

# Genome mosaicism and organismal lineages

Olga Zhaxybayeva, Pascal Lapierre and J. Peter Gogarten

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269, USA

**Some recently published genome-wide analyses of orthologous genes have found plurality consensus support for fully or partially resolved tree topologies. These findings were interpreted to discount the prevalence of horizontal gene transfer and the observed consensus was suggested to reflect organismal history. In this article, we explore the claim that methodological differences might have given rise to the conflicting findings and we show that horizontal gene transfer makes different contributions to the evolution in different classes of prokaryotes.**

Several years ago, it seemed that the concept of tree-like organismal history had failed completely at least with respect to microbial evolution [1,2]. Different approaches to identify transferred genes appeared to reinforce each other in suggesting that horizontal gene transfer (HGT), even among divergent microorganisms, was rampant throughout evolutionary history [3–5]. Comparisons of genomes from closely related organisms seemed to provide the final blow to fell the tree of life. For example, a comparison of completely sequenced genomes of three different strains of *Escherichia coli* revealed that these genomes differ by 778–1860 genes and only have 39.2% of their combined set of genes in common [6]. By contrast, several recent publications appear to resurrect the organismal tree [7–11]. Concatenation of genes was shown to result in statistically well-supported phylogenies and the resulting multigene phylogenies are similar to rRNA phylogenies. Analysis of genomes from closely related organisms indicate that most of the transferred genes persist in the recipient genome for only a short time [12]. In many instances of genome-wide analyses, a plurality of orthologous genes shows support for one fully or partially resolved tree topology [7,11–14]. Do these plurality consensus reveal the true history of organisms, even though each individual gene and/or pathway has their own slightly different history? Or could it be that, in some instances, microbial evolution is dominated by vertical inheritance, whereas in others HGT plays a more crucial role? The spectrum of approaches and opinions on this subject varies between two extremes: one centers on a strong uniform tree-like signal in genomic data [15]; the other focuses on the many genes that contradict the plurality consensus and the role that these genes had in extending the ecological niches of the recipients [16,17].

## What genes are transferred?

Several criteria have been proposed for the detection of horizontally transferred genes (for a summary of the different methods see [5,18]). In 1997, Lawrence and Ochman proposed the use of compositional bias as a criterion for the identification of horizontally transferred genes [19]. Briefly, genes that are acquired horizontally would retain the composition of their previous hosts, which is most probably different from the composition of the new host. However, Daubin and colleagues [12] found that the vast majority of horizontally transferred genes are always more AT-rich than the recipient genome and consist of phage genes, selfish genetic elements and putative open reading frames that do not show any homology to known proteins. These genes are frequently transferred but, apparently, they are not maintained in individual recipient genomes over long periods of time.

A comparison between closely related genomes suggests the transfer of several hundred genes every four million years [6,20,21]. Only a few of these are the typical well-characterized genes that encode proteins involved in metabolism, cell structure or information processing. What rates of HGT would endanger the tree paradigm? Transfer rates as low as one gene every four million years would be sufficient to replace most genes from the most recent common ancestor of all living organisms to a present day genome. Would the replacement of every gene in a genome invalidate the tree concept for ORGANISMAL LINEAGES (see Glossary) [1]?

## Can different ortholog selection schemes explain contradictory findings?

Phylogenetic analysis is often considered to be one of the most reliable ways to investigate the occurrence of HGT; however, conflicting phylogenies can be a result of either artifacts of phylogenetic reconstruction, HGT or unrecognized PARALOGY. Recent articles have reported a surprising level of phylogenetic agreement between different gene families [7,11]. The authors suggested that many of the claims for HGT might be due to the faulty selection of orthologous genes. There is no recipe that guarantees the 'correct' selection of ORTHOLOGS. A commonly used approach is to use circular or reciprocal best BLAST hit relationships. For example, a circular BLAST [22] hit scheme is employed by the Clusters of Orthologous Groups (COG) database (<http://www.ncbi.nlm.nih.gov/COG/>) [23]; it requires only unidirectional, circular best-hit relationships for three of the reference genomes. The strict

Corresponding author: J. Peter Gogarten (gogarten@uconn.edu).

## Glossary

**Organismal lineage:** can be defined as the majority consensus of genes passed on over short time intervals. Provided the time intervals are sufficiently short, this definition only fails in the rare event of two organisms making co-equal contributions to a new line of descent. Gary Olsen (University of Illinois, Urbana-Champaign; <http://www.uiuc.edu/>) used the metaphor of a rope to illustrate this concept. No single cellulose fiber (representing the genes) might persist throughout a rope (representing the organismal lineage) from beginning to end; nevertheless, the rope has continuity.

