

Joel Velasco

The Tree of Life (draft as of 12/14/10)

Common ancestry is one of the pillars of Darwin's theory of evolution. Today, the Tree of Life, which represents how all life is genealogically related, is often thought of as an essential component in the foundations of biological systematics and so therefore of evolutionary theory – and perhaps all of biology itself. It is an iconic representation in biology and even penetrates into popular culture.

Massive amounts of time, effort, and money are being put into understanding and reconstructing the Tree. Yet there are serious debates as to the usefulness and even the very existence of the Tree. Here I will attempt to critically evaluate the merits of some of these worries. In doing so, we will see that questions about the Tree and the foundations of systematics can only be answered in the light of not only a wide range of empirical considerations, but of philosophical considerations as well. An historically informed picture of how and why we got to where we are today is important for understanding these debates, however, here I can give only the briefest of introductions to the history of the Tree as it has been used in systematics. Then we will focus on contemporary discussions, and finally, look to the future.

A POTTED HISTORY

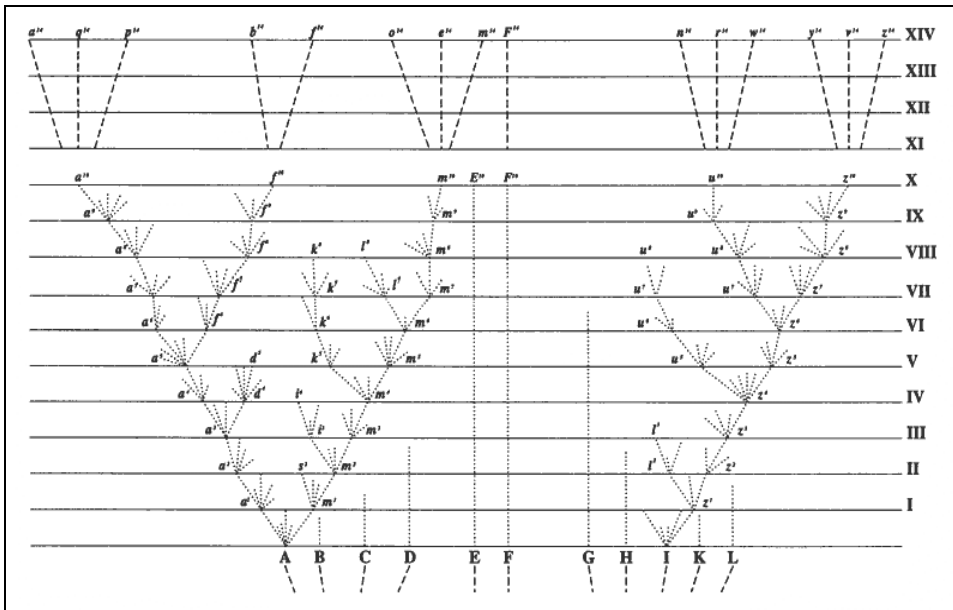
Many authors before Darwin had considered the possibility of, or even promoted the idea that some species were directly genealogically related to each other. Some, including Jean Baptiste Lamarck, had even proposed tree-like structures to capture these

relationships (Ragan 2009). But it was Darwin who revolutionized our understanding of the diversity of life with his *On the Origin of Species* (Darwin 1859). It is in the *Origin* that we first see the importance of genealogy on a grand scale where Darwin convincingly argues that common ancestry explains both the striking similarities between different species and the apparent naturalness of a groups-within-groups hierarchical classification. In the *Origin*, Darwin introduces the metaphor of the Tree of Life which connects all life through common descent:

“The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth.....The green and budding twigs may represent existing species; and those produced during former years may represent the long succession of extinct species.....the great **Tree of Life**...covers the earth with ever-branching and beautiful ramifications.” (129-30, emphasis added).

To help us understand descent with modification, which is essential for his theory of natural selection, Darwin gives us a figure – the only figure in the entire *Origin* – to which he then repeatedly refers (116).

Figure 1. The only figure in *On the Origin of Species*, it represents "descent with modification".



This tree represents real genealogical history and is not simply a classification scheme representing subordination of groups within groups such as the diagrams previously given by Linnaeus, among others.

