



Evolution / Évolution

# Molecular phylogeny: reconstructing the forest

Philippe Lopez<sup>\*</sup>, Eric Bapteste

*UMR 7138 : systématique, adaptation, évolution, Université Pierre-et-Marie-Curie, 75005 Paris, France*

Accepted after revision 9 July 2008

Available online 29 November 2008

Presented by Claude Combes

## Abstract

Phylogeny, be it morphological or molecular, has long tried to explain the extant biodiversity by the Tree of Species, which is a logical consequence of strict Darwinian evolutionary principles. Through constant improvement of both methods and data sets, some parts of this diversity have actually been demonstrated to be the result of a tree-like process. For some other parts, and especially for prokaryotes, different molecular markers have, however, produced different evolutionary trees, preventing the reconstruction of such a Tree. While technical artifacts could be blamed for these discrepancies, Lateral Gene Transfers are now largely held for responsible, and their existence requires an extension of the Darwinian framework, since genetic material is not always vertically inherited from parents to offspring. Through a variety of biological processes, sometimes large parts of DNA are exchanged between phylogenetically distant contemporary organisms, especially between those sharing the same environment. While mainly concerning prokaryotes, Lateral Gene Transfers have been also demonstrated to affect eukaryotes, and even multicellular ones, like plants or animals. Most of the time, these transfers allow important adaptations and the colonisation of new niches. The quantitative and qualitative importance of genetic transfers has thus severely challenged the very existence of a universal Tree of Species, since genetic connections, at least for microbes, seem more reticulated than tree-like. Even traditional biological concepts, like the concept of species, need to be re-evaluated in the light of recent discoveries. In short, instead of focusing on a elusive universal tree, biologists are now considering the whole forest corresponding to the multiple processes of inheritance, both vertical and horizontal. This constitutes the major challenge of evolutionary biology for the years to come. **To cite this article: P. Lopez, E. Bapteste, C. R. Biologies 332 (2009).**

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

## Résumé

**Phylogénie moléculaire : reconstruire la forêt.** La phylogénie, qu'elle soit morphologique ou moléculaire, a longtemps tenté d'expliquer la biodiversité actuelle par l'Arbre des Espèces, qui est une conséquence logique des stricts principes évolutifs darwiniens. Grâce à l'amélioration constante à la fois des méthodes et des données, on a effectivement montré que certaines parties de cette biodiversité étaient le résultat d'un processus arborescent. Pour d'autres parties, en revanche, et spécialement les procaryotes, différents marqueurs moléculaires produisent des arbres différents et empêchent de reconstruire un arbre universel. Bien que divers artefacts méthodologiques puissent être la cause de ces incongruences, les Transferts Horizontaux de Gènes en sont maintenant reconnus responsables et leur existence nécessite l'extension du cadre darwinien, puisque le matériel génétique n'est pas nécessairement transmis de parents à descendants. Toute une gamme de processus biologiques permet l'échange de fragments d'ADN, parfois très grands, entre organismes contemporains phylogénétiquement éloignés, et particulièrement entre ceux qui partagent un

<sup>\*</sup> Corresponding author.

E-mail address: [philippe.lopez@upmc.fr](mailto:philippe.lopez@upmc.fr) (P. Lopez).

même environnement. Bien que concernant surtout les procaryotes, on a montré récemment que les transferts horizontaux affectent aussi les eucaryotes, et même les multicellulaires, comme les plantes ou les animaux. La plupart du temps, ces transferts permettent d'importantes adaptations et la colonisation de nouveaux milieux. L'importance quantitative et qualitative des transferts de matériel génétique a donc très fortement remis en cause l'existence même d'un Arbre des Espèces, dans la mesure où les connections génétiques, du moins pour les microbes, semblent plus réticulées qu'arborescentes. Même les concepts traditionnels de biologie, comme la notion d'espèce, doivent être réévalués à la lumière des découvertes récentes. En somme, plutôt que de se concentrer sur un insaisissable Arbre Universel, les biologistes considèrent maintenant la forêt qui correspond aux multiples processus de transmission, à la fois verticaux et horizontaux. Cela constitue le défi majeur des prochaines années pour la biologie évolutive.

**Pour citer cet article :** P. Lopez, E. Baptiste, C. R. Biologies 332 (2009).

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

**Keywords:** Evolution; Genome fluidity; Lateral gene transfers; Phylogeny

**Mots-clés :** Évolution ; Fluidité du génome ; Transferts latéraux des gènes ; Phylogénie

Traditionally, phylogeneticists assume that the genetic markers they use to reconstruct the Tree of Species are vertically inherited. Recently, however, due to the importance of lateral gene transfer, the development of tools and concepts to transform this classical phylogenetic practice was suggested. Instead of focusing their efforts on the elaboration of an hypothetical single Tree to model the evolutionary course, phylogeneticists could endorse a more pluralistic approach, in order to see the forest of phylogenetic trees (and networks) depicting the different histories of the various evolutionary units present in nature. Here, we review a series of biological cases, evolutionary processes and new evolutionary concepts that justify such a change of perspective.

