# Biased gene transfer in microbial evolution

*Cheryl P. Andam and J. Peter Gogarten*

Abstract | Horizontal gene transfer (HGT) is an important evolutionary process that allows the spread of innovations between distantly related organisms. We present evidence that prokaryotes (bacteria and archaea) are more likely to transfer genetic material with their close relatives than with distantly related lineages. This bias in transfer partners can create phylogenetic signals that are difficult to distinguish from the signal created through shared ancestry. Preferences for transfer partners can be revealed by studying the distribution patterns of divergent genes with identical functions. In many respects, these genes are similar to alleles in a population, except that they coexist only in higher taxonomic groupings and are acquired by a species through HGT. We also discuss the role of biased gene transfer in the formation of taxonomically recognizable natural groups in the tree or net of life.

Phylogenetics
The study of the evolutionary (or natural) relationships of organisms as they change through time. Phylogenies can be strictly furcating (often bifurcating) or can include reticulations.

Genetic material can be transferred between two organisms that do not share an ancestor–descendant relationship in a process called horizontal gene transfer (HGT). This process is a mechanism of evolutionary change, particularly in microorganisms, because it makes a considerable contribution to the rapid creation and spread of biological innovation in many lineages that otherwise would have taken millions of years to proceed[1]. The size of the genetic material that can be moved horizontally ranges from small gene fragments[2–4] to entire operons[5–7] and superoperons that encode complex biochemical pathways[8], and in some cases even to whole chromosomes[9]. The size of the transferred material does not seem to be a barrier to successful transfer, as demonstrated by a recent study in which plasmids up to 100 kb in size released from lysed *Escherichia coli* could be taken up by *Bacillus subtilis*[10]. HGT between highly divergent organisms has serious implications for phylogenetics aimed at reconstructing microbial evolution because it creates convoluted relationships; that is, diverse ancestors can substantially contribute to the genetic repertoire of a lineage.

Gene transfers may occur between distantly related organisms or between close relatives. In the case of horizontal acquisitions from distant relatives, the gene of the recipient taxon would exhibit high similarity to that of the donor taxon, despite the evolutionary distance that separates them. Examples of genes that have been transferred between distantly related organisms are listed in BOX 1. Highways of gene sharing have been identified, for example, between hyperthermophilic bacteria and archaea[11–16], most probably as a result of their common niches. Other cases of HGT that were facilitated by a common environment have been reported[17–19]. Even the highly conserved 'informational genes' (those that are involved in transcription and translation processes) can be transferred through HGT[2,20–25]. This implies that genes in organisms that are genealogically distant can nonetheless be similar and, hence, species can appear to be related even though they did not arise from a recent organismal common ancestor. Other examples of HGT between phylogenetically distant taxa include the genes that encode proteins involved in plant cell wall degradation in plant-parasitic nematodes; these genes originated from different bacterial sources[26]. All of the enzymes involved in the novel methylaspartate cycle for acetyl-CoA assimilation in the members of the order Halobacteriales were acquired through the HGT and subsequent recombination of different genes from various bacterial genomes[27]. This cycle has been referred to as a metabolic patchwork because all the genes were originally involved in various metabolic processes, such as glutamate fermentation and propionate assimilation[27].

Chimeric genomes can trace their ancestry from a myriad of sources, both living and extinct. Thus, relatedness in microbial taxonomy can be a consequence of horizontal acquisition and not of shared ancestry, a fact that blurs the concepts of lineage and genealogy in the microbial world. Defining a lineage must then incorporate this kind of relationship, and therefore a re-evaluation of the definition of a microbial lineage and genealogy is necessary to reflect the genetic relatedness

*Department of Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Road, Storrs, Connecticut 06269-3125, USA. Correspondence to J.P.G. e-mail: jpgogarten@gmail.com*
doi:10.1038/nrmicro2593
Published online 13 June 2011

**Box 1 | Examples of horizontal gene transfer events in evolution**

- The threonyl-tRNA synthetase of the cyanobacterium genus *Prochlorococcus* is most closely related to that of marine gammaproteobacteria[95].
- The tyrosyl-tRNA synthetase of opisthokonts originated from within the haloarchaea[86].
- The ATP synthases of the genera *Thermus* and *Enterococcus* cluster with those of the archaeal genera *Methanosarcina* and *Halobacterium*[96].
- Certain genes that are found in red and green algae and in plants are derived from *Chlamydia* spp.[50,97–99].
- A gene that encodes *N*-acetylneuraminate lyase in the genus *Trichomonas* is of gammaproteobacterial origin[100].
- Two genes that are required for acetoclastic methanogenesis in the archaeal family Methanosarcinaceae are derived from cellulolytic bacteria of the class Clostridia[52].
- Many genes in the order Aquificales: ribosomal RNA phylogenies place Aquificales as a deep-branching lineage, but the phylogenies of several informational and most operational genes place Aquificales with the class Epsilonproteobacteria[15].
- The majority of the genes in the order Thermotogales, a group that is composed mostly of thermophilic bacteria: most genes group with the Gram-positive, sulphite-reducing class Clostridia (which is part of the phylum Firmicutes), but the ribosomal components group at the base of the bacterial domain, and 10% of genes group within the Archaea[16].

that is due to HGT. As HGT causes a topological discrepancy between the gene tree and the organismal (or species) tree[28] — which is often represented by phylogenies that are based on ribosomal RNA, consensus or central tendencies — a tree-like depiction alone is not the most accurate representation of evolution.

In this Analysis article, we present evidence that HGT occurs mainly as biased gene transfer, occurring frequently among closely related individuals and species and rarely between distant relatives. Hence, instead of introducing noise in phylogenetic reconstruction, this bias can reassemble or at the very least reinforce the patterns of evolution that are generated through shared ancestry. We provide examples from phylogenetic analyses of genes encoding three different aminoacyl-tRNA synthetases (aaRSs)[29,30]. We also argue that biased gene transfer does not erode the definition of taxonomic groups and instead may reinforce the observed relationships.

## Preferences for transfer partners

Closely related partners in a group will probably preferentially exchange genes with each other, which may eventually lead to cohesion of the group. A recent study using *Legionella pneumophila* reported that the likelihood of HGT events is a function of the phylogenetic proximity of the partners, as well as their shared ecological niche[31]. Gene exchange between close relatives is greatly enhanced as a consequence of their similar genomic architecture[32] and genetic machineries[22,33] (including their transcription and translation features such as promoters and regulatory sequences), which increases the chance of successful integration and expression of novel genes.

Genes shuttle between cellular chromosomes and mobile elements. Plasmids and viruses that are evolving with their host can bias HGT between closely related species, as was recently shown for archaeal minichromosome maintenance (MCM) helicase genes[34]. Indeed, the biology of HGT mechanisms — transformation, conjugation and transduction (reviewed in REF. 35) — supports this bias, which is determined largely by the genome similarity of the donor and recipient. However, a recent study of HGT agents derived from alphaproteobacterial prophages found that these agents can transfer genes to a surprisingly diverse set of recipients[36]. Another recent study provides quantitative support for biased gene transfer. Using network analyses of 657 prokaryotic (bacterial and archaeal) genomes, the frequency of HGT was shown to be linearly correlated with similarity between the donor and recipient in both genome and proteome sequences, with a high percentage of HGT events (86%) occurring between pairs that possess less than 5% difference in GC content[37]. Another network analysis revealed that 91% of transposase genes that were acquired through HGT are shared between pairs of microbial taxa that are members of the same phylum[38].

As a result of the bias to transfer genes preferentially between more similar organisms, a member of a lineage can use not only the genes in its genomes, but also other genetic material that it can acquire from its relatives and from the mobile gene pool (or mobilome), which includes various kinds of mobile genetic elements such as transposons, plasmids, phages and self-splicing molecular parasites[39–43]. Consequently, the genetic repertoire of a species or higher taxonomic group — also known as the pan-genome — comprises all the genes that are present in the genomes of members of the species or group, including those genes that are found in one or only a few strains[44,45], and so forms a significant repository of genetic material that can be shared among organisms, with genes moving into and out of the genome. The composition and size of the pan-genome introduces an incredible range of variability and adaptability to species.

Transfer of genes across higher-level taxonomic boundaries (such as between domains or phyla) is less likely than HGT at the species, genus or family level. The rate of successful transfer thus relates to the overall genetic similarity of the two partners[37,38,46,47]. Therefore, we expect to observe similar distribution patterns between a specific gene tree and the ribosomal tree (on which most phylogenies are based) instead of incongruent or conflicting tree topologies, as this preference for exchange partners leads to the recovery of an individual-gene phylogenetic signal that mimics the signal which is created by shared ancestry[48]. It was reported that, given the exploration of parameter space that was carried out in one study, biased HGT alone cannot explain the deep split between the Archaea and the Bacteria[46]; nonetheless, the simulations of HGT and vertical inheritance that were carried out in this study do confirm the possibility that biased HGT can create patterns that are similar to vertical inheritance.

Other sources for HGT bias are a shared ecological niche and symbiotic interactions. Transfer from a bacterium to the genome of a multicellular eukaryote is often a consequence of a close symbiotic relationship[49,50], endosymbiotic gene transfer providing a well-documented example[51]. Symbiotic associations might

Biased gene transfer
Horizontal gene transfer between preferred partners (usually close relatives) rather than random transfer between any species. Other factors, such as shared ecological niches or symbiotic relationships, can also create a bias in transfer partners.

Aminoacyl-tRNA synthetases
(aaRSs). A family of enzymes that are responsible for the specific attachment of each amino acid to its cognate tRNA during the translation process.

also play a part in many transfers between bacteria[52]. The resulting highways of gene sharing[13] between divergent organisms can dominate the phylogenetic signal that is retained in genomes, as was suggested for the genomes of the extreme thermophilic bacteria[15,16] (see below for a discussion).