This idea of the Tree which connects all life has been part of the biological literature since Darwin, but it would require twin revolutions in methodology and in the types of data available before serious attempts could be made at building truly universal phylogenies. By the 1950s, despite great advances in the knowledge of the phylogeny of eukaryotes, bacteriologists had generally given up on the idea of it being possible to build a comprehensive phylogeny for most groups of bacteria. Morphological and physiological data just seemed too sparse and often conflicted (Sapp, 2009). But in the early 1960's, Emile Zuckerkandl and Linus Pauling among others, suggested that molecules such as genes, amino acids, or proteins could be used to track phylogenetic

history. Zuckerkandl and Pauling proposed that some changes might occur at a constant rate forming a “molecular clock” which would aid in phylogenetic reconstruction as well as in determining the timing of evolutionary events (Zuckerkandl and Pauling, 1965a, 1965b).

At the same time, Carl Woese was working on the evolution of the genetic system itself. To examine the early evolution of life, one needs to know the broad-scale phylogenetic history of all life. In 1977, after painstakingly cataloguing numerous rRNA sequences (and then searching for further kinds of data to validate their findings), Woese and George Fox announced that they had discovered a third kind of life: what they called the Archaeobacteria. Despite being prokaryotic, the Archaeobacteria lacked the typical signature found in all bacterial rRNA and in addition, also shared many deep similarities with eukaryotes such as the way that they performed transcription and translation. Over the next thirteen years, Woese and colleagues produced the first universal phylogenies (Fox et al. 1980, Pace et al. 1986, Woese 1987), and eventually proposed the three domain model in which the Archaeobacteria were renamed the Archaea, as opposed to the Bacteria and the Eucarya (Woese et al. 1990). Today, the most common representations of the Tree are akin to the phylogenetic tree depicted of Woese et al. in Figure 2. While some of the details of the tree are no longer accepted, this division of life into three great Domains – the Bacteria, the Archaea, and the Eucarya, has been generally (though not universally) accepted as can be seen in Figure 3 taken from the back cover of an evolution textbook (Barton et al. 2007).

Figure 2. The three domain version of the Phylogenetic Tree of Life first produced in Woese et al. (1990). The branch lengths are proportional to the genetic distances based on ribosomal RNA.

Phylogenetic Tree of Life

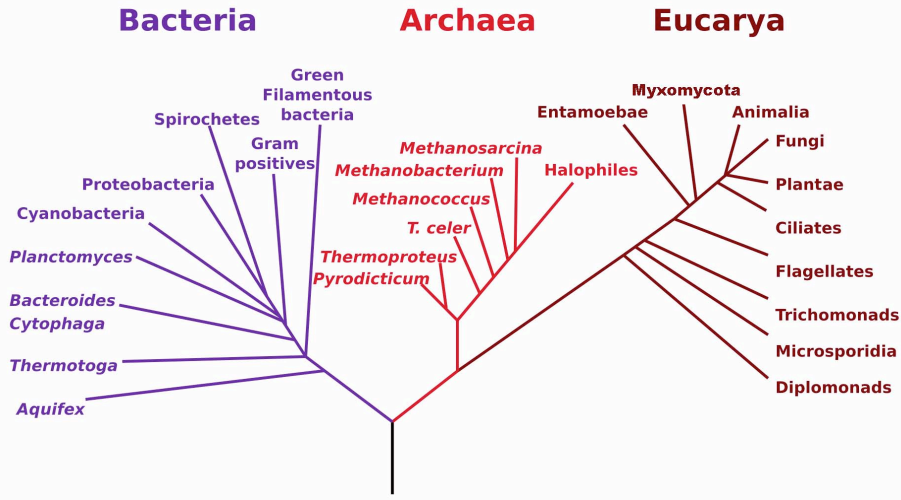
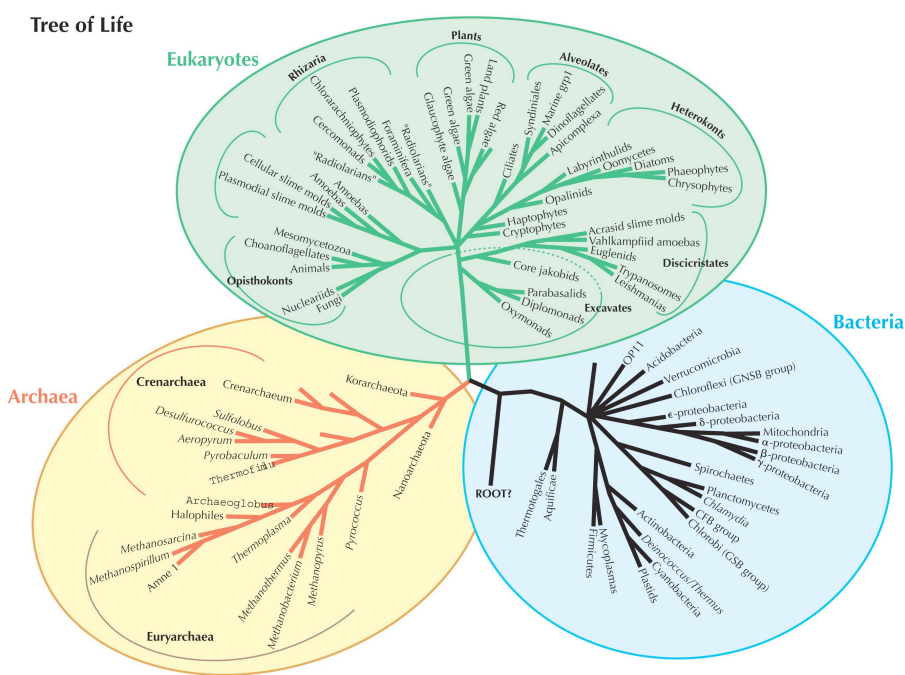


