Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life

Gregory P. Fournier¹, Jinling Huang² and J. Peter Gogarten^{1,*}

¹*Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269-31258, USA*

²*Department of Biology, East Carolina University, Greenville, NC 27858, USA*

Horizontal gene transfer (HGT) is often considered to be a source of error in phylogenetic reconstruction, causing individual gene trees within an organismal lineage to be incongruent, obfuscating the ‘true’ evolutionary history. However, when identified as such, HGTs between divergent organismal lineages are useful, phylogenetically informative characters that can provide insight into evolutionary history. Here, we discuss several distinct HGT events involving all three domains of life, illustrating the selective advantages that can be conveyed via HGT, and the utility of HGT in aiding phylogenetic reconstruction and in dating the relative sequence of speciation events. We also discuss the role of HGT from extinct lineages, and its impact on our understanding of the evolution of life on Earth. Organismal phylogeny needs to incorporate reticulations; a simple tree does not provide an accurate depiction of the processes that have shaped life’s history.

Keywords: horizontal gene transfer; chlamydiae; cyanobacteria; acetoclastic methanogenesis; pyrrolysine; extinct lineages

1. INTRODUCTION

For almost 200 years, furcating, often strictly bifurcating, trees have been used to depict the evolutionary history of species, tracing organismal lineages back to common ancestral groups. First depicted by Lamarck (1809), the tree concept for the evolution of species was also proposed by Darwin (1859) and extended to encompass the tree of life, with the mechanism of speciation events at furcations explained by natural and sexual selection. By logical consequence, he also inferred that all living organisms could therefore be traced to a single living ancestor (Darwin 1859). While Darwin also frequently used the term ‘tree of life’, he suggested that ‘coral of life’ would be more appropriate, as the base of coral is made of extinct, dead organisms (Darwin 1836–1844). Unbeknown to him, this metaphor is especially appropriate, as the dead layers of coral form a richly connected network analogous to horizontal gene transfer (HGT) between species.

Comparison of bacterial genomes from the same species revealed astounding amounts of gene transfer and gene loss. An analysis of three *Escherichia coli* genomes revealed that only 39.2 per cent of their common shared gene pool was present in all, 585, 1623 and 1346 genes, each present in only one of the three genomes (Welch *et al.* 2002); In the case of three *Frankia* strains whose small subunit ribosomal RNA showed less than 3 per cent sequence divergence, only 20 per cent of the common shared gene pool had detectable

homologues in all three genomes (Normand *et al.* 2007). Many of the recently acquired genes are only under weak purifying selection (Daubin & Ochman 2004). Gogarten & Townsend 2005 proposed a nearly neutral theory for gene transfer and suggested that the weak signatures of purifying selection found for these transferred genes could be owing to selection against novel deleterious features that may be acquired by the transferred genes (e.g. protein toxicity), and not owing to a selective advantage that the transferred gene might provide to the host. Most of the transferred genes might be neutral or nearly neutral to the recipient; however, among the huge number of horizontally transferred genes, a few will provide an adaptive advantage to the recipient. The horizontal transfer of these adaptive genes greatly accelerates evolution by transferring inventions made in one part of the tree to other lineages (Jain *et al.* 2003), especially in single-celled microorganisms. While HGT is extensive, it is not so rampant as to create a continuum of phenotypes, as microbes are still clearly recognized as belonging to discrete groups, frequently forming clearly nested clades. For this reason, in many instances, specific HGT events can be identified against a ‘background’ of a plurality signal, often ascribed to vertical descent.

Gene trees furcate in the same way traditionally attributed to species trees. However, recombination, horizontal transfer, fusion of independent lines of descent and lineage sorting can cause differences between a species tree and phylogenies of individual genes represented therein (Margulis 1995; Page & Holmes 1998; Felsenstein 2003). These processes complicate the story of evolution by ensuring that the most recent common ancestors (MRCAs) of individual

* Author for correspondence (gogarten@uconn.edu).

One contribution of 11 to a Theme Issue ‘The network of life: genome beginnings and evolution’.

genes probably did not all coexist within the genome of the most recent common organismal ancestor. A description of lineage fusion and genetic exchange needs to be part of phylogenetic explanations. Restricting phylogenetic reconstruction to furcating trees provides an incomplete, and possibly misleading, description of life's history. In particular, it fails to recognize many of the events that created biological innovations.

While reticulations are a significant contributor to the structure of gene and organismal phylogenies, other independent processes also act in concert to affect the apparent distribution of furcations across a phylogenetic tree. Generally, it has been observed that the number of lineages increases in an exponential fashion across evolutionary time (Martin *et al.* 2004). This distribution is greatly affected by the inclusion of only existing organisms in molecular phylogenies, as genetic information is available from extinct species only in very rare and geologically recent circumstances (Gibbons 2005; Willerslev & Cooper 2005; Donoghue & Spigelman 2006). Applying a steady-state model of extinction and speciation, the deepest branches of a molecular phylogenetic tree cover, on average, half of the time the group is in existence (Zhaxybayeva & Gogarten 2004). Therefore, the long 'empty' branches deep in the tree of life (e.g. the ancestral lineages of the archaeal and bacterial domains) are expected, and any observed deviation actually requires explanation. For example, a rapid set of speciations results in a bush-like 'radiation' within a phylogenetic tree (Rokas & Carroll 2006), suggesting either a new biological invention that opened new ecological niches or a disruptive event, such as a mass extinction, that emptied the available niches has taken place. This scenario is apparent in the early evolution of the bacterial domain, as all of the known ribosomal RNA-defined bacterial phyla diverge over a very small phylogenetic distance, leaving their relationships largely undefined. This pattern could be explained by a mass extinction triggered by the Late Heavy Bombardment (Gomes *et al.* 2005), followed by rapid adaptation and diversification of the few microbial survivors (Raup *et al.* 1989; Gogarten-Boekels *et al.* 1995; Nisbet & Sleep 2001).

