Pre-Proof Version:

Horizontal gene transfer and the formation of groups of microorganisms

David Williams, Cheryl P. Andam, and J. Peter Gogarten*

Department of Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Rd., Storrs, CT 06269-3125, USA

*Author for correspondence
E-mail: gogarten@uconn.edu

In:

Molecular Phylogeny of Microorganisms,

Editors: Aharon Oren and R. Thane Papke,

Caister Academic Press, ISBN: 978-1-904455-67-7

http://www.horizonpress.com/phylogeny

Table of Contents:

Abstract

In this chapter, we discuss the impact of gene transfer on the formation of groups of organisms. We begin by discussing the obvious: gene transfer can make it more difficult to define and determine relationships. In those cases where many genes have been transferred between preferred partners, the majority of genes in a genome may reflect gene acquisition, and as a consequence, if a coherent signal is detected, one nevertheless might not be sure that the signal is due to organismal shared ancestry. In the second part of this chapter we will focus on two positive aspects of gene transfer. The presence of a particular transferred gene was shown, in several cases, to constitute a shared derived character useful in classification. Gene transfer can put together new metabolic pathways that open up new ecological niches, and consequently, the transfer of an adaptive gene might create a new group of organisms.

Problems encountered in phylogenetic reconstruction

Towards a natural taxonomy

Phylogeny asks the question "How did higher taxonomic units come into existence?" According to Hennig, a natural taxonomic system aims for a classification that reflects organismal shared ancestry (Hennig, 1966). In such a natural system of classification the formation of proper taxonomic groups is based on shared derived characters. The distinction between primitive and derived characters is important in a cladistic framework. Only shared derived characters define a monophyletic clade[1], whereas a primitive character defines a paraphyletic group. A paraphyletic group is derived from a common ancestor, but this common ancestor also gave rise to organisms outside the group. A group defined by a primitive character is paraphyletic because those organisms that possess the derived character state also evolved from the same ancestor as those that retained the primitive character state. In molecular phylogenies, derived and primitive character states usually are not, and often cannot, be distinguished. A tree resulting from sequence-based phylogenetic reconstruction is usually unrooted. The use of outgroups is one way to polarize part of a phylogenetic tree, but the placement of the root should be explicitly discussed. In case of doubt it is preferable to avoid cladistic terminology altogether (Wilkinson et al., 2007). This comment is not trivial. Many data suggest that for many molecular systems the root of the net[2] of life is located on the bacterial branch (Zhaxybayeva and Gogarten, 2007; Zhaxybayeva et al., 2005), making the Archaea and Eukaryotes sister domains. However, the root of the bacterial domain is uncertain with respect to the different bacterial phyla. Ribosomal data, including both ribosomal proteins and ribosomal RNA, often place the Aquificae as the deepest branching lineage in the bacterial domain (Woese, 1987), but the outgroup is located on a very long branch, and prudence suggests that the point where the outgroup joins the ingroup should be considered as unresolved. In unrooted ribosomal phylogenies, Aquificae and Thermotogales form a strongly supported clan (*sensu* Wilkinson et al., 2007), but if the traditional rRNA phylogeny were correct, these two organisms would not

[1] Sometimes monophyletic clades *sensu* Hennig (1966) are labeled as holophyletic, and the term monophyletic is re-defined to include holo- and paraphyletic groups (Ashlock, 1971). Here we use the terms as defined by Hennig.

[2] We use the term "net of life" in place of "tree of life" to highlight that evolution has been a reticulate process that should not be reduced to a single bifurcating process.

form a clade, but a paraphyletic grouping. The same uncertainty also permeates placing the root of the eukaryotes (we use this term to denote the nucleocytoplasmic component of the eukaryotic cell) and the relationships between Eukaryotes and Archaea (for discussion see Dagan and Martin, 2007b; Kurland et al., 2006; Martin et al., 2007; Penny and Poole, 1999).

Gene transfer and phylogenetic reconstruction

In addition to the problems generated by the difficulty and uncertainty of phylogenetic reconstruction, horizontal gene transfer (HGT) creates additional complications. Because of the horizontal transfer of genetic information, genes coexisting in a genome frequently have different histories (Doolittle, 1999; Gogarten et al., 2002; Gogarten and Townsend, 2005; Hilario and Gogarten, 1993; Lawrence and Ochman, 1997; Ochman et al., 2000). Based on the frequency with which genes are encountered in related organisms, genomes of microorganisms can be divided into different sets of genes. The extended core includes genes found in all members of the group. The extended bacterial core includes of about 250 gene families, i.e., 250 gene families are found represented in nearly all bacterial genomes. In contrast, the pan-genome of a group, such as a species, genus or larger taxonomic unit, is defined as the non-overlapping set of all gene families found in at least one genome. According to current estimations, the pan-genome of many bacterial groups (Tettelin et al., 2005), and of bacteria as a whole (Lapierre and Gogarten, 2009) is open, *i.e.*, each randomly selected bacterial genome that is sequenced will encode many proteins that do not belong to any previously characterized family. At present, the average number of novel genes per newly sequenced genome is estimated to be over a hundred genes per genome, with no leveling off being detected (Lapierre and Gogarten, 2009). The genes that are encountered only once, or very infrequently in other genomes have been labeled as accessory genes and constitute on average about one forth of each bacterial genome. Most of these genes are not well characterized, they are strain specific, frequently acquired through gene transfer, most do not persist in a lineage for long periods of time, and they are more AT rich and shorter than the average gene (Daubin and Ochman, 2004; Lapierre and Gogarten, 2009; Lawrence and Ochman, 1997). For accessory genes found in closely related strains the rate of non-synonymous to synonymous substitutions reveals a low level of purifying selection (Daubin and Ochman, 2004); however, it was suggested that this low level of selection might be due to mutations generating detrimental properties and not due to an adaptive value of the transferred

gene (Gogarten and Townsend, 2005). The majority of the recently acquired genes might be neutral or nearly neutral in their effect on the fitness of the recipient.

A corollary to the large number of accessory genes per genome is that the common shared gene pool of strains that belong to the same species is surprisingly small[3]. Welch et al. (2002) reported that the common shared gene pool of three *Escherichia coli* genomes was less than 40% of the non-overlapping gene set; in the case of three *Frankia* strains whose small subunit ribosomal RNA showed less than 3 per cent sequence divergence, only 20% of the common shared gene pool had detectable homologues in all three genomes (Normand et al., 2007). The accessory genes illustrate that microbial genomes are changing rapidly, and most of the accessory genes constitute derived characters for the strains in which they are found; however, these genes are not found in other organisms, thus they do not provide a taxonomic marker. In contrast, core genes contain phylogenetic information; however, the debate is ongoing with respect to the presence and meaning of a phylogenetic consensus signal (see Bapteste et al., 2009; and Gogarten-Boekels et al., 2009 for an overview of recent discussions).

The phylogenetic analysis of extreme thermophilic bacteria provides an excellent illustration of the problems created through highways of gene sharing (Beiko et al., 2005). Ribosomal RNA and ribosomal protein-based phylogenies of Thermotogales (Zhaxybayeva et al., 2009b) and Aquificales (Boussau et al., 2008) place them as early branching groups at the base of the bacterial domain; about 10% to 20% of the genes in the genomes of Thermotogales appear to have been acquired by horizontal gene transfer from archaeal donors (Nelson et al., 1999; Nesbø et al., 2001; Zhaxybayeva et al., 2009b); however, a majority of their genes reveal affinities of Thermotogales with Clostridia (Gophna et al., 2005; Zhaxybayeva et al., 2009b) and of Aquificales with Epsilonproteobacteria (Boussau et al., 2008). If we assume that the ribosomal

[3] Assume the following thought experiment: three genomes from three different strains of a species each encode 2000 genes. 600 genes per genome, or 30% of the genes per genome are not found in the other two genomes. 1400 genes are present in each of the three genomes, i.e., they belong to the core. Therefore, in total there are 1400 gene families in the core genome, and 3 time 600 = 1800 genes in the accessory pool. The pan-genome defined as the total non-overlapping gene pool is 1800+1400 = 3200 genes, and the common shared gene pool, in this example, is 1400 genes or 44% of the non-overlapping gene pool.

phylogeny reflects the organismal phylogeny of the Thermotogales, then the vast majority of genes would have to have been transferred from Clostridia.

The Thermotogales contain three distinct phylogenetic signatures in their genome, archaeal genes acquired through HGT, clostridial genes that likely also were acquired through HGT, and the ribosome that groups them with Aquificales and that may reflect the evolutionary history of the translation machinery. Reducing the phylogeny of these organisms to only one of these components misses a major process; in particular, the plurality signal, also known as emerging consensus (Wolf et al., 2002) or the central trend (Puigbo et al., 2009), at least in the case of the Thermotogales and Aquificales appears to be overwhelmed by a highway of gene sharing and might not represent the history of the ribosomal nor the organismal tree as defined by cell divisions.

The complexity of genome evolution cannot be captured in a single tree-diagram without reticulations, nor by a hierarchical classification system that is based on a strictly furcating process. A hierarchical classification system possibly might be founded on a sub-cellular system such as the ribosome, which consists of many interacting parts that encounter higher barriers to HGT than other genes (Sorek et al., 2007). Such a system would fall short of describing organismal evolution and phylogeny; however, phylogenies for such subsystems promise to constitute a good starting point to reconstruct the reticulated history of life (Gogarten, 1995; Swithers et al., 2009).