## 1. Traditional tree-making

In 1859, Charles Darwin proposed two fundamental principles of evolution in *The Origin of Species*: descent with modification and natural selection [1]. If evolution is driven by these two principles only, then genetic information is passed on vertically, from genitors to descendants. A logical consequence of this would be that present biodiversity is the result of one or more tree-like processes. Admittedly, the most popular hypothesis for explaining biodiversity is that there is a unique tree describing the evolutionary relationships of all living organisms: the so-called Tree of Species. Recovering that tree has then been one of the main goals of biology for the last 150 years, spawning a new discipline called phylogeny. This is perfectly understandable since the tree of species could serve three highly desirable purposes. First, it could provide a natural classification of living organisms: all the extant descendants of a given ancestor form a natural group, or clade. Knowing the Tree of Species thus conveniently defines a hierarchical classi-

fication of Life, or “groups within groups”, according to Darwin. Second, the tree could provide insights on the shared properties of organisms belonging to the same clade: all mammals have hair, for example. Third, the tree could be seen as a time machine, since it allows retrodiction. Knowing the tree and the properties of the extant organisms, one can infer the properties of the ancestors, to a certain extent.

However, assuming there is a Tree of Species, Darwinian principles do not say anything about its recoverability, a question that is generally overlooked. Phylogenetic reconstruction is based on the comparison of homologous characters between extant organisms. An hypothesis (or inference) is then produced, through different methods, in order to explain this repartition of character states. When comparing more or less closely related organisms, producing an explanatory tree is generally straightforward, because evolutionary changes have been few, and information is easily recoverable. Yet, recovering the tree becomes increasingly difficult as the compared organisms become more divergent, or as the number of evolutionary changes grows, a problem that obviously culminates when trying to recover the Tree of Life. Reconstructing very ancient evolutionary events requires a lot of information, or a lot of characters, for that matter. Typically, this is where morphological phylogeny reached its limits: unicellular organisms, for example, lacked enough characters for comparative analysis. Even if the unicellular diversity was the result of a tree-like process, that tree was not recoverable through the use of morphological characters.

The discovery of DNA and the development of sequencing techniques brought a new hope to the phylogenetic field, as introduced by Zuckerkandl and Pauling in their pioneering article [2]. For once, phylogeneticists had access to a seemingly unending wealth of homologous characters for every living organism, re-opening

the way to the recoverability of the Tree of Life. In the beginning, molecular trees of life were rather modest in their taxonomical sampling, and obtaining the molecular data was clearly seen as the major limiting factor. Ribosomal RNA progressively became the molecule of choice for deep phylogenies [3–5]. While phylogenetic reconstruction methods were thoroughly improved, sequencing was also thought to be the path to enlightenment until the early 1990s: more sequences of more genes could only improve the depth and resolution of our knowledge of life's history. Molecular phylogenetics scored quite a few successes indeed, like replacing the microsporidia inside the fungi [6] or finding a sister-group relationship for euglenes and trypanosomes [7], for example. However, it became rapidly clear that more genes of more organisms also brought some discrepancies: different phylogenetic markers told different evolutionary stories, especially when dealing with deep phylogenies, such as the tree of eukaryotes for example [8]. As a matter of fact, more than 40 years after the birth of molecular phylogeny, and despite the exponential growth of sequence data banks, the Tree of Life remains elusive at best, even if substantial achievements have been made.

Generally speaking, molecular phylogeny readily identifies the most recent clades: most of the time, a phylogenetic analysis can tell whether a given sequence comes from a ciliate or a proteobacterium, for example. Reconstructing the evolutionary relationships of these groups (or going further back in time, for that matter) proves more challenging, for different reasons. First, the very structure of the tree can be a problem, especially in case of rapid radiations. When many well defined clades appear in a short period of time, the phylogenetic signal about their relationships can be obscured by the ensuing evolution [9]. Yet, radiations seem to be a common process in the tree of life, as exemplified by placental mammals [10] or flowering plants [11]. Second, it has been known for a long time that phylogenetic reconstructions are plagued by an artifact called Long Branch Attraction (LBA) [12] that shows up in case of molecular saturation, when sequences have undergone a lot of multiple substitutions. Very divergent sequences tend to be attracted together, and thus towards the outgroup, irrespectively of their true position. In many cases, LBA and molecular saturation were responsible for producing wrong and/or incongruent deep phylogenies: different molecules are subject to different rates of evolution and thus to different LBA. Microsporidia, which are now known to be highly derived fungi, are arguably emblematic of this artifact, since LBA consistently placed them as the ear-

liest emerging eukaryotic clade in the reference rRNA tree and in some trees based on other molecules [13].

