

Is the microbial tree of life verificationist?

E.K. Lienau and R. DeSalle*

Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th street at Central Park West, New York, NY 10024, USA

Accepted 8 August 2009

Abstract

The field of microbial phylogenetics has questioned the feasibility of using a tree-like structure to describe microbial evolution. This debate centres on two main points. First, because microorganisms are able to transfer genes from one to another in zero generations (horizontal gene transfer, or HGT), the use of molecular characters to perform phylogenetic analyses will yield an erroneous topology and HGT clearly makes the evolution of microorganisms non tree-like. Second, the use of concatenated gene sequences in a total evidence approach to phylogenetic systematics is a verificationist endeavour, the aim of which is to bolster support. However, the goal of the total evidence approach to phylogenetic research is based in the idea of increasing explanatory power over background knowledge through test and corroboration, rather than to bolster support for nodes in a tree. In this context, the testing of phylogenetic data is a falsificationist endeavour that includes the possibility of not rejecting the null hypothesis that there is no tree-like structure in molecular phylogenetic data. We discuss several tests that aim to test rigorously the hypothesis that a tree of life exists for microorganisms. We also discuss the philosophical ramifications of background knowledge and corroboration in microbial studies that need to be considered when suggesting that HGT confounds the tree of life.

© The Willi Hennig Society 2009.

Recent attempts to fell the tree of life

Recently, whole genome prokaryotic and tree of life (ToL) phylogenetics has been viewed as a “useless and inscrutable endeavor” because of the prevalence of horizontal gene transfer (HGT) (Doolittle and Baptiste, 2007; Boucher and Baptiste, 2009; Doolittle and Zhaxybayeva, 2009; Lopez and Baptiste, 2009). [The above phrase comes from a review of one of the current authors’ papers that reconstructs the tree of life using a concatenated amino acid matrix. The reviewer was a microbialist.] In an earlier paper, Baptiste et al. (2008) articulated this idea, claiming: “it is safer to assume a comb-like topology of life”. This “safe” topology would be a polytomy, reflecting the apparent lack of resolution at deep nodes in the ToL due to the inability (because of HGT) of the data at hand to resolve a bifurcating relationship. These and other publications (Creevey et al., 2004; Dagan and Martin, 2007) have led to the attitude amongst many microbiologists that the tree of

life has been felled, and has been called the “microbi-alist” viewpoint by (Dagan and Martin, 2007).

In promulgating this argument, several authors have made the claim that phylogenetic tree building is verificationist. Specifically, Dagan and Martin (2007) suggest that the assumption of a tree-like structure of the relationships of life is a positivist philosophical construct. Baptiste and Boucher (2009) and Bucknam et al. (2006) also erect arguments that lead to the conclusion that the construction of the ToL is verificationist. Although Dagan and Martin (2007) do not offer a philosophical argument for this viewpoint, Bucknam et al. (2006) do. With specific reference to Carnap (see below), they suggest that “total evidence methods have a defined epistemological background: they belong to the *verificationist* toolkit” (Bucknam et al., 2006, p. 7; emphasis added). The alternative approach is the falsificationist approach as articulated by Popper in several publications. We agree with Bucknam et al. (2006) that the falsificationist approach is the most appropriate one for phylogenetics. However, we take issue with their interpretation of what is and what is not falsificationist in phylogenetic analysis.

*Corresponding author:

E-mail address: desalle@amnh.org

As an example of their notion of falsificationism, consider the following statement from Bucknam et al. (2006, p. 7): ‘... instead of aiming to confirm clades *via* a progressive increase in their bootstrap support when more data is added, a falsificationist would more likely be interested first by the rejection of impossible groupings, underlining what can not be a natural group.’

We agree with their assessment of the search for progressive increases of support as being verificationist. Dagan and Martin (2007, p 84) also make this observation by stating that recent studies that use concatenation approaches to construct a ToL (e.g. Ciccarelli et al., 2006) work “by accumulating data that are compatible with the null hypothesis of a common tree without having a chance to refute it”, an approach that is indeed verificationist, if the initial null hypothesis is that a ToL exists. Bucknam et al. (2006) developed two interesting new approaches [progressive reconstruction analysis (PRA) and concatenation analysis (CA)] to reassess the validity of concatenation in ToL inference. However, some, such as Baptiste et al. (2008), see the null hypothesis of phylogenetic systematics as a lack of phylogenetic structure in the data, or a bush (a safe phylogeny) which, if not rejected, could be explained as all organisms being studied being genetically identical, that all organisms are genetically distinct, or that reticulate dissemination of genetic information is the rule.

