RESEARCH ARTICLES

Visualizing and Assessing Phylogenetic Congruence of Core Gene Sets: A Case Study of the γ-Proteobacteria

E. Susko,[1] J. Leigh,† W. F. Doolittle,† and E. Bapteste[1]*

*Genome Atlantic, Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada; and †Canadian Institute for Advanced Research and Genome Atlantic, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

Here, we address a much-debated topic: is there or is there not an organismal tree of γ-proteobacteria that can be unambiguously inferred from a core of shared genes? We apply several recently developed analytical methods to this problem, for the first time. Our heat map analyses of *P* values and of bootstrap bipartitions show the presence of conflicting phylogenetic signals among these core genes. Our synthesis reconstruction suggests that at least 10% of these genes have been laterally transferred during the divergence of the γ-proteobacteria, and that for most of the rest, there is too little phylogenetic signal to permit firm conclusions about the mode of inheritance. Although there is clearly a central tendency in this data set (it is far from random), lateral gene transfers cannot be ruled out. Instead of an organismal tree, we propose that these core genes could be used to define a more subtle and partially reticulated pattern of relationships.

Introduction

Darwin asserted that the tree-like hierarchical pattern of relationships constructed by systematists parallels a real historical process of lineage splitting which may itself be visualized as a tree—indeed as a single "great tree of life," embracing all living forms (Darwin 1859). Many biologists since Darwin have devoted their professional lives to reconstructing the branching pattern of part or all of this tree of life. In general, they have asked only what the structure of the tree is and not whether a tree is the appropriate model (Darwin called it a "simile"). And for the universal tree as a whole, they have focused on the use of gene or protein sequence data because no organism-level morphology, physiology, or behavior is universally distributed.

The availability of many (more than 250) complete prokaryotic genome sequences has provided phylogeneticists with a wealth of genes with which to reconstruct gene phylogenies for this group. Unexpectedly often, these phylogenies have turned out to be in disagreement with each other and with the small subunit ribosomal RNA (SSU rRNA) phylogeny that many had come to accept as the (organismal) tree of life. Much of the time, the disagreement reflects the phylogenetic artifact and/or lack of signal, but sometimes it results from lateral (horizontal) gene transfer (LGT) (Bapteste et al. 2005). Estimates of the fraction of any individual genome that has been transferred can reach 25%–30% (Nelson et al. 1999; Deppenmeier et al. 2002). Yet, it is in the nature of most phylogeny-based estimates that they will fail to detect many transferred genes, especially those that have been exchanged over short phylogenetic distances or early in the divergence of a group. Conceivably, a web-like pattern of LGT might in the long run dominate over the tree-like pattern of vertical descent, and Darwin might be said to have chosen the wrong simile to describe organismal relationships (Doolittle 1999), at least for microbes.

Molecular phylogeneticists seeking the universal "organismal" tree have responded to this challenge by advancing plausibility arguments in favor of the notion that the SSU rRNA genes first chosen as a "universal chronometer" are uniquely immune to transfer and reliably track organismal history (Woese, Kandler, and Wheelis 1990). To support such arguments, they have also attempted to identify a highly conserved set of genes—in addition to those encoding rRNAs—that by their essential nature and fundamental function might also resist both loss and transfer (Jain, Rivera, and Lake 1999). In practical terms, this entails the identification of a "core" of genes shared by the taxa to be related phylogenetically. Thus, for example, one might construct a phylogeny for proteobacteria using genes common to all proteobacteria or a bacterial phylogeny using the rather fewer genes found in all bacterial genomes (the bacterial core). A universal tree of life might be based on the even smaller set ubiquitous among prokaryotes and eukaryotes (the universal core). Such cores have been further refined by exclusion of genes present in multiple copies (paralogs) and those which, although orthologous and ubiquitous among the taxa to be related, have shown unexpected topologies in preliminary phylogenetic analyses (Brochier et al. 2002; Lerat, Daubin, and Moran 2003; Philippe and Douady 2003; Brown and Volker 2004).

In many cases, individual core genes have weak phylogenetic signal, and their sequences have been concatenated in an attempt to enhance the ratio of signal to noise and thus more reliably recover the "true" organismal phylogeny. For this to be an appropriate procedure, however, it must be shown—not just assumed—that core genes do in fact share a common phylogenetic history. This has proven surprisingly difficult (Bapteste et al. 2005), when the test has been node-for-node congruence between trees constructed for different genes, individually. Harris et al. (2003), Teichmann and Mitchison (1999), Nesbo, Boucher, and Doolittle (2001), Raymond et al. (2002), and Wertz et al. (2003) report multiple-conflicting—but poorly supported—phylogenies for genes shared by, respectively, all life, the two prokaryotic domains, Euryarchaeotes, photosynthetic bacteria of five phyla, and enteric bacteria. Most

[1] These authors equally contributed to the present work.

Key words: γ-proteobacteria, phylogeny, heat map, congruence.

E-mail: eric.bapteste@dal.ca.

Mol. Biol. Evol. 23(5):1019–1030. 2006
doi:10.1093/molbev/msj113
Advance Access publication February 22, 2006

recently, Creevey et al. (2004), after examining a 61-genome bacterial data set, concluded that "congruence among gene trees spanning deep relationships is not better than random."

Thus, several investigators have taken the converse approach: assessing whether individual genes of a core gene set have phylogenies that are demonstrably "incongruent" with each other or with a preferred reference tree obtained with their concatenated sequences (or rRNA) (Kumar and Rzhetsky 1996; Baldauf et al. 2000; Bapteste et al. 2002; Lerat, Daubin, and Moran 2003). Statistical approaches such as the Shimodaira-Hasegawa (Shimodaira and Hasegawa 1999) or approximately unbiased (AU) test (Shimodaira 2002) have been used to assess if any single genes reject such reference trees. Genes that reject are secondarily excluded from the core (Lerat, Daubin, and Moran 2003). Genes that fail to reject (usually at the 95% confidence level) are taken to be phylogenetically congruent, and the tree produced by their concatenated sequences is taken as the true organismal phylogeny (Brochier et al. 2002; Matte-Tailliez et al. 2002; Lerat, Daubin, and Moran 2003), because overall it is generally more resolved than any individual phylogeny.

But the fact that constituent genes fail to reject the tree of the concatenate does not prove that they share the same history. "Failure of rejection" is not the same as "support". Genes with little phylogenetic signal (random and short sequences, for instance) may fail to reject many or even all possible trees relating a given set of taxa. Yet, such genes cannot logically be said to support multiple and mutually incompatible topologies. Indeed, it is possible that a robust tree based on concatenated sequences is well supported because different constituent genes contribute strong support to different individual nodes of the tree, without any supporting that tree over all—because their signals are weak or genuinely conflicting. (Indeed, most genes in a core set could be mutationally saturated, with individual cases of recent LGT actually providing the signal at the best-supported nodes, as Teichmann and Mitchison [1999] suggested from their universal core analysis, several years ago.) In such cases, there would be no reason for concluding that the genes shared a common phylogenetic history, no matter how robust the support for their collective tree is (Bapteste et al. 2004). The fact that almost always more than one tree is not rejected when only one can be true means that some wrong trees will not be rejected. Hypothesis tests are designed to give small probabilities that a tree is rejected when it is the true tree. They provide no control over the probability that a wrong tree is not rejected.

To gain more confidence in or information concerning phylogenetic congruence of core genes, it is necessary to perform statistical tests for each gene against as many relevant alternative topologies as feasible, not just that favored by the concatenated data set and a few variants of it or by canonical markers such as SSU rRNA. It is furthermore useful to have some statistical measure of each gene's' compatibility with each tested tree more nuanced than "rejection" or failure of rejection. And it would be most helpful to be able to visualize the relative compatibility of all genes with all trees simultaneously by methods that cluster genes with similar patterns of compatibility. Such visualization would be the first step toward determining how many (if any) common evolutionary histories are exhibited by a core of shared genes.