Two different approaches have been explored in order to overcome these problems: improving the reconstruction methods and improving the data sets. First, highly efficient probabilistic reconstruction methods have made their appearance during last years [14,15], allowing robust analyses of much larger data sets. In the same time, sophisticated models of evolution have been developed, since wrong models are bound to strengthen LBA and thus yield wrong trees. Mixed models, where different data subsets are allowed to evolve under different evolutionary models, seem to be a particularly promising approach [15,16], even if these new models tend to be especially parameter-rich and fall prey to over-parameterization [17]. Second, molecular phylogenetics immensely benefited from the genomics era and the wealth of sequences it produced. In order to reconstruct very ancient speciation events, data sets must harbor characters that have undergone some evolutionary changes during the time of the speciation, and only a few afterwards. Such characters are rare, which explains why increasing the number of characters generally improves the reconstruction. Admittedly, no single phylogenetic marker harbors enough information to reconstruct the Tree of Life, so new methods were developed in order to pool together bits of information from different markers: molecular phylogenetics turned into phylogenomics [18]. In some cases, such as for prokaryotes, complete genome sequences are known, which opens the way to whole-genome phylogenetic approaches, through the comparison of gene order or gene content for example [19]. However, more popular methods consist in aggregating trees that are produced by different markers into a supertree [20], or in aggregating different data sets into a supermatrix [21]. In both cases, the purpose is to maximize the data set for a given problem and it is not uncommon to simultaneously analyze up to 150 markers.

As a whole, phylogenomics greatly improved our knowledge of the Tree of Life and especially pushed back the time limit beyond which evolutionary relationships really become unclear. Parts of the tree of Life are now much more resolved, like, for example, the monophyly of green plants and red algae, a feature that was nearly impossible to reliably recover with single markers [21]. At a larger evolutionary scale, phylogenomics associated to structural analyses also tentatively reduced the eukaryotic diversity to only six super-groups, when using more than a hundred markers [22,23]. Similarly, multi-marker analyses helped to resolve the tree of archaea [24], under the assumption that archaeal genes are

indeed vertically inherited, which will be extensively discussed later on. However, even such trees are not fully resolved. Very ancient speciation events are especially hard to reliably infer, even with hundreds of markers: not only are the relationships of the six eukaryotic supergroups still unknown, but the groups themselves prove somehow unstable [25]. In the case of radiations, like that of placental mammals, much younger events remain difficult to resolve by molecular phylogeny [26].

As a conclusion, phylogenomics certainly helped uncover some parts of the Tree of Species, especially for multicellular organisms. However, other parts still remain elusive, even when using a significant fraction of the proteome as a data set. A reason for this could be that some events are too old to be recoverable: ancient phylogenetic signal has been erased by the following multiple substitutions. In other words, there would be a temporal horizon beyond which molecular phylogeny is unable to see. While generally overlooked, this constitutes a very serious problem for deep phylogenies [27]. A perfect example for that would be the notorious root of the Tree of Life, for which there is clearly no significant phylogenetic signal available [28]. Yet, even if the historical signal may be lost at molecular level for such ancient events, higher scale evolutionary units might still retain some information. For example, some phylogeneticists have been trying to search for rare genetic changes (RGC) in order to reconstruct very ancient events. RGC are highly unlikely genomic events, like insertions/deletions or genes fusions/fissions, that in consequence should exhibit very little homoplasy. This could be called molecular cladistics, since it consists in regrouping genomic shared derived characters. For example, the root of the eukaryotic tree, which molecular phylogeny seems unable to reliably infer, has been tentatively placed between unikonts and bikonts, based on the fusion/fission of two groups of genes [29].

However, an increasing proportion of evolutionary biologists also believes that the Tree of Species remains elusive because it is a mere hypothesis [30]. While some lineages are certainly the result of a tree-like process, like those of multicellular organisms, it might not be the case for the whole picture. If genetic material can be exchanged between contemporary organisms for example, the strictly Darwinian framework has to be extended, thus severely challenging the existence of the universal Tree of Species.

## 2. On the complexity of evolutionary processes

Opposite to the traditional assumption of molecular phylogenetics, there is not one vertical but multiple

processes of gene inheritance. In prokaryotes (bacteria and archaea), biologists now widely acknowledge the fact that not all individuals within a species resemble each other because they clonally descend from a last single ancestor [31]. Prokaryotes can also acquire their genes laterally, from unrelated organisms with which they share their environment. This phenomenon, called lateral gene transfer (LGT), was discovered in the late 1950s, when it was observed that resistance to multiple antibiotics could be transferred simultaneously from *Shigella* to *Escherichia coli*, and that such between-species transfer was probably responsible for the increase in drug resistance among dysentery-causing shigellae in Japan [32]. Subsequent studies of this phenomenon both revealed the frequency and the diversity of the mechanisms of gene transfer.