**Paralogy:** paralogous genes are those that in different, or the same, species are related by a gene duplication event. Paralogs can be mistaken for orthologs particularly in conjunction with gene losses.

**Orthologs:** genes in different species that are related to one another by speciation events.

**Ortholog selection schemes:** let  $A_1, A_2, \dots, A_n$  denote a set of orthologous genes from  $n$  genomes (one gene  $A_i$  from each of  $n$  genomes). Let  $A_i \rightarrow A_j$  denote the best BLAST hit relationship between two genes  $A_i$  and  $A_j$ , where gene  $A_j$  from the genome  $j$  is the best hit in the BLAST search of gene  $A_i$  against genome  $j$ .

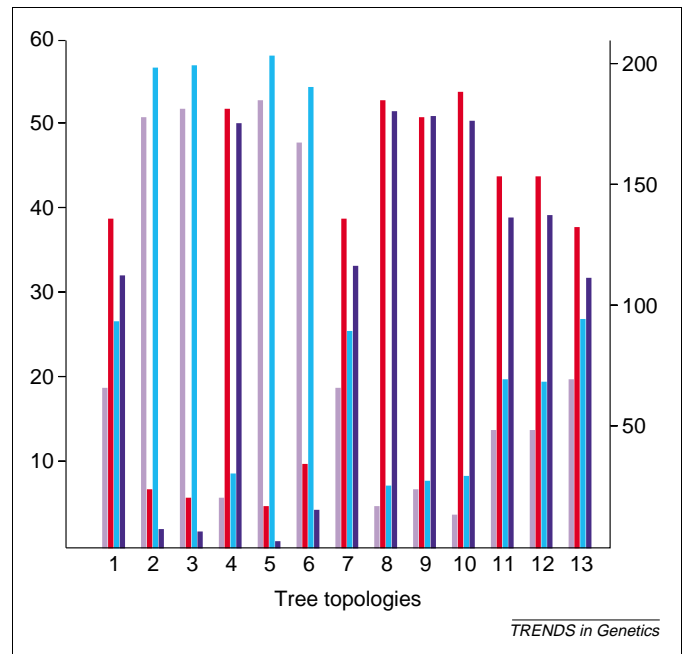
(i) In the circular BLAST hit ortholog selection scheme for each set of selected putative orthologous genes  $A_1, A_2, \dots, A_n$  members are connected through a 'circular' unidirectional best BLAST hit relationships (i.e.  $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1$ ).

(ii) In the reciprocal BLAST hit ortholog selection scheme, a set of genes is considered orthologous if all genes in the set pick each other as a top BLAST hit (i.e.  $A_1 \rightarrow A_2, A_1 \rightarrow A_3, \dots, A_1 \rightarrow A_n, A_2 \rightarrow A_3, \dots, A_2 \rightarrow A_n, \dots, A_{n-1} \rightarrow A_n$ ). This selection scheme is more stringent than the circular BLAST hit ortholog selection scheme.

(iii) In a single BLAST hit ortholog selection scheme, genes from one reference genome are used to search all other genomes and the top-scoring BLAST hits (above a pre-set cut-off point) are merged into a dataset. Additional criteria need to be applied to eliminate paralogs. For example, one can exclude datasets that have more than one representative gene per genome.

reciprocal-best BLAST hit criterion employed (for example, in Refs [24,25]) is more stringent but not perfect. Two of us have previously reported an analysis of 353 quartets of orthologous genes that were assembled under the strict application of the reciprocal-hit criterion [26]. In only two instances was an unexpected phylogenetic relationship a result of unrecognized paralogy.

Daubin *et al.* [11] and Lerat *et al.* [7] use a single best-hit approach (non-reciprocal) but require, as an additional criterion, that no other hit above an arbitrary cut-off point is present in the genome. Using their ortholog selection criterion, Lerat and colleagues analyzed the evolution of 13  $\gamma$ -proteobacteria and observe that the majority of the selected datasets is in agreement with one fully resolved topology. To investigate if this high level of congruence is indeed attributable to the different ORTHOLOG SELECTION SCHEME, we re-analyzed these genomes using their single-best-hit selection scheme and the reciprocal-hit selection scheme. We detected 207 genes using the best-hit selection scheme and 252 genes using the reciprocal-hit selection scheme and found that 198 sets of genes are common between two selection schemes. However, the reciprocal-hit criterion detected 54 additional putative orthologous gene sets. Repeating the Shimodaira–Hasegawa tests [27] that were performed by Lerat *et al.* [7], for the additional 54 putative orthologs, produced the same qualitative results (Figure 1). Clearly, the congruence observed by Lerat *et al.* is not due to an improved ortholog selection criterion. Gene families assembled under either criterion result in the same conclusion. The reciprocal-hit criterion appears to be slightly superior because it detected more orthologs that were missed by the other scheme.