Several methods are used to detect HGT (see Gogarten & Zhaxybayeva 2008 for a recent review), including compositional analysis (Lawrence & Ochman 1997, 1998), unusual phyletic patterns (Snel *et al.* 2002; Kunin & Ouzounis 2003; Mirkin *et al.* 2003) and phylogenetic incongruence (Pennisi 1998; Lerat *et al.* 2003; Beiko *et al.* 2005; Zhaxybayeva *et al.* 2006). Numerous instances of HGT have been identified and discussed in the literature, many of which are useful in understanding evolutionary adaptations, as well as further unravelling the phylogenetic relationships between clades.

2. CHLAMYDIA AND THE ORIGIN OF PRIMARY PLASTIDS

As a group of obligate intracellular bacteria, chlamydiae are not known to exist in any photosynthetic eukaryotes. However, a surprising finding from the early chlamydial genome sequencing projects was

that a disproportional number of chlamydial genes are more similar to plant homologues than to other bacterial sequences (Stephens *et al.* 1998; Horn *et al.* 2004). This finding has been confirmed by several detailed studies on individual genes, including those encoding ATP/ADP translocase (Wolf *et al.* 1999; Greub & Raoult 2003; Schmitz-Esser *et al.* 2004), CMP-KDO synthetase (Royo *et al.* 2000), enoyl-ACP reductase (Ryall *et al.* 2003), L,L-diaminopimelate aminotransferase (McCoy *et al.* 2006) and others. Because chlamydiae-related proteins are often targeted at plastids and because plastids are derived from cyanobacterial endosymbionts, it was suggested that the chlamydial and plant sequence similarity reflected an ancestral relationship between chlamydiae and cyanobacteria (Brinkman *et al.* 2002).

A recent phylogenomic analysis by Huang & Gogarten (2007) identified 21 chlamydiae-related genes in red algae and green plants. These red algal and green plant sequences form a monophyletic group with chlamydial homologues that is distinct from the cyanobacterial homologues. In some instances, homologues are altogether absent in cyanobacteria. Therefore, the overall data suggest that such chlamydiae-related genes are unlikely to be of plastidic (cyanobacterial) origin. This finding is confirmed by two later studies, in which over 50 genes were reported as being of chlamydial origin in plants and other plastid-containing eukaryotes (Becker *et al.* 2008; Moustafa *et al.* 2008). The number of chlamydiae-related genes in plastid-containing eukaryotes is particularly noteworthy, given that 16 genes in the red alga *Cyanidioschyzon* were found to be related to chlamydiae versus only 5 to γ - and β -proteobacteria and that there are 70 times more sequences from γ - and β -proteobacteria in GenBank than from chlamydiae. Because all existing chlamydial species are obligate endosymbionts, a parsimonious explanation would be that these chlamydiae-related genes are derived from an ancient symbiotic relationship between chlamydiae and the ancestor of primary photosynthetic eukaryotes (red algae, green plants and glaucophytes). In this case, the chlamydiae-related genes in other plastid-containing eukaryotes would likely be derived from secondary or tertiary endosymbioses with algal cells.

The hypothesis of an ancient chlamydial endosymbiosis with the ancestral primary photosynthetic eukaryote bears implications for the origin of plastids. Because plastids are derived from cyanobacteria, which are free-living and autotrophic, effective communications with the host cell must be in place in order to enslave and transform cyanobacterium into a photosynthetic organelle. Therefore, the early establishment of the plastid organelle in the ancestral primary photosynthetic eukaryote must have been heavily contingent on the availability of suitable transporters. Because of the nature of chlamydiae as obligate endosymbionts, some of the transporters that were required by cyanobacterial endosymbionts (e.g. ATP/ADP translocase) might have originally evolved in the chlamydial endosymbionts within the same host cell. Indeed, at least four plastid transporter genes in photosynthetic eukaryotes are derived from chlamydiae (Huang & Gogarten 2007; Tyra *et al.*

2007). Therefore, the putative ancient chlamydial endosymbiosis may have contributed to the early evolution of plastids (Huang & Gogarten 2007; Moustafa *et al.* 2008), and provides independent support for the unique origin of primary plastids.

3. CONCERTED GENE RECRUITMENT IN EUKARYOTIC EVOLUTION

Recent data indicate that genetic novelties introduced by HGT events can contribute greatly to the adaptation of recipient organisms. Even for eukaryotes, where HGT identification is often complicated by intracellular transfer from mitochondria and plastids, there is clear evidence for the important role of HGT. For instance, acquisition of the gene encoding *N*-acetylneuraminase lyase from γ -proteobacterial parasites probably contributed to the parasitic lifestyle of *Trichomonas* (de Koning *et al.* 2000). Similarly, gene transfer from pathogenic fungi might have facilitated the evolution of an osmotrophic lifestyle in oomycetes (Richards *et al.* 2006); this also offers a convincing example for HGT among eukaryotes.