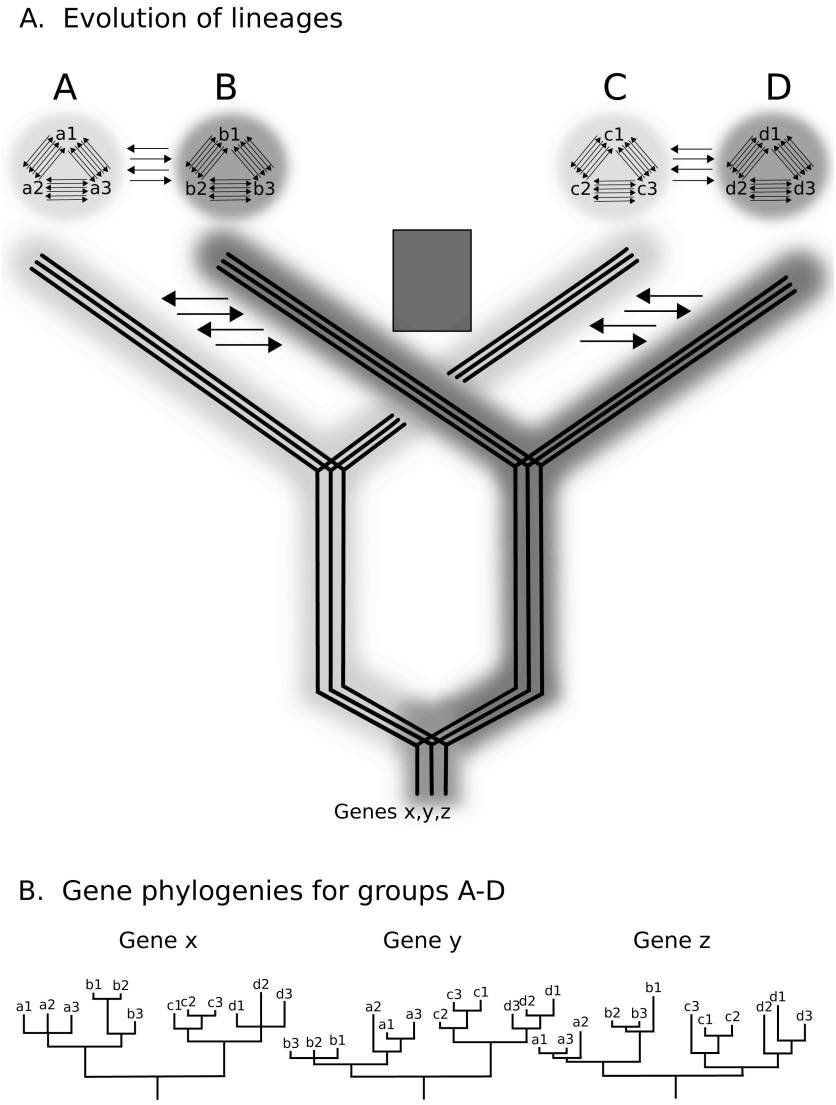

Figure 1. The role of biased gene transfer in the evolution of organismal lineages. Panel A shows four groups of organisms representing lineages, composed of members that frequently swap genes among each other. Within each group, high levels of gene transfer act as a homogenizing force. The organismal tree of these four groups indicates that lineages A and C share a last common ancestor, different from the last ancestor of lineages B and D. However, a barrier to gene transfer (represented by a gray rectangle) between A & B and between C & D is established, halting gene exchange between more closely related organisms. As a consequence, A & B continue to swap genes with each other, and C with D. Without assuming a known organismal phylogeny, we would conclude from the gene phylogenies that lineages A & B (and C & D) are more related to each other and thus will cluster together. Hence, it will be inaccurate to assume that these two larger groups were formed as a result of vertical inheritance; rather these groups were formed as a consequence of high rates of exchange between the lineages. Panel B depicts phylogenetic trees reconstructed fir different genes (x, y, z) for groups A-D shown in the top panel. Each gene tree might not necessarily show similar topology to the organismal tree. At the level of lower taxonomic units (a1, a2, a3; …), constant shuffling of genes occur within each lineage, creating conflicting topological patterns, even though the group as a whole remain consistent. More frequent transfer between specific lineages (between A & B and between C & D) can skew their groupings, reflecting a bias between exchange partners. This bias may not always be attributed to organismal shared ancestry. The figure is based on concepts and diagrams from Gogarten et al. (2002) and Olendzenski et al. (2001).

HGT can create patterns indistinguishable from those created through shared ancestry

In the previous section we considered HGT as a process that can give rise to conflicting patterns of phylogenetic information. In this section we consider HGT as a process that can create phylogenetic patterns that may be indistinguishable from those created through shared ancestry. We suggest that even in the case where only a single phylogenetic pattern is detected, this pattern may have been created or co-created through biased HGT. Relatedness of organisms in phylogenetic reconstruction has always been correlated with a vertical genealogical path originating from a common ancestor. In the same way, taxonomical classification of organisms, including microorganisms, often implies that individuals possess similar properties that they have inherited from their parents. However, if we trace back the history of a species, particularly for prokaryotic evolution, we observe a more complicated story that reveals not just a single evolutionary process. More specifically, the current taxonomic associations of organisms may actually have been shaped by HGT, thus negating the argument that HGT only creates discord in phylogenetic reconstruction.

Figure 1 illustrates how preferential gene transfer can create clusters of organisms possessing similar properties. Within each group, high levels of gene transfer act as a homogenizing force that holds the group together. Under the scenario depicted in Figure 1 molecular phylogenies will reflect these exchange groups, and not shared ancestry of the organisms. An organism's propensity for specific exchange partners may be a consequence of the transcription and translation machineries that resemble its own, facilitating the entry and establishment of novel genes (Gogarten et al., 2002; Lawrence and Hendrickson, 2005; Olendzenski et al., 2001). On the other hand, incompatible genetic apparatus can reduce the likelihood and frequency of successful transfers (Hendrickson and Lawrence, 2006). Ecological isolation of independent lineages can also enhance biased transfer, due to the limited partners available.

Plurality of pattern and process

If members of the group exchange genes, they become more similar. The more genes they share, the more similar they become, and the more bias may be created for future transfers.

Although this may seem circuitous, it only emphasizes the observation that biased transfers can, at the very least, reinforce hierarchical groups formed through vertical inheritance. Later on, transferred genes can lead to the vertical transmission of innovations that were not present in the ancestor of the group (Gogarten and Olendzenski, 1999) and can persist as shared derived characters that unite the descendants of the group (Huang and Gogarten, 2006; Huang et al., 2005). To the extent that a phylogenetic pattern is created through biased HGT, the observed phylogenetic signal has to be considered phenetic and not cladistic. A consensus phylogenetic pattern should not automatically be equated with shared organismal ancestry. Considering that ribosomal RNA genes are also prone to horizontal transfer (Schouls et al., 2003; van Berkum et al., 2003; Wang and Zhang, 2000; Yap et al., 1999), even ribosomal trees, considered to be a reflection of organismal evolution, may well be, in part, a manifestation of HGT bias. Following transfer, ribosomal RNA encoding genes often are integrated into the recipient genome through homologous recombination (Morandi et al., 2005; Yap et al., 1999). In some instances the ribosomal rRNA operon might be a mosaic that integrates over many gene transfer events between related organisms (Gogarten et al., 2002). Similarity of organisms and their relatedness as identified through molecular phylogenies can be created through two processes: shared ancestry and gene transfer between preferred partners. HGT does create distinct patterns in molecular phylogenies; in addition, biased HGT also provides an additional process that can generate the patterns usually ascribed to shared organismal ancestry.

A population genetic point of view for higher level taxonomic units

Exchange groups not only exist at the population and species level. Higher-level taxonomic units (such as order and phylum) also display partiality toward certain HGT partners, creating larger groups that exhibit phenotypic coherence. Barriers to gene exchange, such as geographical, genetic and physiological features, can induce higher-level groups to swap genes with certain other assemblages of organisms and not with others. A pattern emerges in which higher taxonomic groups exhibit behavior that is analogous to that found in populations of a single species. Hence, taxa, at any level, will persist as long as they maintain their propensity toward their exchange partners. A population genetic model is therefore necessary to account for the processes that underlie the genetic structure of lineages, particularly in the microbial realm.

HGT frequencies differ according to the type of gene, group of organism, and ecosystem (Beiko et al., 2005; Jain et al., 1999; Zhaxybayeva et al., 2007). In addition, methods to detect HGT events have complementary sensitivities with respect to phylogentic depth of the transfer (Ragan, 2001), and have different error rates for false positives and false negatives in detecting gene transfer events (Poptsova and Gogarten, 2007; Zhaxybayeva, 2009). Dagan and Martin (2007a) reported a minimum rate of 1.1 HGT events per gene family, and that at least 65% of all gene families have been affected by HGT; Zhaxybayeva et al. (2006) found traces of gene transfer in more than 50% of gene families in cyanobacterial genomes. In some ecosystems and for some closely related organisms in the process of de-speciation or de-generation the gene transfer rates likely are much higher (Sheppard et al., 2008; Zhaxybayeva et al., 2009a). The impact of HGT on organismal phylogeny will not be constant throughout life's history.