**Figure 1.** Histogram for Shimodaira–Hasegawa (SH) tests [27]. The 13 tree topologies that were tested correspond to the topologies analyzed in Ref. [7]. The test was performed on the sets of orthologous genes detected using the scheme employed in Ref. [7] (the third and fourth columns per tree topology) and for an additional 54 datasets that were detected by reciprocal hits in the ortholog selection scheme but not by the former selection scheme (first two columns). Purple and blue (the first and third) columns indicate the number of datasets that do not reject a topology; the red and dark blue (the second and the fourth) columns indicate the datasets that reject a topology at the 5% significance level. The scale bar on the left hand side of the histogram gives the number of datasets for the analysis of the 54 additional gene families, whereas the scale bar on the right hand side gives the numbers for the analyses of gene families using the scheme employed in Ref. [7]. The SH test was performed using TREE-PUZZLE 5.1 [34]. The consensus topology (#5) is rejected by only a few gene families. The results are qualitatively similar for both of the ortholog selection schemes, indicating that the more stringent approach employed in Ref. [7] is not the reason they found a strong phylogenetic consensus signal.

Both selection schemes err on the side of being overly restrictive and produce an unacceptable high number of false negatives (i.e. orthologs that are not detected and, therefore, excluded from the analyses). For example, genes that underwent lineage-specific amplification have a high chance of being excluded under both schemes, even though they are valid orthologs [28] and should be included.

## How to find and analyze plurality signals and conflicts? Partitions versus trees

Lerat *et al.* [7], in their analyses of 13 species, evaluated only 13 candidate topologies out of possible 13 749 310 575 unrooted tree topologies. They used two topologies that were derived from analyses of concatenated datasets, and four from rRNA analyses and slight modifications of these, comprising a total of 13 similar topologies. Although the majority of the 13 749 310 575 possible topologies would not be supported by any dataset, the selection of a limited number of trees biases the analyses. Lerat *et al.* might have missed unusual topologies; specifically, they would miss topologies that correspond to horizontally transferred genes. Despite this bias, the finding that only a few genes reject the consensus tree reveals the scarcity of a conflicting phylogenetic signal.

Evaluation of all possible trees for 13 genomes is currently almost impossible, whereas evaluation of bipartitions is a promising alternative. A bipartition is a division of a phylogenetic tree into two parts that are connected by a single branch (Figure 2). It divides a dataset into two groups but it does not consider the relationships within each of the two groups. A dataset of aligned sequences might not be able to discriminate between two fully resolved tree topologies because one of the bipartitions has only low bootstrap support. However, another part of the tree might be fully resolved and strongly supported. Consideration of bipartitions enables the capture of these strongly supported parts of phylogenies, even if other parts remain unresolved. A dataset of 13 genomes contains 4082 possible bipartitions, many of which will not be supported by any set of orthologous genes. In the case of the  $\gamma$ -proteobacterial genomes analyzed in the work of Lerat *et al.* [7], we found only 35 bipartitions that are supported by 70% or higher bootstrap values in at least one gene (Figure 3a). Eight of these partitions are strongly supported by the majority of the datasets, they are compatible and can be combined into one partially resolved tree. Only three datasets conflict strongly with these partitions (i.e. they support conflicting partitions with >99% bootstrap support; see Figure 3c).

Analyzing partitions instead of trees has advantages. First, the number of possible bipartitions is much smaller than the number of possible tree topologies, which makes it possible to evaluate all possible bipartitions. Second, analyses of bipartitions enables the use of datasets that would be considered as non-informative if only completely resolved tree topologies are evaluated. Third, putative horizontally transferred genes can be detected because they give rise to partitions that conflict significantly with plurality partitions.