Here we explore different approaches to the question of which and how many evolutionary histories might be shared by the 205 core γ-proteobacterial genes (Bapteste et al. 2002; Brochier et al. 2002; Lerat, Daubin, and Moran 2003) using methods that provide greater quantification of the extent of agreement between genes and of the level of support that individual genes provide for individual topologies. These approaches are principal component analysis (PCA) and a recent application of "heat map" clustering methods more commonly used in functional genomics of which we have recently provided a preliminary description (with applications to other test data sets). In this more complete analysis, we show that heat maps methods outperform PCA in congruence tests, although neither can detect all events of LGT. We also explore clustering algorithms that can be used in conjunction with heat maps to recognize cohorts of genes with shared evolutionary histories. We present an application of heat maps to bipartition-based bootstrap methods, which will allow analysis of much larger data sets. Detailed results from some of these studies are presented as supplementary materials (see Supplementary Materials online). Our biological focus here is on the γ-proteobacterial core because it is frequently assumed that these genes share a common evolutionary history. We show that γ-proteobacterial core genes do have some phylogenetic signal and often exhibit "central tendencies," even though gene-by-gene congruence at every node can seldom if ever be shown. But there is also substantial conflict between markers, and it may be generally impossible with existing methods to determine whether this conflict reflects LGT or reconstruction artifacts (such as long-branch attraction). We argue against systematic biases to explain our results, notably by analyzing the impact of long-branch attraction on heat maps and PCA and by refuting possible compositional biases, marker size effects, and problems due to heterogenity of rates of evolution. Finally, we report the first synthesis of γ-proteobacteria, indicating in detail how 18 genes would have been transferred, from which donors to which hosts, while presenting the limited support for their hypothetical tree. We would not be surprised if those who see vertical descent as the dominant evolutionary process affecting core genes over the long term take our inability to prove LGT to be a support for that model, while LGT advocates will be heartened by the lack of robust support for congruence among core genes. As Kuhn has noted, "when paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense" (Kuhn 1962). We would like, however, to avoid this circularity by putting forward a synthesis of γ-proteobacteria which seems to us a more accurate and pragmatic representation of the relatively weak and complex phylogenetic signal that is really contained by genes.

Materials and Methods

Alignments and Preliminary Phylogenetic Analyses

The γ-proteobacterial (205 genes, 13 species) data set was kindly provided by E. Lerat (Lerat, Daubin, and Moran

2003). All these alignments were inspected, manually refined if required, and are available upon request. For all individual markers, preliminary analyses by Neighbor-Joining using MUST 3.0 (Philippe 1993) and maximum likelihood (ML), performed using PROML with the Jones-Taylor-Thornton amino acid substitution matrix, a rate heterogeneity model with Γ-distributed rates over four categories with the α parameter estimated using Tree-Puzzle (Schmidt et al. 2002), global rearrangements, and randomized input order of sequences (10 jumbles), were done to detect potential nonorthologous copies, but no such copy was identified. A concatenation of all these markers was then realized, and the best ML tree was calculated for it as well as for each gene individually using a JTT model, nine categories of sites (http://evolution.genetics.washington.edu/phylip/doc/proml.html).

Calculation of Matrices of *P* Values

A set of 105 rooted topologies consisting in all the possible rearrangements of four paired taxa identified in Lerat, Daubin, and Moran (2003)—*Escherichia coli* + *Salmonella typhimurium*, the two *Yersinia*, *Haemophilus influenzae* + *Pasteurella multocida*, and the two insect symbionts (with *Vibrio cholerae* as a singleton)—was employed (see Bapteste et al. 2004). Intrees were used as user tree in Tree-Puzzle 5.1, option—wsl, with a JTT + Γ 8 + I model of evolution to estimate the likelihood of each site of a given gene and global tree likelihoods for each tree. These likelihood values were used as input for CONSEL (Shimodaira and Hasegawa 2001) to perform the AU test (Shimodaira 2002). This is a statistical test used to determine. It is a statistical test of the hypothesis that the given topology is the correct topology for the taxa under consideration. Differ significantly or not. When the value associated by the AU test to one of the topologies under study is <0.05, this tree can be said significantly different and worse than other topologies, at a threshold of 5%.

Simulated and Control Data Sets

We produced two kinds of random matrices, which do not contain any bona fide phylogenetic signal. First, matrices were generated by shuffling randomly the *P* values of a given matrix of actual data. Random matrices were also generated by randomly shuffling gene sequences between species. We also created artificial LGT events by randomly reassigning the sequence of one gene from one species to another one, as if the latter has just laterally acquired the sequence of the former. After this operation, a gene alignment presents one additional extreme and recent LGT event. We repeated this up to three times per gene, generating up to three additional LGT events in a single alignment.

We also generated a simulated data set free from LGT but evolving, as nearly as possible, with the same parameters as the γ-proteobacterial genes. A tree was chosen as a representative of the unique history of all the genes to be simulated. The branch lengths of this simulation were inferred for each γ-proteobacterial gene, leading to the calculation of 205 trees identical in topology but with distinct branch lengths, empirically derived. These 205 trees were subsequently used to simulate 205 amino acid data sets by pseq-gen (Grassly, Adachi, and Rambaut 1997) using the estimated α parameters for generating Γ distribution and recreating sequences of the same length as those of the marker from which the branch lengths had been calculated. We thus obtained 205 amino acid data sets, all issued from a single tree, but all presenting close evolutionary characteristics to the actual γ-proteobacterial genes.

Heat Map Analyses

All heat maps were generated using the freely available statistical package R (http://www.r-project.org/). R language functions and example scripts for generating them will be made available at http://www.mathstat.dal.ca/~tsusko.

Heat maps of *P* values from the AU test were also used to test that genes support similar topologies. Dark-colored spot indicates low *P* values for a topology tested for a given gene. By contrast, light-colored spot indicates high *P* values, that is, a good support for this topology by a given gene. These spots of color can be further reordered to stress the presence of patterns of support/rejection. This was done by rearranging rows and columns, separately for genes and topologies, so that they correspond to a dendrogram from hierarchical clustering. In this way, clusters of genes (topologies) showing similar patterns of support across topologies (genes) are grouped together and easily seen. Hierarchical clustering dendrograms were obtained using the euclidean distance matrix for the vectors of *P* values.

To utilize the information from tests over a large number of topologies while easing visualization of results, we present heat maps with a restricted set of selected "plausible" topologies. The criterion of selection for such topologies was that the majority (>103) of genes had a *P* value larger than 0.05 with respect to the topology. Note that this criterion is less restrictive than that under test in such studies as that of Lerat, Daubin, and Moran (2003), in which it is assumed that all genes share a common topology but accepted that some may not exhibit this topology, because of weak signal. Under the hypothesis of vertical descent for all genes, *P* values should be uniformly distributed across genes so that 95% of the genes are expected to have *P* values larger than 0.05 for the true topology. Although with 100 independent genes all evolving according to the same topology, it would be unlikely that exactly 95 of them would be larger than 0.05, one can show that the probability is larger than 0.99 that at least 89 out of the 100 genes would have *P* values larger than 0.05 for the correct topology.

