

Horizontal Gene Transfer: Pitfalls and Promises

J. PETER GOGARTEN, RYAN D. MURPHEY, AND LORRAINE OLENDZENSKI

Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06269-3044

Phylogenetic reconstruction from protein or nucleic acid sequence families provides information on the evolution of individual genes. In contrast to the assumed bifurcating, tree-like evolution of genes, organismal evolution is characterized by the exchange of genetic information between organisms and even by the fusion of formerly independent lines of descent (1). The invocation of horizontal gene transfer events is often regarded as a last-ditch attempt by systematists to reconcile conflicting phylogenies constructed from different markers. In general, however, organismal evolution is clearly visible as the majority consensus of a number of molecular phylogenies, and transfer events can be recognized in phylogenies constructed from one or several markers whose topologies deviate from that of the consensus.

Often it is difficult to decide whether conflicts between molecular phylogenies are due to actual events in evolution (horizontal gene transfer or gene duplications [see 2]), or due to artifacts generated during phylogenetic reconstruction. For example, investigation of the maximum likelihood landscape of 18S rRNA and V-ATPase A-subunit phylogenies suggests that the grouping of microsporidia either with the fungi (3, 4, 5) or close to the root of the archaeal domain (*e.g.*, 6, 7, 8) probably represents an artifact in the 18S rRNA data analyses and not another case of horizontal transfer.

Although the recognition of horizontal transfer as a major factor in prokaryotic evolution (*e.g.*, 9, 10, 11, 12) certainly complicates the interpretation of molecular phylogenies, it also allows synchronization of different parts of the universal tree of life, and thus might provide the key to the detection of periods of rapid substitutions. Two examples

will be presented of attempts to extract organismal evolution from conflicting molecular phylogenies. One concerns the finding of archaeal genes in Deinococcaceae, the other the finding of bacterial genes in Archaea.

Thermus and *Deinococcus* both have an archaeal-type coupling factor ATPase (13, 14, 15); in contrast, most of the other molecular markers available clearly indicate that the group comprising *Thermus* and *Deinococcus* species is an early branching lineage within the Bacteria (*e.g.*, 16, 17). Reflecting this majority consensus, *Thermus* and *Deinococcus* can be considered as Bacteria that branch off before the major bacterial lineages evolved in an at present ill-resolved radiation. However, in addition to the archaeal-type coupling factor ATPase, the malic acid dehydrogenases found in Deinococcaceae do not group with the typical bacterial homologs, but closer to the eukaryotes. The interpretation of the malic acid dehydrogenase phylogenies is complicated by potential switches in substrate specificities between lactic and malic acid dehydrogenases. Nevertheless, these results prompted us to look for additional archaeal genes in *Thermus ruber*. A genomic library of *Thermus ruber* was screened with labeled genomic DNA from the archaeon *Thermoplasma acidophilum*. One of the clones that showed stronger cross reactivity to the archaeal probe encodes a prolyl tRNA synthetase. Phylogenetic analysis of this sequence and its homolog from *Deinococcus radiodurans*, obtained through The Institute for Genomic Research at <http://www.TIGR.org/>, reveals that the prolyl tRNA synthetase from Deinococcaceae groups with the eucaryal and archaeal homologs and not with the typical bacterial prolyl tRNA synthetases (Fig. 1). Distance, parsimony, and quartet puzzling analyses show similar tree topologies. These trees suggest two major types of prolyl tRNA synthetases: a bacterial type and a eukaryotic/archaeal type (see ref. 18 for an in-depth comparison of the bacterial and eucaryal types). The majority of bacteria form a group that also includes *Candida* and *Saccharomyces* sequences, which may be of mitochondrial origin. Grouping with the archaeal and eu-

This paper was originally presented at a workshop titled *Evolution: A Molecular Point of View*. The workshop, which was held at the Marine Biological Laboratory, Woods Hole, Massachusetts, from 24–26 October 1997, was sponsored by the Center for Advanced Studies in the Space Life Sciences at MBL and funded by the National Aeronautics and Space Administration under Cooperative Agreement NCC 2-896.