Figure 3. The Phylogenetic Tree of Life as seen in Barton et al. (2007). Notice that the three domains are still present.



Woese was not concerned primarily with classification, but was trying to answer a particular question: what is the correct evolutionary branching sequence for a number of 'major groups' of taxa? More recent reconstructions of the Tree may add more taxa or use more or different kinds of data and may come to different conclusions than Woese, but fundamentally, they are working on the same project. While this certainly seems like a perfectly objective task, it depends on the idea that in fact there is one unique, objectively correct Tree of Life. If there isn't, then what is it for us to infer the Tree?

WHAT IS THE TREE?

A standard way to describe the Tree is that it is a universal phylogenetic tree depicting the genealogical relationships of all species through time. Thus the Tree of Life is meant to be universal, it is meant to be a phylogeny, and it is meant to be a tree. Critics have directly or indirectly attacked each of these three apparently essential features.

What does it mean to say that the Tree of Life is a tree? Modern depictions of the Tree of Life do not look at all like *biological* trees such as in the familiar drawings of Ernst Haeckel, but rather, are *phylogenetic* trees which are trees in the mathematical sense of a special kind of object in graph theory. It is conceptually helpful to think of a tree as a set of directed branches connecting nodes where there is a root node with no parental nodes, interior nodes which have exactly one parent and two (or sometimes more) offspring nodes, and leaf tips which are nodes that have one parent but no offspring nodes. Important features of trees for systematics is that between any two points, there is a unique path on the tree and that each node (other than the root) has exactly one parent. To say that the Tree of Life is a tree is to say that is a phylogenetic tree in this sense.

To say that the Tree is universal implies that the Tree should depict the relationships between all living things. Are viruses alive? Canonical representations of the Tree typically do not mention viruses, but it is worth mentioning viruses in this context, since they place limitations on those wishing to defend the Tree. One must either embrace the idea that the Tree is not universal, or deny that viruses are alive or have the kind of evolutionary history that the Tree is supposed to be tracking. The typical assumption is that the Tree must connect all species, but that viruses (along with mobile genetic elements like transposons and plasmids) don't form species at least in the way relevant for inclusion on the Tree. But do all *organisms* form species in the relevant way? Many bacteriologists as well as systematists and philosophers of systematics, deny that prokaryotes form species (Gevers et al. 2005, Ereshefsky 2010, Lawrence and Retchless 2010). Worse, many of those who do accept that there is a good species concept that applies to prokaryotes will deny that these groups are phylogenetic groups and have branching histories.