The simplest one is probably transformation, where free DNA is acquired by the cells, once it has passed through the permeabilized outer membrane to the cytoplasm. Another mechanism is called conjugation. Segments of DNA can be mobilized by replicating genetic elements known as conjugative plasmids (or chromosomally Integrated Conjugative elements) [33]. Such mobile elements, characterized by their peripatetic nature and their broad-host-range, can carry important quantities of DNA. For instance, the ‘megaplasmids’ are the size of small chromosomes (up to a few Mb) and strains of *Halorubrum* efficiently mobilized chromosomes ensuring Hfr-like transfer of the genes for anaerobic growth in extreme thermophiles. A broad variety of genes, fulfilling a wide set of biological functions, are also mobilized by a recently discovered molecular structure: the integron [34]. Integrons are often found on plasmids and they are present in at least 10% of the bacterial genomes and in archaea. Each integron has its own promoter and recombination toolkit, and contains a few to hundreds of gene-cassettes, which can thus be moved around in a single transfer event, and potentially expressed in the host.

Yet, transduction by phage is probably the most important mode of lateral gene transfer. Phages are indeed the most abundant ( $\sim 10^{30}$  tailed phage particles) and the most rapidly replicating life forms on earth ( $10^{25}$  infections every second) and their genetic diversity is enormous [33]. These entities “only” carry from a few to several hundreds kb of genetic material, yet they cause dramatic effects in the genomic evolution of the bacterial cell they infect [35]. Significantly, physical proximity plays a role in lateral gene transfer. Distantly related lineages that live in the same environment can frequently exchange genes, and not only due to viral transfers. In plants, for instance, illegitimate pollination,

uptake of naked DNA in the soil, fusion of mitochondria [36], epiphytism [37], endophytism [38], endosymbionts [39] or bacterial symbionts [33] are postulated to cause LGT. In phagotrophic protists (unicellular eukaryotes that ingest their preys), feeding habits are often suspected to trigger some lateral gene transfer [40], as the foreign genetic material that is constantly entering the cell via food organisms seems to make its way to the predator genome. It has thus been proposed that some microorganisms could be “what they eat” [41] or “where they live” and that, over the long run, their genomic composition could be influenced by their environment maybe to a greater extent than by their genealogical (vertical) origin.

### 3. Biological consequences of LGT

It is important to appreciate the profound biological impact of these gene transfers to realize that this evolutionary process and its effects should not be dismissed from our reconstruction of the evolutionary history.

In prokaryotes, genome sequencing and MLST studies have revealed that gene transfer between non-mating species is remarkably common. At least 24% of the thermophile bacteria *Thermotoga maritima*, would have been inherited from Archaea [42]. Similarly, both phylogenetic analyses and examination of atypical sequences predict that *in vivo* more than 20% of the *E. coli* genome had been recently introduced by gene transfer [43]. As a result, the genomic composition of *E. coli* strains is highly variable: 3 strains of the same *E. coli* species, K-12, 0157:H7 and CFT073, for instance, share less than 40% of genes in common [44]. Although it might be easier to introduce DNA from a distant relative with different promoter characteristics than from a closer relative [45], and although members of informational gene families (i.e. the transcription and translation machineries) seem to be less easily transferable, interspecies gene transfer is not restricted to special categories of genes [46]. For instance, out of 246,045 genes from 79 different species of prokaryotes, there was no single gene that, along with all its prokaryotic homologs, resisted transfer by way of a plasmid into *E. coli* in laboratory conditions [45]. Similarly, all prokaryotic phyla have experienced lateral transfers between very distantly related species, involving a large diversity of genes: central metabolic functions [47], complete biosynthetic pathways [48–50], portions of the transcription and translation machinery [51–53], ribosomal proteins [54] and ribosomal RNA [43,55]. Overall, the recombination rate is often comparable,

and even (much) higher than mutation rates in prokaryotes [43].

Obviously, such LGT have allowed important adaptations and the colonisation of new niches. Mobile genetic elements can be seen as a large pool of genetic resources, that prokaryotes can try in different niches. These resources can then be improved and passed on to others micro-organisms, much like open source software engineering [33]. For instance, organisms acquiring one pathogenicity island would begin exploring pathogenic niches that were previously unavailable, therefore making the acquisition of subsequent pathogenicity islands far more favorable [43]. Less damaging to the human world, the ability to form intracellular gas-filled compartments, laterally disseminated among unrelated organisms, allows haloarchaea and cyanobacteria to position themselves at a depth in the water column where the amount of oxygen and light is favorable [48]. Similarly, the beta-proteobacteria, *Rubrivivax gelatinosus*, acquired its 31 photosynthetic genes, organized on a super operon, from an alpha-proteobacteria [48]. Another neat example of adaptive LGT concerns the transfer of the AHL quorum sensing system and its role in the regulation of the life of bacterial communities [48]. Through this mechanism, involving two characteristic and conserved proteins: the AHL or autoinducer synthase (LuxI) and the transcription regulator (LuxR), Proteobacteria can indirectly determine their population density by sensing concentration of a signal molecule. This affects multiple phenotypes in Proteobacteria, such as biofilm formation, exo-enzymes, surface motility, antibiotic production, secondary metabolites, virulence, extracellular polysaccharide, gene transfer agent, and conjugation. According to Boucher [48], LuxR and Lux I would have been transferred at the species, genus and class level. In general, a bacterium's niche seems redefined continually by virtue of the constant influx of DNA, rather than occurring only at the time of lineage diversification [43].