## Homeoalleles in prokaryotes

In higher taxonomic groups, prokaryotes can exhibit features in a similar way to populations of a single species. Genes are swapped through HGT similarly to the interbreeding or hybridization of members of a single-species population. In a population, alleles (variant forms of a gene) may be distributed among the members at varying frequencies, and an individual may possess one or the other copy of the gene, or both copies in the case of diploids, depending on the genetic make-up of its parents. In the same way, allele-like copies of genes, called homeoalleles[48], can be exchanged among individuals that are related at higher taxonomic levels. Homeoalleles encode enzymes with identical functions but with dissimilar characteristics that have arisen as a result of gene divergences[48]. Individual lineages can acquire different homeoalleles through HGT, and these homeoalleles can also be lost from the lineage and replaced by other homeoalleles. Moreover, a lineage may possess two homeoalleles at a locus, similarly to heterozygous individuals possessing two gene alleles in a single-species population. Although a single genome usually carries only one homeoallele at a particular locus, both forms may be present within one phylum owing to HGT[48]. The outcome of this gene exchange is that higher-order taxonomic units function as exchange groups that donate and receive genetic material within the group[48,53] in the same way as populations of recombining individuals. The characterized homeoalleles are homologues that arose through ancient divergence. Isofunctional enzymes, the homology of which cannot be established[54], may also function as homeoalleles; however, further investigation of the phylogeny of these enzymes is necessary to test this hypothesis.

***TyrRS phylogeny: mimicking patterns of vertical inheritance.*** The rapid and extensive generation of molecular data in recent years has allowed more genomes to be compared for phylogenetic analyses. The aaRSs link activated amino acids with their cognate tRNAs, and in bacteria two versions of tyrosyl-tRNA synthetases (TyrRSs; encoded by *tyrS* genes) exist[29,48,55]. Each version of TyrRS forms a well-supported clade, with several phyla and/or classes represented in both clades (FIG. 1; see Supplementary information S1 (figure)). Here, we use a larger bacterial data set than that used by Andam and colleagues[48] and compare it with a combined 16S–23S rRNA tree for a more robust reference. Indeed, when observed independently, each of these clades exhibits a phylogenetic pattern that mimics the patterns which are created through shared ancestry, using the combined 16S–23S rRNA phylogeny as a substitute for the species tree. When a gene is acquired through HGT but a divergent homologue with the same function already exists in the recipient, two pathways for gene replacement are possible. The first, and perhaps more common, pathway is the addition of the novel gene through insertion into another part of the chromosome. Following a period of coexistence, one of the two homologues may eventually be lost. The second pathway through which a divergent homologue can be integrated into a genome is through homologous recombination of more-conserved neighbouring genes in a syntenic context, leading to the immediate replacement of the iso-functional homologue (see FIG. 2). Homologous recombination between the two *tyrS* gene types is not likely to occur because they have less than 55% sequence identity[48]. However, *tyrS* (regardless of type) is present in the same gene neighbourhood in many species, indicating that in various species one *tyrS* homeoallele was replaced with another through homologous recombination in their neighbouring genes. In many bacteria, including several gammaproteobacteria[48], the region around the *tyrS* gene is syntenic. For example, in two members of the class Deltaproteobacteria, the genes that surround *tyrS* types A and B are similar (FIG. 2), including the rRNA genes to the right (which have a similarity ranging from 95% to 98%) and considerably similar genes to the left (68–74% identity), both of which regions could act as possible recombination points.

To assess the agreement in phylogeny for each TyrRS type with the rRNA genes, we plotted the pairwise distances in TyrRS sequence against their corresponding distances in 16S–23S rRNA for each pair of taxa (see Supplementary information S2 (figure)). Each TyrRS clade exhibits a phylogenetic pattern that is similar to the rRNA phylogeny, as represented by the strong correlation in the distances between each pair of organisms, indicating that there is a similarity between the phylogenies of each TyrRS clade and the ribosomal tree (see Supplementary information S2 (figure)). We show that positive correlation exists for each clade despite the occurrence of HGT and that there is no obvious trace of HGT that can be detected by phylogenetic conflict within each clade, because the transfers occur between taxa with close phylogenetic affinity. Within each TyrRS type, the distances between pairs of TyrRS homologues and the corresponding rRNA gene pairs show a strong correlation (see Supplementary information S2 (figure); $R^2 = 0.92$ and $P < 2.2 \times 10^{-16}$ for type A TyrRSs (plotted in green), and $R^2 = 0.72$ and $P < 2.2 \times 10^{-16}$ for type B TyrRSs (plotted in red); $R^2$ is the square of the correlation coefficient). These results indicate that within each of the two TyrRS clades, evolution is similar to the expected rRNA phylogeny, but if we consider the TyrRS phylogeny as a whole, we observe a conflicting signal. The distances between the two TyrRS types do not correlate to the corresponding 16S–23S rRNA distances (see Supplementary information S2 (figure); plotted in blue), but rather reflect the ancient divergence between the two TyrRS types.

A likelihood-based evolutionary simulation for inferring HGT events was performed using the LGT3State[56] program to estimate the transition probabilities among the three possible TyrRS states under two

Homeoallele
One of several divergent but functionally identical genes that are swapped within an exchange group (a group of organisms that has a higher rate for within-group gene transfers than for between-group transfers) which contains organisms belonging to different higher-level taxa.

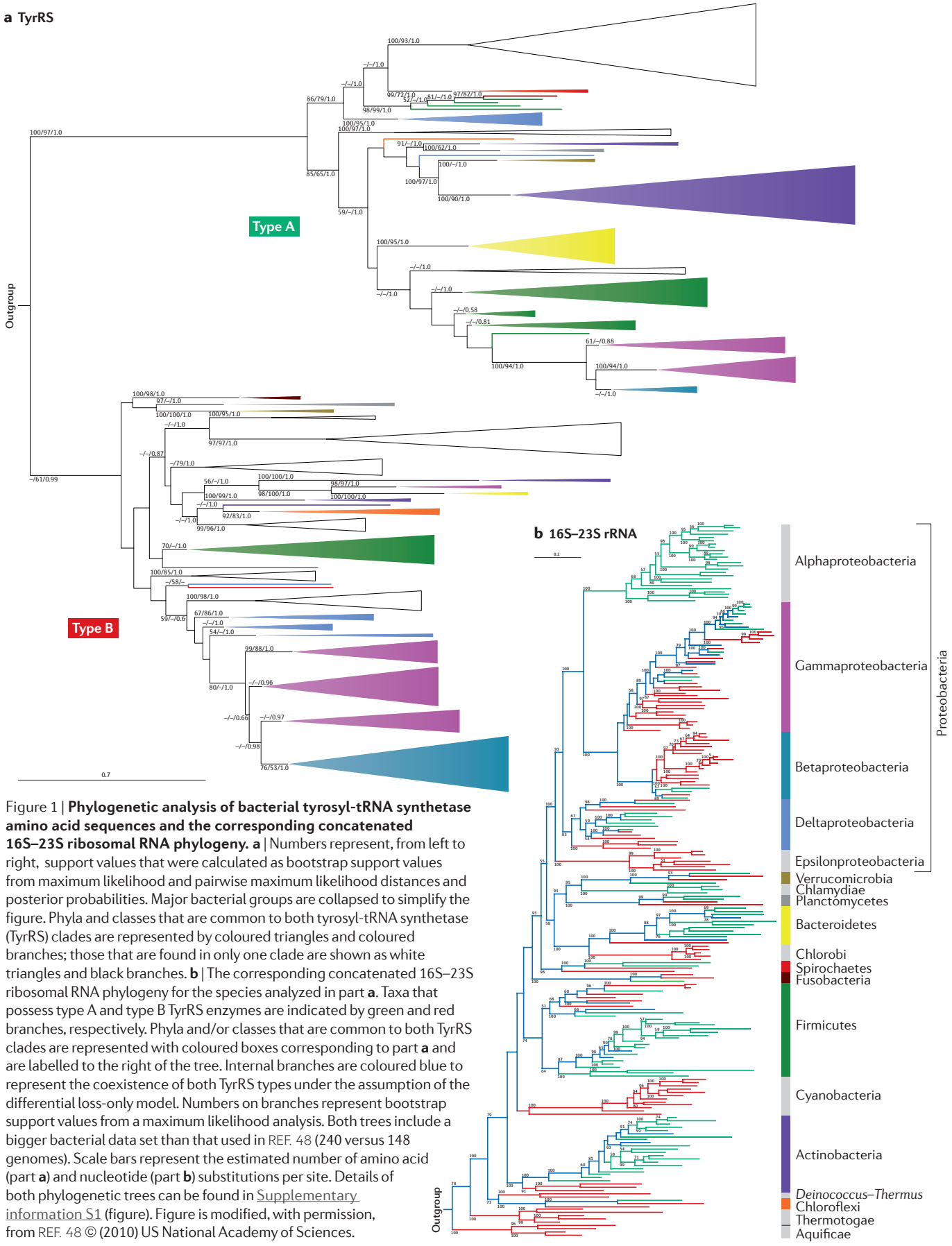



Figure 1 | **Phylogenetic analysis of bacterial tyrosyl-tRNA synthetase amino acid sequences and the corresponding concatenated 16S–23S ribosomal RNA phylogeny. a** | Numbers represent, from left to right, support values that were calculated as bootstrap support values from maximum likelihood and pairwise maximum likelihood distances and posterior probabilities. Major bacterial groups are collapsed to simplify the figure. Phyla and classes that are common to both tyrosyl-tRNA synthetase (TyrRS) clades are represented by coloured triangles and coloured branches; those that are found in only one clade are shown as white triangles and black branches. **b** | The corresponding concatenated 16S–23S ribosomal RNA phylogeny for the species analyzed in part **a**. Taxa that possess type A and type B TyrRS enzymes are indicated by green and red branches, respectively. Phyla and/or classes that are common to both TyrRS clades are represented with coloured boxes corresponding to part **a** and are labelled to the right of the tree. Internal branches are coloured blue to represent the coexistence of both TyrRS types under the assumption of the differential loss-only model. Numbers on branches represent bootstrap support values from a maximum likelihood analysis. Both trees include a bigger bacterial data set than that used in REF. 48 (240 versus 148 genomes). Scale bars represent the estimated number of amino acid (part **a**) and nucleotide (part **b**) substitutions per site. Details of both phylogenetic trees can be found in Supplementary information S1 (figure). Figure is modified, with permission, from REF. 48 © (2010) US National Academy of Sciences.