### The extent of inferred genome mosaicism depends on the analyzed genomes

When analyzing a set of genomes, two of which are closely related, it is unsurprising to find a vast majority of orthologous genes that group the two more 'related' genomes together for most of the genes. We included an example of this in our earlier analyses [29], where we obtained results that were strikingly similar to those reported by Daubin *et al.* [11]. However, the absence of detected conflict only means that when using the particular set of genomes few

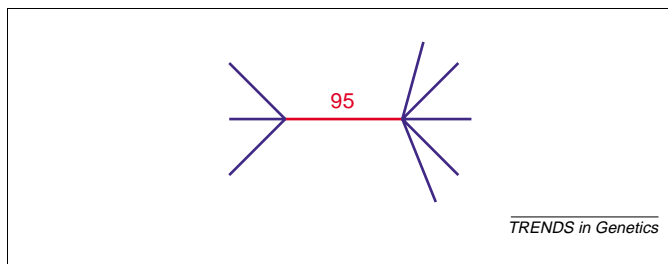
conflicting phylogenetic signals were obtained; it does not mean that the selected genomes are not mosaic. The impact of selected genomes is particularly strong if only four genomes are analyzed. Owing to sparse taxon sampling, many gene transfer events will not result in gene phylogenies with atypical topology; therefore, the amount of genome mosaicism will be underestimated.

The selection of genomes greatly impacts the ability to detect gene transfers using a phylogenetic approach. In addition, different groups of organisms can be expected to have different propensities for gene transfer. Bacteria living in hot environments that are mainly populated by Archaea can be expected to exchange genes mainly with these Archaea [3,30]. Bacteria that live as endosymbionts in a eukaryotic host are unlikely to have received many genes from other prokaryotes. The proteobacteria analyzed by Lerat *et al.* live in close association with multicellular eukaryotes. Can the obtained results be extrapolated to other classes of bacteria? To provide a comparison, we investigated cyanobacteria (Figure 4), a group of photo-autotrophs that, based on 16S rRNA, is less divergent than the  $\gamma$ -proteobacteria analyzed in Figure 3. These have up to 19.8% sequence divergence in their rRNA genes, whereas the ten sequenced cyanobacteria analyzed in Figure 4 are, at the most, 14% divergent. The analysis of the cyanobacteria reveals fewer consensus partitions (Figure 4a) than those found in proteobacteria. In comparing Figure 3 with Figure 4, the most striking finding is that the three most-supported consensus partitions are conflicted strongly by 13 gene families (i.e. these genes support a conflicting partition with >99% bootstrap support). The conflicting proteins include the ribulose biphosphate carboxylase and several enzymes that are involved in chlorophyll biosynthesis (Figure 4c). This might be because the selected cyanobacteria do not include a representative with an extremely streamlined genome. The inclusion of a small genome might exclude weakly selected, and thus more frequently transferred, genes from the analyses [31]. A reason for a less obvious consensus might be smaller and less supported internal branches but this does not explain the strong support for the conflicting partitions. In cyanobacteria more gene families deviate from the consensus, which illustrates that conclusions drawn for one class of bacteria might not be true for others.