Despite the increasing amount of recent data, some critical questions regarding HGT in eukaryotes, such as the overall scope and the impacts of HGT on macroevolution, are largely untapped. These issues are often linked to the frequency of anciently acquired genes in eukaryotes, which are particularly difficult to pinpoint because of the many complications involved, in particular, fading phylogenetic signals over time, differential gene losses and/or gains, and a possible organellar origin. Based on the frequent HGT in extant unicellular eukaryotes and the fact that multicellularity is derived from unicellularity, it has been speculated that ancient HGT was also frequent in eukaryotes (Huang & Gogarten 2006). In an attempt to understand the role of ancient HGT, Huang & Gogarten (2008) identified 39 genes that were probably introduced from prokaryotes to the red alga *Cyanidioschyzon*, prior to the split of red algae and green plants. While many of these anciently acquired genes may represent homologous displacement with a similar function, several others probably introduced novel functions essential to the ancestral plant. Interestingly, the vast majority of these anciently acquired genes are related to the functionality of plastids, the hallmark of photosynthetic eukaryotes, and they were also acquired from different donors. Therefore, ancient cyanobacterial endosymbiosis and HGT from multiple sources in concert drove the establishment of primary plastids. This observation of gene acquisition from multiple sources in the origin and optimization of evolutionary novelties has also been found in several other eukaryotic groups (Andersson *et al.* 2003; Striepen *et al.* 2004; Ricard *et al.* 2006).

Although HGT is a random process, the acquisition of beneficial genes that become fixed in the recipient lineage may *a posteriori* be labelled as recruitment from the perspective of the receiving organisms. The observation of apparently concerted gene recruitment in eukaryotic groups points to an interesting question about how HGT events may be viewed in terms of organismal interactions and natural selection. While

the traditional Darwinian speciation process is built on the basis of competition between organisms for limited resources, gene sharing depends upon innovations resulting from (and promoting) the evolutionary success of other distinct lineages. This could be viewed as 'cooperation' between genomes. However, rather than defining interspecific HGT as a purely cooperative aspect of organismal evolution, this reinforces that Darwinian selection also acts on the level of the gene, with the overall fitness of individual organisms as a higher order derivation thereof. This does not negate the role of cooperation in promoting innovations that increase the fitness of the units of selection. Rather, it clarifies that while HGT may often result from and be a part of cooperative systems, it should be recognized as a distinct evolutionary process.

That being said, the role of cooperation is important considering that much of eukaryotic evolution is contingent on the endosymbiotic origin of mitochondria and plastids, as well as other reticulate relationships such as hybridization. It should be noted that the metabolic diversity and gene pool of prokaryotes was in all likelihood significantly larger than those of eukaryotes during the early phases of eukaryotic evolution. If individual bacterial organisms can be incorporated into eukaryotic cells and transformed into organelles, frequent acquisition of individual genes should not be surprising. In this regard, HGT from multiple sources is shown to be able to generate and optimize novel evolutionary characters in eukaryotes, a likely prelude to the origin and adaptation of some major organismal lineages (e.g. two photosystems in cyanobacteria, evolution of plastids).

4. EVOLUTION OF ACETOCLASTIC METHANOGENESIS IN METHANOSARCINACEAE VIA HGT

Several classes of the archaeal phylum *Euryarchaeota* possess the unique ability to generate energy exclusively via the reduction of single-carbon substrates to methane, a process known as methanogenesis. It is likely that methanogenesis evolved early in the evolution of the euryarchaeotes, making this an extremely ancient microbial lifestyle (Bapteste *et al.* 2005; Gribaldo & Brochier-Armanet 2006). In fact, methanogenic pathways use a large number of genes with no recognizable paralogs in other metabolic processes, as well as a large number of complex cofactors and carrier proteins that appear to be unique to methanogens. This complexity and exclusivity provides a barrier against HGT. In fact, this barrier is so effective that no HGT of a methanogenic pathway to a non-methanogenic lineage has ever been reported (Gribaldo & Brochier-Armanet 2006).

Substrate-specific pathways vary across groups of methanogens, with distinct enzymes and cofactors for the use of carbon dioxide, carbon monoxide, methanol, methylamines, dimethyl sulphide, formate, acetate and other compounds. Each of these pathways feeds into the 'core' of methanogenesis, involving the transfer of a methyl group to coenzyme M (CoM), and its subsequent reduction to methane, generating

an H^+/Na^+ gradient for ATP production (Li *et al.* 2006). Many methanogens contain more than one of these pathways, allowing for a broad spectrum of environmental growth conditions. Acetoclastic methanogenesis (methanogenesis from acetate) is especially interesting since it produces two-thirds of biogenic methane on Earth, occurring in diverse environments such as marine and freshwater sediments, wetlands, soils and the gastrointestinal tract of animals (Ferry 1992; Boone *et al.* 2001; Galagan *et al.* 2002). Only two existing groups of methanogens can use acetate in this way, *Methanosarcinaceae* and *Methanosaeta*, although the relative contribution of each to global methane production is unclear. Both organisms first activate acetate to acetyl-CoA, which then transfers its methyl group to H_4MPT (tetrahydromethanopterin) (or H_4SPT —tetrahydrosarcinapterin) via the activity of the carbon monoxide dehydrogenase enzyme complex, generating CO_2 in the process. The methyl group is then transferred to CoM. While *Methanosaeta* use the enzyme acetate synthase for generation of acetyl-CoA from acetate in a single step, *Methanosarcina* use two enzymes, acetate kinase (AckA) and phosphoacetyltransferase (Pta) (figure 2). Interestingly, neither *ackA* nor *pta* has identified homologues in any other archaeal genome. Both genes are widely distributed throughout the bacterial domain, providing an indication that this pathway may have evolved via interdomain HGT.