HGT and the creation of metabolic pathways

Categorising gene transfer

Homologous and non-homologous recombination

Natural selection requires variation as raw material. Throughout the biosphere this variation is achieved through mutation of genetic sequences (Chao and Cox, 1983; Kimura, 1967; Levins, 1967)[and recombination, the mixing of existing genetic material (Fisher, 1930; Jeffreys and Neumann, 2002; Papke et al., 2007). Numerous mechanisms to increase the amount of both sources of novelty, representing the evolution of evolvability, have been described including mutator phenotypes in bacteria (Gross and Siegel, 1981; Taddei et al., 1997), conjugation in Bacteria (Kuusinen, 1996) and Archaea (Prangishvili et al., 1998; Schleper et al., 1994) as well as the elaborately choreographed homologous recombination between chromosomes in sexually reproducing organisms such as ourselves (Creighton and McClintock, 1931). This last example is found in populations of sexually reproducing individuals among which recombination occurs every generation but which also experience strong barriers to recombination between individuals of different populations. This is the basis for Mayr's Biological Species Definition (Mayr, 1942, 1963) but evidence is building that similar barriers influence diversity within and between microbial lineages (Doolittle and Zhaxybayeva, 2009; Dykhuizen and Green, 1991; Fraser et al., 2009; Papke, 2009). As described above, recombination may occur with sufficient frequency to

create patterns resembling those of vertical inheritance at different taxonomic levels, as delineated by the effectiveness of barriers to recombination. Recombination between closely related microorganisms (sometimes referred to as HGT although in this context has an effect similar to sex) can occur at very high frequencies as in colonizing populations of the human pathogen *Helicobacter pylori* (Suerbaum et al., 1998) and is usually homologous due to high levels of sequence similarity and genome synteny. Between more distant lineages recombination becomes less likely as sequence similarity and synteny decrease (Gogarten et al., 2002; Majewski and Cohan, 1999; Wolf et al., 2001). Integration of transferred genes by non-homologous recombination, causing a decrease in synteny between closely related lineages, has been postulated as part of the speciation process in some microorganisms (Retchless and Lawrence, 2007).

Effects on fitness

The evolutionary impact of HGT can, to some extent, be categorized by the frequency of successful HGT events in conjunction with the apparent evolutionary fitness effects on the recipient. One category includes transfers among an exchange group of closely related lineages as defined above. The sufficiently relaxed recombination barriers among lineages results in discrete, random, one way transfers at a high enough frequency to allow a degree of perceived reciprocity and an allele exchange-like process. Analyses of genetic data have revealed such processes in diverse groups including between pathogenic and commensal strains of the *Neisseria* genus (Feil et al., 1996; Zhou et al., 1997); between *Agrobacterium* biovar 1 'species' (Costechareyre et al., 2009); and between isolates of the free living haloarchaeal genus *Halorubrum* (Papke et al., 2007). Between a closely related donor and recipient ecological and intracellular conditions are likely to be similar so that potential fitness gains by genetic transfer will usually be slight, but enough to allow persistence in and potential subsequent transfer from the recipient. However, stronger selective pressures may cause more abrupt changes in the allele-like population dynamics of transferred genetic material, for example the rapid spread of antibiotic resistance genes among phyla of microbial pathogens (Davies, 1994)., for example among strains of the *Neisseria* genus (Feil et al., 1996; Zhou et al., 1997). Ecological and intracellular conditions may be similar between donor and recipient so that potential fitness gains by genetic

transfer are usually relatively slight but enough to allow persistence in and potential subsequent transfer from the recipient. However, the allele-like population dynamics of transferred genetic material would respond to changes in selective pressure such that stronger pressures cause more abrupt changes, for example the rapid spread of antibiotic resistance in genes among phyla of microbial pathogens (Davies, 1994).

HGT between distant relatives

Despite their comparative rarity, HGT events between distant lineages have also had profound evolutionary impacts. The success of some very large extant clades of organisms appears to be due to HGT-mediated synergism between pre-existing metabolic machinery and the products of the newly acquired genetic material. Examples include the assembly to oxygen producing photosynthesis in cyanobacteria (also referred to as blue-green algae and Cyanophyta; Oren, 2004) which are among the most abundant organisms known and are found throughout the oceans (Chisholm et al., 1988; Waterbury et al., 1979; Zwirglmaier et al., 2008), rivers and lakes (Zwart et al., 2002), hot springs (Papke et al., 2003), hypersaline pools (Sørensen et al., 2005), and in symbiotic niches (Kuusinen, 1996; Lesser et al., 2004). Their oxygenic photosynthetic metabolic pathway is credited with the oxygenation of the atmosphere starting 2.7 billion years ago based on geological evidence (Buick, 1992). Members of the euryarchaeal order Methanosarcinales are capable of acetoclastic methanogenesis, a process that today produces approximately two-thirds of biogenic methane annually (Ferry, 1992), and are widely distributed in marine and freshwater sediments, soils and the gastrointestinal tracts of animals (Donovan et al., 2004; Koizumi et al., 2003). The enzymes used to initiate this process by members of the genus *Methanosarcina* are coded for by genes atypical in Archaea and phylogenetic evidence reveals that they were acquired from cellulolytic clostridia via HGT (Fournier and Gogarten, 2008).

Rather than modifying an existing metabolic function, this category of HGT event leads to more profound innovation but may be rare for two reasons. Firstly, successful chromosomal integration of genetic material between distant relatives is less likely because of several reasons, including the following: low sequence similarity prevents homology dependent recombination (Gogarten et al., 2002; Majewski and Cohan, 1999; Wolf et al., 2001); non-overlapping

specificity of genetic vectors such as phage (Breitbart and Rohwer, 2005), and geographical isolation because of contrasting ecological niches or natural history. Secondly, following chromosomal integration, the fitness effect on the recipient is likely to be near-neutral or negative if the promoter system is incompatible or the gene product interferes with the existing cell chemistry (which is more likely between distantly related, divergent organisms). A decrease in fitness of the recipient is likely to lead to its extinction and leave no trace in the evolutionary record. The frequency of such events is retrospectively undetectable and the impact negligible when considering evolutionary legacies of HGT. Even if HGT products with near-neutral fitness effects are fixed due to genetic drift, such transfers alone do not hold much promise for the recipient's descendants.

HGT and the expansion of metabolic networks

However, the acquisition of a gene that allows the recipient to occupy a new, previously unoccupied niche provides an increase in fitness even if the expression and regulation of the gene is not optimally integrated into the recipients metabolic network (Gogarten et al., 2002). Consequently, it is not surprising that many of the genes successfully transferred across domain boundaries encode enzymes that transport metabolizable substrates into the cell (e.g., Noll and Thirangoon, 2009) and that expand the capabilities at the periphery of the recipient's metabolic network (Pal et al., 2005). Much like winning a state lottery, some of the rare successes have been spectacular as demonstrated by the above examples. The key to this success is in the nature of cell metabolism: the energy transductions and matter transitions, mediated by electron transfer among chemical species, necessary for the proliferation of life. This complex process can be considered a network of discrete nodes or modules linked together as a network (Duarte et al., 2004; Reed et al., 2003). Each node is usually a multimeric protein complex catalyzing a specific chemical reaction or the transfer of a specific chemical species across a membrane. Reagents are transferred between connected nodes for subsequent reactions. If the products of laterally transferred genetic material happen to facilitate one or more chemical reactions under the recipient's internal conditions, and if that chemical reaction fits into the existing metabolic network, a significant change in eco-physiology can occur; as well as the potential for enhancing the competitiveness in a presently occupied ecological niche. Whole new modes of metabolism may become possible, opening up new niches for the HGT recipient and its descendants

(Gogarten et al., 2002; Pal et al., 2005) and even causing globally significant changes in biogeochemistry. The legacies of such rare events include, but are not limited to (Boucher et al., 2003), the widely distributed *Methanosarcina* and their influence on the global methane budget and the even more ubiquitous cyanobacteria as primary producers in many ecosystems and their oxygenation of the atmosphere.

In the following section, we discuss a few selected detailed examples and a fortuitous side effect of such rapidly appearing derived characters with respect to cladistics.

Oxygen producing photosynthesis

Bacteria that are able to gain energy through anoxygenic photosynthetic pathways form a diverse polyphyletic group. Some members of the alpha-, beta- and gamma-proteobacteria use light energy to transfer electrons from sulfide, elemental sulfur or hydrogen to inorganic carbon compounds for subsequent release of energy in anaerobic autophototrophy (purple bacteria). However, when molecular oxygen is available heterotrophic oxidation of organic matter is possible by reversing the flow of electrons through cytochrome *b/c* (Blankenship et al., 1995). Chloroflexi (green filamentous bacteria) possess a reaction center similar to that of the purple bacteria and photosytem II of the cyanobacteria, whilst Chlorobi (green sulfur bacteria) and Heliobacteria (of the Gram positive phylum Firmicutes) each possesses similar reaction centers that contrast with the former and are similar to photosynthetic reaction center I in cyanobacteria (Raymond, 2008; Sadekar et al., 2006). The recent discovery of members of other bacterial phyla (Acidobacteria) capable of photoheterotrophy (Bryant et al., 2007) further increases the wide distribution of anoxygenic photosynthetic metabolism (Blankenship, 2001; Raymond, 2008).