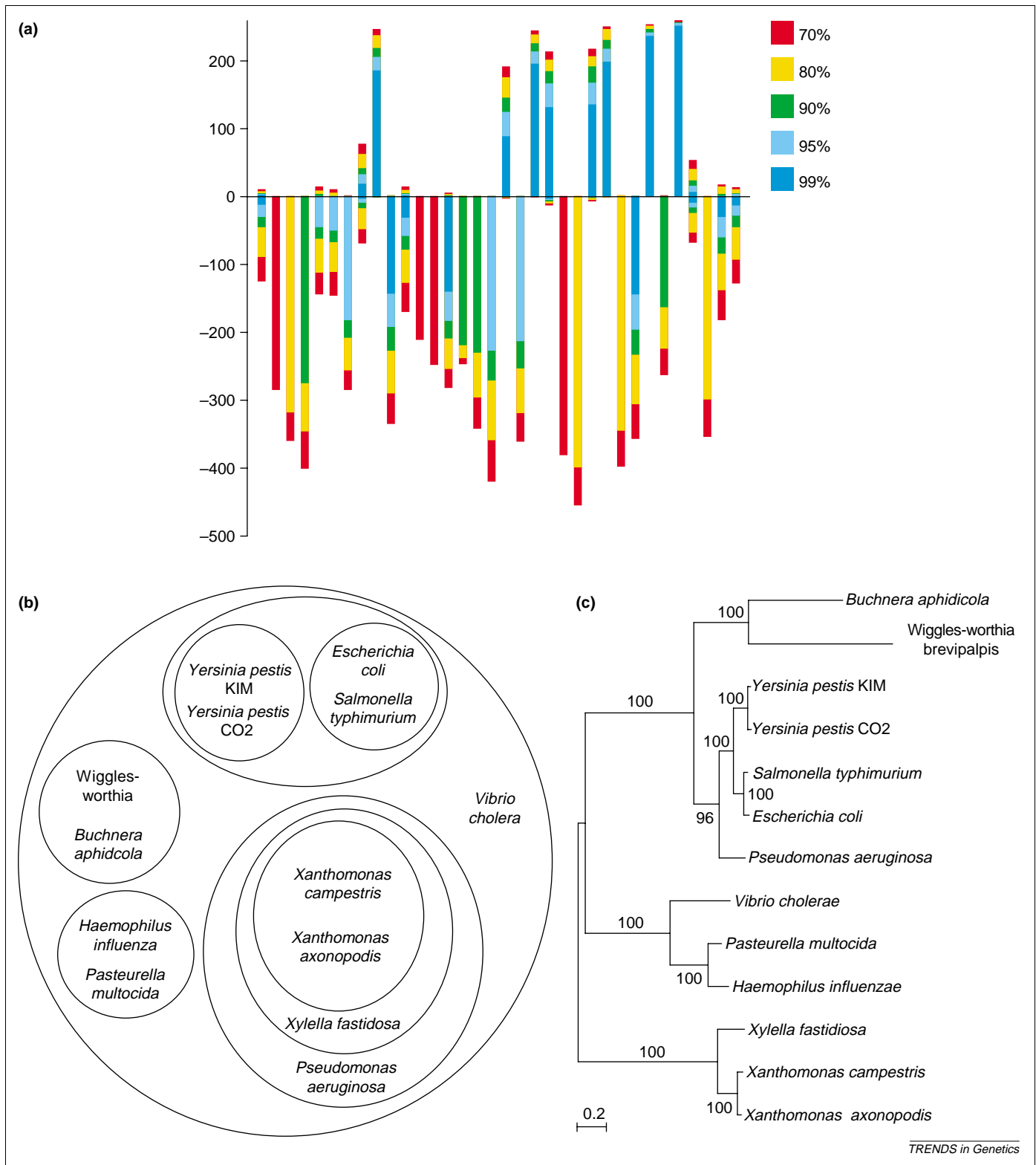
### Group-defining characteristics might be horizontally transferred

Environmental and pathogenicity islands reveal that gene transfer can have an important role in microorganisms evolving into new ecological niches [32]. The importance of HGT in creating new phenotypes is not limited to the recent evolution. In many instances, HGT generated phylum- or class-specific characteristics that define these classes and phyla [13,16,33].