Notwithstanding, Dagan and Martin (2007) interpretation that the goal of total evidence or concatenation is to obtain higher and higher bootstrap support is mistaken. We suggest that the validity of a tree-like structure to represent evolutionary history cannot be tested by any method that purports to test whether the preponderance of data supports this tree over that tree. This leads us to the observation that the microbialist philosophical argument ignores one of the great principles of the philosophy of science as outlined by Popper (2002), namely corroboration. We also suggest that background knowledge as articulated by Popper is an important aspect of phylogenetic science that has not been discussed in any detail by authors addressing the HGT problem. Here we attempt to clarify, based on four approaches, some of the philosophical problems brought on by attempts to fell the ToL. First, we ask, “can we test for a ToL using tree thinking?” Next we attempt to clarify what concatenation does in a philosophical context and third we examine the background knowledge inherent in ToL studies. Finally, we examine the principle of corroboration as a guiding light to a falsificationist, bifurcating ToL.

Three phylogenetic tests for the existence of a vertical ToL

We agree that to deny the existence of HGT and its potential effects on vertical evolution would be

frivolous. However, although some studies have demonstrated large amounts of HGT amongst organisms, to date there have been very few rigorous attempts to test the hypothesis that a vertical ToL (vToL) exists. The jump from the existence of a large amount of HGT to the lack of a vToL is usually made in an anecdotal fashion. We suggest that a proper assessment of the futility of a vToL should include tests of the following hypotheses that will address the reason for the current lack of support at the base of the ToL. First, a rigorous test of whether or not a supermatrix composed of information using as much sequence as possible (regardless of HGT) will give a resolved tree needs to be accomplished. Second, if a resolved tree is produced, a rigorous test of whether the nodes in the tree are robust lends credence to the phylogeny; however, see Kluge (1998) for arguments against the utility of support measures in phylogenetic analysis, an especially interesting concept when considering noisy data (like the ToL). Any analysis that passes these first two tests could be said to have uncovered a most corroborated signal (*sensu* Lienau et al., unpublished data) for the massively concatenated supermatrix of all available sequence data. The revelation behind this signal would be antecedent to the most important test, as passing the first two tests as discussed above does not in and of itself test the existence of a vToL. A final and most rigorous test of whether or not the relationships implied by the resultant tree are representative of organismal branching patterns is needed. These tests are meant to assess whether the polytomy at the base of the ToL is hard (adding data gives no resolution) or soft (adding data gives resolution), and if resolution is achieved (whether it is believably supported).

For the first test, if a poorly resolved tree results from such an analysis, then we would have to conclude, as have Baptiste et al. (2006), that large amounts of sequence information add little to a ToL hypothesis and we would consider the hard polytomy hypothesis. In this case, we would be forced to admit to the approach as misleading and not useful until more data could be added.

But when is enough, enough? That is, it is interesting and timely, given the amount of molecular data that is just becoming available to microbial phylogeneticists, to speculate just what criteria would suffice to test the hard polytomy hypothesis rigorously enough for it to be accepted as well corroborated. Furthermore, we have to consider what a hard polytomy hypothesis would mean in a biological sense for different group of microbes.

First, gathering and testing of data never ends, which is why a scientist can say, “I ain’t often right, but I’ve never been wrong” (from *Scarlet Begonias* by Hunter and Goscia, 1974; Dodd, 2007). However, given a static taxonomic sample, for genome sequences one can make the reasonable claim that fully sequenced genomes

represent the total evidence available for testing a relationship of a fixed number of taxa. Taxonomic ranges, of course, can never be known, as there may always be new discoveries just around the corner that change the entire relationship scheme of a phylogeny. One reasonable way to approach this problem is through the examination of the universality and precision used to make ToL hypotheses (Lienau and DeSalle, 2009). By using more universal taxa choices a systematist can make the best statement possible about the phenomenon being studied.