The oxygenic photosynthetic machinery of cyanobacteria is more complex combining two distinct reaction centers (photosystems) that work in series when electrons are transported from water to NADP (Ferreira et al., 2004). The photosynthetic machinery is encoded in more than 100 genes, whose products form a tightly interacting machinery (Shi et al., 2005). In addition to their electron transport chain, cyanobacteria have evolved many diverse adaptations to the environments that live in, including complex carbon dioxide concentrating centers (Badger and Price, 2003). Some of these adaptations occurred in the cyanobacterial lineage after some

cyanobacteria became endosymbionts that evolved into the plastids of eukaryotes. Structural, functional, and sequence similarity have been noted between each of the cyanobacterial reaction centers and those of the green filamentous bacteria and the purple bacteria respectively (Buttner et al., 1992; Liebl et al., 1993; Sadekar et al., 2006).

Two distinct scenarios have long been discussed (see Woese, 1987, for an early review) on how the more complex electron transport chain came into existence:

A) Internal gene duplications of the reaction centers in the cyanobacterial stem (*i.e.*, the lineage leading to the most recent common ancestor (MRCA) of all extant cyanobacteria), gave rise to the two types of reaction centers now working in series (Mulkidjanian et al., 2006). These reaction centers were then transferred horizontally from the cyanobacteria to other bacterial groups.

B) The different types of reaction centers evolved in different groups of photosynthetic bacteria and were brought together in the cyanobacterial stem though HGT. In this scenario the cyanobacterial stem contained photosynthetic bacteria that were not able to use water as an electron donor in photosynthesis. Before the successful integration of the two photosytems the organisms of the cyanobacterial stem group might have used PSI for cyclic electron transport (as in heterocysts and the bundle sheath cells of some C4 plants); and/or they might have used either of the photosystems with another electron donor.

Several observations strongly favor the latter scenario for the creation of oxygenic photosynthesis through HGT:

- Gene transfers, and not internal gene duplications, are the pathway by which extant bacteria expand their metabolic capabilities (Gogarten et al., 2002; Pal et al., 2005; see the following discussion of proton and sodium pumping ATPases for an illustration).
- HGT of genes encoding the photosynthetic machinery, and in particular components of the reaction center complexes, occur frequently. In some Proteobacteria the genes encoding the photosynthetic machinery are all clustered together in one region of the genome, and phylogenetic analyses revealed that this super cluster was transferred from alpha- to beta-proteobacteria (Igarashi et al., 2001). Marine cyanophages carry genes encoding peptides integral to photosynthetic reaction centers, and the phylogeny of the phage and bacterial

version of this enzyme is intertwined suggesting transfer in both directions (Lindell et al., 2004; Mann et al., 2003; Sharon et al., 2009).

- The plurality consensus phylogeny, the phylogeny of the ribosome, and the phylogeny of enzymes encoding the chlorophyll biosynthetic pathway do not agree with each other, revealing that the photosynthetic machinery did not evolve exclusively through vertical inheritance, but included HGT (Raymond et al., 2003; Raymond et al., 2002; Xiong and Bauer, 2002; Xiong et al., 2000; Zhaxybayeva et al., 2004). Furthermore, the two reaction centers in cyanobacteria are not closely related to each other (Sadekar et al., 2006).

While debate continues on how the reaction center dimers themselves evolved, and in which directions they were transferred, the autochthonous evolution of the two reaction centers within the cyanobacterial stem group without the involvement of HGT is unlikely. The synergistic union in ancestral cyanobacteria of pre-existing metabolic units containing the two reaction centers also is compatible with the rather long branch that connects the cyanobacteria to the other bacterial phyla, a pattern for the number of lineages through time that is not typical for other bacterial phyla (Gogarten and Lapierre, unpublished). The coming together of two different reaction centers in a single cell, and their subsequent evolution to work in series in a single electron transport chain, was a singular event in evolution. After it had occurred, electrons could be transported over a larger electrochemical potential differences allowing for the use of water as electron donor. The ability to use a ubiquitous substrate as electron donor made the cyanobacteria an extraordinary successful and diverse group.

Acetoclastic methanogenesis in *Methanosarcina*

In the case of the acquisition of genes innovating acetoclastic methanogenesis in *Methanosarcina*, the supportive phylogenetic inferences and the direction of the transfer are clearer as methanogenesis is restricted to a relatively narrow group of euryarchaeota (Fournier, 2009). Although this group, consisting of several classes of the phylum Euryarchaeota, is paraphyletic according to phylogenies of gene involved in transcription and translation (Bapteste et al., 2005; Brochier et al., 2005; Woese, 1987), the non-methanogens that would make it monophyletic (Thermoplasmatales, Archaeoglobales and Halobacteriales) are likely to have independently lost the core methanogenic genes. In addition, these core genes have not been

found in other lineages, the evolution of the core methanogenesis pathway appears to have been dominated by vertical descent (Bapteste et al., 2005). In contrast, the history of the enzymes that allow methanogens to use different substrates (methylamines and acetate) involves HGT (Fournier, 2009; Fournier et al., 2009).

Methanosarcina perform the transformation of acetate to acetyl-CoA in two steps using acetate kinase and phosphoacetyltransferase. The genes coding for these two enzymes are not homologous to any known archaeal gene, but are widely distributed among bacteria (Fournier and Gogarten, 2008). In phylogenetic analyses these two enzymes from *Methanosarcina* cluster inside a group of homologs from cellulolytic clostrida, within the clostridia. These data suggest that these two genes were acquired by an ancestor of *Methanosarcina* in a single HGT event from a donor within the group of cellulolytic clostridia. HGT between methanogens and clostridia is not unprecedented (Beiko et al., 2005) and modern representatives of the two groups can be found living in close proximity (Stams, 1994).

Evolution of Na^+ and H^+ pumping ATPases/ATP synthase

In prokaryotes genes that are typically considered as paralogs are often created through gene transfer (Jeffery Lawrence, personal communication; Gogarten et al., 2002). *Sensu stricto*, paralogs are genes that at the root of their phylogenetic relationship are related by a gene duplication event (Fitch, 2000); in contrast, genes related through a speciation are labeled as orthologs. One way for two homologous genes with different function to end up in the same genome is through adaptation to different function in two distinct lineages, followed by HGT of the then functionally distinct version into the same genome. In this scenario, the enzyme acquires a new function in isolation, thus avoiding problems caused through gene conversion and recombination. At this point the genes can be considered diverging orthologs. After one or both of the diverging genes have acquired a different function, the genes are so different in sequence that following an HGT event, they no longer undergo homologous recombination and, following non-homologous recombination, can co-exist in the same genome.

The ATP synthases/ATPases provide an illustration for this scenario. ATP synthases belong to a large group of multi-subunit enzymes that includes the bacterial coupling factor ATP synthase (also labeled as F-ATPases); the vacuolar ATPase found as an exclusively proton

pumping ATPase in Eukaryotes energizing the endomembrane system (e.g., vacuoles and lysosomes), but also on the plasma membrane of specialized cells, such as osteoclasts and in cells of transport epithelia (Gogarten et al., 1992; Harvey and Nelson, 1992); and as coupling factor ATPase in Archaea (Gogarten et al., 1989a; Gogarten et al., 1989b; Gogarten and Taiz, 1992). The latter ATPase type is often labeled as A-type ATPase, because it is mainly found in Archaea, where it has the same function as the F-type ATPase in bacteria (Gogarten et al., 1989a; Gruber and Marshansky, 2008; Ihara and Mukohata, 1991). Both the F- and the A-ATPases are reversible, and for both types different versions of the enzymes are known that transport of protons or sodium ions (Dimroth et al., 1999; Takase et al., 1993; Yokoyama et al., 1998). However, because A-ATPases are in sequence more similar to the V-ATPases, they are often included under the category of vacuolar type ATPases. Further complicating the terminology is the finding that both the bacterial and the archaeal type ATPase have been horizontally transferred between the domains (Hilario and Gogarten, 1993), e.g., *Methanosarcina* possess genes encoding an F-type ATPase in addition to its A-type ATPase (Sumi et al., 1997); the Deinococcaceae only encode an A-type ATPase in their genome (Lapierre et al., 2006; Olendzenski et al., 2000; Tsutsumi et al., 1991).

Other enzymes homologous to the ATP synthase are ATPases that are part of the type III secretion system, the bacterial flagella assembly machinery (Vogler et al., 1991), and the rho transcription termination factor (Richardson, 2002). The latter functions as a helicase that unwinds the DNA RNA heteroduplex. All these ATPases appear to function as rotary motors, consisting of a ring of 6 homologous subunits that rotate a central element, either a protein subunit, a protein substrate, or a nucleic acid strand (Abrahams et al., 1994; Mulkidjanian et al., 2007). Sequence similarity between the ATP binding subunits is sufficient to establish homology unambiguously, but insufficient phylogenetic information has been retained in the sequences to reliably reconstruct the phylogenetic relations between the different enzymes; however, interesting scenarios have been developed that relate ion pumps and protein translocating ATPases (Kibak et al., 1992; Mulkidjanian et al., 2007).