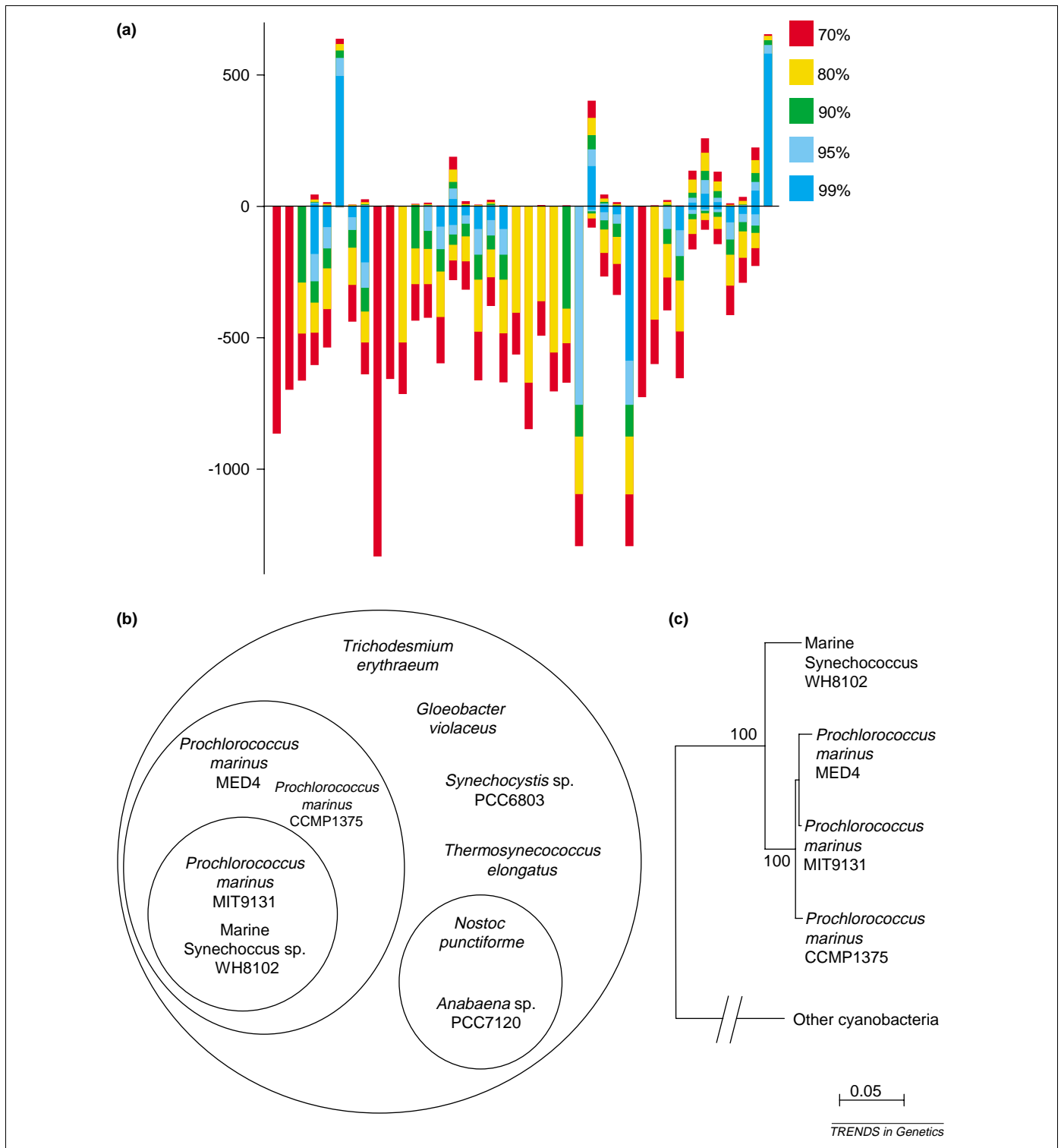
As illustrated by Ref. [8], the three domains of life and the different prokaryotic phyla are recovered reliably from concatenated data. Even the relationship among the phyla, after removing frequently transferred genes from the analyses, appears strikingly similar to the rRNA-based tree of life. In some instances, these patterns



**Figure 2.** An example of a bipartition. A bipartition is a division of a phylogenetic tree into two parts that are connected by a single internal branch. It divides a dataset into two groups but it does not consider the relationships within each of the two groups. In our analyses, we define the support for a bipartition (95 in this particular example) as the bootstrap support of the internal branch. The number of all possible bipartitions for  $N$  genomes is equal to  $(2^{(N-1)} - N - 1)$ .



**Figure 3.** Analyses of bipartitions for  $\gamma$ -proteobacteria. **(a)** A modified Lento plot (after Ref. [35]) for partitions with at least 70% bootstrap support. Each column represents the number of datasets that support (columns that are pointing upwards) or conflict (columns that are pointing downwards) a bipartition. The level of support is color coded. The support for partitions was extracted from the output of the CONSENSE program that was run on 100 Neighbor-joining trees calculated using the NEIGHBOR program [of the PHYLIP (Phylogeny Inference Package) version 3.6 distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle, USA] from TREE-PUZZLE [34] distances determined from 100 bootstrap samples. **(b)** The partitions supported by the majority of the datasets do not conflict with one another and, therefore, can be combined into nested consensus clusters. Only three datasets conflict strongly with these bipartitions. **(c)** One of the conflicting datasets, virulence factor homologues (mviN), is depicted. This topology is not among the 13 topologies that were analyzed in Ref. [7]. The tree was calculated using the NEIGHBOR program from distances calculated with TREE-PUZZLE v.5.1 [34] under the auto-detected substitution model with among site rate variation taken into account (estimated  $\alpha = 0.94$ ). The support values are bootstrap support values from 100 bootstrap samples.