If this is the case in microbial genomic phylogenetics, the interpretation of the hard polytomy can be made in different ways for different groups given background knowledge about the biology of the group. A well-corroborated hard polytomy hypothesis in bacterial and archaeal phylogenetics might best be explained by the hypothesis that horizontal rather than vertical gene transfer is the dominant role of genetic transmission in a particular group or time of bacterial or archaeal evolution. Our results (Lienau et al., unpublished data) suggest that the polytomy observed in many bacterial phylogenies is actually a soft one.

In contrast, unpublished results from phage phi phylogenetics (our group) using the picornoviruses (W.C. Wheeler, personal communication) show that even when all available data are used, a bifurcating relationship for all taxa in a group is sometimes not the most plausible hypothesis. In the case of these viruses, it is most parsimonious to invoke an evolutionary scenario that involves multiple origins of entities, like so much dandruff from their hosts.

However, if a resolved tree results, then the second test would be whether or not the inferences in the resultant tree are robust. If the inferences are not robust then again the inclusion of large amounts of data would add little to resolving the ToL and such an approach would be less than desired. If, however, the tree is robust then the third and final test would be necessary. This third test involves assessing whether or not the tree generated from a supermatrix with as much sequence information as possible is accurate or biologically meaningful.

In a series of studies using whole genome information on over 150 organisms, our group has consistently shown that the ToL hypothesis passes tests one and two. Although the kinds of resolution and degree of support in these studies is complex (Lienau et al., 2006), it is clear that a well-resolved and robust phylogenetic hypothesis results when concatenated gene presence/absence and raw amino acid characters are analysed in a parsimony framework (Lienau, 2007). Hence the third type of test of the ToL is necessary and such a test or tests cannot afford to be circular or weakly constructed. One way to approach this problem would be to take a well-corroborated set of hierarchical relationships based on taxonomy, or “stand on the

shoulders of giants” and test the congruence of the data-based hypothesis. Using the multitude of information gathered by previous studies to test genomic ToL hypotheses provides an independent assessment of the worth of those hypotheses. This approach is usually a good place to start and is one we have taken to assess the accuracy of gene presence/absence (Lienau et al., 2006) amino acid data (Lienau et al., unpublished data) and combined gene presence/absence and amino acid sequence (Lienau et al., unpublished data) data. Using the results of other data provides the opportunity to test the results of a phylogenetic analysis from another point of view (see “Corroboration: vantage points and the bold ToL” below). If a conflict occurs between the two analyses a scientist must decide which of the analyses provides a more rigorous test of relationships. If results concur, corroboration is increased. There are also situations in which conflicting results from multiple previous analyses may be resolved by new data in a manner that explains the conflicting results in a new, previously unanticipated way. For instance, in one of our ToL analyses based on information imbedded in over 300 000 amino acid sequences, we classify the *Epsilonproteobacteria*, along with thermophilic bacteria as a group sister to the *Firmicutes*, a result that resolves many previously conflicting phylogenetic analyses (Lienau et al., unpublished data).

Concatenation, simultaneous analysis and total evidence

Kluge (1989) first suggested the term total evidence and referred to Carnap’s definition, as discussed by Bucknam et al. (2006). Kluge (2004) also recognized that Carnap was a verificationist and inductivist. In a subsequent paper, Grant and Kluge (2004) quote Carnap to clarify their own use of the total evidence term. While Carnap suggests that total evidence is the “rule of the methodology of induction” (from Kluge, 2004, p. 207), Kluge points out that the logic behind his own use of the term is as far from induction—verification as it possibly can be. Specifically, “Kluge did argue for his concept of total evidence in terms of maximizing explanatory power” (Kluge, 2004, p. 207).

In a review of data combination methods and theory, Nixon and Carpenter (1996) recognized potential problems with the total evidence term and the way other researchers had interpreted it and suggested that the term “simultaneous analysis” be used instead. Their argument for the switch is critical to an understanding of data partitioning and concatenation. They state that the term total evidence “is probably not appropriate to contrast the method of separate analyses of partitioned data followed by consensus of results with the method of simultaneous analysis of multiple combined data sets” (Nixon and Carpenter, 1996, p. 223). This statement

follows because both separate analyses of all the available data partitions and simultaneous concatenated analyses could be argued to use all available data (Miyamoto and Fitch, 1995) and hence both would be total evidence.

