The Emergence of Darwinian Evolution

David B. Wells

December 14, 2006

Abstract

Horizontal gene transfer (HGT) is seen to have been a pervasive phenomenon in the early history of life, necessitating a new evolutionary model for the deep past. We find that in this early time, HGT is the dominant evolutionary mechanism, in fact completely obscuring vertical descent at the time of the universal ancestor—a time of communal evolution out of which Darwinian evolution was to emerge.

1 Introduction and Background

Since Darwin's theory of evolution was first formulated, biologists' view of the process of evolution has been predictably "Darwinian", i.e. the evolutionary dynamic is one of vertical inheritance: traits are passed from mother to daughter; occasional copying errors are more occasionally advantageous from a survival standpoint; and the species evolves. The advent of molecular biology placed the impetus on genotypes rather than phenotypes, but the basic idea of evolution remained essentially vertical.

Until quite recently, this remained the case. A universal phylogenetic tree based on ribosomal RNA (rRNA) was produced, and the "holy grail" of evolutionary biology seemed found [1, 2] (See Fig. 1).

With the arrival of the genome era in the 1990s, however, the picture became murky once again. Many sequences became known, and phylogenetic trees from sources other than rRNA became possible, with the expectation that they would be more or less congruent with the rRNA tree. This turned out not to be the case [3].

The initial, knee-jerk question was, "Which one is right?" This is, of course, the wrong question to ask, as it fails to question the fundamental

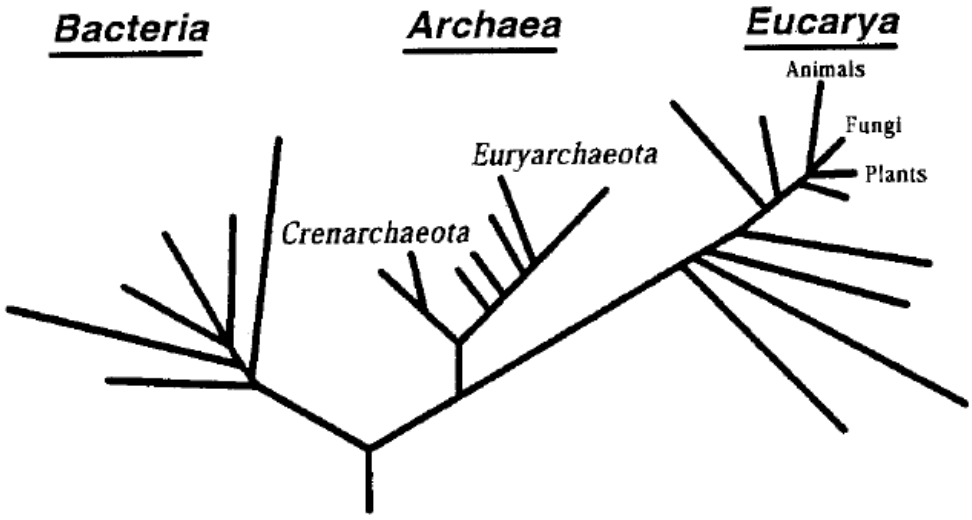

Figure 1: The universal phylogenetic tree, determined using ribosomal RNA. Reproduced from Ref. 2.

underlying assumption, namely the assumption of vertical, Darwinian inheritance as the primary evolutionary player. In questioning this assumption, we are naturally led to "fill the void", as it were, by asking a new question: What mechanism can account for these diverse phylogenetic tress? This is the question explored by this paper.

As we shall see, the theory explored here presents a novel example of an emergent phenomenon, with the onset of Darwinian evolution as a sort of phase transition, insofar as it represents the spontaneous appearance of a new, more organized, and more sophisticated level of life.

This is an interesting topic because it is so fundamental to our existence. It begins developing ideas which can help us understand the origin of life, for it truly pushes the "event horizon", so to say, further into the past than Darwin could have dreamed. Indeed, the ideas and results presented here represent not an overturning of Darwinian theory, but rather an exposing of its limitations: just as Newtonian mechanics are inadequate for describing near-light-speed particles, so Darwinian evolution cannot explain the deep roots of the universal phylogenetic tree.

2 Methods

Genome sequences

Complete genomes catalyzed the rethinking of Darwinian evolution, and as such, genome sequences represent the bulk of the raw experimental data [3, 4]. Ref. 4 is a particularly important review article detailing phylogenetic trees and other comparisons determined for a certain type of protein, the aminoacyl-tRNA synthetases, whose importance will be elucidated below.