In eukaryotes too, LGT from non-organellar sources (as well as from organellar sources, known as IGT, for internal genetic transfer) have a significant biological impact. LGT from bacteria appear to be an ongoing process in various protists [56–60], which lack a sequestered germline and often engulf their prey, releasing DNA near the nucleus [41]. There would be at least 50 transferred genes in Kinetoplastids [60], 84 in the diplomonad *Spironucleus salmonicida* [60], 96 in the parasite *Entamoeba histolytica* [61], and so on.

However, it is not so much the quantity (limited) than the nature of the transferred genes that makes LGT a relevant theme in protist evolution. In *Enta-*

*moeba histolytica*, 58% of the LGT genes encode a variety of metabolic enzymes, contributing significant enhancements to its metabolism, based on a patchwork of genes of multiple phylogenetic origins (clearly arising from the Cytophaga–Flavobacterium–Bacteroides (CFB) group of the phylum Bacteroidetes, which are abundant in the human digestive tract). Such a mosaic configuration is not odd. For example, while all eukaryotes share a conserved glycolytic pathway, some of its enzyme components have been replaced in various eukaryotic lineages by their eubacterial counterparts, through gene transfer [62]. For instance, out of the ten enzymatic steps which appear to be universal amongst eukaryote glycolytic pathways, the anaerobic flagellate *Trimastix pyriformis* present at least four cases where the relationship of the *Trimastix* genes to homologs from other species differs from accepted organismal relationships: FBA, GAPDH, PGK, and PK have been acquired from a bacterium by LGT, and the phylogenies of two more enzymes (PGAM and PPK) suggest additional LGTs. In other words, about half of the glycolytic enzymes of *Trimastix* were acquired by lateral gene transfer events, likely from different bacterial donors. A comparable situation is observed in the human pathogen *Cryptosporidium parvum*, with the problematic consequence that laterally acquired genes prevent the success of antiparasitic chemotherapy by classical treatments such as antifolates. This parasite is entirely dependent on salvage from the host for its purines and pyrimidines nucleotides, the basic building blocks of DNA and RNA, as well as crucial components of other metabolic processes and of the nucleotide biosynthetic pathways [63]. The loss of pyrimidine *de novo* synthesis is compensated for by possession of three salvage enzymes, unique to *C. parvum* within the phylum Apicomplexa: two of them, the uridine kinase-uracil phosphoribosyltransferase and the thymidine kinase are laterally acquired from either algal or plants and from an alpha- or a gamma-proteobacteria, respectively.

In other protists, such as the chlorarachniophyte *Bigeloviella natans*, LGT have been suggested in significant proportions [40]. These amoebflagellate algae acquired photosynthesis secondarily by engulfing a green alga (likely of chlorophyte origin, i.e. related to *Chlamydomonas reinhardtii*) and retaining its plastid (chloroplast), now surrounded by 4 membranes. Interestingly, the actual chlorarachniophyte plastid proteome is, however, a mosaic derived from various organisms rather than a clone of the ancestral chlorophyte plastid. Out of the 78 *B. natans* genes encoding plastid-targeted proteins, twenty-one percent have been laterally transferred from various organisms: bacteria, red algae,

streptophytes, and algae with red algal endosymbionts. It is even more impressive that, despite the number of membranes isolating the plastid and the nucleus of *B. natans*, nuclear genes encoding plastid-targeted proteins have successfully moved from the nucleus of the endosymbiont (the nucleomorph) to the host nucleus. More surprising still, two recently acquired *B. natans* plastid-targeted proteins were not related to plastid sequences at all, but were instead bacterial proteins. The Calvin cycle enzyme ribulose-5-phosphate 3-epimerase and the GAPDH have been laterally acquired from Pseudomonadaceae and proteobacteria/Gram-positive bacteria, respectively, and thus the success of their transfer in *B. natans* required no less than the addition to these sequences of (a) a signal peptide for the proteins to be directed to the endomembrane system, and (b) a transit-peptide to be targeted to the plastid [64]!

In plants, many cases of transfer from bacteria are also well documented, and probably as fascinating. One of the most ecologically and agriculturally important elemental transformations on the planet – symbiotic nitrogen fixation – is indeed mediated by plasmid-encoded genes of the genus *Rhizobium*. Very large (>250 kb) conjugative plasmids in this genus carry genes for the invasion and the conversion of host-plant root cells into factories that convert atmospheric dinitrogen to ammonia, which meets the nitrogen needs of the plant. Regularly, during pathogenesis, *Agrobacterium* transforms its host with several plasmid-encoded genes, with LGT as a natural consequence [33]. For instance, *Agrobacterium rhizogenes* has donated genes, some of them functional, to members of its host genus *Nicotiana* [65]. Additional putative cases of bacterium-to-plant nuclear genome LGT include the acquisition of aquaglyceroporins from a eubacterium 1200 million years ago [66] and of glutathione biosynthesis genes from an alpha-proteobacterium [67].