Phylogenetic analyses of bacterial and methanogen AckA and Pta homologues consistently support the methanogen homologues rooting within a group of cellulolytic bacteria of the class Clostridia, strongly supporting a HGT event occurring between these two clades (Fournier & Gogarten 2008) (figure 1). Since all major bacterial phyla are represented in this analysis, the polarity of the transfer is clear; if the genes had been transferred from a member of the *Methanosarcinaceae* to all other bacterial lineages, they would be 'deeply rooting', as the transfer must have occurred before the diversification of existing bacterial groups (followed by vertical inheritance). Rampant independent transfer events between bacterial phyla also cannot explain the observed topology, given the enormous coincidence required for these events to result in a tree reflecting the major accepted relationships within and between bacterial groups.

Another observation supporting the inferred transfer polarity (and the transfer itself) is gene synteny. While *ackA* and *pta* are adjacent in all *Methanosarcinaceae* genomes, the same is not true for most bacterial genomes. However, cellulolytic clostridia all contain adjacent copies of *ackA* and *pta* in the same order as *Methanosarcinaceae*. This reinforces the described HGT scenario, and suggests that this transfer probably occurred as a single biological event, capturing two genes at once. In this way, a functional and selectively advantageous pathway for the activation of acetate would immediately be present, avoiding an evolutionarily useless 'intermediate' stage caused by stepwise transfer (figure 2). In addition, close physical interactions (consortia) between members of *Methanosarcinaceae* and cellulolytic clostridia are commonplace in aquatic environments (Stams 1994), and extensive

gene transfer between these groups has been documented (Beiko *et al.* 2005).

It is probable that global biogenic methane production was significantly lower before this transfer event, primarily resulting from other less productive forms of methanogenesis. As this may have had significant consequences for the global climate, from a geobiophysical perspective it becomes critical to know when in geological time this transfer event likely occurred. From the phylogenetic reconstruction, it is clear that the HGT occurred after the cellulolytic Clostridia were already well established and diversified. This suggests the existence of a diverse, cellulose-rich freshwater environment, presumably rich in terrestrial plant biomass. As such, environments could only have existed after land plants had become established, and the transfer event could not have occurred earlier than the mid-Ordovician, approximately 475 Myr ago (Wellman *et al.* 2003). Consequently, the diversification of *Methanosarcinaceae* could also not have occurred before this time.

5. MODIFICATION AND THERMOSTABILITY OF *THERMOTOGA* SSU RRNA

Organisms under strong purifying selection due to extreme environmental conditions can potentially benefit immediately from certain horizontal transfer events. For example, *Thermotoga maritima* is a bacterium native to geothermal vents on the sea floor, with a growth optimum of 80°C, and a maximum growth temperature of 90°C (Hubert *et al.* 1986). To thrive in this environment this organismal lineage, like other hyperthermophiles, has several adaptations, including modification of G + C content, codon usage and amino acid composition of proteins (Singer & Hickey 2003). These organisms also often modify the nucleosides of their ribosomal RNAs to increase the stability of their interactions, and the efficiency of ribosomal assembly and protein translation in extreme physiological conditions. Using high-throughput liquid chromatography coupled to electrospray ionization mass spectroscopy, the locations of modified nucleosides in rRNA have been mapped for several organisms, including *T. maritima* (Decatur & Fournier 2002, 2003; Guymon *et al.* 2006, 2007). It was discovered that *T. maritima* shows extensive nucleoside modification, including four pseudouridine positions likely to be involved with increasing thermostability, and an unknown modification of cytidine at 16S rRNA position 1404 (Guymon *et al.* 2007). While being unique among known bacterial maps, the mass and location of this modified residue precisely match a modified position within the 16S rRNA of the archaeon *Haloferax volcanii*, a halophilic mesophile (Gupta *et al.* 1983). The aberrant phylogenetic signal of this character is suggestive of HGT of an RNA modification enzyme. Until the specific protein responsible for this modification is identified, however, comparative genomic studies and phylogenetic analysis cannot be used to confirm this hypothesis. Furthermore, until the specific structure of the modified base is known, it will be impossible to infer whether it would increase the

Figure 1. Consensus trees of Pta and AckA homologues within the Clostridia. Branches containing horizontal gene transfers to *Methanosarcinaceae* are indicated by an asterisk. Homologues from *Fusobacterium nucleatum* were used as outgroups to root the trees. The numbers associated with each clade indicate bootstrap values for maximum likelihood and neighbour joining analyses and the posterior probability from Bayesian inference. Methodologies for sequence alignment and tree reconstructions are published elsewhere (Fournier & Gogarten 2008). Figure adapted from Fournier & Gogarten (2008). © American Society for Microbiology, *J. Bacteriol.* 190, 1124–1127.

Figure 2. Transfer of the acetoclastic metabolic pathway to *Methanosarcinaceae*. Bold arrows indicate path of reduced single-carbon unit through the methanogenic/acetogenic pathways. Cofactors and phosphate substrate identical to those in *Methanosarcinaceae* are omitted from the cellulolytic Clostridia pathway for clarity. Multiple letters in the enzymatic descriptions indicate enzymatic complexes made of multiple subunits.

Figure 3. Phylogenetic relationships of class II aminoacyl-tRNA synthetases. Numbers associated with selected clades indicate bootstrap values for maximum likelihood and neighbour joining, and posterior probability from Bayesian inference. Methodologies for sequence alignment and tree reconstructions are identical to those used in Fournier & Gogarten 2008. Shaded regions indicate distinct aaRS families predominantly inherited vertically within the labelled clades. PyIRS homologues are not shaded, to reflect their lack of vertical inheritance and uncertain evolutionary history (as indicated by the '?' at the root of the PyIRS tree). Dashed lines represent phylogenetic relationships with poor bootstrap support. MRCA nodes (circles) refer to positions in the protein family tree, which correspond to the MRCA within the organismal tree.

stability of 16S rRNA under both thermophilic and halophilic conditions. The physiochemical demands of both conditions are similar in some aspects, suggesting that such a transfer could be fixed under purifying selection. As a large number of genes have been shown to be horizontally transferred from various archaeal species to the Thermotogales, this scenario would not be without precedent (Nelson *et al.* 1999; Nesbo *et al.* 2006).