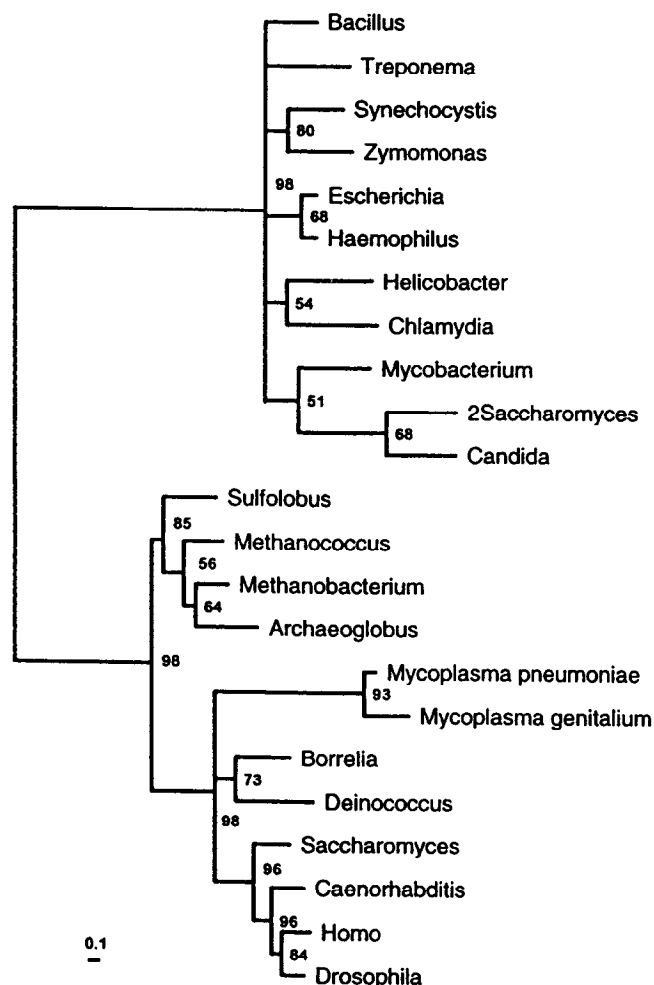


Figure 1. Quartet puzzling analysis (21) of prolyl tRNA synthetase amino acid sequences. Sequences were aligned using the profile alignment option of ClustalX with default parameters (22). Numbers reflect quartet puzzling support values for each node. Nodes with values of less than 50% have been collapsed. Analysis was corrected for among site rate variation using the Gamma distribution ($\Gamma=1.97$). All sequences were obtained from GenBank except those of *Deinococcus* and *Treponema*, which were obtained through BLAST search from The Institute for Genomic Research, and *Sulfolobus*, which was kindly provided by Mark Regan and Christoph Sensen, Institute for Marine Biosciences, Dalhousie University, Halifax, Nova Scotia.

karyotic sequences are the bacteria *Deinococcus*, *Borrelia*, and *Mycoplasma*. All three of these bacteria appear to have acquired a eukaryotic/archaeal-type prolyl tRNA synthetase. The finding that the gene encoding prolyl tRNA synthetase in *Sulfolobus sulfotaricus* is located immediately downstream from the operon encoding the archaeal coupling factor ATPase (Mark Regan and Christoph Sensen, Dalhousie University; pers. comm.) suggests that these genes could have been transferred simultaneously before the evolutionary divergence of *Deinococcus* and *Thermus*. Interestingly, *Borrelia* and *Treponema*, both spirochetes, each have a different prolyl tRNA synthetase type. *Borrelia*

possesses a eukaryotic/archaeal type, whereas *Treponema* has the type found more commonly in other bacteria.

One noteworthy exception from using the majority consensus as backbone for the organismal phylogeny are the Archaea. Although many characters clearly characterize the Archaea as a group that is distinct from the Bacteria (2), recent complete genome analyses confirm earlier observations that many (44%) archaeal genes are very similar to their bacterial homologs, whereas only 13% are more similar to their eukaryotic homologs (19). If gene families are analyzed in more detail, the archaeal genes are found to group within the bacterial domain, and the different archaeal genes do not even cluster together (2). Thus the majority consensus does not show Archaea and Bacteria to be sister groups and would dissolve both the Archaea and the Bacteria into paraphyletic groups. In contrast, most genes involved in genome structure, transcription, translation, and chemiosmotic coupling clearly set the Archaea apart from the Bacteria. In this case the backbone of our preliminary organismal phylogeny reflects those markers less prone to horizontal gene transfer, and explains the presence of many genes in the Archaea that are bacterial in character as the result of horizontal gene transfer. The fact that so many genes found in the archaeal genomes appear to be bacterial in character raises the question of how these genes were transferred. Initially, a single horizontal transfer event of a major portion of the genome was considered (1, 19); however, comparison of individual gene families suggests at least several transfer events, each involving a subset of the total complement of genes transferred (20).

Acknowledgments

This work was supported through the NASA-Exobiology program.