Most of the documented cases of transfer in plants, however, seem to involve the mitochondria. For instance, Won and Renner [68] showed that an intron-containing portion of the mitochondrial *nad1* gene had been recently transferred (2–5 million years ago) from an angiosperm (asterid) to a single Asian clade within Gnetum (gymnosperm). A similar example is found for the endophytic holoparasites *Rafflesiaceae*. These organisms, producing the largest flowers in the world, lack leaves, stems, and roots, and rely entirely for their nutrition on their host plants, species of *Tetrastigma* (Vitaceae), with who they live lives as “an almost mycelial haustorial system” [38]. These plants would have acquired part of their mitochondrial genome (such as *nad1B-C*) via LGT from their hosts. Still, the most

emblematic case is offered by *Amborella trichopoda*, a subcanopy shrub, endemic to the South Pacific island of New Caledonia, which lives covered with diverse epiphytes, including mosses and other bryophytes. Its mitochondrial genome may contain more foreign than native DNA [37]. *Amborella trichopoda* is considered to have acquired, via LGT, one or several full-length copies of 20 of its 31 mitochondrial genes. In total, *Amborella trichopoda* presents at least 26 transferred genes in its mitochondria (including 8 pseudogenes). These 26 foreign genes were acquired from a broad range of plant donors: 7 genes were acquired from mosses (cox2, nad5 and nad7 comes from 3 different lineages of moss donors!) and the other 19 from angiosperms (especially eudicots) [36]. As put by Bergthorsson et al., “one wonders how many other *Amborella*-type situations exist among the  $\approx 255,000$  species of flowering plants” [37].

In Metazoa, cases of LGT have also been reported, although more anecdotally. Cyst nematodes, root-knot nematodes and migratory endo-nematodes would have acquired genes from bacteria and fungi, opening them the niche of plant parasitism. More precisely, these nematodes would use cellulases and pectinases of foreign origins that can degrade two major components of plant cell walls, with, for *Meloidogyne*, 12 other laterally inherited proteins: 4 with highest similarity to genes in rhizobia–nitrogen-fixing soil bacteria that nodulate plant roots and 8 with putative functions that might be directly related to the ability of these nematodes to parasitize plants [69]. Another most striking history of transfer from bacteria to *Drosophila* reported to date concerns *Wolbachia pipientis*, a bacterium present in developing gametes of various metazoan. The *Wolbachia* genome was almost entirely transferred to the fly nuclear genome, as evidenced by the presence of PCR-amplified products from 44 of 45 physically distant *Wolbachia* genes in cured strains of *D. ananassae*. Some of these inserted *Wolbachia* genes (2 per cent) are even transcribed within eukaryotic cells lacking endosymbionts. Therefore, heritable lateral gene transfer occurs into eukaryotic hosts from their prokaryote symbionts, potentially providing a mechanism for acquisition of new genes and functions. Interestingly, significant evidence of *Wolbachia*-host transfer are also described in the bean beetle *Callosobruchus chinensis*, in the filarial nematodes *Onchocerca* spp. and *Brugia malayi*, in the mosquito *Culex pipiens quinquefasciatus*, in the wasps *N. giraulti*, *N. longicornis*, *N. vitrepennis*, as well as in the flies *Drosophila simulans* and *Drosophila willis-toni* [70]. Thus it is maybe relevant to be cautious when searching for potential LGT in completely sequenced

eukaryotic genomes, yet not finding much. The whole genome sequencing projects routinely exclude bacterial sequences on the assumption that these represent contamination. Yet, one might wonder whether such a practice could not lead to overlook some *bona fide* bacterial LGT in Metazoans.

Outside the prokaryotes to eukaryotes LGT, transfers between eukaryotes have also been reported (several dozens are identified between protists [60]). For instance, the virulence factors ToxA would have been exchanged from one species of fungal pathogen (*Stagonospora nodorum*) to another one (*Pyrenophora tritici-repentis*), leading to the emergence of a new damaging disease of wheat, shortly after 1941 [71]. The transfer of genes between the filamentous fungi and the oomycetes, two phyla that are amongst the most economically costly plant pathogens, is even more striking. These two phyla are very distantly related, since fungi are sister group to the animals and oomycetes are part of the Chromalveolata, alongside with photosynthetic algae. Yet, convergent evolution had them share their osmotrophic growth habit: both produce thread-like hyphae and secrete enzymes that break down complex nutrients. The resulting simple sugars and amino acids are then recovered by osmotrophy. This feature has been allowed by the transfer of a sugar transporter, a permease, an enzyme degrading lignin derivatives and an enzyme involved in lactose metabolism from these fungi to these oomycetes [72].