A large number of methods are available for determining the optimal number of clusters in clustered data. Some of these, such as tests of the number of components in a mixture model, are specific to the clustering methods used. Others apply generally to any clustering method. Surveys of methods for determining the number of clusters are available in Milligan and Cooper (1985) and in Gordon (1999). The ones that we considered here are the maximizer of the CH index (Calinski and Harabasz 1974), the maximizer of the KL index (Krzankowski and Lai 1985), and the smallest cluster size such that gap statistic index is greater than 0 (Tibshirani, Walther, and Hastie 2001). The CH index

was one of the better performers in the extensive simulation studies comparing methods for the determination of the number of clusters. The CH and KL indices estimate the optimal number of clusters k, for $k = 2, 3, \ldots$. The more recent gap statistic method includes $k = 1$ in its search for the optimal number of clusters and thus gives an indication as to whether any clustering is required at all.

All these approaches, which can be used to cluster topologies as well as genes, utilize the between, $B(k)$, and within, $W(k)$, sums of squares for k clusters, $k = 1, 2 \ldots$ For a set of k clusters, to calculate $W(k)$, one first determines, for each given gene and topology, the difference between its P value and the average P value for the given topology, averaged across all genes in the same cluster as the given gene. If clustering is present, P values for topologies should be similar for genes in the same cluster and so these differences should be small. The within sum of squares, $W(k)$, is then the sum of all the squares of these differences and should be small for a good choice of the number of clusters, k. In a similar way, $B(k)$ is a sum of squared differences, where the differences are between the P values for clusters and the average P values across all genes. If clustering is present, average P values within clusters should be different and thus differ from the overall mean P value; $B(k)$ should be large for a good choice of the number of clusters, k. Different methods differ in their manner of determining what small values of $W(k)$ are and, where applicable, what large values of $B(k)$ are. In the present analysis, phylogenetic congruence should produce a single cluster, and incongruence should produce additional clusters that differ from this main cluster. Although clustering may sometime appear obvious on inspection, final decisions on the true number of clusters will always have a subjective component. Ideally, the true number of clusters will be that chosen by several independent methods, should agreement be found between them.

Synthesis Reconstruction

The "synthesis" of γ-proteobacteria was inferred from the analyses of 205 ML trees. Individual ML trees were calculated using PROML. Options were global rearrangements, randomized input order of sequences (10 jumbles), JTT amino acid substitution matrix, and a rate heterogeneity model with Γ-distributed rates over four categories, with the α parameter estimated using Tree-Puzzle. Bootstrap support values represent a consensus (obtained using CONSENSE) of 100 Fitch-Margoliash distance trees (obtained using PUZZLEBOOT and FITCH) from pseudoreplicates (obtained using SEQBOOT) of the original alignment. The settings of PUZZLEBOOT (http://bioinformatics.ubc.ca/resources/tools/index.php?name=puzzleboot) were the same as those used for PROML, except that global rearrangements and randomized input order of sequences are not available in this program. The clades supported with more than 50% bootstrap support in these 205 gene trees were compared to the concatenated tree from Lerat, Daubin, and Moran (2003) using two programs: Horizstory and Lumbermill (MacLeod et al. 2005). Briefly, Horizstory allows the inference of the most parsimonious scenarios involving LGT and vertical descent to explain the common features and the discrepancies between the organismal tree and each of the 205 trees. Lumbermill draws the synthesis by mapping the outcomes of these scenarios onto the reference tree and calculates all the estimates presented here. A strict consensus option was applied, meaning that only the relationships supported or inferred in 100% of the evolutionary scenarios resulting from the comparison between the reference and a given tree were considered in this drawing.

Results and Discussion

PCA

PCA is a widely used technique for dimension reduction (Mardia, Kent, and Bibby 1979; Venables and Ripley 2002) that has already been employed to examine the congruence of individual genes in a concatenated set, for instance see Brochier et al. (2002) and Matte-Tailliez et al. (2002). The method has great visual and heuristic appeal, but serious limitations, as we discuss in this section.

If P values are used as the measure of support for topologies, principal components are weighted sums of the P values, summed over topologies, for the genes. In PCA studies, each gene is often represented as a point in a two-dimensional space, the coordinates for each gene being the first two principal components. The weights for the first principal component are chosen to maximize the sample variance of this weighted sum of the P values, over all genes. Successive principal components are chosen as those that maximize variation subject to being uncorrelated with previous principal components. Because principal components maximize variation, if the underlying cause of variation in P values is due to clusters or individual genes showing very different patterns of support across topologies, it is expected that this will show up as clustering patterns that can easily be visualized in a plot of the first two principal components. Most of the time, these analyses have produced central clouds containing most markers, leading to claims that such genes share common evolutionary histories or at least display some central tendency (Brochier et al. 2002; Matte-Tailliez et al. 2002). For our PCA of the γ-proteobacterial core, P values were obtained by the AU test for the likelihoods of each of 205 aligned gene data sets giving 105 P values for each gene (one for each of the 105 tested topologies described above). The plot of the first two principal components in figure 1 shows a cloud with scattered outliers; the two components encompassed 25% of the variance in the data. This value is made lower by our inclusion within the analysis of several manipulated data sets, described below (see *Materials and Methods*). The first two components encompass 37.5% of the variance if the randomized data sets are not included and 42.2% if the simulated LGT data sets are also eliminated.

Although genes with a common phylogeny should produce such a cloud, clustering in this way does not require a common phylogeny. The weights for the principal components are chosen to maximize variation and are highly dependent on the choice of topologies for P value calculation. If a substantial source of variation is in support for overall poorly supported topologies, the weights for the P values for these topologies will be large. The cloud in this case reflects the lack of support for a number of toplogies

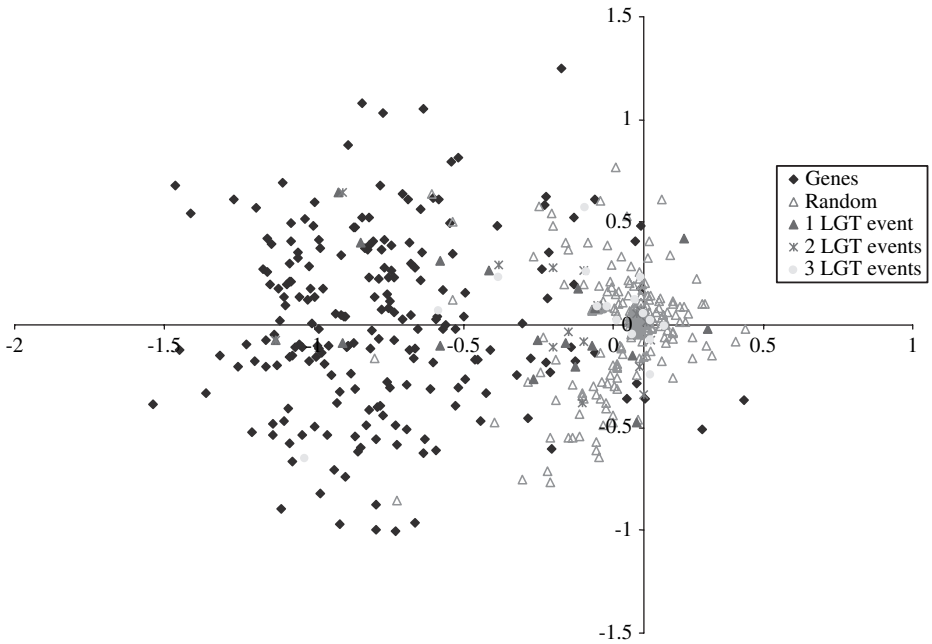


FIG. 1.—PCA for γ-proteobacteria. *P* values were obtained by the AU test for the likelihoods of each of 205 aligned gene data sets (blue diamond) given each of 105 trees, see text. Purple triangles corresponds to data sets with a completely scrambled phylogenetic signal. Markers with artificial single, double, and triple LGT events are represented by red triangles, green crosses, and yellow dots, respectively.

rather than support for a common phylogeny. Even genes with very strongly conflicting signal (due for instance to recent LGT) might cluster together on a PCA, if their most common shared characteristic is a similar pattern of rejection for poorly supported topologies. We can assess only the extent of clustering: we do not know what central tendencies such clustering bespeaks. Finally, with PCA we do not know which (or how many) trees are favoured or rejected by any group of genes, or which (or how many) genes are consistent with any group of trees.