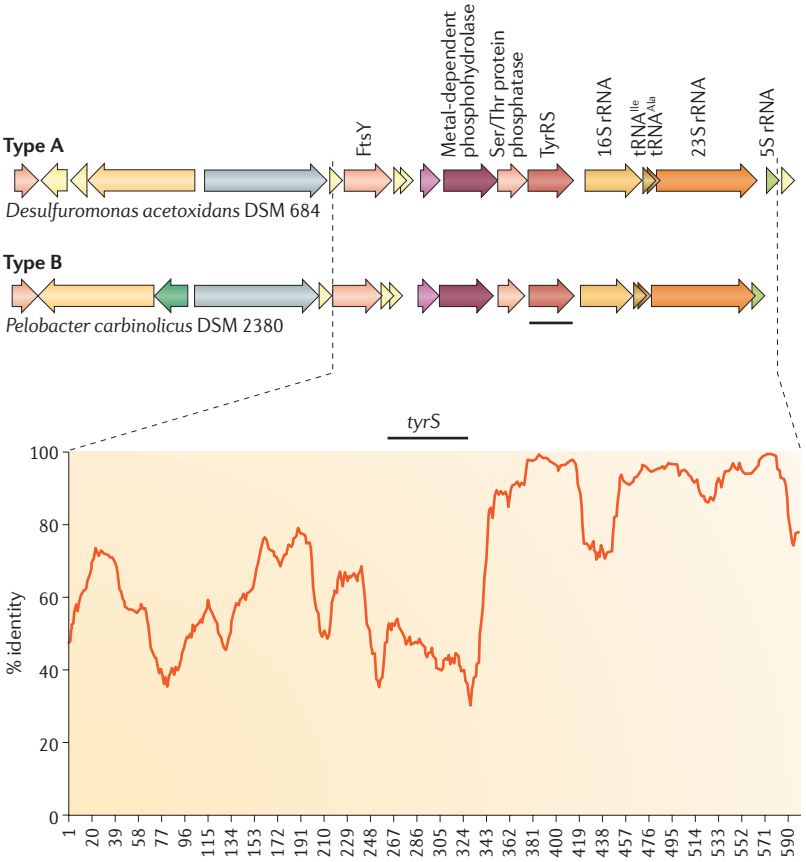



Figure 2 | **Gene neighbourhood of the two tyrosyl-tRNA synthetases in two members of the Deltaproteobacteria.** The two tyrosyl-tRNA synthetases (TyrRSs; encoded by *tyrS* genes) belong to different types. Genes with the same function are given in the same colour, with hypothetical genes in light yellow, and the encoded proteins are annotated above. A sliding-window analysis (with a window size of 300 nucleotides and a step size of 20 nucleotides) of the aligned DNA sequence shows regions of high identity that constitute possible recombination points flanking the *tyrS* genes. rRNA, ribosomal RNA.

alternative models. The null model assumes that only gene loss can explain the distribution pattern observed for TyrRS proteins and that the ancestral state is characterized by the presence of both TyrRS types, whereas the alternative model assumes that HGT occurs. The program then calculates the maximum log-likelihood values for each scenario. Results indicate that both gene loss and gene gain through HGT, and not gene loss alone, best explain the TyrRS distribution in the Bacteria[48]. Simulations of unbiased transfers (transfers in which the phylogenetic relatedness of the organisms is not taken into account) led to a decrease in the correlation with the 16S rRNA distances, indicating that such transfers erode the phylogenetic signal in the original data set; this erosion was not detected in biased transfers[48]. Additional lines of evidence for HGT within larger groups (such as the phylum or class) are the fact that topological conflicts occur at taxonomic units below the phylum or class levels, while the phylum or class as a whole remains cohesive, and the fact that inter-phylum transfers of *tyrS* were rarely detected by strong phylogenetic conflict[48].

HGT has been generally considered as noise in phylogenetic analyses, as it results in conflicting relationships between gene (or protein) and species phylogenies. However, biased HGT maintains and perhaps creates phylogenetic patterns that are indistinguishable from those generated through vertical descent. In the case of homeoalleles, transfers can be inferred through the switching between homeoallele types, rather than through significantly supported phylogenetic conflict within one type. The phylogenetic detection of HGT events is based on statistically well-supported conflicts between the gene and reference phylogenies[57]. Often, the rRNA phylogeny, or an average phylogenetic signal extracted from the analysed genomes, is used as a reference. Transfer between close relatives usually does not result in significant conflicts, whereas transfer between different phyla or domains results in notable differences with the reference phylogeny. It is likely that biased HGT is not restricted to replacing homeoallele types; however, transfers that lead to the replacement of genes with similar homologues are not easily detectable through phylogenetic approaches.

***Multiple homeoalleles in TrpRS phylogeny.*** In some cases, homeoalleles exist in more than two types, in the same way as there can be multiple alleles for a single locus (such as the alleles for the ABO blood types). For example, three divergent versions of the gene encoding tryptophanyl-tRNA synthetase (TrpRS) are present in the domain Bacteria (FIG. 3; see Supplementary information S3 (figure)). Similar to what was found for TyrRS, several bacterial phyla or classes are represented in each clade of the TrpRS tree (labelled types A, B and C; FIG. 3a and Supplementary information S2 (figure)), and within each TrpRS clade there is a strong correlation between the pairwise rRNA and TrpRS distances between pairs of taxa (see Supplementary information S2 (figure)), indicating that evolution within each TrpRS clade seems to be similar to the expected rRNA phylogeny (FIG. 3). In some cases, we observe a number of pairwise distances that depart from the expected trend, as in the case of TrpRS type A (see Supplementary information S2 (figure), the green dots located above the correlation line). This can be interpreted as phylogenetic evidence for HGT of TrpRS type A-encoding genes between divergent organisms.

A small number of genomes encode two versions of the TrpRS enzyme, in different combinations. For example, TrpRS types A and B are found in *Desulfitobacterium hafniense*, types A and C coexist in *Hahella chejuensis*, and types B and C are detected in *Kribbella flavida*. In rare cases, two copies of the homeoallele encoding a particular enzyme type are present in a single genome, as is the case in *Deinococcus radiodurans*, which harbours two copies of the homeoallele encoding TrpRS type A, and *Saccharopolyspora erythraea*, which has two copies of the homeoallele encoding TrpRS type B. The proteins encoded by the two TrpRS gene copies in *D. radiodurans* and *S. erythraea* do not group together, and therefore the acquisition of a second copy, whether of the same type or of a different type, probably occurred through HGT and not through duplication.