Instead Nixon and Carpenter (1996) suggest: “by pooling the data and determining the most parsimonious solution for all the data, simultaneous analysis seeks the cladogram that is best supported and maximally explanatory for all the data. This approach maximizes information content and corroboration of the resulting hypothesis.” (p. 225). Data concatenation, the term used by most genomicists (e.g. Rokas et al., 2003) to describe data combination, is the process of combining data prior to simultaneous analysis, partitioned analysis or core gene analysis (Ciccarelli et al., 2006). Because of the more precise nature of the term “simultaneous analysis”, we prefer it over the terms “total evidence” and “concatenation”.

We suggest that microbialists (not to be confused with the microbiological community as a whole) have confused the logical basis of simultaneous analysis with the endeavours of concatenating data and assessing support. The view of microbialists that the purpose of concatenation is to obtain a more and more robust phylogenetic hypothesis is misleading. Rather, as Kluge (1989, 2004) points out, total evidence (which we qualify by suggesting simultaneous analysis be used in this discussion) is accomplished so that we can obtain the best explanation for the relationships of organisms we encounter. Kluge (1989, 2001, 2004) is clear that in the face of not being able to discover the truth, the much prized best explanation can only be arrived at when all possible data at hand are included in an analysis. Arguing from this viewpoint we do not see phylogenetics, total evidence or simultaneous analysis as verificationist. Rather, we see simultaneous analysis from the point of view that phylogenetic systematics as a science is based on the testing of relationships and evolutionary history of character transformations, which in turn are based on the Popperian principals of background knowledge and corroboration.

Background knowledge: verticality

Another part of the microbialist belief that phylogenetic analysis is verificationist stems from the argument that tree thinking is circular. In other words, if we assume that the relationships of organisms are tree-like and therefore use tree-building algorithms to analyse data, we will always get a tree to represent those relationships (but see the three tests for a vertical tree section above). However, we must examine the basis for making this assumption of tree-like diversification of organisms. In essence, this examination becomes a discussion of back-

ground knowledge in phylogenetic analysis. Background knowledge is an important Popperian principle; without it, scientific analyses cannot be anchored in reality. Despite recent arguments about background knowledge in phylogenetic analysis, descent with modification has been argued as one of the most basic elements of background knowledge in phylogenetics (Kluge, 1997).

Only if descent with modification is metaphysical and unfalsifiable can we eliminate it as background knowledge (Farris, 1983). If it were unfalsifiable, this would render descent with modification to the realm of metaphysics, and therefore unusable in a scientific context. We suggest that 150 years of systematic research has determined that descent with modification is eminently falsifiable and not in the least metaphysical even in the microbial realm, despite claims of large amounts of HGT.

Furthermore, there is a difference between what Kluge (1999) calls phylogenetic parsimony and operational parsimony. The latter is determining a pattern without background knowledge, and the former is what most systematists hope to do with the aid of background knowledge. Baptiste et al. (2006) confuse “phylogenetic tree based analysis” with “operational tree analysis” and so their critique about background knowledge is similarly mistaken.

Background knowledge: horizontality

As already stated, it is frivolous to ignore HGT completely. It is important to note that HGT does not necessarily make it impossible to recover vertical history. We can examine the question of the nature of prokaryotic evolution from an alternative vantage point, namely horizontality. That is, we can challenge the rationale (HGT) for felling (falsifying) Darwin's ToL for prokaryotes. In order for a ToL hypothesis to be falsified for any group of organisms, it would have to be shown that HGT is the dominant force in the generation of variation in prokaryotic organisms, or that there are more, overall, horizontal transmission events of genetic material than the vertical transfer of genetic information from parent to offspring. Furthermore, those HGT events would have to be directed and not stochastic in nature.

However, microbialists have not met this burden of proof. Microbiologists (not to be confused with microbialists) have observed that the horizontal transfer of genes in the laboratory (transformation) occurs much less frequently than the vertical transfer of genes (i.e. cell division). Although a recent estimate of the percentage of horizontally transferred genes in the genomes of microorganisms reached an average of 80% in some lineages (Dagan and Martin, 2007), most estimates of the composition of bacterial genomes indicate that the majority of