**Figure 4.** Analyses of bipartitions for cyanobacteria. **(a)** A modified Lento plot (after Ref. [35]) for partitions with at least 70% bootstrap support. Each column represents the number of datasets that support (columns that are pointing upwards) or conflict (columns that are pointing downwards) a bipartition. The level of support is color coded. The support for partitions was extracted from the output of the CONSENSE program that was run on 100 Neighbor-joining trees calculated using NEIGHBOR program [of the PHYLIP (Phylogeny Inference Package) version 3.6 distributed by J. Felsenstein, Department of Genetics, University of Washington, Seattle, USA] from TREE-PUZZLE [34] distances determined from 100 bootstrap samples. There are 678 orthologous genes detected by the reciprocal-hit scheme in ten cyanobacterial genomes. **(b)** The nested consensus clusters for the three most supported partitions. **(c)** The phylogeny of ribulose biphosphate carboxylase large subunit significantly conflicts with the consensus. Only bootstrap support values >99% are shown. Other genes that are in conflict with the consensus at >99% bootstrap support are: genes that encode cell division protein FtsH, translation initiation factor IF-2, ferredoxin, geranylgeranyl hydrogenase, amidophosphoribosyltransferase, photosystem II protein D2, photosystem II CP43 protein, photosystem II CP47 protein, photosystem I core protein A2, photosystem I core protein A1, photosystem II manganese-stabilizing protein and 5'-methylthioadenosine phosphorylase. Compared with the  $\gamma$ -proteobacteria the cyanobacteria reveal fewer unanimously supported bipartitions and there are many more conflicting gene families.

### Outstanding questions

- Does the consensus tree reflect organismal history or patterns of horizontal gene transfer (HGT)? The answer to this question might not be the same for different groups of organisms, different genes included in the consensus and different taxonomic levels. Understanding why this is the case will provide new insights into microbial evolution.
- How much of HGT would endanger the tree paradigm for the tree of life? The Ship of Theseus paradox ([http://en.wikipedia.org/wiki/Ship\\_of\\_Theseus](http://en.wikipedia.org/wiki/Ship_of_Theseus)) is frequently invoked to illustrate this point (F. Doolittle and C. House, pers. commun.). Even moderate levels of gene transfer will make it impossible to reconstruct the genomes of early ancestors; many researchers, nevertheless, feel that an organismal tree, even if not a single gene survived throughout a lineage, will be useful and might be re-constructible.
- Can the observations on the evolution of one class of prokaryotes be generalized for the other classes? The limited analyses performed to date suggest a negative answer. For some groups of microorganisms, gene transfer has become an integral part to regulate population structure in response to a changing environment, whereas other groups live in isolated ecological niches where the exchange of genes has only a limited role.

might be due to vertical inheritance and reflect shared ancestry. However, this might not be the case for all groups. If HGT is not random, it can produce similar phylogenies for different molecular markers [3,4]. For example, the presence of extreme thermophilic bacteria at the root of the bacterial domain might be because these groups are ancient bacterial lineages or it might indicate that these organisms live in environments where the only partners for genetic exchange are Archaea. The genes invented by and shared between mesophilic bacteria are only transferred infrequently into the genomes of extreme thermophiles. Even in those cases where a strong consensus phylogeny exists, one should consider the possibility that the shared signal might be due to preferential HGT. To date, none of the findings prove that the agreement that is sometimes found between genome content trees, 16S rRNA phylogenies and phylogenies that are calculated from concatenated datasets is due to a signal generated by vertical inheritance, reflecting a natural taxonomy of prokaryotes. This does not mean that genome-content trees or trees based on the concatenation of multiple genes might not reflect a classification based on common descent; however, the extent to which this is the case remains to be determined.

### Concluding remarks

Genes and organisms are not equal in their propensity for horizontal transfer. Although it is probably true that all types of genes, including rRNA encoding genes, were transferred at some point during their history, the genes most frequently transferred are not the typical protein-coding housekeeping genes. In addition, contributions of vertical inheritance and HGT are not the same across the tree of life. In some instances the contributions of HGT might be negligible and shared ancestry might be readily inferred from individual or concatenated gene families. In other cases, the signal created through shared ancestry has decayed as a result of HGT or was overwritten by more-recent substitutions. Another possibility is that the

readily recognizable signal might be due to preferential gene transfer and not shared ancestry. The completion of >165 published microbial genomes projects to date, enables the exploration of these gene-, organism- and group-specific differences in more detail. Some of the philosophical questions will remain tantalizing for years to come; however, comparisons between different groups of microorganisms will help to draw a clearer picture of the many evolutionary mechanisms operating in the microbial world (Box 1).

### Acknowledgements

This work is supported by the NASA Exobiology Program and NSF Microbial Genetics Program. We thank Jason Raymond for stimulating discussions and the anonymous reviewers for helpful comments on the manuscript.