6. HGT FROM EXTINCT LINEAGES

There is no reason to doubt that HGT has been a factor in evolution since the origin of genes themselves (Woese 1998). Since it is indisputable that the vast majority of species that have ever existed are now extinct, it directly follows that the genomes of today's living organisms contain a large number of genes that originated in extinct lineages. The more ancient the transfer event, the less likely the donor lineage is still in existence, and the more divergent the transferred genes will be compared with existing homologues. In many cases, no recognizable homologues may even exist in living organisms, causing these

transfers to appear as 'orphan' gene families. Such genes would represent a 'genetic life raft', as the only surviving genetic legacy of entire clades of extinct organisms (Gogarten *et al.* 2008). In other cases, the transferred gene(s) may root even deeper than the divergence of bacterial and archaeal orthologues, if the lineage in which they evolved diverged even before the time of the MRCA. Such deep transfers can also help resolve the dilemma of the large amount of biological innovation that occurred between the origin of life and the MRCA, in a relatively short period of evolutionary time (Woese 2002). A large number of divergent lineages undergoing extensive HGT would act as a 'parallel evolution processor', with the subsequent ancestor of existing organisms inheriting a genome encoding biological systems that could have taken billions of years to evolve in a purely vertical, Darwinian biological world.

A potential example of such an ancient transfer is the pyrrolysine (Pyl) system (Fournier *in press*). As previously mentioned, methanogenesis consists of several sub-pathways for the use of many single-carbon substrates. Within *Methanosarcina*, these pathways include methanogenesis from methylamines,

Phil. Trans. R. Soc. B (2009)

single transposon family; Zhang *et al.* 2005) in that each requires Pyl in its catalytic site, enabling the transfer of a methyl group to the cobalt atom of the corrinoid carrier protein (Krzycki 2004). Pyl is a non-canonical amino acid, in that (i) the vast majority of existing organisms do not have it as part of their genetic code and (ii) it requires special machinery for insertion. Unlike selenocysteine (Sec), the other known non-canonical amino acid, Pyl has its own unique aminoacyl tRNA synthetase (aaRS) for the aminoacylation of tRNA^{Pyl}, and does not rely on any tRNA-dependent steps for its biosynthesis (Polycarpo *et al.* 2004). However, like Sec, Pyl-encoding UAG codons are associated with a downstream stem-loop structure (PYLIS) that facilitates the incorporation of Pyl (Zhang *et al.* 2005).

In order for an organism to use encoded Pyl, it must have an intact Pyl biosynthesis pathway, a specific tRNA^{Pyl}, Pyl-aaRS (PylRS) and, often, PYLIS elements. Evolution of a system to use an entirely novel amino acid would require the kind of positive selection that would both require and result in extensive utilization. Yet, this amino acid is only used for a single type of enzymatic reaction within a handful of enzymes, none of which are essential (Galagan *et al.* 2002). Usually, biological traits with this kind of narrow/sparse phylogenetic distribution are assumed to be derived, having evolved relatively recently within the lineage in question, and perhaps subject to very recent HGT. However, the molecular phylogeny of class II aaRS proteins places PylRS as a deep branching lineage within the IIb subfamily, indicating that the evolution of PylRS predates the evolution of the bacterial and the archaeal domains (figure 3). To be explained only by vertical inheritance with lineage-specific losses, the Pyl usage machinery (as well as genes that use Pyl) must have been present, like other aaRS genes, in the MRCA. Then, for unknown reasons, these genes were lost in all lineages except the direct ancestors of the clades that currently possess the system. Aside from requiring an enormous number of independent losses to produce the observed PylRS distribution, this explanation also fails to address the fact that both archaea and bacteria only use Pyl in the very same protein homologues. Given this model, either the ancestral Pyl system only ever evolved to be used in this one set of enzymes (which is highly unlikely) or Pyl was originally used in many families of proteins, all of which have been lost except for, in three independent cases, methylamine methyltransferases (perhaps even more unlikely).

HGT provides a more plausible explanation. It has been shown experimentally that the Pyl incorporation machinery is clustered as a 'cassette' that permits the synthesis and usage of Pyl in any genome to which it is transferred, as is also evidenced by its usage in *D. hafniense* (Longstaff *et al.* 2007). Since PylRS is such a deeply branching protein lineage, the donor of the transfer would have to be a member of a very ancient group of organisms, one that diverged before the MRCA of the bacterial and archaeal domains (figure 4). As previously discussed, coalescence theory strongly suggests that such deeply branching lineages existed, and presumably were engaging in HGT with the ancestors of known organisms (Zhaxybayeva & Gogarten 2004). This solution also explains the absence of other Pyl-utilizing genes, as, even though these probably evolved in the donor lineage, the transfer of only a single Pyl-utilizing gene would be sufficient for the retention of the Pyl system.

HGT implies the contemporaneous existence of both the recipient and donor lineages. Therefore, transfer from extinct lineages must have occurred sometime in the biological past. However, if the transfer event/extinction is too early, one must again invoke massive lineage-specific losses of the Pyl system to explain the observed phylogenetic distribution, as the transfer would predate the divergence of the observed recipient clades. The only plausible explanation would therefore be a 'balanced' scenario, with ancient

transfer to the ancestors of distinct groups of extant organisms (e.g. Clostridia, δ -proteobacteria, Euryarchaeotes), followed by limited independent losses within each lineage.