Mulkidjanian et al. (2008) argue, based on phylogentic reconstruction and protein structure, that the bacterial and archaeal coupling factor type ATPases first evolved to transport

sodium ions, and that the lineages independently evolved specificity for protons (protons usually have a much lower concentration than sodium ions, thus a higher discrimination between substrates is needed for enzymes to become specific for protons). In agreement with this hypothesis some lineages that often are recovered as deep branching lineages in molecular phylogenies contain sodium ion translocating ATPases; for example, the Thermotogales (Ludwig et al., 1998; Mulkidjanian et al., 2008; Woese, 1987) use sodium ions to energize their plasma membrane (Galperin and Koonin, 1997) and possess a typical F-type ATPase (Ludwig et al., 1998), whereas most Bacteria and Archaea use protons in energy coupling for ATP synthesis.

Some organisms, including the firmicute *Enterococcus hirae* possess both an A-type ATPase with specificity for sodium (Takase et al., 1994), and an additional F-ATPase that transports protons and is involved in pH regulation (Shibata et al., 1992). A similar situation is found in *Thermotoga neapolitana* (Iida et al., 2002), although the enzymes in *T. neapolitana* have not been characterized as well as in the case of *E. hirae.* Analysis of molecular phylogenies and of gene presence absence data reveals that these F and A- type ATPase coexisting in the same cell did not evolve by gene duplication in the *E. hirae* or *T. neapolitana* lineages, respectively (Lapierre, 2007; Mulkidjanian et al., 2008). Rather, these enzymes are homologs that have a long independent evolutionary history, and that were brought together in the same lineage through HGT. The presence of two homologous enzymes with different substrate specificity allows the recipient to use one or both of these ATPases for ion transport, and the other for ATP synthesis.

Transferred genes as shared derived characters

It is perhaps counterintuitive that transferred genes also aid the reconstruction of organismal phylogeny. If a gene is transferred from a divergent donor, and incorporated into the recipient in a way that places the gene under purifying selection, then the presence of the transferred gene can be considered a shared derived character for the recipient and its descendent (Huang and Gogarten, 2006). Only a small minority of transferred genes will attain the status of a shared derived character. Many transferred genes will be neutral or nearly neutral in the recipient (Gogarten and Townsend, 2005), or be under selection only for a short period of time (Lawrence

and Roth, 1996), and may persist in the recipient lineage for less than a few hundred million years (Lawrence and Ochman, 2002). For a gene to survive in the recipient lineage on the long run, it must become essential for the recipient, either because it provides or contributes to a function that became essential for the recipient lineage (e.g., photosynthesis; Huang and Gogarten, 2007), or because it replaces the gene that was originally present in the recipient (e.g., the ATP synthase in Deinococcaceae; Olendzenski et al., 2000). The assembly of new complex traits frequently uses genes that were acquired by gene transfer (Huang and Gogarten, 2008).

HGT of aminoacyl tRNA synthetases

Aminoacyl tRNA synthetases (aaRS), *i.e.*, the enzymes that charge the tRNAs with their cognate amino acids, often provide useful derived characteristics. These enzymes are essential in all living organisms, they are sufficiently conserved to allow phylogenetic reconstruction at the level of the whole net-of-life (Wolf et al., 1999), and they often exist in different versions that have the same basic function (Farahi et al., 2004), but might differ in sensitivity to antibiotic resistance. The fact that aaRSs only make specific interactions with few other macromolecules (i.e., the tRNAs), and that they are frequent targets of antibiotics (Brown et al., 2003) might make them prone to HGT between divergent organisms (Woese et al., 2000).

A possible complication of using genes as phylogenetically informative characters that resulted from orthologous replacement is that the replacement process is not instantaneous. The transferred gene is so divergent from its ortholog that it cannot be integrated into the recipient genome by homologous recombination. Consequently, two versions of the gene coexist for some time in the recipient lineage, until one or the other is deleted (Stern et al., in press). The ultimate deletion of one or the other gene likely occurs because functional redundancy reduces the level of purifying selection acting on the individual copy. Lineage splitting events that occurred during the time when both of the genes were still present in the recipient might not be correctly resolved (Huang and Gogarten, 2009).

The presence of a haloarchaeal tyrRS in all opisthoconts supports the animals and fungi as a monophyletic clade (Huang et al., 2005). A thrRS in *Prochlorococcus* that was transferred from within the Gammaproteobacteria, supports the notion of *Prochlorococcus* as a

monophyletic group (Huang and Gogarten, 2009; Zhaxybayeva et al., 2006), which is in contrast to many phylogenies that are based on an average of the phylogenetic information contained in the genomes, which surprisingly recover *Prochlorococcus* as paraphyletic (e.g., Beiko et al., 2005; Zhaxybayeva et al., 2006; for further see discussion see Zhaxybayeva et al., 2009a). The presence of proline and alanine RSs in diplomonads and parabasalids that group with the homolog from *Nanoarchaeum equitans* suggests a closer than expected relationship between these two protist lineages (Andersson et al., 2005).

The origin of primary plastids

Chamydial type genes were detected in several plant and algal lineages (Becker et al., 2008; Brinkman et al., 2002; Huang and Gogarten, 2007; Moustafa et al., 2008). These gene often have functions in plastids (Brinkman et al., 2002). Different scenarios were proposed, when these genes were first identified:

- Everett et al. (2005) suggested that these genes might have been acquired by plants through HGT from Chlamydiae via plant feeding insects; however, these genes are also found in red algae, and the split between red and green algae predates the existence insects. Furthermore, the plant homologs group within the Chlamydiae, often specifically with homologs from Protochlamydia, the chlamydial homologs do not group within the plants (Huang and Gogarten, 2007).
- Brinkman et al. (2002) suggested that these genes revealed a close relationship between Cyanobacteria and Chlamydiae, and that these genes were contributed to the plant and algal cells via the cyanobacterial endosymbiont that evolved into the plastid. However, when more of the chlamydial type genes present in plants were analyzed, cyanobacterial homologs were identified that did not group with the chlamydial homologs (Huang and Gogarten, 2007), illustrating that the chlamydial and plant homologs are distinct from the cyanobacterial ones.
- Huang and Gogarten (2007) suggested that a chlamydial parasite might have played a crucial role in establishing the cyanobacterial symbiosis that let to the evolution of the primary plastid.

To date over 50 genes of chlamydial origin were reported in plants, red algae and glaucocystophytes. In addition to these chlamydial genes, additional bacterial genes were found

in red and green algae, plants, and in some instances glaucocystophytes (Huang and Gogarten, 2006, 2008; Tyra et al., 2007). The presence of these transferred genes in all lineages that possess primary plastids supports a single origin of the primary plastid and suggests that these lineages form a monophyletic clade, the Archaeplastida (Adl et al., 2005)

Correlating evolution across the domains of life

In addition to defining clades in the net-of-life, genes transferred between divergent organisms also provide a means to correlate evolutionary events that occurred in different groups. The donor and recipient of an HGT event had to live either at the same time, or in the case of intermediate vectors or carriers, the donor might have existed before the recipient. Applied to the HGT of tyrRS from *Haloarcula* to the opisthokonts, we can conclude that the halophilic Archaea had already diversified into different groups including *Haloarcula* as distinct from *Halobacterium* before the opisthocont lineage split into animals and fungi. The grouping of mitochondrial and cyanobacterial proteins with the alphaproteobacterial and cyanobacterial lineage (Gray, 1993), respectively, and the presence of chlamydial type genes in all Archaeplastida (Adl et al., 2005) reveals that he bacteria were already diversified into different phyla before the different kingdoms and phyla of eukaryotes evolved.

Conclusions

Horizontal gene transfer is an innovative force in evolution that has changed the face of planet Earth. HGT impacts the creation and recognition of groups in different ways: gene transfer biased towards more related partners maintains and possibly creates groups of recognizably related organisms, but the same process also has the potential to create a phylogenetic signal that may be different from organismal shared ancestry; gene transfer between divergent organism can create new metabolic pathways and extend existing metabolic networks, and it can it can provide characters useful for a natural taxonomic classification. The examples discussed in this chapter illustrate that the pathways created through HGT have shaped Earth's biosphere. A more detailed mapping of HGT events will help to unravel the intertwined histories of organismal evolution and the evolution of biochemical and regulatory networks.

Acknowledgments

The authors thank the National Science Foundation Assembling the Tree of Life program (DEB 0830024) for support.

References

Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. Nature *370*, 621-628.

Adl, S.M., Simpson, A., Farmer, M., Andersen, R., Anderson, O., Barta, J., Bowser, S., Brugerolle, G., Fensome, R., Fredericq, S., James, T.Y., Karpov, S., Kugrens, P., Krug, J., Lane, C.E., Lewis, L.A., Lodge, J., Lynn, D.H., Mann, D.G., McCourt, R.M., Mendoza, L., Moestrup, O., Mozley-Standridge, S.E., Nerad, T.A., Shearer, C.A., Smirnov, A.V., Spiegel, F.W., and Taylor, M.F. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. *5*, 399-451.

Andersson, J.O., Sarchfield, S.W., Roger, A.J., Sjogren, A.M., Davis, L.A., and Embley, T.M. (2005). Gene transfers from Nanoarchaeota to anancestor of Diplomonads and Parabasalids - phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Mol. Biol. Evol. *22*, 85-90.

Ashlock, P.D. (1971). Monophyly and associated terms. Syst. Zool. *20*, 63-69.

Badger, M.R., and Price, G.D. (2003). CO_2 concentrating mechanisms in cyanobacteria: Molecular components, their diversity and evolution. J. Exp. Bot. *54*, 609-622.