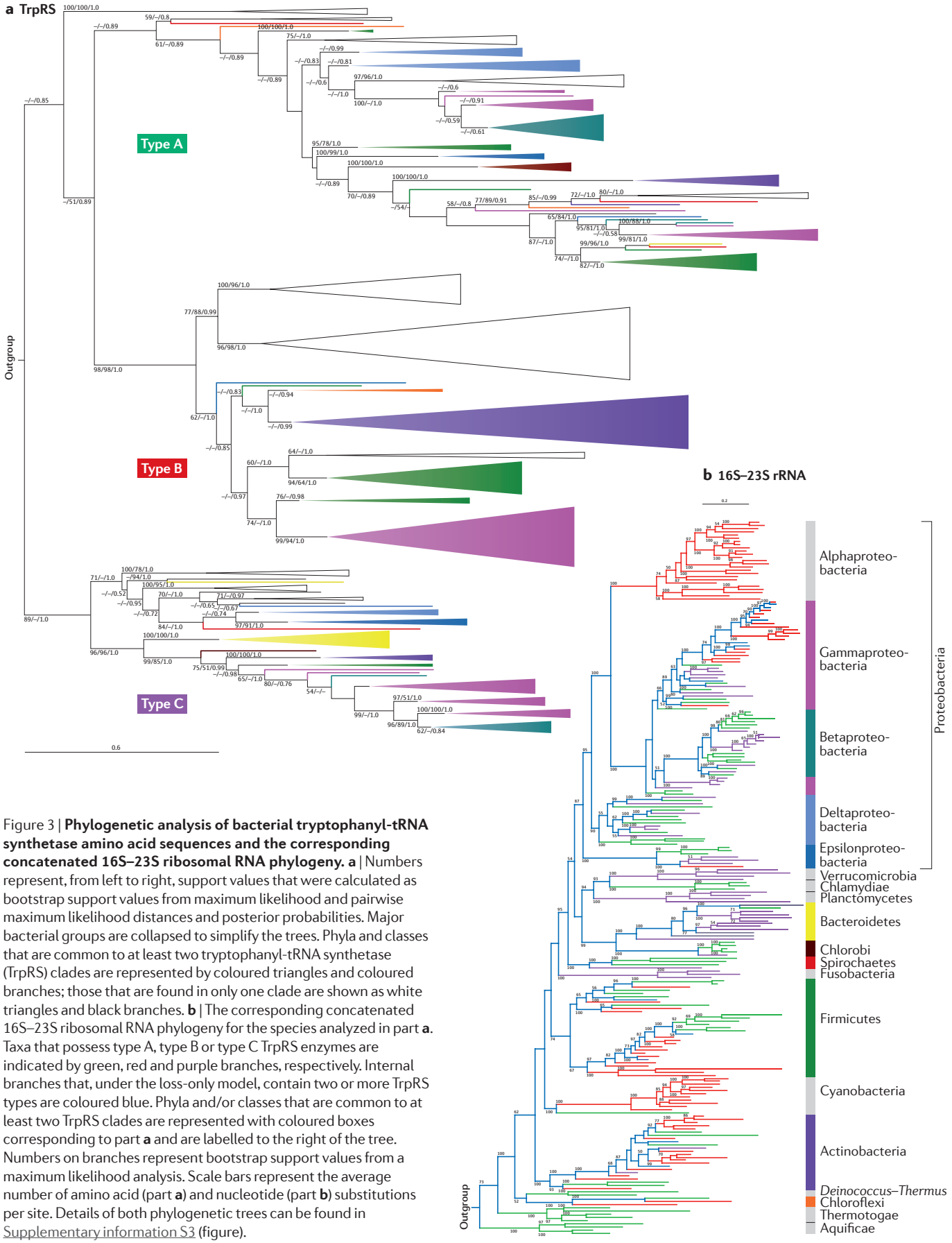



Figure 3 | **Phylogenetic analysis of bacterial tryptophanyl-tRNA synthetase amino acid sequences and the corresponding concatenated 16S–23S ribosomal RNA phylogeny. a** | Numbers represent, from left to right, support values that were calculated as bootstrap support values from maximum likelihood and pairwise maximum likelihood distances and posterior probabilities. Major bacterial groups are collapsed to simplify the trees. Phyla and classes that are common to at least two tryptophanyl-tRNA synthetase (TrpRS) clades are represented by coloured triangles and coloured branches; those that are found in only one clade are shown as white triangles and black branches. **b** | The corresponding concatenated 16S–23S ribosomal RNA phylogeny for the species analyzed in part **a**. Taxa that possess type A, type B or type C TrpRS enzymes are indicated by green, red and purple branches, respectively. Internal branches that, under the loss-only model, contain two or more TrpRS types are coloured blue. Phyla and/or classes that are common to at least two TrpRS clades are represented with coloured boxes corresponding to part **a** and are labelled to the right of the tree. Numbers on branches represent bootstrap support values from a maximum likelihood analysis. Scale bars represent the average number of amino acid (part **a**) and nucleotide (part **b**) substitutions per site. Details of both phylogenetic trees can be found in Supplementary information S3 (figure).

The acquisition of an additional copy of a homeoallele can provide a selective advantage to the recipient. For example, two coexisting forms of a gene coding for the same tRNA synthetase in a single genome have been reported, including the presence of *lysS* and *lysU* (encoding lysyl-tRNA synthetases) in *E. coli*[58], *thrSv* and *thrS2* (encoding threonyl-tRNA synthetases) in *B. subtilis*[59], *ileS1* and *ileS2* (encoding isoleucyl-tRNA synthetases) in *Pseudomonas fluorescens*[60,61], and *metS1* and *metS2* (encoding methionyl-tRNA synthetases) in *Streptococcus pneumoniae*[62]. The presence of two versions of an aaRS in an organism can be a benefit to the organism, as the newly acquired gene may provide protection against naturally produced antibiotics[63–65]. Imported genes that encode enzymes which are already present in the cell can also confer a gene dosage advantage to a species[66–68]. For example, the ability of an organism to synthesize proteins faster by utilizing different tRNA synthetases might enhance its potential to switch between different growth forms[59], differential expression at different temperatures[58] or differential sensitivity to inhibitors[62]. Alternatively, the existence of two divergent genes encoding the same enzyme may be a transitional phase involving the transfer and replacement of one or the other[69], and may be a case of transfer being nearly neutral to the recipient[70].

## Ancient origin of homeoalleles

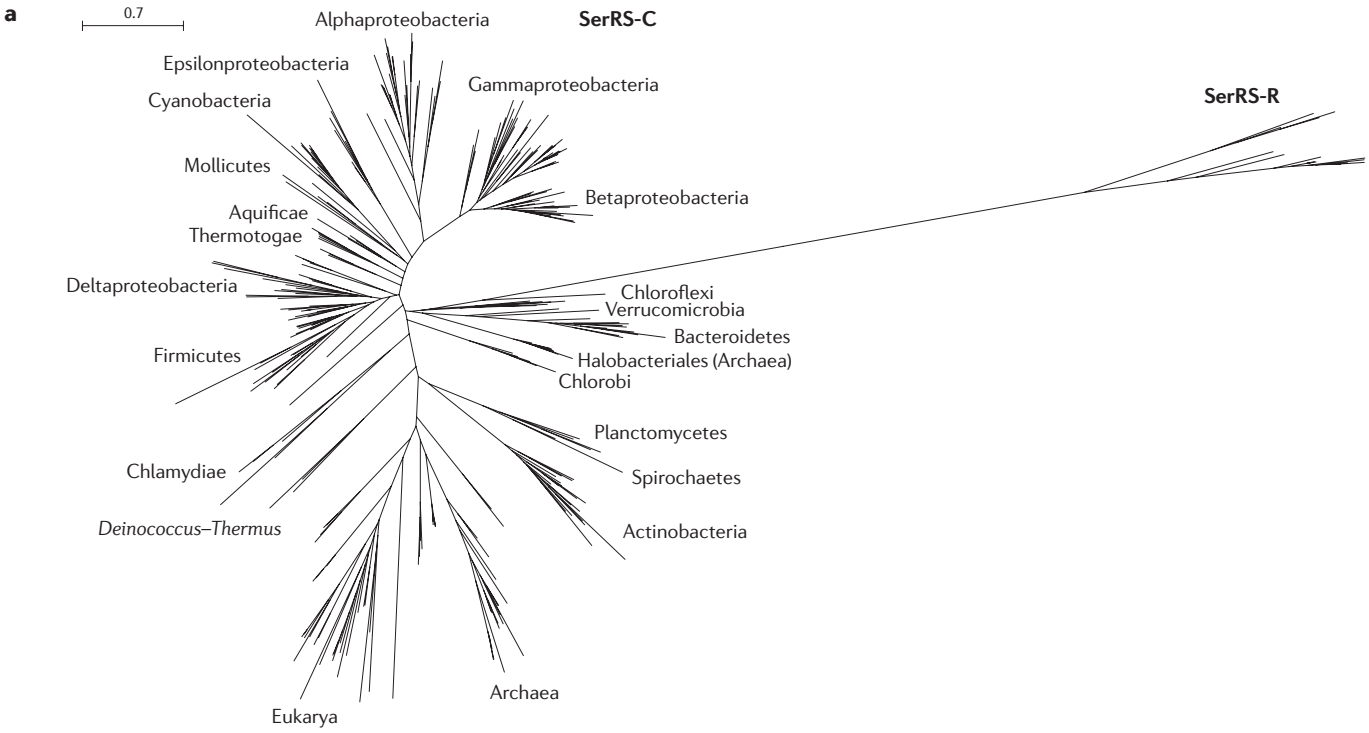
The magnitude and impact of HGT early in the history of life remains obscure because of the difficulty in reconstructing the evolutionary history of early life. This is particularly true with respect to the time before the last universal common ancestor (LUCA; also known as the cenancestor[71] and the organismal most resent common ancestor (MRCA)) of all extant life forms. It has been suggested that early life was dominated by extensive HGT, creating a primordial community, the members of which constantly swapped genetic material with each other[72,73]. Coalescence theory suggests that many other lineages coexisted with the LUCA, and that in the presence of even moderate amounts of HGT, the genes that are shared by the three domains do not trace back to the LUCA, but rather ancestral genes existed in different lineages and at different times[74]. The ancestral cell from which the three domains were derived (which we call the LUCA) is a member of a population of ancient cellular organisms that existed in the past. As a consequence of HGT, most genes of this cellular LUCA have been replaced in some or all of the cellular lineages that diverged from it, by genes from other cellular lineages that are now themselves extinct. In other words, most modern genes are derived from molecular most recent common ancestors that were present in cellular or viral entities other than the LUCA.