7. CONCLUDING REMARKS

HGT is an ongoing evolutionary process that imposes a web-like structure on the tree/coral of life. Rather than artefacts that confound our understanding of evolution, HGTs are phylogenetic characters in their own right, revealing much about the relationships between diverse groups of organisms, and the selective pressures that have shaped the metabolism, physiology and ecology of the modern living world. Realizing the significance of HGT also allows for a better understanding of the early evolution of life on Earth. Methods for the detection of HGT should continue to be applied and expanded to further explore the genetic contribution of extinct lineages to existing genomes, in an effort to unravel some of the deepest and most challenging problems in evolutionary biology.

We thank Kristen Swithers and Juliana Mastrorunzio for their helpful suggestions in the preparation of this manuscript. The reported work was supported in part through the NASA Applied Information System Research (NNG04GP90G), Exobiology (NNX07AK15G; NNX08AQ10G) and NSF Assembling the Tree of Life (DEB 0830024) Programs, and by a Research and Creative Activity Award from the East Carolina University to J.H.

REFERENCES

- Andersson, J. O., Sjogren, A. M., Davis, L. A., Embley, T. M. & Roger, A. J. 2003 Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* **13**, 94–104. (doi:10.1016/S0960-9822(03)00003-4)
- Bapteste, E., Brochier, C. & Boucher, Y. 2005 Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. *Archaea* **1**, 353–363.
- Becker, B., Hoef-Emden, K. & Melkonian, M. 2008 Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol. Biol.* **8**, 203. (doi:10.1186/1471-2148-8-203)
- Beiko, R. G., Harlow, T. J. & Ragan, M. A. 2005 Highways of gene sharing in prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 14332–14337. (doi:10.1073/pnas.0504068102)
- Boone, D., Whitman, W. & Koga, Y. 2001 Order III. Methanosarcinales. In *Bergey's manual of systematic bacteriology*, vol. 1 (eds D. Boone, G. Castenholz & G. Garrity), pp. 268–294. New York, NY: Springer-Verlag.
- Bose, A., Pritchett, M. A. & Metcalf, W. W. 2008 Genetic analysis of the methanol- and methylamine-specific methyltransferase 2 genes of *Methanosarcina acetivorans* C2A. *J. Bacteriol.* **190**, 4017–4026. (doi:10.1128/JB.00117-08)
- Brinkman, F. S. *et al.* 2002 Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. *Genome Res.* **12**, 1159–1167. (doi:10.1101/gr.341802)
- Burke, S. A., Lo, S. L. & Krzycki, J. A. 1998 Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J. Bacteriol.* **180**, 3432–3440.