Bapteste, E., Brochier, C., and Boucher, Y. (2005). Higher-level classification of the Archaea: evolution of methanogenesis and methanogens. Archaea *1*, 353-363.

Bapteste, E., O'Malley, M.A., Beiko, R.G., Ereshefsky, M., Gogarten, J.P., Franklin-Hall, L., Lapointe, F.J., Dupré, J., Dagan, T., Boucher, Y., and Martin, W. (2009). Prokaryotic evolution and the tree of life are two different things. Biology Direct *4*, 34.

Becker, B., Hoef-Emden, K., and Melkonian, M. (2008). Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. BMC Evol. Biol. *8*, 203.

Beiko, R.G., Harlow, T.J., and Ragan, M.A. (2005). Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. USA *102*, 14332-14337.

Blankenship, R.E. (2001). Molecular evidence for the evolution of photosynthesis. Trends Plant Sci. *6*, 4-6.

Blankenship, R.E., Madigan, M.T., and Bauer, C.E. (1995). Anoxygenic photosynthetic bacteria (Dordrecht: Kluwer Scientific).

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E., Nesbo, C.L., Case, R.J., and Doolittle, W.F. (2003). Lateral gene transfer and the origins of prokaryotic groups. Annu. Rev. Genet. *37*, 283-328.

Boussau, B., Gueguen, L., and Gouy, M. (2008). Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of Aquificales in the phylogeny of Bacteria. BMC Evol. Biol. *8*, 272.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol. *13*, 278-284.

Brinkman, F.S., Blanchard, J.L., Cherkasov, A., Av-Gay, Y., Brunham, R.C., Fernandez, R.C., Finlay, B.B., Otto, S.P., Ouellette, B.F., Keeling, P.J., Rose, A.M., Hancock, R.E., Jones, S.J., and Greberg, H. (2002). Evidence that plant-like genes in *Chlamydia* species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. Genome Res. *12*, 1159-1167.

Brochier, C., Forterre, P., and Gribaldo, S. (2005). An emerging phylogenetic core of Archaea: phylogenies of transcription and translation machineries converge following addition of new genome sequences. BMC Evol. Biol. *5*, 36.

Brown, J.R., Gentry, D., Becker, J.A., Ingraham, K., Holmes, D.J., and Stanhope, M.J. (2003). Horizontal transfer of drug-resistant aminoacyl-transfer-RNA synthetases of anthrax and Gram-positive pathogens. EMBO Reports *4*, 692-698.

Bryant, D.A., Garcia Costas, A.M., Maresca, J.A., Chew, A.G.M., Klatt, C.G., Bateson, M.M., Tallon, L.J., Hostetler, J., Nelson, W.C., Heidelberg, J.F., and Ward, D.M. (2007). *Candidatus* Chloracidobacterium thermophilum: An aerobic phototrophic acidobacterium. Science *317*, 523-526.

Buick, R. (1992). The antiquity of oxygenic photosynthesis: Evidence from stromatolites in sulfate-deficient archaean lakes. Science *255*, 74-77.

Buttner, M., Xie, D.L., Nelson, H., Pinther, W., Hauska, G., and Nelson, N. (1992). Photosynthetic reaction center genes in green sulfur bacteria and in photosystem 1 are related. Proc. Natl. Acad. Sci. USA *89*, 8135-8139.

Chao, L., and Cox, E.C. (1983). Competition between high and low mutating strains of *Escherichia coli*. Evolution *37*, 125-134.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature *334*, 340-343.

Costechareyre, D., Bertolla, F., and Nesme, X. (2009). Homologous recombination in *Agrobacterium*: Potential implications for the genomic species concept in bacteria. Mol. Biol. Evol. *26*, 167-176.

Creighton, H.B., and McClintock, B. (1931). A correlation of cytological and genetical crossing-over in *Zea mays*. Proc. Natl. Acad. Sci. USA *17*, 492-497.

Dagan, T., and Martin, W. (2007a). Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. Proc. Natl. Acad. Sci. USA *104*, 870-875.

Dagan, T., and Martin, W. (2007b). Testing hypotheses without considering predictions. Bioessays *29*, 500-503.

Daubin, V., and Ochman, H. (2004). Bacterial genomes as new gene homes: The genealogy of ORFans in *E. coli*. Genome Res. *14*, 1036-1042.

Davies, J. (1994). Inactivation of antibiotics and the dissemination of resistance genes. Science *264*, 375-382.

Dimroth, P., Wang, H., Grabe, M., and Oster, G. (1999). Energy transduction in the sodium F-ATPase of *Propionigenium modestum*. Proc. Natl. Acad. Sci. USA *96*, 4924-4929.

Donovan, S.E., Purdy, K.J., Kane, M.D., and Eggleton, P. (2004). Comparison of Euryarchaea strains in the guts and food-soil of the soil-feeding termite *Cubitermes fungifaber* across different soil types. Appl. Environ. Microbiol. *70*, 3884-3892.

Doolittle, W.F., and Zhaxybayeva, O. (2009). On the origin of prokaryotic species. Genome *19*, 744-756.

Doolittle, W.F. (1999). Phylogenetic classification and the universal tree. Science *284*, 2124-2129.

Duarte, N.C., Herrgard, M.J., and Palsson, B.O. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. Genome Res. *14*, 1298-1309.

Dykhuizen, D.E., and Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. *173*, 7257-7268.

Everett, K.D., Thao, M., Horn, M., Dyszynski, G.E., and Baumann, P. (2005). Novel chlamydiae in whiteflies and scale insects: endosymbionts *'Candidatus* Fritschea bemisiae' strain Falk and *'Candidatus* Fritschea eriococci' strain Elm. Int. J. Syst. Evol. Microbiol. *55*, 1581-1587.

Farahi, K., Pusch, G.D., Overbeek, R., and Whitman, W.B. (2004). Detection of lateral gene transfer events in the prokaryotic tRNA synthetases by the ratios of evolutionary distances method. J. Mol. Evol. *58*, 615-631.

Feil, E., Zhou, J., Smith, J.M., and Spratt, B.G. (1996). A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: Evidence for extensive interspecies recombination within adk. J. Mol. Evol. *43*, 631-640.

Ferreira, K.N., Iverson, T.M., Maghlaoui, K., Barber, J., and Iwata, S. (2004). Architecture of the photosynthetic oxygen-evolving center. Science *303*, 1831-1838.

Ferry, J.G. (1992). Methane from acetate. J. Bacteriol. *174*, 5489-5495.

Fisher, R.A. (1930). Sexual reproduction and sexual selection. In: The Genetical Theory of Natural Selection (Oxford: Clarendon Press), pp. 121-145.

Fitch, W. (2000). Homology a personal view on some of the problems. Trends Genet. *16*, 227-231.

Fournier, G. (2009). Horizontal gene transfer and the evolution of methanogenic pathways. Meth. Mol. Biol. *532*, 163-179.

Fournier, G.P., and Gogarten, J.P. (2008). Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic clostridia. J. Bacteriol. *190*, 1124-1127.

Fournier, G.P., Huang, J., and Gogarten, J.P. (2009). Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. Phil. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2229-2239.

Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., and Hanage, W.P. (2009). The bacterial species challenge: Making sense of genetic and ecological diversity. Science *323*, 741-746.

Galperin, M.Y., and Koonin, E.V. (1997). A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. Protein Sci. *6*, 2639-2643.

Gogarten, J.P. (1995). The early evolution of cellular life. Trends Ecol. Evol. *10*, 147-151.

Gogarten, J.P., and Olendzenski, L. (1999). Orthologs, paralogs and genome comparisons. Curr. Opin. Genet. Dev. *9*, 630-636.

Gogarten, J.P., and Taiz, L. (1992). Evolution of proton pumping ATPases: Rooting the tree of life. Photosynth. Res. *33*, 137-146.

Gogarten, J.P., and Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. Nature Rev. Microbiol. *3*, 679-687.

Gogarten, J., Kibak, H., Dittrich, P., Taiz, L., Bowman, E., Bowman, B., Manolson, M., Poole, R., Date, T., Oshima, T., Konishi, J., Denda, K., and Yoshida, M. (1989a). Evolution of the vacuolar H^+-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA *86*, 6661-6665.

Gogarten, J.P., Rausch, T., Bernasconi, P., Kibak, H., and Taiz, L. (1989b). Molecular evolution of H^+-ATPases. I. *Methanococcus* and *Sulfolobus* are monophyletic with respect to eukaryotes and eubacteria. Z. Naturforsch. *44*, 641-650.

Gogarten, J.P., Starke, T., Kibak, H., Fishman, J., and Taiz, L. (1992). Evolution and isoforms of V-ATPase subunits. J. Exp. Biol. *172*, 137-147.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. *19*, 2226-2238.

Gogarten-Boekels, M., Olendzenski, L., and Gogarten, J.P. (2009). Horizontal Gene Transfer - Genomes in Flux (New York: Humana Press).

Gophna, U., Doolittle, W.F., and Charlebois, R.L. (2005). Weighted genome trees: refinements and applications. J. Bacteriol. *187*, 1305-1316.

Gray, M.W. (1993). Origin and evolution of organelle genomes. Curr. Opin. Genet. Dev. *3*, 884-890.

Gross, M.D., and Siegel, E.C. (1981). Incidence of mutator strains in *Escherichia coli* and coliforms in nature. Mutation Res. *91*, 107-110.