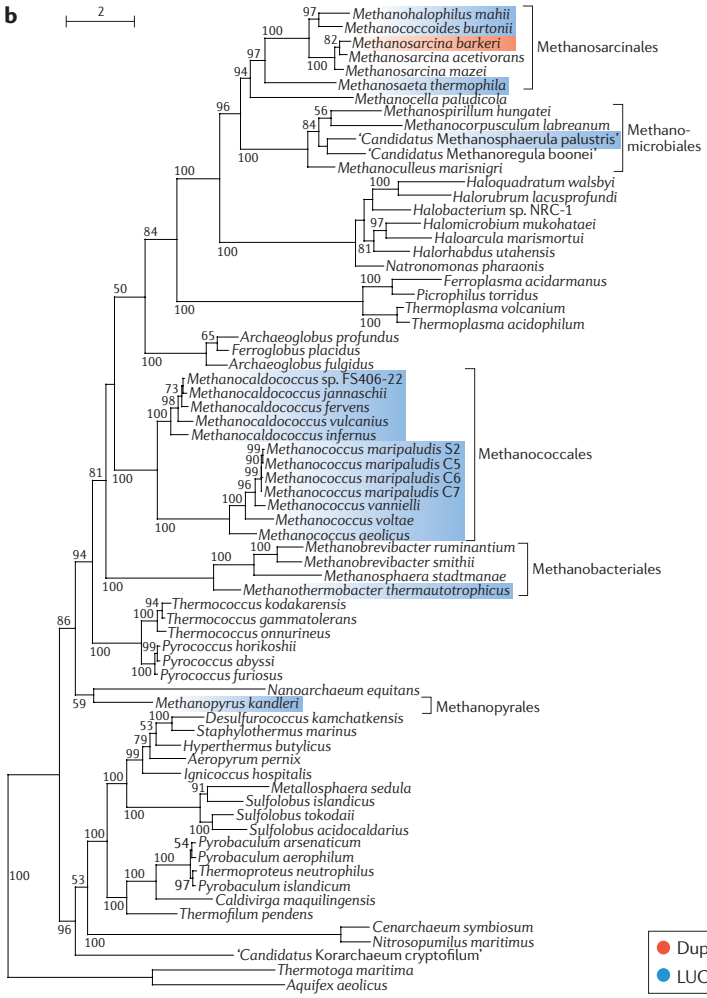
Although genetic information from extinct species has remained largely unexplored owing to the difficulty of characterizing extinct microbial organisms, HGT events from unknown extinct lineages did occur, and the genes that were transferred from those lineages still exist within the genomes of extant organisms. These ancient genes therefore allow extrapolation backwards in time and so may provide information about the genetic composition of early life. One example of an enzyme that was probably present in lineages that are now extinct is pyrrolysyl-tRNA synthetase (PylRS). Two distinct, unrelated classes of aaRSs exist (classes I and II), with each class being defined by unique domain structures and sequence homologies[75]. In the class II aaRS phylogeny, PylRS constitutes a deeply branching lineage that diverged before the evolution of the bacterial and archaeal domains[76]. Its phylogenetic position in relation to other aaRS enzymes and its sparse distribution among extant organisms indicate that HGT, instead of independent gene losses, is the most likely explanation for its existence today. Because the PylRS proteins diverge from homologous aaRS enzymes before these diverge into archaeal and bacterial homologues, the donor lineage would have been present before the organismal LUCA from which the three domains emerged[76]. Similarly, phylogenetic evidence suggests that the gene encoding the rare form of seryl-tRNA synthetase (SerRS), which is found in some archaea, emerged through HGT from an unknown, possibly extinct, deep-branching lineage to a few extant taxa. We propose that some homeoalleles, as in the case of the two forms of SerRS, have an ancient origin, appearing before the existence of the LUCA.

***Phylogenetic distribution of rare SerRS.*** As found in previous analyses[77–79], two distinct forms of SerRS can be detected (FIG. 4a), a common form (SerRS-C) that is found throughout the Bacteria and the Eukarya as well as in most archaea, and a rare form (SerRS-R) that is present in some methanogenic archaea within the phylum Euryarchaeota. These two forms have only ~25% identity, and thus homologous recombination between them is unlikely. Basic local alignment search tool (BLAST) searches revealed that only 18 taxa, belonging to eight genera and five orders (the Methanobacteriales, Methanococcales, Methanomicrobiales, Methanopyrales and Methanosarcinales), possess this rare form of SerRS. We generated a maximum-likelihood phylogenetic tree based on aligned amino acid SerRS sequences that were sampled to include representatives from all domains of life (FIG. 4a). Phylogenetic analyses show high support for the distant relationship of SerRS-R and SerRS-C.

Mapping the occurrence of SerRS-R onto the concatenated 16S–23S rRNA phylogeny of the Archaea, used here as an approximation for the archaeal species relationships, reveals a dispersed distribution of this rare enzyme (FIG. 4b). Most members of the archaeal orders Methanobacteriales, Methanomicrobiales and Methanosarcinales encode the rare form of the enzyme, but these orders also contain taxa that encode SerRS-C. This suggests that the taxa with the rare form acquired SerRS-R through independent HGT events and subsequently lost SerRS-C, or that the ancestors of these archaeal orders possessed both forms of the enzyme and the descendants lost one of the two forms. In the case of the Methanococcales, all member species and strains possess SerRS-R, which suggests that the extant members of this group retained SerRS-R since

Last universal common ancestor
(LUCA). The most recent organism (or organisms) from which all organisms that are now living on Earth descend. It is necessary to distinguish between the organismal LUCA and the most recent common ancestors of molecules and genes.



◀ Figure 4 | **Phylogenetic analysis of archaeal seryl-tRNA synthetase amino acid sequences. a** | A maximum likelihood phylogenetic tree showing the phylogeny of the common (SerRS-C) and rare (SerRS-R) forms of seryl-tRNA synthetase (SerRS) from all three domains of life. Branch lengths and topologies were calculated with PhyML v3.0 (REF. 101) using the LG[102] amino acid substitution model, estimated portions of invariable sites, estimated Γ-distribution shape parameter, four substitution rate categories, estimated amino acid frequencies, an SPR (subtree pruning and regrafting) tree search method and 20 random starting trees. The scale bar indicates the average number of amino acid substitutions per site. See part **b** (red and blue boxes) for names of the taxa labelled as SerRS-R. **b** | A maximum likelihood phylogenetic tree of concatenated 16S–23S ribosomal RNA of the archaeal domain. The tree is rooted using sequences from *Thermotoga maritima* and *Aquifex aeolicus* as the outgroup. Phylogenetic tree reconstruction and bootstrapping were performed using PhyML v3.0 (REF. 101), with estimated portions of invariable sites, four substitution rate categories, estimated transition/transversion ratio, estimated Γ-distribution shape parameter, estimated amino acid frequencies, a BioNJ starting tree, 100 bootstrap replicates and using the GTR[103] nucleotide substitution model. The scale bar represents the average number of nucleotide replacements per site. Taxa in blue boxes possess SerRS-R only, that in the red box carries both types of SerRS, and those not in a box are those with SerRS-C only. The orders that carry the rare form are indicated on the right of the tree. **c** | A phylogenetic tree of representatives from the class II aminoacyl-tRNA synthetases, with node support generated using PhyML v3.092 and MrBayes v3.1.2 (REF. 104). The scale bar represents the average number of amino acid replacements per site. The orange circle represents the divergence between the common and rare forms of SerRS, and blue circles represent nodes that have eukaryotic, archaeal and bacterial descendants and that can tentatively be identified as nodes representing the last universal common ancestor (LUCA). The choice of substitution models was determined using ProtTest[105] and JModelTest v.0.1.1 (REF. 106). AsnRS, asparaginyl-tRNA synthetase; AspRS, aspartyl-tRNA synthetase; LysRS, lysyl-tRNA synthetase; PheRS, phenylalanyl-tRNA synthetase; ProRS, prolyl-tRNA synthetase; ThrRS, threonyl-tRNA synthetase.

their last common ancestor. Similar to the TyrRS[48] and TrpRS phylogenies, we also observe a case in which both types of SerRS are present in a single genome: that of *Methanosarcina barkeri* (FIG. 4b). Although both types of SerRS enzyme have the same function and recognize $tRNA^{Ser}$ in a similar manner, SerRS-R uses a zinc-dependent serine recognition mechanism that is different to that used by SerRS-C[78,80].

***Ancient duplication and HGT from an extinct lineage.*** To account for the rare occurrence and dispersed distribution of SerRS-R, we propose that the gene encoding this rare form diverged early from that encoding SerRS-C, with subsequent HGT from an unknown ancient lineage that has probably gone extinct or is undiscovered. SerRS is a member of the class II family of aaRSs that have a different mode of aminoacylation and are unrelated to the class I aaRSs in terms of their sequence and structure[75,81]. Phylogenetic reconstruction based on representatives from several class II aaRS enzymes shows that an ancient divergence gave rise to the two SerRS forms. This divergence seems to have occurred before the most recent common gene ancestor encoding SerRS-C, from which most of the bacterial and archaeal SerRS-encoding genes descended (FIG. 4c). It is likely that the lineage which initially possessed SerRS-R existed along with the organismal LUCA as well as other lineages, and that HGT among these ancient lineages facilitated the transmission of this rare enzyme to a number of extant lineages. As discussed above, the last common ancestors for each modern gene did not all coexist in the same organismal ancestor. The SerRS phylogeny provides an illustration for this conclusion. The last common ancestor of all SerRS-encoding genes seems to be more ancient than the last common ancestor of SerRS-C-encoding genes (FIG. 4c).

In some instances, members of the same archaeal family carry different forms of SerRS, and for these families, the most likely explanation is that the introduction of SerRS occurred through HGT after the divergence of the domain Archaea. The distribution of SerRS-R on the 16S–23S rRNA phylogenetic tree indicates that several HGT events have occurred in different archaeal orders. Most of these transfers will have occurred between lineages with extant representatives; however, at least one transfer would have had to transmit the rare form of SerRS from the deep-branching lineage to the Euryarchaeota. An alternative scenario is that an ancient gene duplication event took place and was followed by differential gene loss, but as only a minority of organisms encode SerRS-R, this scenario is less likely. If the last common ancestor of all organisms, including the extinct lineages, carried both forms of the enzyme as a result of gene duplication, then many independent gene loss events must have occurred; explaining the sparse distribution of SerRS-R in the archaeal domain (FIG. 4b) under this scenario requires 67 gene loss events. Furthermore, the gene loss scenario implies that the two SerRS forms coexisted over long periods of time, followed by late gene loss events, and therefore losses occurred after the divergence of the different archaeal phyla. Using parametric bootstrapping as implemented in LGT3State[56], the loss-only model is rejected with $P < 0.001$ (see Supplementary information S4 (figure) for further explanation and illustration); it is very unlikely that the loss-only model gave rise to the observed distribution of SerRS-R and SerRS-C. These results suggest that gene gain through HGT contributed to the evolution of the archaeal SerRS.

The phylogeny of class II aaRS enzymes (FIG. 4c) suggests that the lineage which possessed the gene encoding SerRS-R diverged early, before the LUCA, and continued to exist until the time of the divergence of the Archaea. Although the donor organism for SerRS-R remains obscure, our findings suggest that its lineage overlapped in time with the emergence of methanogens within the Euryarchaeota.