Literature Cited

- Gogarten, J. P. 1995. The early evolution of cellular life. *Trends Ecol. Evol.* 10: 147–151.
- Gogarten, J. P., E. Hilario, and L. Olendzenski. 1996. Gene duplications and horizontal gene transfer during early evolution. Pp. 267–292 in *Evolution of Microbial Life*, Society for General Microbiology, D. McL. Roberts, P. Sharp, G. Alderson, and M. Collins, eds., University Press, Cambridge, UK.
- Edlind, T. D., J. Li, G. S. Visvesvara, M. H. Vodkin, G. L. McLaughlin, and S. K. Katiyar. 1996. Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. *Mol. Phylogenet. Evol.* 5: 359–367.
- Keeling, P. J., and W. F. Doolittle. 1996. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* 13: 1297–1305.
- Müller, M. 1997. What are the Microsporidia? *Parasitol. Today* 13: 455–456.
- Sogin, M. L., J. H. Gunderson, H. J. Elwood, R. A. Alonso, and D. A. Peattie. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243: 75–77.

7. Sogin, M. L. 1991. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**: 457–463.
8. Kamaishi, T., T. Hashimoto, Y. Nakamura, F. Nakamura, S. Murata, N. Okada, K. Okamoto, M. Shimizu, and M. Hasegawa. 1996. Protein phylogeny of translation elongation factor EF-1 alpha suggests microsporidians are extremely ancient eukaryotes. *J. Mol. Evol.* **42**: 257–263.
9. Lorenz, M. G., and W. Wackernagel. 1994. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**: 563–602.
10. Lawrence, J. G., and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**: 383–397.
11. Herrick, J. B., K. G. Stuart-Keil, W. C. Ghiorse, and E. L. Madson. 1997. Natural horizontal transfer of a naphthalene dioxygenase gene between bacteria native to a coal tar-contaminated field site. *Appl. Environ. Microbiol.* **63**: 2330–2337.
12. Lawrence, J. G., and J. R. Roth. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843–1860.
13. Yokoyama, K., T. Oshima, and M. Yoshida. 1990. *Thermus thermophilus* membrane-associated ATPase; indication of a eubacterial V-type ATPase. *J. Biol. Chem.* **265**: 21946–21950.
14. Tsutsumi, S., K. Denda, K. Yokoyama, T. Oshima, T. Date, and M. Yoshida. 1991. Molecular cloning of genes encoding major two subunits of a eubacterial V-Type ATPase from *Thermus thermophilus*. *Biochim. Biophys. Acta* **1098**: 13–20.
15. Gogarten, J. P., T. Starke, H. Kibak, J. Fichmann, and L. Taiz. 1992. Evolution and isoforms of V-ATPase subunits. *J. Exp. Biol.* **172**: 137–147.
16. C. R. Woese. 1987. Bacterial evolution. *Microbiol. Rev.* **51**: 221–271.
17. Hensel, R., W. Demharter, O. Kandler, M. Kroppenstedt, and E. Stackebrandt. 1986. Chemotaxonomic and molecular-genetic studies of the genus *Thermus*: evidence for a phylogenetic relationship of *Thermus aquaticus* and *Thermus ruber* to the genus *Deinococcus*. *Int. J. Syst. Bacteriol.* **36**: 444–453.
18. Stehlin, C., B. Burke, F. Yang, H. Liu, K. Shiba, and K. Musier-Forsyth. 1998. Species-specific differences in the operational RNA code for aminoacylation of tRNA^{Pro}. *Biochemistry* **9**: 8605–8613.
19. Koonin, E. V., A. R. Mushegian, M. Y. Galperin, and D. R. Walker. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* **25**: 619–637.
20. Olenzenski L., E. Hilario, and J. P. Gogarten. 1998. Horizontal gene transfer and fusing lines of descent: the Archaeobacteria—a chimera? Pp 349–362 in *Horizontal Gene Transfer*, M. Syvanen and C. Kado, eds., Chapman and Hall, London.
21. Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**: 964–969.
22. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.

Discussion

ELLINGTON: It seems that phylogeneticists assume that all genes work the same way in all organisms. Is there a possibility that different biochemical usages, even of something as common as malate dehydrogenase, could lead to convergent evolution, and that convergence would appear as if it were horizontal transfer?

GOGARTEN: At least in theory this seems a viable alternative. However, if one examines actual data, one rarely finds instances of convergent evolution. There are instances of convergent substitutions in lineages that independently evolved in extreme environments, for example, high-salt environments (1). A more serious problem for phylogeneticists is that if an enzyme evolves to perform a different function, one often observes faster substitution rates. That does not constitute convergent evolution in itself, but the accelerated rates of evolution can lead to long-branch attraction artifacts (2), which might look very much like convergent evolution. An example might be the ATPase catalytic subunit in *Borrelia*. When this enzyme was transferred from one domain to another, it had to adjust to the different cellular environment and therefore might have had much higher substitution rates.