- Darwin, C. 1836–1844 *Charles Darwin's notebooks transcription published in 1987*, p. B26. Ithaca, NY: Cornell University Press.
- Darwin, C. 1859 *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, UK: John Murray.
- Daubin, V. & Ochman, H. 2004 Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res.* **14**, 1036–1042. (doi:10.1101/gr.2231904)
- de Koning, A. P., Brinkman, F. S., Jones, S. J. & Keeling, P. J. 2000 Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**, 1769–1773.
- Decatur, W. A. & Fournier, M. J. 2002 rRNA modifications and ribosome function. *Trends Biochem. Sci.* **27**, 344–351. (doi:10.1016/S0968-0004(02)02109-6)
- Decatur, W. A. & Fournier, M. J. 2003 RNA-guided nucleotide modification of ribosomal and other RNAs. *J. Biol. Chem.* **278**, 695–698. (doi:10.1074/jbc.R200023200)
- Donoghue, H. D. & Spigelman, M. 2006 Comment. Pathogenic microbial ancient DNA: a problem or an opportunity? *Proc. Biol. Sci.* **273**, 641–642; discussion 643. (doi:10.1098/rspb.2005.3261)
- Felsenstein, J. 2003 *Inferring phylogenies*. Sunderland, MA: Sinauer.
- Ferry, J. G. 1992 Methane from acetate. *J. Bacteriol.* **174**, 5489–5495.
- Fournier, G. P. 2008 Evolution of the genetic code: computational methods and inferences. In *Computational methods for understanding bacterial and archaeal genomes*, vol. 7 (eds Y. Xu & J. P. Gogarten), pp. 75–97. London, UK: Imperial College Press.
- Fournier, G. P. in press HGT and the evolution of methanogenic pathways. In *Horizontal gene transfer—genomes in flux* (eds M. Gogarten-Boekels, L. Olendski & J. P. Gogarten). New York, NY: Humana Press.
- Fournier, G. P. & Gogarten, J. P. 2008 Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic clostridia. *J. Bacteriol.* **190**, 1124–1127. (doi:10.1128/JB.01382-07)
- Galagan, J. E. *et al.* 2002 The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**, 532–542. (doi:10.1101/gr.223902)
- Gibbons, A. 2005 Ancient DNA. New methods yield mammoth samples. *Science* **310**, 1889.
- Gogarten, J. P. & Townsend, J. P. 2005 Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687. (doi:10.1038/nrmicro1204)
- Gogarten, J. P. & Zhaxybayeva, O. 2008 Horizontal gene transfer: its detection and role in microbial evolution. In *Computational methods for understanding bacterial and archaeal genomes*, vol. 7 (eds Y. Xu & J. P. Gogarten). London, UK: Imperial College Press.
- Gogarten, J. P., Fournier, G. & Zhaxybayeva, O. 2008 Gene transfer and the reconstruction of life's early history from genomic data. *Space Sci. Rev.* **135**, 115–131. (doi:10.1007/s11214-007-9253-8)
- Gogarten-Boekels, M., Hilario, E. & Gogarten, J. P. 1995 The effects of heavy meteorite bombardment on the early evolution—the emergence of the three domains of life. *Orig. Life Evol. Biosph.* **25**, 251–264. (doi:10.1007/BF01581588)
- Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. 2005 Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469. (doi:10.1038/nature03676)
- Greub, G. & Raoult, D. 2003 History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a Chlamydiales ancestor to plants 1 billion years ago. *Appl. Environ. Microbiol.* **69**, 5530–5535. (doi:10.1128/AEM.69.9.5530-5535.2003)
- Gribaldo, S. & Brochier-Armanet, C. 2006 The origin and evolution of Archaea: a state of the art. *Phil. Trans. R. Soc. B* **361**, 1007–1022. (doi:10.1098/rstb.2006.1841)
- Gupta, R., Lanter, J. M. & Woese, C. R. 1983 Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaeobacterium. *Science* **221**, 656–659. (doi:10.1126/science.221.4611.656)
- Guymon, R., Pomerantz, S. C., Crain, P. F. & McCloskey, J. A. 2006 Influence of phylogeny on posttranscriptional modification of rRNA in thermophilic prokaryotes: the complete modification map of 16S rRNA of *Thermus thermophilus*. *Biochemistry* **45**, 4888–4899. (doi:10.1021/bi052579p)
- Guymon, R., Pomerantz, S. C., Ison, J. N., Crain, P. F. & McCloskey, J. A. 2007 Post-transcriptional modifications in the small subunit ribosomal RNA from *Thermotoga maritima*, including presence of a novel modified cytidine. *RNA* **13**, 396–403. (doi:10.1261/rna.361607)
- Herring, S., Ambrogelly, A., Polycarpo, C. R. & Soll, D. 2007 Recognition of pyrrolysine tRNA by the *Desulfotobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic Acids Res.* **35**, 1270–1278. (doi:10.1093/nar/gkl1151)
- Horn, M. *et al.* 2004 Illuminating the evolutionary history of chlamydiae. *Science* **304**, 728–730. (doi:10.1126/science.1096330)
- Huang, J. & Gogarten, J. P. 2006 Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* **22**, 361–366. (doi:10.1016/j.tig.2006.05.004)
- Huang, J. & Gogarten, J. P. 2007 Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* **8**, R99. (doi:10.1186/gb-2007-8-6-r99)
- Huang, J. & Gogarten, J. P. 2008 Concerted gene recruitment in early plant evolution. *Genome Biol.* **9**, R109. (doi:10.1186/gb-2008-9-7-r109)
- Hubert, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B. & Stetter, K. O. 1986 *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* **144**, 324–333.
- Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. 2003 Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602. (doi:10.1093/molbev/msg154)
- Krzycki, J. A. 2004 Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. *Curr. Opin. Chem. Biol.* **8**, 484–491. (doi:10.1016/j.cbpa.2004.08.012)
- Kunin, V. & Ouzounis, C. A. 2003 GeneTRACE—reconstruction of gene content of ancestral species. *Bioinformatics* **19**, 1412–1416. (doi:10.1093/bioinformatics/btg174)
- Lamarck, J.-B. 1809 *Philosophie zoologique, Part II*, pp. 463. Paris, France: Dentu, et l'Auteur. p.
- Lawrence, J. G. & Ochman, H. 1997 Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**, 383–397. (doi:10.1007/PL00006158)
- Lawrence, J. G. & Ochman, H. 1998 Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417. (doi:10.1073/pnas.95.16.9413)
- Lerat, E., Daubin, V. & Moran, N. A. 2003 From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol.* **1**, E19.
- Li, Q., Li, L., Rejtar, T., Lessner, D. J., Karger, B. L. & Ferry, J. G. 2006 Electron transport in the pathway of acetate conversion to methane in the marine archaeon *Methanosarcina acetivorans*. *J. Bacteriol.* **188**, 702–710. (doi:10.1128/JB.188.2.702-710.2006)