genes have been inherited vertically (Beiko et al., 2005; Gressmann et al., 2005; Choi and Kim, 2007; Fall et al., 2007; Lercher and Pal, 2008; Tamames and Moya, 2008) and, at least in one case, HGT events often seem to be ancient (Comas et al., 2007). Furthermore, recent experimental evidence suggests that barriers to HGT, even among closely related genera, such as the gammaproteobacterial groups *Shewanella* and *Escherichia*, prevent over 75% of genes tested from being transferred (Sorek et al., 2007). Therefore, we suggest that it is reasonable to assume that, for almost all speciation events, the vast majority of genes that are passed into new species are inherited by vertical descent. Even taking into account the likelihood that the ToL contains HGT hubs (Gogarten and Townsend, 2005) such as *Fusobacterium* (Mira et al., 2004), and that there may be massive directed HGT events (as discussed in Lienau et al., unpublished data), the most corroborated phylogenetic signal (MCS) that is common to all characters should be interpreted as the vToL until evidence shows otherwise, especially when one considers that horizontally transferred genes are inherited vertically after transfer.

Corroboration: vantage points and the bold ToL

“A science needs points of view, and theoretical problems.” (Popper, 2002, p.88).

It is the seeming inability (because of HGT) of most phylogenomic studies to test reliably the vToL hypothesis that poses the greatest problem to microbial ToL analyses. We propose that the application of Karl Popper’s philosophy of science, “Logic of Scientific Discovery” (Popper, 2002), to ToL phylogenetic analyses through a simultaneous analysis approach (Kluge, 1989, 1998; Kluge and Wolf, 1993) can illuminate methods that may be able to solve this problem. A recent ToL study that invoked Popper’s philosophy (Bucknam et al., 2006) searched for clades that were not supported by their data, a method based on Rieppel’s (2003) assertion that phylogenetic analyses are not subject to the asymmetrical types of tests that can only falsify, but to symmetrical tests that “support a certain hypothesis of relationships or falsify [it]”.

We suggest, by contrast, that because falsification is never absolute, no test is truly asymmetrical, or definitive, but can only lend temporary preference to certain hypotheses over others. This comparison and tentative acceptance of hypotheses (including phylogenies) is Popper’s unending “game” of science in which a “... theory [that] withstands severe tests and is not superseded by another theory in the course of scientific progress ... has ‘proved its mettle’ or ... is corroborated” (Popper, 2002). In other words, science is the continuing espousal and critical test of bold, falsifiable hypotheses. A hypothesis is bold, or more testable, when it is

universal (says a lot about nature), precise, (predicts specific outcomes of potential tests), and contains more information than, or directly conflicts with, background knowledge. In addition to boldness, hypotheses need ample and relevant evidence to provide rigorous tests of their predictions to become corroborated scientific theories.

The “safe” rake-like ToL hypothesis is not bold, because it does not give information over background knowledge. This is because the hypothesis that there is no ToL has not been adequately tested. The HGT problem makes it more difficult to test hypotheses of evolutionary history for prokaryotes because it is difficult to assess the reliability of the evidence at hand concerning organisms.

Popper wrote that evidence is conjectural: “Every test of a theory, whether resulting in its corroboration or falsification, must stop at a basic statement which we *decide to accept*.” (Popper, 2002, p. 86; emphasis added). Popper’s concept of evidence as basic statements about reality, or conjectures, that we tentatively accept as fact for the purposes of doing science provides a good solution to the HGT problem; when we accept that evidence is conjectural we can arrange our observational statements in such a way as to test our specific question. Phylogenetic character observations may be arranged in many different ways, some relevant to test some groupings, and others relevant to test other groupings. For instance, in ToL analyses gene presence/absence data may be able to levy more severe tests on ancient relationships, while translated gene sequence data may be better suited for testing more recent relationships. With specific regard to the HGT problem, we have utilized an empirical approach to determine the vantage points of observation (similarity threshold values for making homology statements) that yield gene presence/absence (Lienau et al., 2006) and amino acid sequence (Lienau et al., unpublished data) ToL matrices that retain as much information as possible while producing highly consistent, totally resolved and reasonably well-supported phylogenies that are, for the most part, consistent with accepted classification systems and other recent ToL studies.

Conclusion

We hold that, in the absence of a better-supported hypothesis of the nature of relationships of all life, we should retain Darwin’s tree of forms that have descended from common ancestors through the process of natural selection, even for prokaryotes. We base this conclusion on four arguments that outlined herein. First, we present a series of tests that strongly suggest that a bifurcating ToL is an empirical reality. This first point in and of itself is not sufficiently convincing. We therefore argue three philosophical points relevant to

systematic analysis. First, we point out that the microbialist view of the philosophical basis of total evidence is flawed. Total evidence (or better simultaneous analysis) exists as means to better explanatory power. Second, we argue that the background knowledge that microbialists want to disregard is an important and viable part of phylogenetic analysis that cannot be discarded because it is neither metaphysical nor untestable. Finally, we argue that microbialists have either missed or disregarded one of Popper's guiding lights of scientific endeavour, namely corroboration.