Finally, we cannot conclude this section about the biological impact of transfers without evoking gene transfer between organites of a same cell, or IGT. They too can dramatically modify the chromosomic genome of eukaryotes. To pursue with some of our aforementioned case-studies, massive endosymbiotic gene transfers from eukaryote to eukaryote have been observed from the nucleomorph (the reduced nucleus of the eukaryotic endosymbiont) to the nucleus of the mixotrophic alga *Bigeloviella natans* [40]. In addition, IGT from the mitochondrial and the chloroplastic genome would occur relatively frequently in flowering plants [73]. Chloroplast-derived sequences are commonly found in plant mtDNAs. Most impressively, 18% of the nuclear genome of *Arabidopsis thaliana* was interpreted as derived from its plastids. A lower, yet still significant, percentage of IGT (9.1%) was also proposed for the glaucophyte *Cyanophora paradoxa* [74].

To ignore these many events, whose impact varies across lineages and organisms, but is in many cases biologically significant, would be to ignore an important part of the evolutionary history.

#### 4. Consequences of LGT on evolutionary biology

If genome sequencing has taught us anything, it would be that homology-dependent recombination and LGT are much more important, in quantity and quality, than we previously thought [31]. The Darwinian–Mendelian model of parent-to-offspring (‘vertical’) gene flow is thus severely challenged, at least for microbes [75]. Typically, the scientific community has raising doubts regarding the explanatory power, the meaning and the existence of a Tree of Life [76], that was, however, for a long while hiding the phylogenetic forest. While in practice and theory, many prokaryotes are identified and described according to key physiological adaptations, such as photosynthesis, respiration, nitrogen fixation, or sulfur metabolism, vital physiological processes (photosynthesis, methylotrophy, etc.), basic adaptive strategies (halophily, thermophily, etc.) (or more original ones), these ones do not map simply to the SSUrRNA tree or to any unique Tree at all [48]. The traditional naming approach, based on a phenotype that could itself be transferred encounters problems. For example, *E. coli* and *Salmonella enterica* – perhaps the best studied pair of sister species – are distinguished by features that were acquired via LGT (e.g., pathogenicity in *Salmonella* or lactose utilization in *E. coli*) or by gene loss in one lineage [31]. Furthermore, there are in fact very little data for which one can confidently assess their strictly vertical transmission, and, for this reason, universal species trees are often built on a ridiculously small amount of information. Dagan and Martin made this particularly clear when they qualified an apparently magnificent Tree of Life based on 34 core genes (presumably non-transferred) of miserable “tree of one per cent” [77], indicating its utility to generalize about a species’ and lineage’s genomic and genetic evolution was likely close to nil at a broad evolutionary scale.

The role of a unique Tree in the classification and indexation of organisms is also questioned. Genetic connections seems, for some organisms at least, more reticulated than tree-like. Incongruent gene phylogenies (really conflicting topologies) are expected, due to the many independent lateral gene transfers. Trying to reconcile their genuine heterogeneity and to ignore their real disagreement in the name of a unique species tree raises many conceptual issues. Whether we should produce a single tree for a single evolutionary unit (i.e. the species) rather than report the diversity of evolutionary histories out there has become a relevant question for phylogenetics [78], even now, more so that the notion of species itself is now strongly debated in prokaryotes.

For instance, since the middle of the 1970s, Sonea et al. have claimed that LGT forces us to consider all prokaryotes together as a single large species or a global super-biosystem [32]. Lawrence et al. defend a close although much more subtle model, a fragmented view of the prokaryotic species, acknowledging at best, a “fuzzy” species boundary, where ecological distinctiveness counter-selects recombination at some loci, but not at others [79]. Since small DNA fragments are sometimes exchanged between strains during bacterial recombination, different sets of niche-specific genes may be maintained in populations that freely recombine at other loci. Therefore, genetic isolation may be established at different times for different chromosomal regions during speciation as recombination at niche-specific genes is curtailed. Thus, a named species (such as *E. coli*) contains in fact multiple biological species at once. To acknowledge the fact that, in prokaryotes, “*data clearly show that the strategy to sequence one or two genomes per species, which has been used during the first decade of the genomic era, is not sufficient and that multiple strains need to be sequenced to understand the basics of bacterial species*” [80], the fuzzy concepts of pan-genome was proposed as a substitute of the notion of species. The two notions are incommensurable, since “*given that the number of unique genes is vast, the pan-genome of a bacterial species might be orders of magnitude larger than any single genome*” [80]. Typically, species that colonize multiple environments and have multiple ways of exchanging genetic material have an open pan-genome, in which the gene content changes according to the strains, to the environments, even according to the individuals and “*mathematical modelling predicts that new genes will be discovered even after sequencing hundreds of genomes per species*” [80]. However, and most importantly, we have no ecological or evolutionary theory for how species demarcations should relate to the extent of gene sharing between organisms. The recommendation to delineate species using a 70% DNA–DNA binding criterion does not correspond to a theory-based concept of what properties a species should have, but was calibrated empirically to yield many of the phenotype-based species already recognized at the time of its inception [81]. As a result of these issues and alternative propositions, there is currently no unique general notion of species that applies in prokaryotes. According to F. Doolittle [82], the use of the species concept remains pragmatical and thus context-dependent (ecologists are interested in eco-types, traditional phylogeneticists are interested in phylo-types, microbiologists are interested in bio-types ... while none of these categories neces-



sarily matches the others, nor is one more natural than others). What the so-called universal Tree of Species truly classifies when dealing with prokaryotic DNA sequences is thus very arguable.