Genetic material from unknown, deep-branching extinct lineages persists today largely as a result of the horizontal dissemination of genes, as in the case of PylRS[76] and the rare form of SerRS, and such material constitutes biological novelties that would have been lost forever from the living world without HGT. These rare genes represent a 'genetic life raft', as they form the only enduring legacy of extinct or ganisms that may have been prevalent in the past[82]. Transfers from extinct lineages, the identities of which may never be resolved, provide support for the view that the most recent common ancestors of molecules and organisms may be different from each other and may not even coincide in time[74].

## Biased gene transfer and prokaryotic taxonomy

An essential aspect of microbial taxonomy is the organization of strains according to the similarity of their

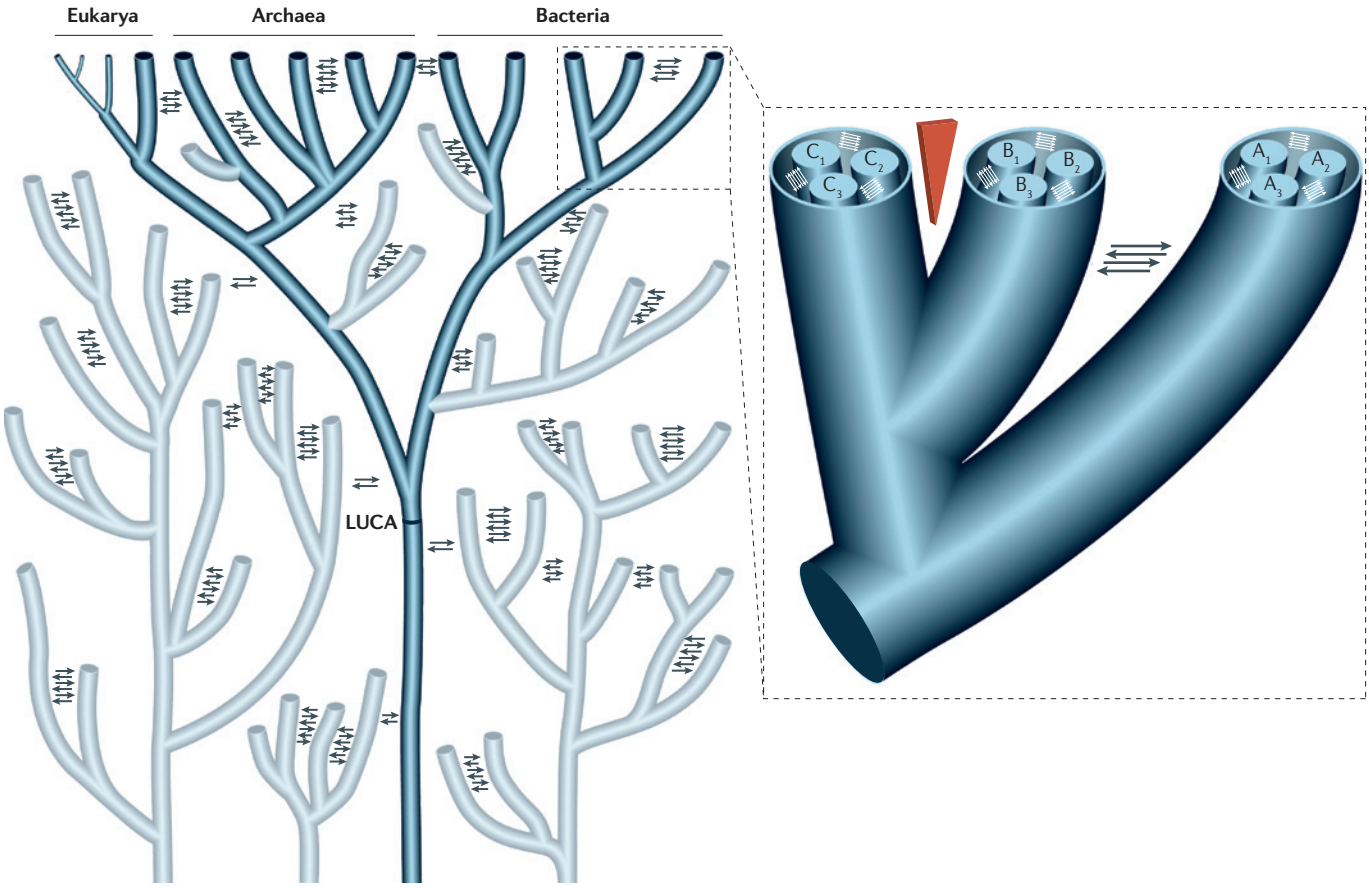



Figure 5 | **Model of the tree or net of life incorporating extinct lineages, shared ancestry and exchange groups.** Varying levels of horizontal gene transfer (HGT) within and between groups (arrows) can reinforce the assembly of taxonomically recognizable natural groups. Closely related organisms can exhibit higher frequencies of transfer (represented by numerous arrows among A1, A2 and A3), and fewer transfers occur between these groups (A and B) than between members of an individual group. A barrier to transfer (for example, the type III-like restriction endonuclease in *Staphylococcus aureus*[107], or geographical barriers as in the case of *Burkholderia pseudomallei*[108]) exists between some groups, in this case between B and C (red triangle). Hence, phylogenetic reconstruction of any typical gene will show a close affinity between groups A and B, and group C will appear to be more distantly related, even though B and C are each other's closest relatives based on an organismal tree representing the majority of genes that are passed on vertically over short periods of time[94]. HGT also occurred between extinct and extant lineages. The dark blue branches represent the parts of the tree of life that are commonly recognized, composed of the three domains emerging from an organismal last universal common ancestor (LUCA). The branches in gray represent extinct lineages, including those that evolved from the LUCA and those that diverged before the LUCA. The further one moves back in time, the more likely it is that HGT events occurred with now-extinct lineages as donors. Genes from lineages that diverged pre-LUCA have thus found their way into extant lineages through HGT.

characteristics, including biochemical, physiological, genetic and morphological features[83]. In taxonomic studies, lineages and biological groups are usually assumed to have been formed through vertical inheritance, in which an organism gives rise to descendants through cell division. Hence, members of extant lineages can trace their history back to a common ancestor, the LUCA, and they possess features that are similar to those of their ancestor because of the genes that they have inherited from it. Despite the general agreement that HGT can influence prokaryotic evolution, the classification of microorganisms into groups at different hierarchical levels still conveys a tacit assumption that vertical inheritance is the main determinant of relationships. However, the observed monophyly in the prokaryotic phylogeny may in some instances be the result of the high rates of transfer among the members of the group and does not necessarily reflect a common ancestry. Therefore, the current view of prokaryotic taxonomy that is based on ancestral relationships needs re-evaluation. A pluralistic approach[84] that embraces both cladistics and phenetics[53] may better capture the ancestry of the Bacteria and the Archaea.

The presence of a laterally acquired gene constitutes a shared characteristic that can be used to identify a particular group, including all the descendants that emerge from the taxon that initially carried the foreign gene[85,86]. Hence, the transferred gene can be used as a unique feature that reflects the monophyletic nature of the group[85,86], and this gene can then be passed on to succeeding generations through vertical inheritance. As

**Monophyly**
A phylogenetic characteristic of a group of organisms, such that the group contains all the descendants of the recent common ancestor of the group members.

**Cladistics**
A method of classifying organisms into clades based on shared derived characteristics that arise from a common ancestor.

**Phenetics**
A method of classifying organisms based on their overall similarity.

Tree of life
The tree-like representation of the history of all extant and extinct organisms.

Net of life
The depiction of evolutionary history that integrates both vertical ancestry and horizontal gene transfer events.

Bifurcating scheme
A phylogenetic tree in which all internal nodes have exactly two descendants.

seen in the Methanococcales (FIG. 4b), the ancestral species of this order acquired the gene encoding SerRS-R, and this gene was retained in all of the descendants of this ancestor through either vertical inheritance or biased transfer between close relatives. Considering all of the archaeal taxa that carry the recently acquired SerRS-R, the monophyly of this group is due to the similarity of their SerRS proteins, thus making the presence of the rare SerRS form a shared derived characteristic that defines the group. However, this relationship is not caused by vertical inheritance of the transferred gene from an original recipient to its progeny. Instead, multiple transfers of the same gene to different lineages have generated cohesion among the recipients and mark them as a group that is monophyletic only with respect to SerRS.

A consequence of HGT is that the source organism for the transferred gene will show higher levels of similarity to the recipient lineage and will group with this lineage in a phylogenetic tree[28,70]. Without knowing the history of the organisms, one would conclude that any pair of taxa which exhibit similar features and cluster together in a tree arose from the same ancestor. In the case of microorganisms that are known to have the ability to exchange genes even over a broad phylogenetic range, caution is necessary, as various evolutionary mechanisms aside from vertical inheritance are also at work. For example, the majority of the genes in the order Thermotogales, an order that is composed mostly of thermophilic bacteria, group with genes from the Gram-positive, sulphite-reducing class Clostridia (which is part of the phylum Firmicutes). Only the ribosomal components group the Thermotogales at the base of the bacterial domain, and about 10% of the genes from this order group within the Archaea[16]. The most likely explanation is that members of the Thermotogales exchange genes with thermophilic members of the Clostridia and the Archaea, with which they share ecological niches, and that therefore only a minority of the genes in members of the Thermotogales reflects the organismal ancestry of these species. A similar observation was made for the order Aquificales, another group of thermophilic bacteria[15]. In unrooted phylogenies of bacterial rRNA, Aquificales groups with Thermotogales, and in rooted rRNA phylogenies these two orders are found close to the root of the Bacteria; however, the phylogenies of several informational and operational genes place Aquificales with the class Epsilonproteobacteria. Another example of the impact of HGT is the frequent grouping of the haloarchaea at the base of the archaeal domain[87,88]. These organisms were named as the class Halobacteria before the Archaea was recognized as a distinct domain of life. For rRNA and many conserved proteins, this class groups within the phylum Euryarchaeota as a sister group to the class Methanomicrobia[89,90]. The placement of the haloarchaea at the base of the archaeal domain in some phylogenomic analyses is probably the result of the high number of genes that have been transferred between the two prokaryotic domains[13,91].