It is sometimes thought that a simple change in how we describe the Tree can solve this problem. The Tree of Life shows how *organisms* (or perhaps genomes instead) are genealogically related. But many organisms are not related to each other in a tree-like hierarchy of descent. Rather, they form a reticulated network. This is even more clear for genomes where recombination is present. The defender of the Tree needs to say something about how at the appropriate level of description (perhaps talking about populations or lineages or clades of organisms directly) these entities can form a tree. It is not clear how this can be done and the burden of proof is surely on the defender of the Tree here.

HYBRIDS

If we do manage to muddle through the species problem and say that the Tree can connect all species, we then have the empirical question of just how tree-like this evolutionary history is. There is massive reticulation in the form of gene flow through hybridization and introgression between species. While some have attempted to minimize the problem, we now know that even in the best behaved groups (plants and animals) hybrids regularly form. Mallet (2005) surveys a variety of studies on hybrids and concludes that at least 25% of plant and 10% of animal species form hybrids with other species in nature. This usually leads to introgression and therefore gene flow between species. This problem is far worse with populations at the tips (now any migration is reticulation) or any kind of lineages.

Of course, like Darwin, we can allow that some hybridization is consistent with the Tree. But how much reticulation is it reasonable to allow? This is a difficult question and can only be realistically answered in a context where we know what the purpose of the Tree is. If the Tree is supposed to allow us to make inferences about genetic history, similarity, biogeography, etc. then it is okay if it sometimes leads to errors – any possible model will do that – but it must have a good balance of simplicity, explanatory power, predictive power, and perhaps other less easily describable virtues. If systematists were aided in their research by using the Tree, that would count in its favor. If they were positively misled, that would count against its use. Exactly how these have to be balanced against each other is a perennial question in the philosophy of science and one

that is unlikely to have a general answer but rather, needs to be examined carefully in the particular case at hand.

LATERAL GENE TRANSFER

The problem of reticulation might plausibly be thought to be manageable in eukaryotes, but when we generalize to all forms of reticulation, we face what is arguably the most serious problem for the tree: the phenomenon of lateral gene transfer. Lateral gene transfer (LGT), also called horizontal gene transfer, is the name for any instance of a variety of processes where genetic material moves from one organism to another by some process other than reproduction. This includes transformation, transduction, and conjugation.

It is now widely agreed that LGT has been, and still is, a major force in evolutionary history (Dagan et al. 2008, Gogarten et al. 2002). The epistemological question of what can be inferred about genetic history is a serious one given that genes do not in general track the same history and that as we go deeper in time, any trace of signal may be lost. But the metaphysical question is serious as well - what could the Tree be tracking since clearly the history of ALL genes is not a single tree. It is not clear exactly what this means for the Tree since different proposals about what the tree is will be affected differently. For arguments that widespread LGT undermines the Tree concept and possibly traditional phylogenetics as a whole, see Baptiste, et al. (2004), Baptiste et al. (2005), and Baptiste E, Boucher Y (2008). Before looking at different responses to lateral transfer, we will first look at more potential problems for the tree.

ENDOSYMBIOSIS

Another source of problems for the Tree is endosymbiosis. In endosymbiosis, one organism comes to live inside another and eventually its descendants become obligate symbiotes. Over evolutionary time, they reach the point where they are so tightly interconnected, often because of extensive LGT between host and symbiote, that it is appropriate to think of the host+symbiote as one integrated organism. For example, most eukaryotic cells contain many mitochondria in the cytoplasm surrounding the nucleus of the cell. Mitochondria are clearly functional parts of the cells today and are not organisms in their own right. But mitochondria have their own genomes and it is now clear that historically, they are closely related to various groups of the alpha proteobacteria. Likewise, the chloroplasts which give plants and other organisms such as some algae the ability to photosynthesize were once free living cyanobacteria. A natural way to depict these genealogical relationships is with a fusion of lineages of very distant branches on the Tree as in Figure 4. Endosymbiotic events have occurred a number of times in the history of life (Lane and Archibald 2008) but even with a strict understanding of “new lineage”, then while in some respects these events might be rare, they could hardly be more important. If there is any sense to be made of ‘key’ events in evolutionary history, the origins of mitochondria and of chloroplasts surely count. Any purported universal phylogeny which fails to represent these events is lacking in a very important respect.

THE ROOT OF THE TREE

A major feature of the Tree is its root. The root is typically thought to represent LUCA: the last universal common ancestor. Understanding the root is essential for studying the evolution of various ancient biological features such as the genetic code, protein synthesis, cellular membranes, and indeed, the cell itself.