MARGULIS: When you moved microsporidians like *Nosema* and others in with the fungi and showed maximum likelihood data, you said that *Giardia* is then no longer the deepest branch (oldest ancestor). What is the deepest branching eukaryotic lineage?

GOGARTEN: The tree that according to a distance matrix analysis best described the V-ATPase catalytic subunit evolution had *Giardia* as the deepest and the microsporidia as the second deepest branch within the protists. When we moved the microsporidia into the fungi, we found that the resulting tree actually had a higher likelihood than the starting tree. However, when we moved *Giardia* away from the bottom of the tree, the likelihood dropped significantly. Thus, based on the V-ATPase sequences, *Giardia* still appears to be the deepest branch (3). We also looked at the 18S rRNA tree, and found that both the microsporidia and *Giardia* are highly supported as deep branches within the eukaryotic domain. However, when we moved the branch leading to the microsporidia throughout the eukaryotic domain, we found a second relative likelihood maximum when the microsporidia are part of the eukaryotic crown group. Thus, although the absolute maxima in the likelihood landscape are found with very different tree topologies, we found a dual affinity for the microsporidia for both the ATPase catalytic subunits and the 18S rRNAs. This suggests that ribosomal rRNAs and proteins had the same phylogenetic history. The reason that we find two very different tree topologies is probably due to problems associated with phylogenetic reconstruction.

KUHLMAN: You said that spirochetes have an archaeal-type ATPase. It was reported earlier in this meeting that they also have

this archaeal-type lysyl tRNA synthetase. Is there any evidence from other genes that the spirochetes and the *Deinococcaceae* are really a natural group? Am I correct that this suggests there is one transfer?

GOGARTEN: I agree with you that one possible solution is to group the *Deinococcaceae* and the spirochaetes together. One then needs to postulate only a single horizontal transfer event. However, if one takes a closer look at the different gene trees, things are not so straightforward. Spirochaetes and *Deinococcaceae* both have the archaeal/vacuolar-type ATPase; in the case of the prolyl tRNA synthetase, the *Deinococcaceae* and *Borrelia* have the archaeal/eucaryal-type enzyme, but *Treponema* has the bacterial type. In the case of the lysyl tRNA synthetase (4) the spirochaetes have the archaeal type of enzyme, but *Deinococcus* does not. One possible explanation for these observations is that a single inter-domain transfer took place just before the split between spirochaetes and *Deinococcaceae*, and that different subsets of genes were completely integrated into the genomes of the different lines of descent. The radiation that gave rise to the different bacterial kingdoms appears as a very rapid event. I think we are fooling ourselves if we believe that 16S rRNA or ATPases would reliably resolve this radiation. If you construct a tree, some groups are slightly closer together for one marker and further apart for another. I wouldn't be surprised if spirochaetes and *Deinococcaceae* actually turn out to group together.

CAVALIER-SMITH: I agree with your final comment. On the other hand, I think it is even worse than you have implied. The logic of rooting the tree by reciprocal rooting of duplicate proteins is good,

but in practice, it's probably a lot more difficult than is often thought. In general, the outgroup is a very long branch that is likely to introduce systematic bias into the rooting. Therefore I am more skeptical than most that we really can root the tree of life from the molecular trees.

GOGARTEN: We only have data which suggest that the root is placed between all of the eubacteria on one side and the archaea and eukaryotes on the other. The branches are long, and it could be that there are some problems with long branches. Based on our simulations, I don't have the impression that artifacts due to long branches would be so big as to pull out the root from inside the bacterial domain. However, this is clearly an important question that warrants further careful analysis. The reason I placed the root between the eubacteria and the archaea/eukaryotes is that this is the place suggested by the currently available data.

Literature Cited

1. **Gandhir, M., I. Rasched, P. Marliere, and R. Mutzel. 1995.** Convergent evolution of amino acid usage in archaeobacterial and eubacterial lineages adapted to high salt. *Res. Microbiol.* **146**: 113–120.
2. **Felsenstein, J. 1978.** Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401–410.
3. **Hilario, E., and J. P. Gogarten. 1998.** The prokaryote to eukaryote transition reflected in the evolution of the V/F/A-ATPase catalytic and proteolipid subunits. *J. Mol. Evol.* **46**: 703–715.
4. **Ibba, M., J. L. Bono, P. A. Rosa, and D. Söll. 1997.** Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. USA* **94**: 14383–14388.