## Concluding remarks

Most regions of the tree of life or the net of life can be viewed as having been created through a combination of shared ancestry and HGT that was biased towards related organisms (FIG. 5), although the relative contributions of shared organismal ancestry and HGT vary considerably in different lineages and at different phylogenetic depths. Other mechanisms, such as symbiosis (which is one cause for a highway of gene sharing), hybridization, introgression, lineage sorting[92] and systematic artefacts of phylogenetic reconstruction[93] may dominate some of the inferred relationships. Mechanisms and processes that give rise to natural variation are not uniform[84], and generalizations about the evolution of all living organisms based on only a small group of these organisms are therefore difficult to make. The often hierarchically nested nature of the living world is therefore a consequence of multiple mechanisms and processes. Classification schemes for microbial lineages and the reconstruction of their evolutionary history should consider the diversity of evolutionary mechanisms[83,84].

Gene transfer between close relatives is difficult to ascertain using phylogenetic approaches. Therefore, the true extent of this type of gene transfer remains unknown. The replacement between types of homeoalleles only reveals that transfer occurs less frequently between distantly related taxa than between more closely related organisms; the study of homeoalleles does not tell us whether genes for which no homeoalleles exist are transferred more or less frequently. Although biased HGT can give rise to phylogenetic patterns that are similar to those created through shared ancestry, in nature both processes occur simultaneously. Over few generations, vertical inheritance will dominate over biased HGT, and owing to this dominance of vertical inheritance over short periods of time, organismal lineages can be described by the majority of gene histories over this period[94]. If gene transfer is biased towards close relatives, and as a recent shared ancestry results in closely related organisms, then most of the time the patterns that are created through biased HGT will reinforce the patterns that are due to shared ancestry. Both processes contribute to the creation of phylogenetic patterns to varying degrees, but in most instances the pattern that is created through these processes will be the same. Only in cases for which highways of gene sharing exist between divergent organisms are conflicting patterns created that cannot be forced into a bifurcating scheme.


1. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602 (2003).
2. Yap, W. H., Zhang, Z. & Wang, Y. 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 (1999).
3. Chan, C. X., Darling, A. E., Beiko, R. G. & Ragan, M. A. Are protein domains modules of lateral genetic transfer? *PLoS ONE* **4**, e4524 (2009).
4. Wang, Y. & Zhang, Z. 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 (2000).

5. Olendzenski, L. *et al.* Horizontal transfer of archaeal genes into the Deinococcaceae: detection by molecular and computer-based approaches. *J. Mol. Evol.* **51**, 587–599 (2000).
6. Castillo-Ramirez, S., Vazquez-Castellanos, J. F., Gonzalez, V. & Cevallos, M. A. Horizontal gene transfer and diverse functional constrains within a common replication-partitioning system in *Alphaproteobacteria*: the *repABC* operon. *BMC Genomics* **10**, 536 (2009).
7. Omelchenko, M., Makarova, K., Wolf, Y., Rogozin, I. & Koonin, E. Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol.* **4**, R55 (2003).
8. Igarashi, N. *et al.* Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.* **52**, 333–341 (2001).
9. Akagi, Y., Akamatsu, H., Otani, H. & Kodama, M. Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryot. Cell* **8**, 1732–1738 (2009).
10. Kaneko, S. & Itaya, M. Designed horizontal transfer of stable giant DNA released from *Escherichia coli*. *J. Biochem.* **147**, 819–822 (2010).
11. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R. & Koonin, E. V. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends Genet.* **14**, 442–444 (1998).
12. Nelson, K. E. *et al.* Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**, 323–329 (1999).
13. Beiko, R. G., Harlow, T. J. & Ragan, M. A. Highways of gene sharing in prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 14332–14337 (2005).
14. Nesbo, C. L. *et al.* The genome of *Thermosipho africanus* TCF52B: lateral genetic connections to the Firmicutes and Archaea. *J. Bacteriol.* **191**, 1974–1978 (2009).
15. Boussau, B., Gueguen, L. & Gouy, M. Accounting for horizontal gene transfers explains conflicting hypotheses regarding the position of aquificales in the phylogeny of Bacteria. *BMC Evol. Biol.* **8**, 272 (2008).
16. Zhaxybayeva, O. *et al.* On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl Acad. Sci. USA* **106**, 5865–5870 (2009).
17. Nguyen, K. T., Piastro, K., Gray, T. A. & Derbyshire, K. M. Mycobacterial biofilms facilitate horizontal DNA transfer between strains of *Mycobacterium smegmatis*. *J. Bacteriol.* **192**, 5134–5142 (2010).
18. Chafee, M. E., Funk, D. J., Harrison, R. G. & Bordenstein, S. R. Lateral phage transfer in obligate intracellular bacteria (*Wolbachia*): verification from natural populations. *Mol. Biol. Evol.* **27**, 501–505 (2010).
19. Saisongkorh, W., Robert, C., La Scola, B., Raoult, D. & Rolain, J. M. Evidence of transfer by conjugation of Type IV secretion system genes between *Bartonella* species and *Rhizobium radiobacter* in amoeba. *PLoS ONE* **5**, e12666 (2010).
20. Brochier, C., Philippe, H. & Moreira, D. The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome. *Trends Genet.* **16**, 529–533 (2000).
21. Schouls, L. M., Schot, C. S. & Jacobs, J. A. Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. *J. Bacteriol.* **185**, 7241–7246 (2003).
22. Iyer, L. M., Koonin, E. V. & Aravind, L. Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. *Gene* **335**, 73–88 (2004).
23. Makarova, K. S., Ponomarev, V. A. & Koonin, E. V. Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol.* **2**, RESEARCH 0033 (2001).
24. Boucher, Y., Douady, C. J., Sharma, A. K., Kamekura, M. & Doolittle, W. F. Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* **186**, 3980–3990 (2004).
25. Kanhere, A. & Vingron, M. Horizontal gene transfers in prokaryotes show differential preferences for metabolic and translational genes. *BMC Evol. Biol.* **9**, 9 (2009).
26. Danchin, E. G. *et al.* Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl Acad. Sci. USA* **107**, 17651–17656 (2010).
27. Khomyakova, M., Bukmez, O., Thomas, L. K., Erb, T. J. & Berg, I. A. A methylaspartate cycle in haloarchaea. *Science* **331**, 334–337 (2011).
28. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
29. Woese, C. R., Olsen, G. J., Ibba, M. & Soll, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236 (2000).
30. O'Donoghue, P. & Luthey-Schulten, Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573 (2003).
31. Coscolla, M., Comas, I. & Gonzalez-Candelas, F. Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol. Biol. Evol.* **28**, 985–1001 (2010).
32. Hendrickson, H. & Lawrence, J. G. Selection for chromosome architecture in bacteria. *J. Mol. Evol.* **62**, 615–629 (2006).
33. Olendzenski, L., Zhaxybayeva, O. & Gogarten, J. P. in *Horizontal Gene Transfer* (eds Syvanen, M. & Kado, C. I.) 427–435 (Academic, New York, 2002).
34. Krupovic, M., Gribaldo, S., Bamford, D. H. & Forterre, P. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* **27**, 2716–2732 (2010).
35. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Rev. Microbiol.* **3**, 711–721 (2005).
36. McDaniel, L. D. *et al.* High frequency of horizontal gene transfer in the oceans. *Science* **330**, 50 (2010).
37. Popa, O., Hazkani-Covo, E., Landan, G., Martin, W. & Dagan, T. Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* **21**, 599–609 (2011).
38. Hooper, S., Mavromatis, K. & Kyrpides, N. Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome Biol.* **10**, R45 (2009).
39. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nature Rev. Microbiol.* **3**, 722–732 (2005).
40. Sobecky, P. A. & Hazen, T. H. Horizontal gene transfer and mobile genetic elements in marine systems. *Methods Mol. Biol.* **532**, 435–453 (2009).
41. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of *Staphylococcus aureus*. *Cell. Mol. Life Sci.* **67**, 3057–3071 (2010).
42. Siefert, J. L. Defining the mobilome. *Methods Mol. Biol.* **532**, 13–27 (2009).
43. Wozniak, R. A. & Waldor, M. K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Rev. Microbiol.* **8**, 552–563 (2010).
44. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl Acad. Sci. USA* **102**, 13950–13955 (2005).
45. Lapierre, P. & Gogarten, J. P. Estimating the size of the bacterial pan-genome. *Trends Genet.* **25**, 107–110 (2009).
46. Puigbo, P., Wolf, Y. I. & Koonin, E. V. The tree and net components of prokaryote evolution. *Genome Biol. Evol.* **2**, 745–756 (2010).
47. Lawrence, J. G. & Retchless, A. C. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. *Methods Mol. Biol.* **532**, 29–53 (2009).
48. Andam, C. P., Williams, D. & Gogarten, J. P. Biased gene transfer mimics patterns created through shared ancestry. *Proc. Natl Acad. Sci. USA* **107**, 10679–10684 (2010).
49. Thomas, V. & Greub, G. Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. *Intervirology* **53**, 254–267 (2010).
50. Huang, J. & Gogarten, J. P. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? *Genome Biol.* **8**, R99 (2007).
51. Martin, W., Brinkmann, H., Savonna, C. & Cerff, R. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl Acad. Sci. USA* **90**, 8692–8696 (1993).
52. Fournier, G. P. & Gogarten, J. P. Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. *J. Bacteriol.* **190**, 1124–1127 (2008).
53. Andam, C. P., Williams, D. & Gogarten, J. P. Natural taxonomy in light of horizontal gene transfer. *Biol. Philos.* **25**, 589–602 (2010).
54. Omelchenko, M. V., Galperin, M. Y., Wolf, Y. I. & Koonin, E. V. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol. Direct* **5**, 31 (2010).
55. Farahi, K., Pusch, G. D., Overbeek, R. & Whitman, W. B. Detection of lateral gene transfer events in the prokaryotic tRNA synthetases by the ratios of evolutionary distances method. *J. Mol. Evol.* **58**, 615–631 (2004).
56. Stern, A. *et al.* An evolutionary analysis of lateral gene transfer in thymidylate synthase enzymes. *Syst. Biol.* **59**, 212–225 (2010).
57. Zhaxybayeva, O. Detection and quantitative assessment of horizontal gene transfer. *Methods Mol. Biol.* **532**, 195–213 (2009).
58. Clark, R. & Neidthardt, F. Roles of the two lysyl-tRNA synthetases of *Escherichia coli*: analysis of nucleotide sequences and mutant behavior. *J. Bacteriol.* **172**, 3237–3243 (1990).
59. Putzer, H., Brakhage, A. A. & Grunberg-Manago, M. Independent genes for two threonyl-tRNA synthetases in *Bacillus subtilis*. *J. Bacteriol.* **172**, 4593–4602 (1990).
60. Yanagisawa, T. & Kawakami, M. How does *Pseudomonas fluorescens* avoid suicide from its antibiotic pseudomonic acid? Evidence for two evolutionarily distinct isoleucyl-tRNA synthetases conferring self-defense. *J. Biol. Chem.* **278**, 25887–25894 (2003).
61. Thomas, C. M., Hothersall, J., Willis, C. L. & Simpson, T. J. Resistance to and synthesis of the antibiotic mupirocin. *Nature Rev. Microbiol.* **8**, 281–289 (2010).
62. Gentry, D. R. *et al.* Variable sensitivity to bacterial methionyl-tRNA synthetase inhibitors reveals subpopulations of *Streptococcus pneumoniae* with two distinct methionyl-tRNA synthetase genes. *Antimicrob. Agents Chemother.* **47**, 1784–1789 (2003).
63. Brown, J. R. *et al.* Horizontal transfer of drug-resistant aminoacyl-transfer-RNA synthetases of anthrax and Gram-positive pathogens. *EMBO Rep.* **4**, 692–698 (2003).
64. Vecchione, J. J. & Sello, J. K. A novel tryptophanyl-tRNA synthetase gene confers high-level resistance to indolmycin. *Antimicrob. Agents Chemother.* **53**, 3972–3980 (2009).
65. Vecchione, J. J. & Sello, J. K. Regulation of an auxiliary, antibiotic-resistant tryptophanyl-tRNA synthetase gene via ribosome-mediated transcriptional attenuation. *J. Bacteriol.* **192**, 3565–3573 (2010).
66. Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, RESEARCH0008 (2002).
67. Couturier, E. & Rocha, E. P. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* **59**, 1506–1518 (2006).
68. Zeng, Y., Roy, H., Patil, P. B., Ibba, M. & Chen, S. Characterization of two seryl-tRNA synthetases in albomycin-producing *Streptomyces* sp. strain ATCC 700974. *Antimicrob. Agents Chemother.* **53**, 4619–4627 (2009).
69. Doolittle, W. F. *et al.* How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Phil. Trans. R. Soc. B* **358**, 39–58 (2003).
70. Gogarten, J. P. & Townsend, J. P. Horizontal gene transfer, genome innovation and evolution. *Nature Rev. Microbiol.* **3**, 679–687 (2005).
71. Fitch, W. M. & Upper, K. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 759–767 (1987).
72. Kandler, O. in *Early Life on Earth, Nobel Symposium 84* (ed. Bengston, S.) 152–160 (Columbia Univ. Press, New York,1994).
73. Woese, C. The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859 (1998).
74. Zhaxybayeva, O. & Gogarten, J. P. Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet.* **20**, 182–187 (2004).
75. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206 (1990).