The appearance of a cloud suggests some coherent phylogenetic signal within these genes. We confirmed this by comparison to the result obtained when the signal was completely scrambled by randomly reassigning (without replacement) gene sequences to taxon names in each aligned gene data set. Such randomized data (purple triangles in fig. 1) form a separate and seemingly symmetric cloud around the origin in this projection, within which approximately 10% of the real (unscrambled) data points also fell. As an additional control, we created artificial single, double and triple LGT events by randomly assigning one taxon's gene sequence to another in randomly selected gene data sets (so that in the aligned data set there would be one, two, or three pairs of identical sequences). These markers with artefactual LGT events, especially those with only one such event, overlap with the cloud of real data. In supplementary materials (see Supplementary Materials online), we provide two quantitative estimates of the degree of overlap. (Note, of the two genes identified by Lerat, Daubin, and Moran [2003], as LGTs, one [MviN] falls close to the random genes [−0.211, 0.232] but the other [BioB] is in the center of the cloud [−0.744, 0.112]).

We concluded from this exercise that many of the 205 genes of the γ-proteobacterial core do retain phylogenetic signal but that real single and multiple LGT events affecting some and possibly many of the genes might easily escape detection and that some smaller fraction of the genes could have experienced quite extensive scrambling during evolution. Moreover, PCA allows only an estimation of the extent of clustering: we do not know what central tendencies such clustering bespeaks. That is, we neither know which (or even how many) trees are favored or rejected by any group of genes nor which (or even how many) genes are consistent with any group of trees.

Heat Maps

A method for visualizing data of a fundamentally similar sort has become popular in functional genomics but has yet to be routinely employed in comparative or evolutionary genomics. This involves the generation of heat maps through hierarchical or partitional clustering, as we have briefly described elsewhere, with application to other data sets (Bapteste et al. 2005). In functional genomics, heat maps allow simultaneous display of all combinations of genes and test conditions together with simultaneous clustering of both genes and conditions, according to response values (Alon et al. 1999; Getz, Levine, and Domany 2000; Getz et al. 2003; Somogyi et al. 2004). Thus, genes that have the most similar responses to conditions, and conditions that are the most similar in terms of the responses they evoke from genes, can be independently identified. In transcriptomics, the major area of heat map application, the conditions are usually alterations in growth regimen (or disease state), and the responses are most often alterations in levels of mRNA. For our parallel application here to the problem of the phylogenies of core genes, "conditions" are topologies and

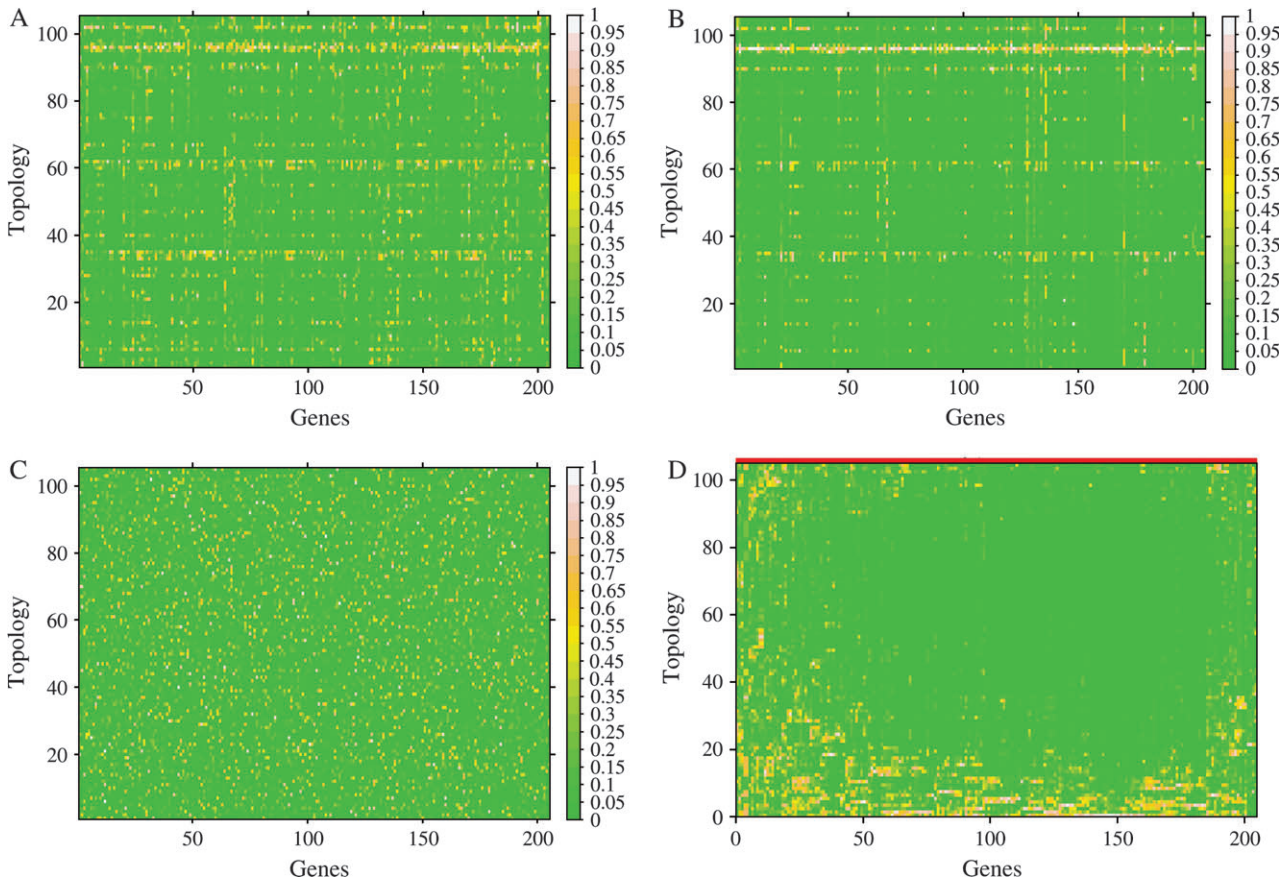


FIG. 2.—(*A*) Heat map for the *P* values for each of the 205 γ-proteobacterial genes assessed against 105 test topologies, before clustering of either genes or topologies. Lighter colors indicate large *P* values and darker colors indicate small *P* values (stronger rejection). (*B*) Heat map for the *P* values for each of the 205 simulated genes, having evolved from a unique tree assessed against 105 test topologies, before clustering of either genes or topologies. (*C*) Heat map for a randomization of the same data as in (*A*). (*D*) Heat map of *P* values after clustering genes and topologies for the γ-proteobacterial data set. The red line indicates real data that have been clustered.

"responses" are *P* values for a test that the topology (condition) is the correct one for that gene. Clustering of genes allows identification of one or more sets of genes in a core that might share a common evolutionary history. Clustering of topologies allows us to identify, for a given gene data set, which trees are equally or nearly equally supported, and thus to assess how many distinct "best trees" there might be.