We do, however, understand that a tree can be used without allusion to the passage of time as an unrooted network that describes the graded similarity that may or may not represent the prokaryotic world. We admire the boldness (based on differences in background knowledge) of the hypothesis that prokaryotic evolution proceeds in a manner drastically different than do other forms of life. However, "playing it safe", by postulating only a rake as our hypothesis of the evolutionary history of prokaryotic forms, or simply saying "these organisms are related" but not in what specific way, gives us no more information than background knowledge without the advent of Darwin's tree, which is the assumption that genetic material is transferred from one cell to another (vertically or horizontally).

To account for the existence of HGT in the ToL, we have recently argued that the ToL may most conveniently be interpreted as a scaffold that represents the MCS of life (Lienau et al., unpublished data). The decision as to whether the MCS represents vertical history in all nodes is dependent upon interpretation and further testing. In this forum, we have discussed phylogenetic methods of testing if the MCS represents vertical or horizontal history. However, all of these methods of testing verticality rely on pitting one tree or classification system against another in order to see if there is topological and or statistical consensus.

In contrast to the Baptiste et al. (2006) and Bucknam et al. (2006) conclusion that the lack of resolution in ToL studies is evidence that there is no bifurcating history of life for prokaryotes, or a hard polytomy, we assert that whenever the result of lack of resolution occurs it is best explained by a soft polytomy, or lack of resolution, due the paucity of data used in ToL analyses. This is not a difficult position to defend, given that the vast majority of information available at the genomic level for most of the studies claiming rake-like structure was not used. We point out that the rush to fell Darwin's ToL has been based on less than whole genomes (the "1%" of Dagan and Martin, 2007), and that the number of genomes has more than tripled since these conclusions were published.

Furthermore, we find that the assertion that HGT is the major force behind the generation of variation in prokaryotic organisms to be weakly supported at best, and have pointed out several studies that support the

opposite conclusion, that vertical dissemination of genetic material is the major force in prokaryotic evolution. Even if the degree of horizontal transfer is as high as some have suggested (Dagan and Martin, 2007), such HGT needs to be shown to obliterate the majority of phylogenetic signal in genomic information. We suggest that such a demonstration has not been offered. Based on these arguments and on our own success in recovering fully resolved and reasonably well-supported ToL hypotheses using more than two orders of magnitude more sequence data than that used to support the idea that there is no ToL for prokaryotes, we suggest that the axe being brought to bear on Darwin's tree is nowhere near sharp enough to cut it down.

Acknowledgements

We would like to thank Paul J. Planet, Al Phillips, Mark Siddall, and Hannah Salomons for insightful discussion and fresh thinking. We would also like to thank the Korein Family Foundation, the Lewis and Dorothy Cullman Program in Molecular Systematics at the AMNH, and the Sackler Institute for Comparative Genomics at the AMNH.

References

- Baptiste, E., Boucher, Y., 2009. Epistemological impacts of horizontal gene transfer on classification in microbiology. *Methods Mol. Biol.* 532, 55.
- Baptiste, E., Susko, E., Leigh, J., Ruiz-Trillo, I., Bucknam, J., Doolittle, W.F., 2008. Alternative methods for concatenation of core genes indicate a lack of resolution in deep nodes of the prokaryotic phylogeny. *Mol. Biol. Evol.* 25, 83–91.
- Beiko, R.G., Harlow, T.J., Ragan, M.A., 2005. Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci. USA* 102, 14332.
- Boucher, Y., Baptiste, E., 2009. Revisiting the concept of lineage in prokaryotes: a phylogenetic perspective. *BioEssays* 31, 526.
- Bucknam, J., Boucher, Y., Baptiste, E., 2006. Refuting phylogenetic relationships. *Biol. Direct* 1, 26.
- Choi, I.G., Kim, S.H., 2007. Global extent of horizontal gene transfer. *Proc. Natl. Acad. Sci. USA* 104, 4489.
- Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved ToL. *Science* 311, 1283.
- Comas, I., Moya, A., Gonzalez-Candelas, F., 2007. Phylogenetic signal and functional categories in Proteobacteria genomes. *BMC Evol. Biol.* 7 (Suppl. 1), S7.
- Creevey, C.J., Fitzpatrick, D.A., Philip, G.K., Kinsella, R.J., O'Connell, M.J., Pentony, M.M., Travers, S.A., Wilkinson, M., McInerney, J.O., 2004. Does a tree-like phylogeny only exist at the tips in the prokaryotes? *Proc. Biol. Sci.* 271, 2551.
- Dagan, T., Martin, W., 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl. Acad. Sci. USA* 104, 870.
- Dodd, D., 2007. The Annotated "Scarlet Begonias". The Annotated Grateful Dead Lyrics. <http://arts.ucsc.edu/gdead/agdl/scarlet.html>.
- Doolittle, W.F., Baptiste, E., 2007. Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad. Sci. USA* 104, 2043.