76. Fournier, G. P., Huang, J. & Gogarten, J. P. Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Phil. Trans. R. Soc. B* **364**, 2229–2239 (2009).
77. Kim, H. S., Vothknecht, U. C., Hedderich, R., Celic, I. & Soll, D. Sequence divergence of seryl-tRNA synthetases in archaea. *J. Bacteriol.* **180**, 6446–6449 (1998).
78. Korencic, D., Polycarpo, C., Weygand-Durasevic, I. & Soll, D. Differential modes of transfer $RNA^{Ser}$ recognition in *Methanosarcina barkeri*. *J. Biol. Chem.* **279**, 48780–48786 (2004).
79. Bilokapic, S., Korencic, D., Soll, D. & Weygand-Durasevic, I. The unusual methanogenic seryl-tRNA synthetase recognizes $tRNA^{Ser}$ species from all three kingdoms of life. *Eur. J. Biochem.* **271**, 694–702 (2004).
80. Jaric, J. *et al.* Identification of amino acids in the N-terminal domain of atypical methanogenic-type seryl-tRNA synthetase critical for tRNA recognition. *J. Biol. Chem.* **284**, 30643–30651 (2009).
81. Ibba, M., Curnow, A. W. & Soll, D. Aminoacyl-tRNA synthesis: divergent routes to a common goal. *Trends Biochem. Sci.* **22**, 39–42 (1997).
82. Gogarten, J., Fournier, G. & Zhaxybayeva, O. Gene transfer and the reconstruction of life's early history from genomic data. *Space Sci. Rev.* **135**, 115–131 (2008).
83. Gevers, D. *et al.* Stepping stones towards a new prokaryotic taxonomy. *Phil. Trans. R. Soc. B* **361**, 1911–1916 (2006).
84. Bapteste, E. *et al.* Prokaryotic evolution and the tree of life are two different things. *Biol. Direct* **4**, 34 (2009).
85. Huang, J. & Gogarten, J. P. Ancient horizontal gene transfer can benefit phylogenetic reconstruction. *Trends Genet.* **22**, 361–366 (2006).
86. Huang, J., Xu, Y. & Gogarten, J. P. The presence of a haloarchaeal type tyrosyl-tRNA synthetase marks the opisthokonts as monophyletic. *Mol. Biol. Evol.* **22**, 2142–2146 (2005).
87. Wolf, Y. I., Rogozin, I. B., Grishin, N. V. & Koonin, E. V. Genome trees and the tree of life. *Trends Genet.* **18**, 472–479 (2002).
88. Korbel, J. O., Snel, B., Huynen, M. A. & Bork, P. SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* **18**, 158–162 (2002).
89. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
90. Brochier, C., Forterre, P. & Gribaldo, S. Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox. *Genome Biol.* **5**, R17 (2004).
91. Mongodin, E. F. *et al.* The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl Acad. Sci. USA* **102**, 18147–18152 (2005).
92. Degnan, J. H. & Rosenberg, N. A. Discordance of species trees with their most likely gene trees. *PLoS Genet.* **2**, e68 (2006).
93. Herbeck, J. T., Degnan, P. H. & Wernegreen, J. J. Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the Enterobacteriales (γ-Proteobacteria). *Mol. Biol. Evol.* **22**, 520–532 (2005).
94. Zhaxybayeva, O., Lapierre, P. & Gogarten, J. P. Genome mosaicism and organismal lineages. *Trends Genet.* **20**, 254–260 (2004).
95. Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* **16**, 1099–1108 (2006).
96. Hilario, E. & Gogarten, J. P. Horizontal transfer of ATPase genes — the tree of life becomes a net of life. *Biosystems* **31**, 111–119 (1993).
97. Horn, M. *et al.* Illuminating the evolutionary history of chlamydiae. *Science* **304**, 728–730 (2004).
98. Becker, B., Hoef-Emden, K. & Melkonian, M. Chlamydial genes shed light on the evolution of photoautotrophic eukaryotes. *BMC Evol. Biol.* **8**, 203 (2008).
99. Moustafa, A., Reyes-Prieto, A. & Bhattacharya, D. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. *PLoS ONE* **3**, e2205 (2008).
100. de Koning, A. P., Brinkman, F. S., Jones, S. J. & Keeling, P. J. Lateral gene transfer and metabolic adaptation in the human parasite *Trichomonas vaginalis*. *Mol. Biol. Evol.* **17**, 1769–1773 (2000).
101. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704 (2003).
102. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).
103. Lanave, C., Preparata, G., Saccone, C. & Serio, G. A new method for calculating evolutionary substitution rates. *J. Mol. Evol.* **20**, 86–93 (1984).
104. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
105. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
106. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
107. Corvaglia, A. R. *et al.* A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proc Natl Acad. Sci. USA* **107**, 11954–11958 (2010).
108. Pearson, T. *et al.* Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol.* **7**, 78 (2009).



### Acknowledgements
The authors thank P. Lapierre, D. Williams, T. Harlow, V. Kask and the Biotechnology Bioservices Center of the University of Connecticut, Storrs, USA, for technical support. This work was supported by a US National Science Foundation Grant (DEB 0830024).


### Competing interests statement
The authors declare no competing financial interests.

### FURTHER INFORMATION
**J. Peter Gogarten's homepage:** http://gogarten.uconn.edu/

### SUPPLEMENTARY INFORMATION
**See online article:** S1 (figure) | S2 (figure) | S3 (figure) | S4 (figure)

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**