Several heat map analyses were conducted with this same γ-proteobacterial core gene data set. Figure 2*A* displays the *P* values for each of the 205 genes assessed against the same set of 105 topologies, before clustering of either genes or topologies. Lighter colors indicate a higher *P* value of the data given the tree (i.e., stronger support), while darker colors indicate lower *P* values (stronger rejection). (We will henceforth designate as support a *P* value of 0.5 or more and as "strong rejection" a *P* value of 0.05 or less: intermediate values will be described as "compatible.") Even without clustering, such a display makes it clear that (1) some topologies are fairly well supported by many genes, (2) even these supported topologies are strongly rejected by some of the genes, (3) some topologies are strongly rejected by many genes, and (4) most genes reject most of this subset of topologies.

Interpretation of the Heat Maps

To better interpret the previous heat map, we generated a control with no conflicting signal: figure 2*B* represents the heat map for a simulated data set of 205 genes free of LGT. This data set was generated from a single topology considered the most likely candidate for the tree under the hypothesis of vertical descent. This tree had the largest number of genes supporting it and corresponds to topology 5 in Lerat, Daubin, and Moran (2003)—a tree made with the concatenated data set. For each gene in the original data set, sequences of the same length were generated by pseq-gen from this topology under a JTT substitution process. Branch lengths and alpha parameters for the Γ rates-across-sites distribution were estimated from the sequences for the gene in the original data set. The fact that this simulation shows the same two principal alternative topologies as the true data set indicates that it could be phylogenetic noise, not genuinely conflicting signal, that produces this pattern. The fact that the simulated set produces a pattern that is overall "cleaner" than the real data (the simulated set shows far more rejections of topologies and fewer topologies that are broadly supported) suggests that noise is not the only reason for the greater complexity of

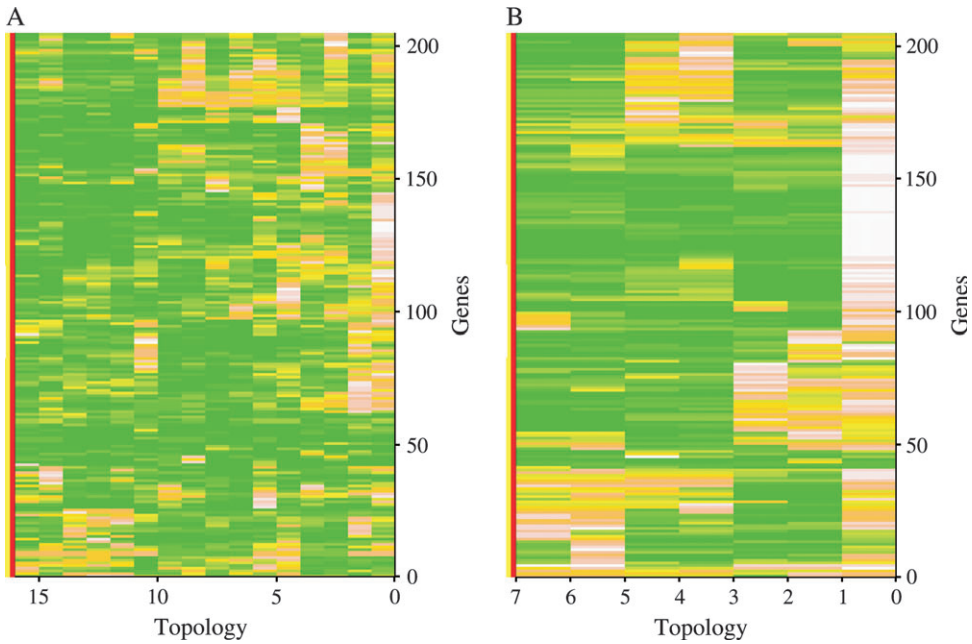


FIG. 3.—(*A*) Heat map of *P* values after clustering genes and topologies for the γ-proteobacterial data set with a restricted set of plausible topologies for which a majority of genes had a *P* value larger than 0.05. (*B*) Same heat map for a control data set where gene sequences were generated from a single topology considered the most plausible vertical descent topology. Colour codes are as in figure 2*A*.

the pattern in figure 2*A*. There is likely some conflicting phylogenetic signal. That said, the situation is clearly not one of maximum incongruence, as represented in figure 2*C*. Here, *P* values obtained for each real gene data set have been randomly assigned to topologies. Figure 2*D* shows the same data as figure 2*A*, but clustered by topologies and by genes, to emphasize the apparent structure in this data.

To completely summarize patterns of support for topologies, it would be necessary to include all possible topologies. This is impractical for the data sets here and even the set of a priori plausible topologies included makes visualization difficult. In order to utilize the information from tests over a large number of topologies while easing visualization of results, we present in figure 3*A* heat maps with a restricted set of selected 16 plausible topologies for which a majority of genes had a *P* value larger than 0.05. This set of plausible topologies is thus constructed under the hypothesis of interest—that genes should share support for a single topology due to their common vertical descent—and focuses attention on genes that are incongruent with this hypothesis. Again, the real data appear to show less structure than simulated genes with a common topology but more than randomized data (not shown). Quantitative statements about structure in this data are possible but depend on the assessment of the number of true clusters exhibited by the data.

Clustering the Heat Maps

One value of our heat map approach, not discussed previously, is its potential to identify and enumerate clusters. It would seem reasonable to assume that there is a limited number of true clusters of genes each with its own true shared topology represented in the γ-proteobacterial data set and that an appropriate statistical treatment could reduce the noisiness of this pattern. Several methods for identifying true clusters have been described (Calinski and Harabasz 1974; Krzankowski and Lai 1985; Milligan and Cooper 1985; Gordon 1999; Tibshirani, Walther, and Hastie 2001), and we used three of them: KL index, CH index, and gap statistics. While these methods have been shown to work well in a number of settings, they gave contradictory results for γ-proteobacterial genes, and at least two aspects of these data make it particularly difficult. The first is the large numbers of topologies. These methods have been tested primarily in situations where the number of features (topologies) for each individual (gene) of interest is much smaller (less than 10). Data sparsity increases dramatically as the dimensionality of the space for clustering increases, requiring many more observations (individuals) for conclusive determination of clusters. The second feature of the present problem that makes determining the appropriate number of clusters difficult is the presence of many small clusters. It is easier to determine the number of clusters if the number of genes within each cluster is relatively large. Still clustering did seem to be present. For the γ-proteobacterial data set, the gap statistic gave 14 as an estimate of the number clusters. The CH and KL indices only allow estimation of cluster sizes greater than 1. Their estimates were 2 and 46. For the data set that included simulated LGT events, the cluster size estimates were 23, 35, and 5. The wide variation in these estimates reflects the inherent difficulties with large numbers of topologies alluded to above, a fact that is further illustrated by the more similar estimates in the case that attention was restricted to 16 plausible topologies. In this case, the estimates were 3 (CH), 9 (KL), and 5 (gap) for the γ-proteobacterial data.