Gruber, G., and Marshansky, V. (2008). New insights into structure-function relationships between archaeal ATP synthase (A_1A_0) and vacuolar type ATPase (V_1V_0). Bioessays *30*, 1096-1109.

Harvey, W.R., and Nelson, N. (1992). V-ATPases. J. Exp. Biol., Vol 172 (Cambridge, U.K.: The Company of Biologists).

Hendrickson, H., and Lawrence, J.G. (2006). Selection for chromosome architecture in bacteria. J. Mol. Evol. *62*, 615-629.

Hennig, W. (1966). Phylogenetic Systematics (Urbana: University of Illinois Press).

Hilario, E., and Gogarten, J.P. (1993). Horizontal transfer of ATPase genes - the tree of life becomes a net of life. Biosystems *31*, 111-119.

Huang, J., and Gogarten, J.P. (2006). Ancient horizontal gene transfer can benefit phylogenetic reconstruction. Trends Genet. *22*, 361-366.

Huang, J., and Gogarten, J.P. (2007). Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol. *8*, R99.

Huang, J., and Gogarten, J.P. (2008). Concerted gene recruitment in early plant evolution. Genome Biol. *9*, R109.

Huang, J., and Gogarten, J.P. (2009). Ancient gene transfer as a tool in phylogenetic reconstruction. Meth. Mol. Biol. *532*, 127-139.

Huang, J., Xu, Y., and Gogarten, J.P. (2005). The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. Mol. Biol. Evol. *22*, 2142-2146.

Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., and Nagashima, K.V. (2001). Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. J. Mol. Evol. *52*, 333-341.

Ihara, K., and Mukohata, Y. (1991). The ATP synthase of *Halobacterium salinarium* (*halobium*) is an archaebacterial type as revealed from the amino acid sequences of its two major subunits. Arch. Biochem. Biophys. *286*, 111-116.

Iida, T., Inatomi, K., Kamagata, Y., and Maruyama, T. (2002). F- and V-type ATPases in the hyperthermophilic bacterium *Thermotoga neapolitana*. Extremophiles *6*, 369-375.

Jain, R., Rivera, M.C., and Lake, J.A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA *96*, 3801-3806.

Jeffreys, A.J., and Neumann, R. (2002). Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nature Genet. *31*, 267-271.

Kibak, H., Taiz, L., Starke, T., Bernasconi, P., and Gogarten, J.P. (1992). Evolution of structure and function of V-ATPases. J. Bioenerg. Biomembr. *24*, 415-424.

Kimura, M. (1967). On the evolutionary adjustment of spontaneous mutation rates. Genet. Res. *9*, 23-24.

Koizumi, Y., Takii, S., Nishino, M., and Nakajima, T. (2003). Vertical distributions of sulfate-reducing bacteria and methane-producing archaea quantified by oligonucleotide probe hybridization in the profundal sediment of a mesotrophic lake. FEMS Microbiol. Ecol. *44*, 101-108.

Kurland, C.G., Collins, L.J., and Penny, D. (2006). Genomics and the irreducible nature of eukaryote cells. Science *312*, 1011-1014.

Kuusinen, M. (1996). Cyanobacterial macrolichens on *Populus tremula* as indicators of forest continuity in Finland. Biol. Conserv. *75*, 43-49.

Lapierre, P. (2007). The Impact of Horizontal Gene Transfers on Prokaryotic Ggenome Evolution. PhD Thesis (Storrs: University of Connecticut).

Lapierre, P., and Gogarten, J.P. (2009). Estimating the size of the bacterial pan-genome. Trends Genet. *25*, 107-110.

Lapierre, P., Shial, R., and Gogarten, J.P. (2006). Distribution of F- and A/V-type ATPases in *Thermus scotoductus* and other closely related species. Syst. Appl. Microbiol. *29*, 15-23.

Lawrence, J.G., and Hendrickson, H. (2005). Genome evolution in bacteria: order beneath chaos. Curr. Opin. Microbiol. *8*, 572-578.

Lawrence, J.G., and Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. *44*, 383-397.

Lawrence, J.G., and Ochman, H. (2002). Reconciling the many faces of lateral gene transfer. Trends Microbiol. *10*, 1-4.

Lawrence, J.G., and Roth, J.R. (1996). Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics *143*, 1843-1860.

Lesser, M.P., Mazel, C.H., Gorbunov, M.Y., and Falkowski, P.G. (2004). Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. Science *305*, 997-1000.

Levins, R. (1967). Theory of fitness in a heterogenous environment. VI. The adaptive significance of mutation. Genetics *56*, 163-178.

Liebl, U., Mockensturm-Wilson, M., Trost, J.T., Brune, D.C., Blankenship, R.E., and Vermaas, W. (1993). Single core polypeptide in the reaction center of the photosynthetic bacterium *Heliobacillus mobilis*: Structural implications and relations to other photosystems. Proc. Natl. Acad. Sci .USA *90*, 7124-7128.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc. Natl. Acad. Sci. USA *101*, 11013-11018.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M., and Schleifer, K.H. (1998). Bacterial phylogeny based on comparative sequence analysis. Electrophoresis *19*, 554-568.

Majewski, J., and Cohan, F.M. (1999). DNA sequence similarity requirements for interspecific recombination in *Bacillus*. Genetics *153*, 1525-1533.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003). Bacterial photosynthesis genes in a virus. Nature *424*, 741.

Martin, W., Dagan, T., Koonin, E.V., Dipippo, J.L., Gogarten, J.P., and Lake, J.A. (2007). The evolution of eukaryotes. Science *316*, 542-543; author reply: Science *316*, 542-543.

Mayr, E. (1942). Systematics and the Origin of Species (New York: Columbia University Press).

Mayr, E. (1963). Animal species and evolution (Cambridge, MA: Harvard University Press).

Morandi, A., Zhaxybayeva, O., Gogarten, J.P., and Graf, J. (2005). Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. J. Bacteriol. *187*, 6561-6564.

Moustafa, A., Reyes-Prieto, A., and Bhattacharya, D. (2008). Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS ONE *3*, e2205.

Mulkidjanian, A.Y., Koonin, E.V., Makarova, K.S., Mekhedov, S.L., Sorokin, A., Wolf, Y.I., Dufresne, A., Partensky, F., Burd, H., Kaznadzey, D., Haselkorn, R., and Galperin, M.Y. (2006). The cyanobacterial genome core and the origin of photosynthesis. Proc. Natl. Acad. Sci. USA *103*, 13126-13131.

Mulkidjanian, A.Y., Makarova, K.S., Galperin, M.Y., and Koonin, E.V. (2007). Inventing the dynamo machine: the evolution of the F-type and V-type ATPases. Nature Rev. Microbiol. *5*, 892-899.

Mulkidjanian, A.Y., Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2008). Evolutionary primacy of sodium bioenergetics. Biology Direct *3*, 13.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., and Fraser, C.M. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature *399*, 323-329.

Nesbø, C.L., L'Haridon, S., Stetter, K.O., and Doolittle, W.F. (2001). Phylogenetic analyses of two "archaeal" genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. Mol. Biol. Evol. *18*, 362-375.

Noll, K.M., and Thirangoon, K. (2009). Interdomain transfers of sugar transporters overcome barriers to gene expression. Meth. Mol. Biol. *532*, 309-322.

Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D., Valverde, C., Wall, L.G., Wang, Y., Medigue, C., and Benson, D.R. (2007). Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome Res. *17*, 7-15.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299-304.

Olendzenski, L., Liu, L., Zhaxybayeva, O., Murphey, R., Shin, D.G., and Gogarten, J.P. (2000). Horizontal transfer of archaeal genes into the Deinococcaceae: Detection by molecular and computer-based approaches. J. Mol. Evol. *51*, 587-599.

Olendzenski, L., Zhaxybayeva, O., and Gogarten, J. (2001). What's in a tree? Does horizontal gene transfer determine microbial taxonomy? In: Cellular Origin and Life in Extreme Habitats, J. Seckbach, ed. (Dordrecht: Kluwer Academic Publishers), pp. 67-78.

Oren, A. (2004). A proposal for further integration of the cyanobacteria under the bacteriological code. Int. J. Syst. Evol. Microbiol. *54*, 1895-1902.

Pal, C., Papp, B., and Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nature Genet. *37*, 1372-1375.

Papke, R.T. (2009). A critique of prokaryotic species concepts. Meth. Mol. Biol. *532*, 379-395.

Papke, R.T., Ramsing, N., Bateson, M., and Ward, D.M. (2003). Geographical isolation in hot spring cyanobacteria. Environ. Microbiol. *5*, 650-659.

Papke, R.T., Zhaxybayeva, O., Feil, E.J., Sommerfeld, K., Muise, D., and Doolittle, W.F. (2007). Searching for species in haloarchaea. Proc. Natl. Acad. Sci. USA *104*, 14092-14097.

Penny, D., and Poole, A. (1999). The nature of the last universal common ancestor. Curr. Opin. Genet. Dev. *9*, 672-677.

Poptsova, M.S., and Gogarten, J.P. (2007). The power of phylogenetic approaches to detect horizontally transferred genes. BMC Evol. Biol. *7*, 45.

Prangishvili, D., Albers, S.V., Holz, I., Arnold, H.P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A., Kristjansson, J.K., and Zillig, W. (1998). Conjugation in Archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. Plasmid *40*, 190-202.