Protein comparison

The theory presented here draws heavily on phylogenetic trees, as such trees to a great extent exposed the problems with a strict vertical inheritance model of evolution. This and other methods of protein comparison measures can be found in Ref. 4.

Limitations

The greatest limiting factor for these types of studies is data availability. However, this limitation is rapidly becoming less important as more genomes are sequenced. We are then limited by our models, and how the data are to be interpreted, because evolutionary biology is essentially an observational science: Nature has already run the experiment [4]. We cannot expect to uncover ancient DNA archaeologically, so therefore we must view the record through the distorting lens of time. As we will see below, however, there is much there to be seen.

3 Results and Discussion

Multiple proteins, multiple trees

As mentioned above, the need for a new perspective on evolution was shown largely by the incongruent phylogenetic trees derived for different proteins. Whereas some proteins might be "canonical"—i.e. show the same basic relationships among the Archaea, Bacteria, and Eucarya shown in Fig. 1—others may be decidedly non-canonical, perhaps grouping Bacteria with Archaea, or even splitting traditional groups [4].

The immediate solution is known as horizontal gene transfer (HGT). HGT is a mechanism whereby genes are passed not along a branch from mother to daughter (*vertical* gene transfer, in this nomenclature), but rather between different branches—from brother to brother of another mother [4].

This was not an entirely new concept, but it had always been assumed to be a relatively innocuous phenomenon. Now, however, it appeared to have had much more evolutionary impact, calling into question the very nature of the universal phylogenetic tree, and especially of the universal ancestor represented by its base [5].

AARSs

Aminoacyl-tRNA sythetases, or AARSs, provide a unique look at the forces of HGT at work. These proteins are enzymes which catalyze the charging of transfer RNA (tRNA) with the appropriate amino acid. As discussed in more detail below, HGT affects genes whose corresponding protein is relatively autonomous. Because of the universality of the genetic code, and the fact that AARSs operate independently of the translation machinery, AARSs would seem ideal candidates for HGT, providing a tidy explanation of their diverse phylogenetic distribution [2, 4]. To accept this, however, calls into question much of traditional evolutionary thought.

The meaning of species

A new picture now starts to emerge. First of all, it appears that the universal phylogenetic tree as presented above is by no means without value, and is still deserving of its name [2, 5]. However, the rampant HGT evidenced by the genetic record of the AARSs (see Fig. 2) means that we must reinvent exactly what this tree represents. The Darwinian picture of the tree is that each branch represents a distinct species with a distinct set of genes. But we now see that, near the base of the tree at least, many genes were not identified with a particular branch, but instead could roam among different branches, crossing major taxonomic divides in the process. Thus, we see that as we descend the tree, the very concept of *species* becomes less and less meaningful, describing a smaller and smaller core of genes around which the maelstrom of HGT swirls [2, 5, 6].

This raises the natural question: What does this mean for the root of the tree, for the universal ancestor? What does it mean for whatever predates