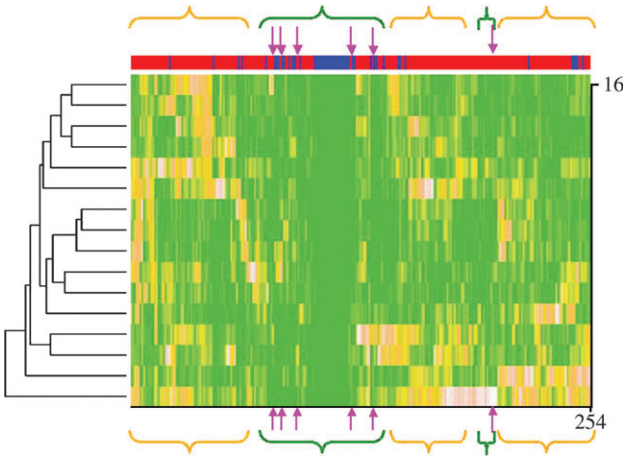


FIG. 4.—Heat map including two kinds of markers: γ-proteobacterial genes, indicated by a red rectangle at the left of the heat map, and artificial (simulated) markers with extreme LGT (see main text), indicated in blue. Simulated markers are based on a set of plausible topologies (see main text). The number of genes and topologies in the analysis are indicated on the heat map. These heat maps are double clustered by genes and by topologies. The hierarchical cluster on the left of the figure represents a tree of topologies along the heat map. In the left band, the relative distribution of red and blue rectangles reflects the presence/absence of clustering of actual markers with artificial ones. Colour codes are as in figure 2*A*. At the top and bottom of the heat map, the orange parentheses indicate regions containing markers with a weak discriminatory power; the green parentheses indicate regions containing markers with a stronger discriminatory power. Among the markers with a stronger phylogenetic signal, pink arrows point to some instances of conflicting signal in actual markers. They indicate different columns displaying a contrasting pattern of color and contradictory *P* values for several orthologues in a data set.

Highlighting the Conflicting Signal of the Heat Maps

We tested whether LGT genes would exhibit similar patterns of support to the actual γ-proteobacterial genes by simulating such events. Starting from the γ-proteobacterial markers and using the same 16 plausible topologies as in figure 3*A*, we simulated artificial genes that had undergone up to three recent LGTs by randomly assigning the sequence of one species to another (creating up to three identical rows in our alignments). As with the PCA (see fig. 1), heat maps could not discriminate firmly between the phylogenetic signal of γ-proteobacterial genes and the phylogenetic signal present in the artificially generated markers (fig. 4). When clustered by phylogenetic affinity, the markers with recent simulated LGT (indicated by a blue line in the band above the heat map) were not separated from the genuine γ-proteobacterial markers (in red in the upper band). Instead, they were interspersed with each other, producing a "bar-code" appearance. Regardless which estimate of the number of clusters is used—9 (CH), 9 (KL), or 4 (gap)—LGT-simulated genes cluster together with actual genes. We thus cannot safely conclude the absence of LGT from this data set. Notably, there is a cluster of artificial genes with LGT, rejecting all the plausible topologies, which also encompasses two true genes, themselves likely to have been transferred (the virulence factor MviN-like protein and of the biotin synthetase, as previously reported [Lerat, Daubin, and Moran 2003]).

However, it is possible to draw new and biologically relevant conclusions about the true signal that is present. Heat maps, unlike the PCA, provide explicit information about the groupings of genes through the display of the color pattern of support/rejection. If we look at figure 4 in more detail, we observe that there are two main categories of γ-proteobacterial markers. First, there are markers with a limited phylogenetic signal, which are in the region indicated by orange parentheses. These genes support or are compatible with multiple different topologies. The majority of the γ-proteobacterial genes belong to this category, and we must remain agnostic about their actual phylogenetic history at this taxonomic level. Second, some markers contain strong phylogenetic signal, being only compatible with only one or two of the plausible topologies and rejecting all others. (It must be recalled that it is rejection, not support, of many topologies that identifies genes with strong signal.) This second category of genes is indicated in regions encompassed by green parentheses. Only those genes can be used to define a set of candidates in which the presence of LGTs could be tested.

Ideally, if there is only one true tree and this tree is among the tested topologies, all the genes with a strong phylogenetic signal should choose it as the best tree to the exclusion of other topologies. Conversely, incompatible patterns of support/rejection among such genes must indicate conflicting signal in the data set. This is clearly observed here. As an example, we have used pink arrows to indicate several individual γ-proteobacterial markers that carry an incompatible phylogenetic signal (see the details in Supplementary Materials online). It should be recalled that here we are examining only the 16 topologies deemed most plausible because the majority of genes accept them ($P > 0.05$). There will be other topologies that are strongly supported by one or more of the genes in this data set but rejected by the majority. To state a provisional conclusion, conservatively: there is no reason for confidence that these 205 γ-proteobacterial genes or even some large fraction of them have an identical phylogeny.

Additional Heat Map Applications: Bootstrap Support

An appealing additional application of the heat map approach not presented before is to data sets larger than a few dozens taxa, using a bipartition (or splits) approach. Bipartition analyses reduce large data sets to two-way splits corresponding to all possible nodes on all possible trees. For a given 13-taxon tree, there are only 10 nonterminal (having at least two taxa in each partition) splits (see Supplementary Material online). Because more than one tree is favored by the γ-proteobacterial data set, there are 364 splits that receive positive bootstrap support for at least one of the 205 genes. Among these only 13 had bootstrap support larger than 5% for a majority of genes. The bootstrap support for these 13 splits is given with two-way clustering in figure 5*A*. One can see that there are eight splits that have large bootstrap support for the vast majority of genes. These splits are compatible and suggest that a large number of trees are in agreement with some patterns of vertical descent. However, a total of 10 nonterminal compatible splits are required to define the γ-proteobacterial tree.

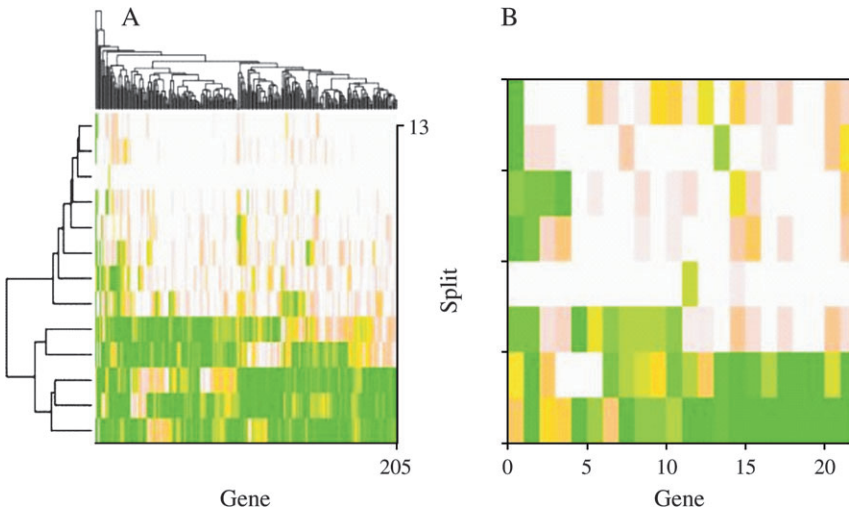


FIG. 5.—(*A*) Heat map of the bootstrap support for splits receiving at least 5% bootstrap support for a majority of genes. Colour codes are as in figure 2*A*. (*A*) Heat map for all of the 205 genes with dendrograms indicating the clustering of genes and splits. (*B*) Heat map for 22 of the genes, clustering together in the left of *A*, for the eight splits that are well supported for the majority of genes in *A*.

Thus, while many of the genes are congruent with some portion of the tree, there is considerable disagreement about where some subtrees should be placed. Clusters of genes support differing groups of some of the additional splits 9–13 that received bootstrap support 5% for a majority of genes. While a majority of the genes support the first eight splits, in figure 5*B*, which considers 22 of the genes on the left side of figure 5*A*, we see that a substantial number of genes do not uniformly support these splits either.

