>]

Additional File 2

Prokaryotic intraspecific genome signature comparisons (the δ^ scores are available from the large matrix at <http://deltarho.amc.nl>).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-84-S2.pdf>]

Additional File 3

Prokaryotic intragenomic genome signature comparisons. In these matrices; the 16S rDNA sequence identity (in %) is given in the top-right half, and the δ^* ($\times 1000$) are given in the lower-left part.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-84-S3.pdf>]

Additional File 4

Prokaryotic intragenomic genome signature comparisons for species that carry more than 1 chromosome. The numbers correspond to the comparisons depicted in Figure 4.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-6-84-S4.pdf>]

Acknowledgements

MWJvP, EEK and TB have received funding from the Renewal Fund of the Royal Netherlands Academy of Arts and Sciences (RNAAS – KNAW). Also, the authors would like to thank the two reviewers for their helpful comments.

References

- Coenye T, Gevers D, Van de Peer Y, Vandamme P, Swings J: **Towards a prokaryotic genomic taxonomy.** *FEMS Microbiol Rev* 2005, **29(2)**:147-167.
- Konstantinidis KT, Tiedje JM: **Genomic insights that advance the species definition for prokaryotes.** *Proc Natl Acad Sci U S A* 2005, **102(7)**:2567-2572.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J: **Opinion: Re-evaluating prokaryotic species.** *Nat Rev Microbiol* 2005, **3(9)**:733-739.
- Kuramae EERVSBWMBT: **Phylogenomics reveal a robust fungal tree of life.** *FEMS Yeast Research* 2006;(doi: 10.1111/j.1567-1364.2006.00119.x).
- Coenye T, Vandamme P: **Extracting phylogenetic information from whole-genome sequencing projects: the lactic acid bacteria as a test case.** *Microbiology* 2003, **149(Pt 12)**:3507-3517.
- Karlin S, Burge C: **Dinucleotide relative abundance extremes: a genomic signature.** *Trends Genet* 1995, **11(7)**:283-290.
- Karlin S: **Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes.** *Trends Microbiol* 2001, **9(7)**:335-343.
- Sandberg R, Branden CI, Ernberg I, Coster J: **Quantifying the species-specificity in genomic signatures, synonymous codon choice, amino acid usage and G+C content.** *Gene* 2003, **311**:35-42.
- Hsiao W, Wan I, Jones SJ, Brinkman FS: **IslandPath: aiding detection of genomic islands in prokaryotes.** *Bioinformatics* 2003, **19(3)**:418-420.
- van Passel MW, Luyf AC, van Kampen AH, Bart A, van der Ende A: **{delta}{rho}-Web, an online tool to assess composition similarity of individual nucleic acid sequences.** *Bioinformatics* 2005, **21(13)**:3053-3055.
- van Passel MW, Bart A, Luyf AC, van Kampen AH, van der Ende A: **Compositional discordance between prokaryotic plasmids and host chromosomes.** *BMC Genomics* 2006, **7(1)**:26.
- van Passel MW, Bart A, Thygesen HH, Luyf AC, van Kampen AH, van der Ende A: **An acquisition account of genomic islands based on genome signature comparisons.** *BMC Genomics* 2005, **6(1)**:163.
- van Passel MWJ, Bart A, van der Ende A: **Default taxonomy and the genomics era.** *Microbiology* 2005, **151(9)**:2818-2820.
- Campbell A, Mrazek J, Karlin S: **Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA.** *Proc Natl Acad Sci U S A* 1999, **96(16)**:9184-9189.
- Gentles AJ, Karlin S: **Genome-scale compositional comparisons in eukaryotes.** *Genome Res* 2001, **11(4)**:540-546.
- Dehnert M, Plaumann R, Helm WE, Hutt MT: **Genome phylogeny based on short-range correlations in DNA sequences.** *J Comput Biol* 2005, **12(5)**:545-553.
- Pupo GM, Lan R, Reeves PR: **Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics.** *Proc Natl Acad Sci U S A* 2000, **97(19)**:10567-10572.
- Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto: **Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis--one species on the basis of genetic evidence.** *Appl Environ Microbiol* 2000, **66(6)**:2627-2630.
- Cherif A, Brusetti L, Borin S, Rizzi A, Boudabous A, Khyami-Horani H, Daffonchio D: **Genetic relationship in the 'Bacillus cereus group' by rep-PCR fingerprinting and sequencing of a Bacillus anthracis-specific rep-PCR fragment.** *J Appl Microbiol* 2003, **94(6)**:1108-1119.
- National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/Genomes/>].
- Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M: **Gain and loss of multiple genes during the evolution of Helicobacter pylori.** *PLoS Genet* 2005, **1(4)**:e43.
- Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P: **Toward automatic reconstruction of a highly resolved tree of life.** *Science* 2006, **311(5765)**:1283-1287.
- Gophna U, Doolittle WF, Charlebois RL: **Weighted genome trees: refinements and applications.** *J Bacteriol* 2005, **187(4)**:1305-1316.
- Dewhirst FE, Shen Z, Scimeca MS, Stokes LN, Boumenna T, Chen T, Paster BJ, Fox JG: **Discordant 16S and 23S rRNA gene phylogenies for the genus Helicobacter: implications for phylogenetic inference and systematics.** *J Bacteriol* 2005, **187(17)**:6106-6118.
- Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kampfer P, Maiden MC, Nesme X, Rossello-Mora R, Swings J, Truper HG, Vauterin L, Ward AC, Whitman WB: **Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology.** *Int J Syst Evol Microbiol* 2002, **52(Pt 3)**:1043-1047.
- van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Moran F, Moya A: **Reductive genome evolution in Buchnera aphidicola.** *Proc Natl Acad Sci U S A* 2003, **100(2)**:581-586.
- Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O: **DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.** *Nature* 2000, **406(6795)**:477-483.
- Chen CY, Wu KM, Chang YC, Chang CH, Tsai HC, Liao TL, Liu YM, Chen HJ, Shen AB, Li JC, Su TL, Shao CP, Lee CT, Hor LI, Tsai SF: **Comparative genome analysis of Vibrio vulnificus, a marine pathogen.** *Genome Res* 2003, **13(12)**:2577-2587.
- Foerstner KU, von Mering C, Hooper SD, Bork P: **Environments shape the nucleotide composition of genomes.** *EMBO Rep* 2005, **6(12)**:1208-1213.
- Xia X, Wei T, Xie Z, Danchin A: **Genomic changes in nucleotide and dinucleotide frequencies in Pasteurella multocida cultured under high temperature.** *Genetics* 2002, **161(4)**:1385-1394.
- Delta-rho web and the Compare_Islands tool** [<http://delta-rho.amc.nl/>].
- Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5(2)**:150-163.
- Campanella JJ, Bitincka L, Smalley J: **MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences.** *BMC Bioinformatics* 2003, **4**:29.