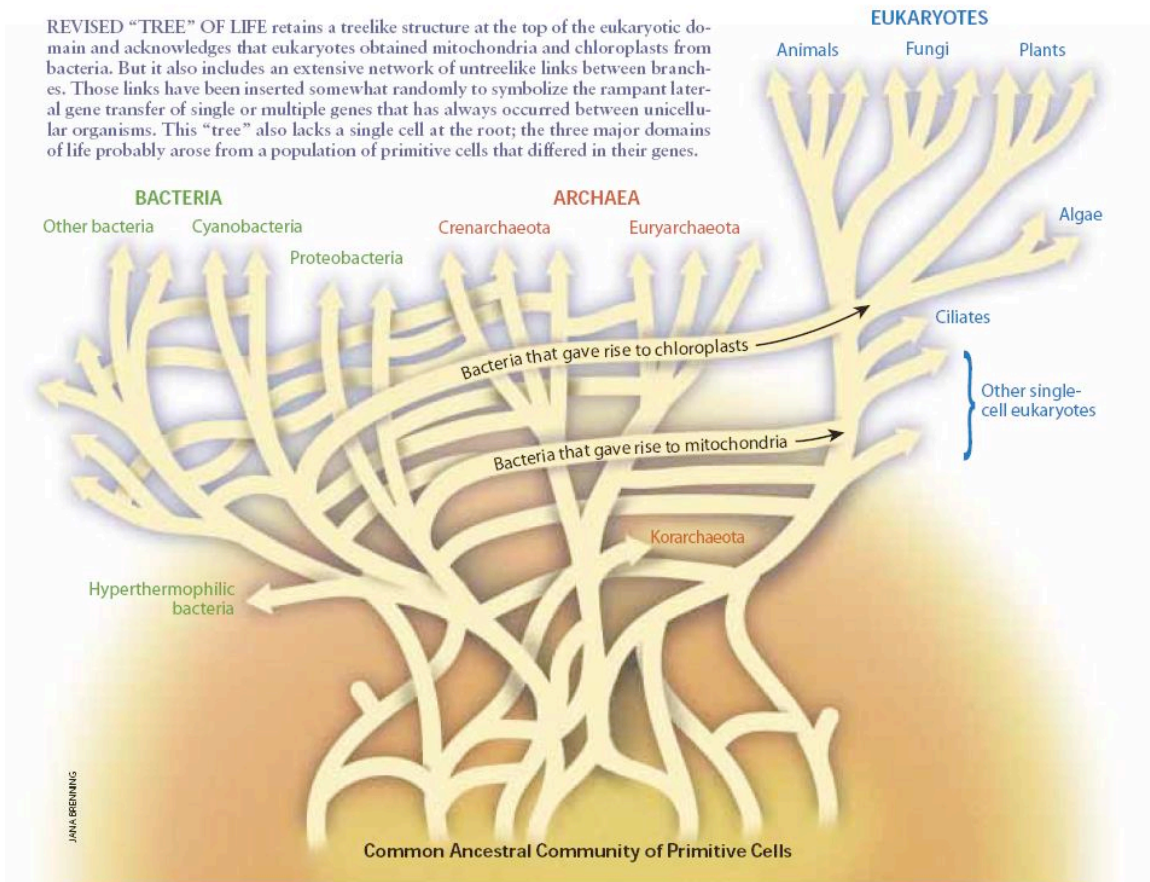
As with the Tree, different authors have a different conception of what would count as a LUCA and different conceptions lead to different conclusions about its existence. In phylogenetics with trees, it is assumed that each descendent node gets its traits through common descent with modification. Thus allowing for mutational or other changes, the genes present in organisms today would have to have their ancestors in LUCA. But if LUCA is a single organism with a single genome, this leads to the absurd conclusion that LUCA contained genes for nearly all types of biochemical reactions known in bacteria and archaea today and had a genome larger than any known prokaryotic genome today. This is what Doolittle et al. (2003) termed “the genome of Eden”. Such an entity surely never existed.