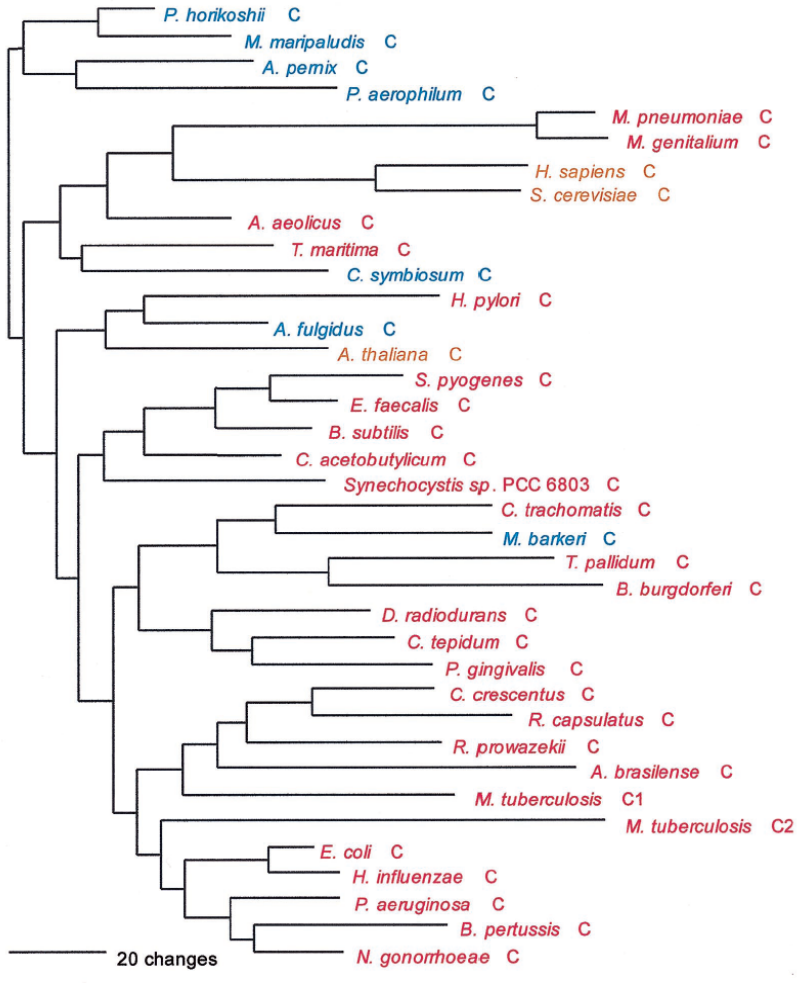

Figure 2: Phylogenetic tree of CysRS, reproduced from Ref. 4. Bacteria are in red, Archaea in blue, and eukaryotes in yellow. Note the very strong differences compared to the rRNA tree in Fig. 1.

the tree?

The universal ancestor

First, we must ask what we are left with of the old theory. While we are not discarding the universal phylogenetic tree, we are certainly modifying what

it means. It describes not species, but rather gene pools, groups of cells which have some subset of genes, however small, in common [2]. Thus, when we descend beyond the initial bifurcation, we cross into a domain where *all cells share a common gene pool.* The initial bifurcation represents a fundamental change in the evolutionary dynamic, and has acquired the moniker "Darwinian Threshold", in a sense marking the true origin of species [6].

This is a radically different notion of the universal ancestor [5]. This ancestor was not an organism in the normal sense, but rather a communal group of countless individual genetic lines. It is doubtful even that these "cells" possessed cell walls, as evidence points to cell wall development after the Darwinian Threshold [7]. It is thus very difficult to even imagine what these proto-cells constituted besides bare DNA and an ancestral ribosome. Woese envisioned it colorfully as a "bag of semi-autonomous genetic elements", a collection of mini-chromosomes [5]. Whatever it was, we can say at least that its defining feature was HGT on a scale that made any type of individual Darwinian vertical descent literally suicidal, replaced instead by communal evolution [5].

Sub-cellular evolution

The era predating the Darwinian Threshold may in a sense be regarded as a time of sub-cellular evolution—perhaps the Darwinian Threshold even identifies with the emergence of the cell. Before the Darwinian Threshold, many basic cellular functions had not been invented/discovered. It was a time when successful evolution required a much more expansive search of genetic phase space, and thus might be thought of as parallel evolution, wherein numerous genetic "ideas" are tried out simultaneously, the most successful (in some sense) of which spread via HGT throughout the entire community. This has been referred to loosely as "high genetic temperature" [5]. To extend the analogy somewhat, the pre-Darwinian era was defined by a genetic symmetry within the extant lifeforms, a symmetry whose breaking coincides with the Darwinian Threshold.

A number of things charaterized this time in evolutionary time period. First, strong resemblances between the ribosomes of the three domains of life indicate well-developed translational machinery *before* the Darwinian Threshold [6, 8]. This of course agrees with the fact that DNA indeed functioned as a code (in need of translation) in its pre-Darwinian context—attested to by its universality. However, this machinery, initially, was much

simpler and less sophisticated by necessity, and certainly much less accurate than today. This would have limited protein size, given the mistranslation rate [5]. Conversely to translation, the evidence indicates that whatever genome replication system was in place (obviously some system must have existed) was so rudimentary that there is little if any resemblance between the bacterial and archaeal/eukaryotic systems today [6, 8]. This again argues for small proteins, and also supports the previously mentioned idea of a pre-Darwinian genome composed of mini-chromosomes, each carrying a single gene [5]. In addition to compartmentalizing the damage of genome mutation, this would also facilitate HGT.

The inaccuracy of the proto-cell's information processing systems limited the sophistication of the proteins it could produce, but also vastly increased the genetic diversity of the community. And because "survival" did not really have meaning on the genome level, some cell lines might run out of control ("error catasotrophe") and nevertheless benefit the community by producing genetic novelty [5].