### References

- 1 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129
- 2 Hilario, E. and Gogarten, J.P. (1993) Horizontal transfer of ATPase genes—the tree of life becomes a net of life. *BioSystems* 31, 111–119
- 3 Lawrence, J.G. and Hendrickson, H. (2003) Lateral gene transfer: when will adolescence end? *Mol. Microbiol.* 50, 739–749
- 4 Gogarten, J.P. *et al.* (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238
- 5 Ragan, M.A. (2002) Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* 10, 1–4
- 6 Welch, R.A. *et al.* (2002) Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17020–17024
- 7 Lerat, E. *et al.* (2003) From gene trees to organismal phylogeny in prokaryotes: the case of the  $\gamma$ -proteobacteria. *PLoS Biol.* 1, E19
- 8 Brown, J.R. *et al.* (2002) Horizontal gene transfer and the universal tree of life. In *Horizontal Gene Transfer* (Syvanen, M. and Kado, C.I., eds), pp. 305–349, Academic Press
- 9 Brown, J.R. *et al.* (2001) Universal trees based on large combined protein sequence data sets. *Nat. Genet.* 28, 281–285
- 10 Brochier, C. *et al.* (2002) Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet.* 18, 1–5
- 11 Daubin, V. *et al.* (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* 301, 829–832
- 12 Daubin, V. *et al.* (2003) The source of laterally transferred genes in bacterial genomes. *Genome Biol.* 4, R57
- 13 Raymond, J. *et al.* (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* 298, 1616–1620
- 14 Rokas, A. *et al.* (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804
- 15 Kurland, C.G. *et al.* (2003) Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9658–9662
- 16 Boucher, Y. *et al.* (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* 37, 283–328
- 17 Jain, R. *et al.* (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* 20, 1598–1602
- 18 Lawrence, J.G. and Ochman, H. (2002) Reconciling the many faces of lateral gene transfer. *Trends Microbiol.* 10, 1–4
- 19 Lawrence, J.G. and Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383–397
- 20 Perna, N.T. *et al.* (2001) Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409, 529–533
- 21 Hayashi, T. *et al.* (2001) Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22
- 22 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
- 23 Tatusov, R.L. *et al.* (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28
- 24 Montague, M.G. and Hutchison, C.A. 3rd (2000) Gene content

- phylogeny of herpes viruses. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5334–5339
- 25 Zhaxybayeva, O. and Gogarten, J.P. (2002) Bootstrap, Bayesian probability and maximum likelihood mapping: Exploring new tools for comparative genome analyses. *BMC Genomics* 3, 4
- 26 Zhaxybayeva, O. and Gogarten, J.P. (2003) An improved probability mapping approach to assess genome mosaicism. *BMC Genomics* 4, 37
- 27 Shimodaria, H. and Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic interference. *Mol. Biol. Evol.* 16, 1114–1116
- 28 Fitch, W.M. (2000) Homology: a personal view on some of the problems. *Trends Genet.* 16, 227–231
- 29 Raymond, J. *et al.* (2003) Evolution of photosynthetic prokaryotes: a maximum-likelihood mapping approach. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 223–230
- 30 Koonin, E.V. (2003) Horizontal gene transfer: the path to maturity. *Mol. Microbiol.* 50, 725–727
- 31 Lawrence, J.G. and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143, 1843–1860
- 32 Hacker, J. and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679
- 33 Zhaxybayeva, O. *et al.* (2004) Visualization of phylogenetic content of five genomes with dekapentagonal maps. *Genome Biol.* 5, R20
- 34 Schmidt, H.A. *et al.* (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504
- 35 Lento, G.M. *et al.* (1995) Use of spectral analysis to test hypotheses on the origin of pinnipeds. *Mol. Biol. Evol.* 12, 28–52

## Cellular and molecular basis of regeneration

### Euroconference on the molecular pathways leading to regeneration

San Feliu de Guixols, Spain, 18–23 September 2004

Chair: Elly M. Tanaka, Dresden, Germany

Vice Chair: Jonathan M.W. Slack, Bath, U.K.

The conference will focus on the basic mechanisms of epigenetic reprogramming, embryonic stem cells, somatic stem cells and mammalian tissue regeneration, model organisms of epimorphic regeneration – hydra, planaria and amphibia – to consider the basic mechanisms underlying regeneration and how they can be applied towards mammalian tissue repair.

Deadline for applications: 10 June 2004

Scientific programme and online application: <http://www.esf.org/euresco/04/lc04177>

European Science Foundation, EUESCO office, 1 quai Lezay-Marnésia, BP 90015,

67080 Strasbourg Cedex, France

Telephone +33 388 76 71 35 Fax +33 388 36 69 87

Email: [euresco@esf.org](mailto:euresco@esf.org)