Different genes have genealogical histories which coalesce in the past at vastly different times. In describing his view of early life, Woese says, “The universal ancestor is not an entity, not a thing. It is a process characteristic of a particular evolutionary stage” from which he believes multiple communities and independent lineages emerged (Woese 1998: 6858). Theobald (2010) argues for universal common ancestry and the existence of LUCA, but clarifies what he means, saying, “Rather, the last universal common ancestor may have comprised a population of organisms with different genotypes that lived in different places at different times” (Theobald 2010: 220). While

Doolittle and others don't consider this "population" dispersed in space and time worthy of being called an "ancestor" of anything, even granting that we should call it LUCA, it is clear that it would be inappropriate to depict this as a single node (the root) on the Tree of Life. If we attempted to use this Tree as we would any other phylogenetic tree (say the tree of primates) we would be led to make mistaken inferences about evolutionary history. As the critics would say, a tree without a root is no tree at all.

If we attempt to represent lateral gene transfer, endosymbiosis, and the base of the tree all on the same diagram, even a very conservative picture will look something like Figure 4 which is taken from Doolittle (2000). This is certainly not a phylogenetic tree, but whether that is an essential feature of anything appropriately called "The Tree of Life" is not clear.

Figure 4. The revised Tree of Life as depicted in Doolittle (2000). The lateral branches represent both lateral gene transfer and endosymbiosis.



SAVING THE TREE?

Given the multitude of problems described above, clearly defenders of the Tree must deny that the Tree must represent the history of species, be universal, and be fully branching with no reticulations at all. But given that it isn't that, it is no longer clear exactly what the Tree is. A first pass might be as simple as saying that the Tree is an idealization that isn't perfect but still gets it mostly right and is extremely useful. This

view makes the existence of the tree dependent on things like the extent of actual reticulation. I would guess that this is the most common view of practicing biologists – especially those who work on eukaryotes. But the problems above are quite serious and the extent of reticulation, especially in prokaryotes, far too great to simply idealize away.

But what of those who work explicitly on reconstructing the Tree including prokaryotes? What exactly are they reconstructing? Not a phylogenetic tree that just idealizes out a few reticulations. Here, a number of different, incompatible views of the Tree have been proposed.

One idea might be that as a practical matter, we simply need some kind of reference tree from which to base our classifications and to locate clades so that we can make sense of things like such as lateral gene transfer in the first place. The 16S SSU rRNA gene is often used this way. We know that many gene histories disagree. Why not just pick one for a reference tree? For a great many taxa, we have sequenced the 16S gene and so it has become the default classification tool for prokaryotes. To say that something is an alpha-proteobacteria or a haloarchaen is just to say that its RNA sequence fits in at a certain place in the universal 16S tree. While this may be practical for classification, it is no defense of anything like the Darwinian hypothesis that there is a unique Tree. Which tree is the Tree, can't depend on conventional choices by us. A plausible Tree could be one of two things – a tree that is as reliable for phylogenetic inferences as possible or a tree that represents the actual genealogical history of some kind or other but not the full genetic history.

Galtier and Daubin (2008) explicitly stick to the idea that the Tree is a tree of species. On the view, LGT isn't obviously a problem metaphysically, though it would be

if it meant there were no species (Lawrence 2002). Similarly, if we are building a tree of organisms it might seem that LGT is only an epistemological issue. Organisms come from other organisms. This history is often referred to as ‘the tree of cells’ since it tracks the cellular history and not necessarily the history of the genes inside the cells. Ciccarrelli et al. (2006) claim that the Tree is a tree of organisms. Given this, it is easy to see why they simply remove from their data set genes which they have reason to believe have been transferred. If they don’t represent the organism’s genealogy, why include this misleading data?

But even if there is some genuine tree of species or a tree of cells, this leaves totally open just how useful it is to reconstruct it and makes it a serious question to what extent the Tree could play the foundational role it is sometimes claimed that it plays which may be relevant to whether this object is properly called “The Tree of Life”. For example, as Galtier and Daubin (2008) themselves point out, their species tree may in fact not be consistent with *any* single gene’s history. Ciccarrelli et al. (2006) are criticized by Dagan and Martin (2006) for producing “The tree of one percent” since their tree is based on only 31 genes which are consistent with at most 1% of the typical prokaryotic genome of 3,000+ genes.