Toward the Darwinian Threshold

As genetic innovations accumulate, the systems start gaining complexity. One can imagine each innovation in translation leading, because of reduced error rates, to longer proteins capable of more complex structure and properties; these in turn lead to general improvements of all systems, including translation; and the cycle repeats [5, 6].

All the while, HGT carries on, each innovation passing to every proto-cell in the gene pool. The increases in complexity, however, begin to reduce the autonomy of the proteins; the proteins become more specific and interdependent. At a certain point, some system or systems evolve to the point where HGT can no longer spread that technology. Returning to the genetic temperature analogy, the system has crystallized [5]. This system is then no longer updated through the parallel, horizontal mechanism; the design must instead be tweaked without wholesale replacement of any of its constituent proteins. In other words, it must evolve in Darwinian fashion [2, 5, 6].

This did not necessarily occur simultaneously for the three domains of life. Rather, it appears that the Bacteria were the first to crystallize out of the universal ancestor stage, which only later gave rise to the Archaea and eukaryotes [6]. Such a scenario was also hinted at by Kandler on the basis of cell wall composition [7]. We must be careful, though, to remember the mean-

ing of the lowest parts of the universal phylogenetic tree: the crystallization of the Bacteria need only represent the crystallization of one subsystem—in fact, the simultaneous crystallization of multiple subsystems seems rather unlikely. (Exactly which subsystem crystallized is unclear to me, although a genome replication system seems a likely candidate, given the drastic differences between the Bacterial and Archaeal/eukaryotic versions [8]. The transcription apparatus is another possibility, perhaps a stronger one given that a rudimentary version of the current systems seems to have existed at the time of the initial bifurcation [8].) Therefore, one would expect continued communal evolution of the Bacteria as a whole, until, as the genetic temperature drops further, more and more subsystems crystallize, to the point where HGT becomes the relatively benign evolutionary force we see today [2, 5, 6]. Likewise, the Archaea/eukaryotes continued to evolve communally, until some system crystallized to distinguish the two (whether the crystallizing group was the Archaea or eukaryotes is debatable, but Archaea seems a popular opinion [6,7].) Meanwhile, HGT among non-crystallized systems in the newly-formed Bacteria, Archaea, and Eucarya is still expected, and indeed manifests itself in the distributions of the aminoacyl-tRNA sythetases [4].

4 Summary and Conclusion

The genomes now available have shown compelling evidence of pervasive HGT in the early history of the universal phylogenetic tree. Such rampant HGT defied interpretation within the classical Darwinian framework, and forced a complete rethinking of what such trees actually represent [2]. Instead of species, wherein a core set of genes may change or diverge through reproductive/hereditary effects, this core is now seen to be ever so slight near the primary branchings of the tree, holding together diverse gene pools which still evolve communally. In hindsight, this can be seen as successfully resisting the temptation to apply ideas (Darwinian evolution) to situations they were never intended to explain.

There are multiple ways forward from here. First, there is much to be done "straightforwardly", that is, to keep working through as many genomes as possible. The possibility of somehow recovering extinct DNA notwithstanding, the genomes of extant species represent our only window on the genetic past. The flip side of this is continued work on evolutionary mod-

els. The works reviewed in this paper make testable predictions, as well as explore many ideas and speculations in need of development.

There are also more novel approaches. Computer simulations have been shown to "predict" (after the fact, naturally) certain aspects of the ordering of the genetic code, as well as its universality [9]. As we are able to understand the more and more distant genetic past, the utility of computer simulation seems greater and greater.

In conclusion, technology has enabled a great leap forward in available genomic data, thus allowing much more to be inferred about the history of life on Earth. With so much yet unknown, there is no doubt that such progress will continue unabated. The insights of Carl Woese and others have pushed our understanding into the deep past, and new ideas will doubtless push us further.

References

[1] Woese, C. R., Kandler, O., & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.

[2] Woese, C. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8392–8396.

[3] Brown, J. R. & Doolittle, W. F. (1997) *Microbiol. Mol. Biol. Rev.* **61**, 456–502.

[4] Woese, C. R., Olsen, G. J., Ibba, M., & Soll, D. (2000) *Microbiol. Mol. Biol. Rev.* **64**, 202–236.

[5] Woese, C. R. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6854–6859.

[6] Woese, C. R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 8742–8747.

[7] Kandler, O. (1994) *J. Biol. Phys.* **20**, 165–169.

[8] Olsen, G. J. & Woese, C. R. (1996) *Trends Genet.* **12**, 377–379.

[9] Vetsigian, K., Woese, C. R., & Goldenfeld, N. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 10696–10701.