Puigbo, P., Wolf, Y.I., and Koonin, E.V. (2009). Search for a 'Tree of Life' in the thicket of the phylogenetic forest. J. Biol. *8*, 59.

Ragan, M.A. (2001). On surrogate methods for detecting lateral gene transfer. FEMS Microbiol. Lett. *201*, 187-191.

Raymond, J. (2008). Coloring in the tree of life. Trends Microbiol. *16*, 41-43.

Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y., and Blankenship, R.E. (2002). Whole-genome analysis of photosynthetic prokaryotes. Science *298*, 1616-1620.

Raymond, J., Zhaxybayeva, O., Gogarten, J.P., and Blankenship, R.E. (2003). Evolution of photosynthetic prokaryotes: a maximum-likelihood mapping approach. Phil. Trans. R. Soc. Lond. B Biol. Sci. *358*, 223-230.

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. *4*, R54

Retchless, A.C., and Lawrence, J.G. (2007). Temporal fragmentation of speciation in bacteria. Science *317*, 1093-1096.

Richardson, J.P. (2002). Rho-dependent termination and ATPases in transcript termination. Biochim. Biophys. Acta *1577*, 251-260.

Sadekar, S., Raymond, J., and Blankenship, R.E. (2006). Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. Mol. Biol. Evol. *23*, 2001-2007.

Schleper, C., Roder, R., Singer, T., and Zillig, W. (1994). An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous galactosidase gene. Mol. Gen. Genet. *243*, 91-96.

Schouls, L.M., Schot, C.S., and Jacobs, J.A. (2003). Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. J. Bacteriol. *185*, 7241-7246.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R.Y., Partensky, F., Koonin, E.V., Wolf, Y.I., Nelson, N., and Béjà, O. (2009). Photosystem I gene cassettes are present in marine virus genomes. Nature *461*, 258-262.

Sheppard, S.K., McCarthy, N.D., Falush, D., and Maiden, M.C. (2008). Convergence of *Campylobacter* species: implications for bacterial evolution. Science *320*, 237-239.

Shi, T., Bibby, T.S., Jiang, L., Irwin, A.J., and Falkowski, P.G. (2005). Protein interactions limit the rate of evolution of photosynthetic genes in cyanobacteria. Mol. Biol. Evol. *22*, 2179-2189.

Shibata, C., Ehara, T., Tomura, K., Igarashi, K., and Kobayashi, H. (1992). Gene structure of *Enterococcus hirae* (*Streptococcus faecalis*) F_1F_0-ATPase, which functions as a regulator of cytoplasmic pH. J. Bacteriol. *174*, 6117-6124.

Sørensen, K., Canfield, D., Teske, A., and Oren, A. (2005). Community composition of a hypersaline endoevaporitic microbial mat. Appl. Environ. Microbiol. *71*, 7352-7365.

Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P., and Rubin, E.M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. Science *318*, 1449-1452.

Stams, A.J. (1994). Metabolic interactions between anaerobic bacteria in methanogenic environments. Antonie van Leeuwenhoek *66*, 271-294.

Stern, A., Mayrose, I., Penn, O., Shaul, S., Gophna, U., and Pupko, T. (In press). An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. Syst. Biol.

Suerbaum, S., Maynard Smith, J., Bapumia, K., Morelli, G., Smith, N.H., Kunstmann, E., Dyrek, I., and Achtman, M. (1998). Free recombination within *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA *95*, 12619-12624.

Sumi, M., Yohda, M., Koga, Y., and Yoshida, M. (1997). F_0F_1-ATPase genes from an archaebacterium, *Methanosarcina barkeri*. Biochem. Biophys. Res. Commun. *241*, 427-433.

Swithers, K.S., Gogarten, J.P., and Fournier, G.P. (2009). Trees in the web of life. J. Biol. *8*, 54.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P.H., and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. Nature *387*, 700-702.

Takase, K., Yamato, I., and Kakinuma, Y. (1993). Cloning and sequencing of the genes coding for the A and B subunits of vacuolar-type Na^+-ATPase from *Enterococcus hirae*. Coexistence of vacuolar- and F_0F_1-type ATPases in one bacterial cell. J. Biol. Chem. *268*, 11610-11616.

Takase, K., Kakinuma, S., Yamato, I., Konishi, K., Igarashi, K., and Kakinuma, Y. (1994). Sequencing and characterization of the *ntp* gene cluster for vacuolar-type Na^+-translocating ATPase of *Enterococcus hirae*. J. Biol. Chem. *269*, 11037-11044.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., and Durkin, A.S. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. USA *102*, 13950-13955.

Tsutsumi, S., Denda, K., Yokoyama, K., Oshima, T., Date, T., and Yoshida, M. (1991). Molecular cloning of genes encoding major two subunits of a eubacterial V-Type ATPase from *Thermus thermophilus*. Biochim. Biophys. Acta *1098*, 13-20.

Tyra, H.M., Linka, M., Weber, A.P., and Bhattacharya, D. (2007). Host origin of plastid solute transporters in the first photosynthetic eukaryotes. Genome Biol. *8*, R212.

van Berkum, P., Terefework, Z., Paulin, L., Suomalainen, S., Lindstrom, K., and Eardly, B.D. (2003). Discordant phylogenies within the rrn loci of rhizobia. J. Bacteriol. *185*, 2988-2998.

Vogler, A.P., Homma, M., Irikura, V.M., and Macnab, R.M. (1991). *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F_0F_1, vacuolar, and archaebacterial ATPase subunits. J. Bacteriol. *173*, 3564-3572.

Wang, Y., and Zhang, Z. (2000). Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology *146*, 2845-2854.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature *277*, 293-294.

Welch, R.A., Burland, V., Plunkett, G., 3rd, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., and Blattner, F.R. (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA *99*, 17020-17024.

Wilkinson, M., McInerney, J.O., Hirt, R.P., Foster, P.G., and Embley, T.M. (2007). Of clades and clans: terms for phylogenetic relationships in unrooted trees. Trends Ecol. Evol . *22*, 114-115.

Woese, C.R. (1987). Bacterial evolution. Microbiol. Rev. *51*, 221-271.

Woese, C.R., Olsen, G.J., Ibba, M., and Söll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. *64*, 202-236.

Wolf, Y.I., Aravind, L., Grishin, N.V., and Koonin, E.V. (1999). Evolution of aminoacyl-tRNA synthetases - analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Res. *9*, 689-710.

Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V. (2001). Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res. *11*, 356-372.

Wolf, Y.I., Rogozin, I.B., Grishin, N.V., and Koonin, E.V. (2002). Genome trees and the tree of life. Trends Genet. *18*, 472-479.

Xiong, J., and Bauer, C.E. (2002). Complex evolution of photosynthesis. Annu. Rev. Plant. Biol. *53*, 503-521.

Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M., and Bauer, C.E. (2000). Molecular evidence for the early evolution of photosynthesis. Science *289*, 1724-1730.

Yap, W.H., Zhang, Z., and Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. J. Bacteriol. *181*, 5201-5209.

Yokoyama, K., Muneyuki, E., Amano, T., Mizutani, S., Yoshida, M., Ishida, M., and Ohkuma, S. (1998). V-ATPase of *Thermus thermophilus* is inactivated during ATP hydrolysis but can synthesize ATP. J. Biol. Chem. *273*, 20504-20510.

Zhaxybayeva, O. (2009). Detection and quantitative assessment of horizontal gene transfer. Meth. Mol. Biol. *532*, 195-213.

Zhaxybayeva, O., and Gogarten, J.P. (2007). Horizontal gene transfer, gene histories and the root of the tree of life. In: Astrobiology and the Origins of Life, R.E. Pudritz, P.G. Higgs, and J. Stone, eds. (Cambridge, UK: Cambridge University Press).

Zhaxybayeva, O., Hamel, L., Raymond, J., and Gogarten, J. (2004). Visualization of the phylogenetic content of five genomes using dekapentagonal maps. Genome Biol. *5*, R20.

Zhaxybayeva, O., Lapierre, P., and Gogarten, J.P. (2005). Ancient gene duplications and the root(s) of the tree of life. Protoplasma *227*, 53-64.

Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., and Papke, R.T. (2006). Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. Genome Res. *16*, 1099-1108.

Zhaxybayeva, O., Nesbø, C., and Doolittle, W.F. (2007). Systematic overestimation of gene gain through false diagnosis of gene absence. Genome Biol. *8*, 402.

Zhaxybayeva, O., Doolittle, W.F., Papke, R.T., and Gogarten, J.P. (2009a). Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. Genome Biol. Evol. *1*, 325–339.

Zhaxybayeva, O., Swithers, K., Lapierre, P., Fournier, G., Bickhart, D., DeBoy, R., Nelson, K., Nesbo, C., Doolittle, W., Gogarten, J., and Noll, K.M. (2009b). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. USA *106*, 5865-5870.

Zhou, J., Bowler, L.D., and Spratt, B.G. (1997). Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. Mol. Microbiol. *23*, 799-812.

Zwart, G., Crump, B.C., Kamst-van Agterveld, M.P., Hagen, F., and Han, S.K. (2002). Typical freshwater bacteria: An analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat. Microb. Ecol. *28*, 141-155.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., Not, F., Massana, R., Ulloa, O., and Scanlan, D.J. (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environ. Microbiol. *10*, 147-161.