In order to preserve the idea that the Tree is supposed to represent something like a dominant pattern, Koonin et al. (2009) examine whether there is a “statistically significant trend” in the Forest of Life which represents all genetic history. They argue that there is and that this could plausibly be called a tree of life. Wu et al. (2009) construct a genome tree based on a concatenation of all the gene data they have and compare this to known the rRNA tree. Although they don’t explicitly present it this way,

one could reasonably say that the Tree of Life is this genome tree which represents something like an average signal which may not be the actual signal of any particular gene.

If we wanted the Tree to play the role of representing the history of all genes or genomes then it is now clear that no such tree can play that role. Some have proposed alternate names and conceptualizations to play the “represent everything” role such as Koonin et al.’s “Forest of Life”. Other suggestions include the “Ring of Life” (Rivera and Lake 2004), the “Net of Life” (Kunin et al. 2005), or a “Web”, “Coral” or “Potato of Life” (Olendzenski and Gogarten 2009). If they are right, then it seems that the Darwin’s hypothesis was wrong.

THE FUTURE OF THE TREE

So should we continue to talk of the Tree of Life and attempt to reconstruct it? There are two kinds of factors relevant to this question. One set of factors relies on the empirical facts. Just how common is LGT and what kinds of patterns does it produce? What was the early evolution of life like? Was there some single universal common ancestor? What is the extent of hybridization and introgression between sexual species? The answers to each of these questions can tell us something about the utility of talk of the Tree and the importance of reconstructing it.

But another equally important set of factors are questions about what the Tree is supposed to represent, about how we do and ought to use the Tree in biological inferences. Does defending the Tree entail defending a particular history of life and perhaps even a particular view about what counts as life? Or does it mean defending a

particular set of practices? Or perhaps defending the explanatory power and heuristic uses of a particular model? Depending on how these questions are answered, empirical facts like the extent and pattern of LGT may or may not dictate abandoning the Tree.

Critics of the Tree are certainly right that the Tree has been used for many different, and sometimes inconsistent purposes. Some practices, such as assuming that one gene will have the same broad-scale genealogy of another are straightforwardly bad practices. A weak reading of the pluralism defended by pattern pluralists such as Doolittle and Baptiste (2007) is surely right: just as there are a multitude of evolutionary processes besides natural selection, there are a multitude of genealogical patterns besides the single tree pattern. But this is consistent with the Tree being one of those patterns – and perhaps a very important one at that. But the critics of the Tree want to claim something stronger – that the history and usage of the phrase “The Tree of Life” dictates that it is the unique pattern or at least a very special kind of universal pattern. Further, any specific way of understanding the Tree, such as the tree of cells, simply fails to have the power to play the role that the Tree was supposed to play.

Is this stronger view correct? This depends on answers to questions like those I raised above. Phylogenetic trees really are of central importance in a variety of contexts and so trees will continue to be built and the phrase “tree of life” has a special kind of importance (it is in the Bible and in Darwin. How’s that for sacred!). But it is now clear that different understandings of what the Tree is supposed to be and how it can be used come apart and so the phrase must be used more carefully and only in restricted contexts. What about the prospects for ‘universal’ tree-building? Research programs investigating questions about the origin of life, the genetic code, the cell, the eukaryotic cell, and the

connections between apparently very disparate forms of life will continue. It is clear that this research will involve phylogenetic trees as well as investigating patterns other than trees and processes other than vertical descent. What this research will uncover is unknown to us now, but we can be certain that it will be a fascinating story of the deep evolutionary connections between all humans, the *Escherichia coli* in my gut, the archaea living in hydrothermal vents deep in the ocean, the roses in my garden, and the penguins in the antarctic. At least in this respect, we can surely claim that Darwin was right.

REFERENCES

Baptiste, E., Y. Boucher, J. Leigh, and W. F. Doolittle. 2004. Phylogenetic reconstruction and lateral gene transfer. *Trends in Microbiology* **12**, 406-11.

Baptiste, 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.

Baptiste, E., and Y. Boucher. 2008. Lateral gene transfer challenges principles of microbial systematics. *Cell* **16**: 200-7.