- Longstaff, D. G., Larue, R. C., Faust, J. E., Mahapatra, A., Zhang, L., Green-Church, K. B. & Krzycki, J. A. 2007 A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proc. Natl Acad. Sci. USA* **104**, 1021–1026. (doi:10.1073/pnas.0610294104)
- Margulis, L. 1995 *Symbiosis in cell evolution: microbial communities in the Archean and Proterozoic eons*. New York, NY: W H Freeman & Co.
- Martin, A. P., Costello, E. K., Meyer, A. F., Nemergut, D. R. & Schmidt, S. K. 2004 The rate and pattern of cladogenesis in microbes. *Evolution* **58**, 946–955.
- McCoy, A. J., Adams, N. E., Hudson, A. O., Gilvarg, C., Leustek, T. & Maurelli, A. T. 2006 L,L-diaminopimelate aminotransferase, a trans-kingdom enzyme shared by *Chlamydia* and plants for synthesis of diaminopimelate/lysine. *Proc. Natl Acad. Sci. USA* **103**, 17909–17914. (doi:10.1073/pnas.0608643103)
- Mirkin, B. G., Fenner, T. I., Galperin, M. Y. & Koonin, E. V. 2003 Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol. Biol.* **3**, 2. (doi:10.1186/1471-2148-3-2)
- Moustafa, A., Reyes-Prieto, A. & Bhattacharya, D. 2008 Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* **3**, e2205. (doi:10.1371/journal.pone.0002205)
- Nelson, K. E. et al. 1999 Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329.
- Nesbo, C. L., Doolittle, W. F., Mongodin, E. F. & Nelson, K. E. 2006 Outside forces helped shape the *Thermotoga* Metagenome. *Microbe* **1**, 235–241.
- Nisbet, E. G. & Sleep, N. H. 2001 The habitat and nature of early life. *Nature* **409**, 1083–1091. (doi:10.1038/35059210)
- Normand, P. et al. 2007 Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. *Genome Res.* **17**, 7–15. (doi:10.1101/gr.5798407)
- Page, R. D. M. & Holmes, E. C. 1998 *Molecular evolution: a phylogenetic approach*. Chichester, UK: Blackwell Publishing.
- Pennisi, E. 1998 Genome data shake tree of life. *Science* **280**, 672–674. (doi:10.1126/science.280.5364.672)
- Polcarpo, C., Ambrogelly, A., Berube, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L. & Soll, D. 2004 An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl Acad. Sci. USA* **101**, 12450–12454. (doi:10.1073/pnas.0405362101)
- Raup, D. M. 1989 The case for extraterrestrial causes of extinction. *Phil. Trans. R. Soc. B* **325**, 421–431; discussion 431–435.
- Ricard, G. et al. 2006 Horizontal gene transfer from Bacteria to rumen Ciliates indicates adaptation to their anaerobic, carbohydrates-rich environment. *BMC Genomics* **7**, 22. (doi:10.1186/1471-2164-7-22)
- Richards, T. A., Dacks, J. B., Jenkinson, J. M., Thornton, C. R. & Talbot, N. J. 2006 Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr. Biol.* **16**, 1857–1864. (doi:10.1016/j.cub.2006.07.052)
- Rokas, A. & Carroll, S. B. 2006 Bushes in the tree of life. *PLoS Biol.* **4**, e352. (doi:10.1371/journal.pbio.0040352)
- Royo, J., Gimez, E. & Hueros, G. 2000 CMP-KDO synthetase: a plant gene borrowed from gram-negative eubacteria. *Trends Genet.* **16**, 432–433. (doi:10.1016/S0168-9525(00)02102-8)
- Ryall, K., Harper, J. T. & Keeling, P. J. 2003 Plastid-derived type II fatty acid biosynthetic enzymes in chromists. *Gene* **313**, 139–148. (doi:10.1016/S0378-1119(03)00671-1)
- Schmitz-Esser, S., Linka, N., Collingro, A., Beier, C. L., Neuhaus, H. E., Wagner, M. & Horn, M. 2004 ATP/ADP translocases: a common feature of obligate intracellular amoebal symbionts related to chlamydiae and rickettsiae. *J. Bacteriol.* **186**, 683–691. (doi:10.1128/JB.186.3.683-691.2004)
- Singer, G. A. & Hickey, D. A. 2003 Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene* **317**, 39–47. (doi:10.1016/S0378-1119(03)00660-7)
- Snel, B., Bork, P. & Huynen, M. A. 2002 Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* **12**, 17–25. (doi:10.1101/gr.176501)
- Stams, A. 1994 Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie Van Leeuwenhoek* **66**, 271–294. (doi:10.1007/BF00871644)
- Stephens, R. S. et al. 1998 Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**, 754–759. (doi:10.1126/science.282.5389.754)
- Striepen, B., Pruijssers, A. J., Huang, J., Li, C., Gubbels, M. J., Umejiego, N. N., Hedstrom, L. & Kissinger, J. C. 2004 Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl Acad. Sci. USA* **101**, 3154–3159. (doi:10.1073/pnas.0304686101)
- Tyra, H. M., Linka, M., Weber, A. P. & Bhattacharya, D. 2007 Host origin of plastid solute transporters in the first photosynthetic eukaryotes. *Genome Biol.* **8**, R212. (doi:10.1186/gb-2007-8-10-r212)
- Welch, R. A. et al. 2002 Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 17 020–17 024. (doi:10.1073/pnas.252529799)
- Wellman, C. H., Osterloff, P. L. & Mohiuddin, U. 2003 Fragments of the earliest land plants. *Nature* **425**, 282–285. (doi:10.1038/nature01884)
- Willerslev, E. & Cooper, A. 2005 Ancient DNA. *Proc. Biol. Sci.* **272**, 3–16. (doi:10.1098/rspb.2004.2813)
- Woese, C. 1998 The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859. (doi:10.1073/pnas.95.12.6854)
- Woese, C. R. 2002 On the evolution of cells. *Proc. Natl Acad. Sci. USA* **99**, 8742–8747. (doi:10.1073/pnas.132266999)
- Wolf, Y. I., Aravind, L. & Koonin, E. V. 1999 Rickettsiae and chlamydiae: evidence of horizontal gene transfer and gene exchange. *Trends Genet.* **15**, 173–175. (doi:10.1016/S0168-9525(99)01704-7)
- Zhang, Y. & Gladyshev, V. N. 2007 High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. *Nucleic Acids Res.* **35**, 4952–4963. (doi:10.1093/nar/gkm514)
- Zhang, Y., Baranov, P. V., Atkins, J. F. & Gladyshev, V. N. 2005 Pyrrolysine and selenocysteine use dissimilar decoding strategies. *J. Biol. Chem.* **280**, 20 740–20 751. (doi:10.1074/jbc.M501458200)
- Zhaxybayeva, O. & Gogarten, J. P. 2004 Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet.* **20**, 182–187. (doi:10.1016/j.tig.2004.02.004)
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. 2006 Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**, 1099–1108. (doi:10.1101/gr.5322306)