- Doolittle, W.F., Zhaxybayeva, O., 2009. On the origin of prokaryotic species. *Genome Res.* 19, 744.
- Fall, S., Mercier, A., Bertolla, F., Calteau, A., Gueguen, L., Perriere, G., Vogel, T.M., Simonet, P., 2007. Horizontal gene transfer regulation in bacteria as a “spandrel” of DNA repair mechanisms. *PLoS ONE* 2, e1055.
- Farris, J.S. 1983. The logical basis of systematic analysis. In: Platnick, N.I., Funk, V.A. *Advances in Cladistics, Volume 2, Proceedings of the Second Meeting of the Willi Hennig Society.* Columbia University Press, New York, pp. 1–47.
- Gogarten, J.P., Townsend, J.P., 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679.
- Grant, T., Kluge, A.G., 2004. Transformation series as an ideographic character concept. *Cladistics* 20, 23.
- Gressmann, H., Linz, B., Ghai, R., Pleissner, K.P., Schlapbach, R., Yamaoka, Y., Kraft, C., Suerbaum, S., Meyer, T.F., Achtman, M., 2005. Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.* 1, e43.
- Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38, 7.
- Kluge, A.G., 1997. Sophisticated falsification and research cycles: consequences for differential character weighting in phylogenetic systematics. *Zool. Scr.* 26, 349.
- Kluge, A.G., 1998. Total evidence or taxonomic congruence: Cladistics or consensus classification. *Cladistics* 14, 151.
- Kluge, A.G., 1999. The science of phylogenetic systematics: explanation, prediction, and test. *Cladistics* 15, 429.
- Kluge, A.G., 2004. On total evidence: for the record. *Cladistics* 20, 205–207.
- Kluge, A.G., Wolf, A.J., 1993. Cladistic: what’s in a word? *Cladistics* 9, 183.
- Lercher, M.J., Pal, C., 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* 25, 559.
- Lienau, E.K. 2007. *Horizontal Gene Transfer and the Tree of Life: An Empirical Approach to Systematics in the Genome Age: Horizontal Gene Transfer and the Tree of Life: An Empirical Approach to Systematics in the Genome Age.* New York University, New York.
- Lienau, E.K., DeSalle, R., 2009. Evidence, content and corroboration and the Tree of Life. *Acta Biotheor.* 57, 187.
- Lienau, E.K., DeSalle, R., Rosenfeld, J.A., Planet, P.J., 2006. Reciprocal illumination in the gene content ToL. *Syst. Biol.* 55, 441.
- Lopez, P., Baptiste, E., 2009. Molecular phylogeny: reconstructing the forest. *C. R. Biol.* 332, 171.
- Mira, A., Pushker, R., Legault, B.A., Moreira, D., Rodriguez-Valera, F., 2004. Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evol. Biol.* 4, 50.
- Miyamoto, M.M., Fitch, W.M., 1995. Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol.* 12, 503.
- Nixon, K.C., Carpenter, J.M., 1996. On consensus, collapsibility, and clade concordance. *Cladistics* 12, 305.
- Popper, K.R. 2002. *The Logic of Scientific Discovery.* Routledge, New York.
- Rieppel, O., 2003. Popper and systematics. *Syst. Biol.* 52, 259.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P., Rubin, E.M., 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* 318, 1449.
- Tamames, J., Moya, A., 2008. Estimating the extent of horizontal gene transfer in metagenomic sequences. *BMC Genomics* 9, 136.