Barton, N. H., D. E. G. Briggs, J. A. Eisen, D. B. Goldstein, and N. H. Patel. 2007. *Evolution*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

- Ciccarelli, F., T. Doerks, C. von Mering, C. Creevey, B. Snel, and P. Bork. 2006. Towards automatic reconstruction of a highly resolved tree of life. *Science* **311**: 1283-7.
- Dagan, T., and W. Martin. 2006. The tree of one percent. *Genome Biol* **7** no. 10: 118.
- Dagan, T., Y. Artzy-Randrup, and W. Martin. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci* **105**: 10039–44.
- Darwin, C. 1859. *The origin of species by means of natural selection*, First Edition. John Murray, London.
- Doolittle, W. F. 2000. Uprooting the tree of life. *Scientific American* **282** no. 2: 90-5.
- Doolittle, W. F., Y. Boucher, C. L. Nesbø, C. J. Douady, J. O. Andersson, and A. J. Roger. 2003. How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philosophical Transactions B*, **358**: 39-58.
- Doolittle, W. F., and E. Baptiste. 2007. Pattern pluralism and the tree of life hypothesis. *Proc Acad Nat Sci U S A* **104**: 2043-9.
- Ereshefsky, M. 2010. Microbiology and the species problem. *Biol Philos* **25**: 553–68.
- Fox, G. E., E. Stackebrandt., R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta,

L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese. 1980. The phylogeny of prokaryotes. *Science* **209**: 457–63.

Galtier, N., and V. Daubin. 2008. Dealing with incongruence in phylogenetic analyses. *Phil Trans R Soc Lond B Biol Sci* **363**: 4023–9.

Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson, and J. Swings. 2005. Re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**: 733-9.

Gogarten, J. P., W. F. Doolittle, and J. G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226-38.

Koonin, E. V., Y. I. Wolf, and P. Puigbo. 2009. The phylogenetic forest and the quest for the elusive tree of life. *Cold Spring Harbor Symp Quant Biol* **74**: 205-13.

Kunin, V., L. Goldovsky, N. Darzentas, and C. A. Ouzounis. 2005. The net of life: reconstructing the microbial phylogenetic network. *Genome Res* **15**: 954–9.

Lane, C. E. and J. M. Archibald. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. *Trends Ecol Evol* **23**: 268–75.

Lawrence, J. G. 2002. Gene transfer in bacteria: speciation without species. *Theor Popul Biol* **61**: 449–60.

Lawrence, J. G., and A. Retchless. 2010. The myth of bacterial species and speciation. *Biol Phil.* **25**: 569–88.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, **20**: 229–37.

Olendzenski, L. and J. P. Gogarten. 2009. Evolution of genes and organisms. *Natural Genetic Engineering and Natural Genome Editing. Ann. N.Y. Acad. Sci.* **1178**: 137–45.

Pace, N., G. J. Olson, and C. R. Woese. 1986. Ribosomal rna phylogeny and the primary lines of evolutionary descent. *Cell* **45**: 325-6.

Ragan, M. 2009. Trees and networks before and after Darwin. *Biol Direct* **4**: 43.

Rivera, M. C., and J. A. Lake. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**: 34-7.

Sapp, J. 2009. *The New Foundations of Evolution: On the Tree of Life*. Oxford University Press, Oxford.

Theobald, D. L. 2010. A formal test of the theory of universal common ancestry. *Nature* **465**: 219-22.

Woese, C. R. 1987. Bacterial evolution. *Microbiology and Molecular Biology Reviews*, **51** no. 2. 221-71.

Woese, C. 1998. The universal ancestor. *Proceedings of the National Academy of Sciences*, **95** no. 12: 6854-9.

Woese, C. R., and G. E. Fox. 1977. The concept of cellular evolution. *Journal of Molecular Evolution*, **10** no. 1: 1-6.

Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, **87**: 4576-9.

Wu, D., P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. Ivanova, V. Kunin, L.

Goodwin, M. Wu, B. J. Tindall, *et al.* 2009. A phylogeny-driven genomic encyclopedia of Bacteria and Archaea. *Nature* **462**: 1056-60.

Zuckerkandl, E. and L. Pauling. 1965a. Molecules as documents of evolutionary history, *Journal of Theoretical Biology* **8**: 357-6.

Zuckerkandl, E. and L. Pauling. 1965b. Evolutionary divergence and convergence in

proteins. *Evolving genes and proteins*. Editors V. Bryson and H. J. Vogel, 97–166. New York, NY: Academic Press.