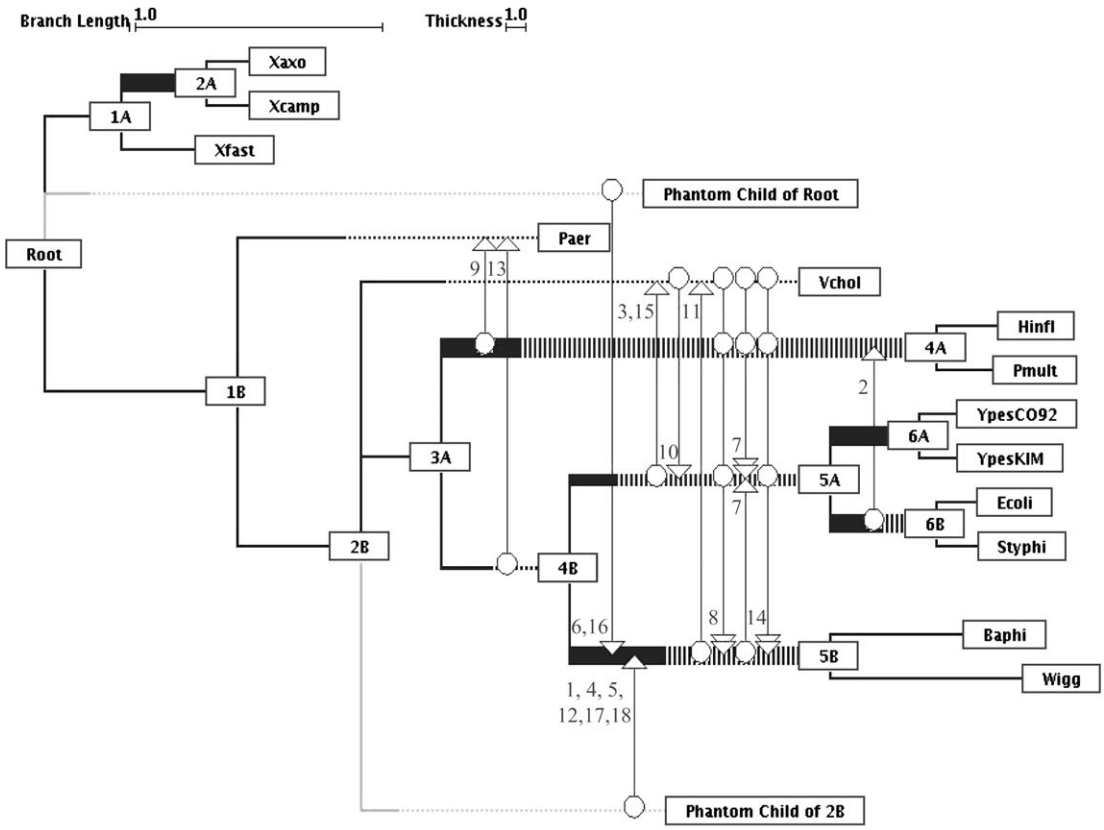
The Synthesis of γ-Proteobacteria

We summarized the safe phylogenetic information present in a synthesis, which allows explicit display of signal erosion, hidden paralogy, and or LGT. In this graph, partly treelike and partly weblike (fig. 6), 26 vertical branches are visible as well as 11 lateral connections. The comparison of the total support for the horizontal and vertical branches indicates that the vertical signal is about 13 times more important than the horizontal signal. However, this evidence for vertical descent is confined to seven of the trees' nodes: the remaining six (as indicated by the minimal thickness of the blue lines) are supported (bootstrap value >50% by none of the genes, and node 4B is supported at this level by less than 50 genes). This synthesis shows us clearly that we simply do not know what the history of most genes is, for most of the nodes. Almost certainly, there are 18 genes (rpl17, rpl8p, rpl18, rpl27, rpl32p, rpl34, rps6, glutamyl-tRNA synthetase, seryl-tRNA synthetase, riboflavin kinase, an outer membrane antigen protein, dihydrolipoamide acetyltransferase, putative deaminase protein, uridine diphosphate-*N*-acetylmuramate:alanine ligase, tRNA pseudouridine synthase B, methionine adenosyltransferase, a thioredoxin-like protein, and an hypothetical protein) that have undergone LGT. Phylogenetically conclusive genes are however not numerous enough to allow us to generalize about LGT in γ-proteobacteria. We can note however that 72.7% of these transfers correspond to local rearrangements of the backbone tree. For instance, the ribosomal protein rpl32 would have been transmitted from the ancestor of *H. influenzae* and *P. multocida* (4A) to *Pseudomonas aeruginosa*. Adding more and more markers could certainly lead us to a better description of the complex relationships between γ-proteobacterial lineages and provide information about the relative importance of lateral/vertical genetic exchanges occurring between them. Likely, a graph incorporating the phylogenetic information from the 20 or more times as many γ-protebacterial genes which are not part of the γ-protebacterial core and which one can assume to be more LGT prone will be less treelike.

Conclusion

Our study of the γ-proteobacterial core genes led to a conclusion similar to that in Bapteste et al. (2005) on other core genes: we do not really know their history. Here, we show that not all the γ-proteobacterial core genes have a similar phylogenetic signal, and they rarely favor only a single tree. From the detailed investigation of support/rejection patterns of some genes with a stronger phylogenetic signal, we suggest that this core likely comprises genes with several contradictory histories. There is thus an easy but important message in this analysis: although there is clearly a central tendency in the data set (it is far from random), LGT cannot be ruled out. Claims (Flintoft 2003) for a single tree for γ-proteobacteria, when at least 2 and up to 14 strong contradictory signals are present in their core gene, should be accompanied by the disclaimer that the tree is somehow an average of the histories carried by these markers, rather than the organismal tree *sensu stricto*.

We have identified among these core genes 18 candidates for LGT. Although each should be investigated more thoroughly at several biological levels, we suggest that this is a minimum number for genes with truly conflicting signal. Nevertheless, differences in evolutionary rates such as



Genes with suggested transfers:
1. a thioredoxin-like protein, 2. dihydrolipoamide acetyltransferase, 3. methionine adenosyltransferase, 4. rpl18, 5. tRNA pseudouridine synthase B, 6. rpl27, 7. an hypothetical deaminase, 8.hypothetical protein, 9. rpl32p, 10. UDP-N-acetylmuramate:alanine ligase, 11. seryl-tRNA synthetase, 12. rpl17, 13. rpl34, 14. rps6, 15. outer membrane antigen protein, 16. glutamyl-tRNA synthetase, 17. riboflavin kinase and 18. Rpl8.

FIG. 6.—Synthesis of 205 γ-proteobacterial genes. The proposed vertical-inheritance backbone is shown in dark blue, with the line thickness of an internal branch corresponding to the frequency of its support across the whole data set. Support was considered significant when clades received >50% bootstrap support. Putative LGT events are in red, connecting donors (circles) with recipients (arrowheads); where there are multiple possible donor candidates, these converge onto a double arrowhead. This happens when the clade founded by a past LGT donor may have subsequently had its species membership obscured by later exchanges of genetic material, yielding a nonreference assemblage of species labels in a presumed lineage. Where the apparent donor of a gene falls outside of the taxa included in the analysis, one is created as a basal group taxon, indicated in light blue. In order to avoid graphical congestion, branches in the tree may be artificially extended, as dotted segments. Baphi, *Buchnera aphidicola*; E. coli, *Escherichia coli*; Hinfl, *Haemophilus influenzae*; Paer, *Pseudomonas aeruginosa*; Pmult, *Pasteurella multocida*; Styphi, *Salmonella enterica*; Vchol, *Vibrio cholerae*; Wigg, *Wigglesworthia glossinidia*; Xaxo, *Xanthomonas axonopodis*; Xcamp, *Xanthomonas campestris*; Xfast, *Xylella fastidiosa*; YpesC092, *Yersinia pestis* CO92; and YpesKIM, *Yersinia pestis* KIM.

long-branch attraction rather than LGT could explain a part of these results. Hence, possible systematic biases were also tested in the supplementary materials (see Supplementary Materials online). Disproving absolutely that long-branch attraction is responsible for the mosaic heat maps produced here might not be possible though, and in any case was not our goal. One could probably argue endlessly (and legitimately) that evolutionary models are flawed, so that any difference between two trees would ultimately be claimed to be of methodological origin. Molecular phylogenetics has entered a self-critical phase, in which many authors call into question the validity of conclusions about the history of life based on misspecified models of the evolutionary process, even when LGT can be eliminated as a cause of conflict. We suggest that further examination of congruence in core gene sets should be undertaken in this context and not the metatheoretical one of whether or not there is a tree of life. In particular, the tendency to require that evidence for LGT should be proven but not the null hypothesis of vertical descent and that when present LGT should be considered as "noise" instead of "signal" (Kurland, Canback, and Berg 2003; Lake and Rivera 2004; Snel, Huynen, and Dutilh 2005) should be resisted. As Paul Feyerabend (1975) has noted, "the consistency condition which demands that new hypotheses agree with accepted theories is unreasonable because it preserves the older theory, and not the better theory. Hypotheses contradicting well-confirmed theories give us evidence that cannot be obtained in any other way. Proliferation of theories is beneficial for science, while uniformity impairs its critical power." Hence, "the consistency condition … eliminates a theory not because it is in disagreement with the facts, but because it is in disagreement with another theory …. It thereby makes the as yet untested part of that theory a measure of validity."

Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank D. Walsh for critical reading of our manuscript. E.B. was supported by a Canadian Institutes of Health Research grant MOP4467.

Literature Cited

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D., and A. J. Levine. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci USA **96**:6745–6750.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

Bapteste, E., Y. Boucher, J. Leigh, and W. F. Doolittle. 2004. Phylogenetic reconstruction and lateral gene transfer. Trends Microbiol. **12**:406–411.

Bapteste, E., H. Brinkmann, J. A. Lee et al. (11 co-authors). 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. Proc. Natl. Acad. Sci. USA **99**:1414–1419.

Bapteste, E., E. Susko, J. Leigh, D. MacLeod, R. L. Charlebois, and W. F. Doolittle. 2005. Do orthologous gene phylogenies really support tree-thinking? BMC Evol. Biol. **5:**33.

Brochier, C., E. Bapteste, D. Moreira, and H. Philippe. 2002. Eubacterial phylogeny based on translational apparatus proteins. Trends Genet. **18**:1–5.

Brown, J. R., and C. Volker. 2004. Phylogeny of gamma-proteobacteria: resolution of one branch of the universal tree? Bioessays **26**:463–468.

Calinski, R. B., and J. Harabasz. 1974. A dendrite method for cluster analysis. Commun. Stat. **3**:1–27.

Creevey, C. J., D. A. Fitzpatrick, G. K. Philip, R. J. Kinsella, M. J. O'Connell, M. M. Pentony, S. A. Travers, M. Wilkinson, and J. O. McInerney. 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? Proc. R. Soc. Lond. B Biol. Sci. **271**:2551–2558.

Darwin, C. 1859. On the origin of species. John Murray, London.

Deppenmeier, U., A. Johann, T. Hartsch, et al. (20 co-authors). 2002. The genome of Methanosarcina mazei: evidence for lateral gene transfer between bacteria and archaea. J. Mol. Microbiol. Biotechnol. **4**:453–461.

Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. Science **284**:2124–2129.

Feyerabend, P. 1975. Against method. Verso, London.

Flintoft, L. 2003. Family tree for gamma-proteobacteria. Nat. Rev. Microbiol. **1:**1.

Getz, G., H. Gal, I. Kela, D. A. Notterman, and E. Domany. 2003. Coupled two-way clustering analysis of breast cancer and colon cancer gene expression data. Bioinformatics **19**: 1079–1089.

Getz, G., E. Levine, and E. Domany. 2000. Coupled two-way clustering analysis of gene microarray data. Proc. Natl. Acad. Sci. USA **97**:12079–12084.

Gordon, A. 1999. Classification. Chapman-Hall, London.

Grassly, N. C., J. Adachi, and A. Rambaut. 1997. PSeq-Gen: an application for the Monte Carlo simulation of protein sequence evolution along phylogenetic trees. Comput. Appl. Biosci. **13**:559–560.

Harris, J. K., S. T. Kelley, G. B. Spiegelman, and N. R. Pace. 2003. The genetic core of the universal ancestor. Genome Res. **13**:407–412.

Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. Proc. Natl. Acad. Sci. USA **96**:3801–3806.

Krzankowski, W. J., and Y. T. Lai. 1985. A criterion for determining the number of groups in a data set using sum of squares clustering. Biometrics **44**:22–34.

Kuhn, T. 1962. The structure of scientific revolutions. University of Chicago Press, Chicago, Ill.

Kumar, S., and A. Rzhetsky. 1996. Evolutionary relationships of eukaryotic kingdoms. J. Mol. Evol. **42**:183–193.

Kurland, C. G., B. Canback, and O. G. Berg. 2003. Horizontal gene transfer: a critical view. Proc. Natl. Acad. Sci. USA **100**:9658–9662.

Lake, J. A., and M. C. Rivera. 2004. Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. Mol. Biol. Evol. **21**:681–690.

Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-proteobacteria. PLoS Biol. **1:**E19.

MacLeod, D., R. L. Charlebois, F. Doolittle, and E. Bapteste. 2005. Deduction of probable events of lateral gene transfer through comparison of phylogenetic trees by recursive consolidation and rearrangement. BMC Evol. Biol. **5:**27.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. Multivariate analysis. Academic Press, London.

Matte-Tailliez, O., C. Brochier, P. Forterre, and H. Philippe. 2002. Archaeal phylogeny based on ribosomal proteins. Mol. Biol. Evol. **5**:631–639.

Milligan, G. W., and M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika **50**:159–179.

Nelson, K. E., R. A. Clayton, S. R. Gill, et al. (20 co-authors). 1999. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of Thermotoga maritima. Nature **399**:323–329.

Nesbo, C. L., Y. Boucher, and W. F. Doolittle. 2001. Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. J. Mol. Evol. **53**:340–350.

Philippe, H. 1993. MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res. **21**:5264–5272.

Philippe, H., and C. J. Douady. 2003. Horizontal gene transfer and phylogenetics. Curr. Opin. Microbiol. **6**:498–505.

Raymond, J., O. Zhaxybayeva, J. P. Gogarten, S. Y. Gerdes, and R. E. Blankenship. 2002. Whole-genome analysis of photosynthetic prokaryotes. Science **298**:1616–1620.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. **51**:492–508.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114–1116.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246–1247.

Snel, B., M. A. Huynen, and B. E. Dutilh. 2005. Genome trees and the nature of genome evolution. Annu Rev Microbiol **59:**191–209.

Somogyi, R., S. Fuhrman, G. Anderson, C. Madill, L. D. Greller, and B. Chang. 2004. Systematic exploration and mining of gene expression data provides evidence for higher-order, modular regulation. Pp. 202–221 *in* Ga G. P. W. Schlosser, ed. Modularity in development and evolution. University of Chicago Press, Chicago.

Teichmann, S. A., and G. Mitchison. 1999. Is there a phylogenetic signal in prokaryote proteins? J. Mol. Evol. **49**:98–107.

Tibshirani, R., G. Walther, and T. Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. B:411–423.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S-PLUS. Springer-Verlag, New York.

Wertz, J. E., C. Goldstone, D. M. Gordon, and M. A. Riley. 2003. A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J. Evol. Biol. **16**:1236–1248.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. Proc. Natl. Acad. Sci. USA **87**:4576–4579.

William Martin, Associate Editor

Accepted January 30, 2006