The appeal to a unique phylotypic classification of life is not convincing either. Why and how to assign a unique phylogenetic position based on its 18SrRNA molecule (for instance) to organisms who carry multiple and potentially divergent copies of this molecule (up to 1000 for a single *Glomus intraradices* cell) [83]! Although this case is clearly an extreme situation, it is nonetheless estimated that any so-called prokaryotic species harbour 4 divergent rRNA copies on average. Moreover, a unique phylotype is also quite insufficient to deal with the classification of the bacterial communities and symbiotic associations. In nature, complexes of unrelated microbes are often the real units of selection and constitute disparate composite evolutionary units, comprising smaller evolutionary units of different origins. It is likely that what has been termed “the great plate count anomaly”, the massive disparity between the number of microbes observed under a microscope in pure culture and those that can be recovered on nutritive agar plates, is due to the existence of such complexes of unrelated microbes [84]. Such a relationship was uncovered, for instance, between closely associated methanotrophic archaea and sulfate-reducing bacteria found in anoxic marine sediments. In this case, the archaeal partner metabolizes methane and the bacteria uses a resulting metabolite as an electron source [85]. Unlike pure isolates, these composite evolutionary unit with important biological properties do not present a single location in the tree, and they can not be classified easily at a single place in a taxonomical hierarchy [78].

Finally, the last task of the Tree (retrodiction) is also challenged. Although ancient LGT could possibly help understanding early evolution [86], their main effect is to put retrodiction attempts under the highest scrutiny. Two very different sorts of lineages – at least – are expected to result from the process of replication, depending on the sources of genetic variation (mutation or lateral transfer) over evolutionary time. First, there are lineages whose genomes have accumulated over time a majority of genes of foreign origins, and for which a majority of the constitutive genes do not show a common pattern of vertical inheritance. We call these lines issued from multiple origins ‘open’ lineages, as they have proved to be highly flexible in their genomic composition. Second, there are lineages whose genomes are in the majority evolving by adapting a basal set of lineage-specific vertically transmitted genes. By contrast with the open lineages, we name such lineages

‘closed’ [87]. Inferences regarding the past of these lineages can not be studied the same way or they will inevitably be misled by imposing an irrelevant model on one of these two very different natural categories [87]. Typically, the closed lineages are amenable to classical phylogenetic analyses. They can be put onto a tree, and their evolution can be in majority thought using the tree-thinking logic. To a certain extent, the history of these “pure” lineages can be inferred, and hypotheses regarding their ancient characteristics can be proposed with reason. However, since open lineages do not evolve in a tree-like manner, their evolution cannot be modelled accurately nor relevantly using the classical model. There is no such thing as a last common ancestor for them, but rather populations of diverse ancestors, which contributed multiple genes to the lineage over time. For these open lineages, it would seem hardly conceivable that classical long term retrodiction should even be attempted: the biology of the X’s of the past had little to do, if anything, with the biology of the X’s of the present (X being for instance a cyanobacterial phylum [88]) or the *Pseudomonas aeruginosa* lineage [89]. Often the question of the origin of a microbe should be replaced by the question of the origins of its many constitutive elements (the various smaller evolutionary units it is made of). This distinction could prefigure a real conceptual and fascinating challenge, a revolution in our way of looking at the history of the living world.

## 5. Conclusion

Traditional phylogenetics is worth it ... to a certain extent. Simply put, its relevance can not be universal. What is true of the elephant evolution is not necessarily so of the *E. coli* one. For micro-organisms, seeing the phylogenetic forest behind the tree seems quite a logical consequence of acknowledging the existence of the multiple processes of inheritance, both vertical and horizontal. Although it challenges traditional phylogenetics up to its roots, it is going to offer us a much more accurate and complete view of the biodiversity and its complex evolution. Importantly, such an open point of view (reckoning the existence of mobile adaptive genes and the variable flexibility of organismal genomic make-up) brings forth large perspectives. One of them is bioremediation, of which we will only detail one example: the remediation of hazardous mixed-waste sites, particularly those co-contaminated with heavy metals and radionuclides, one of the most costly environmental challenges today. While a number of microbes can carry out reductive precipitation of radionuclides, their metal sensitivity suggests that

the acquisition of metal resistance traits (e.g., P-type ATPases that regulate the transport of heavy metals) might be necessary to facilitate and/or enhance microbial metabolism during subsequent biostimulation activities in metal and radionuclide-contaminated subsurface environments. Interestingly, studies of the Field Research Center have determined the presence of PIB-type ATPase genes on mobile genetic elements (i.e., plasmids and transposons) in both gram-positive bacteria and gram-negative bacteria. An experiment was designed, starting from 50 lead-resistant (Pbr) subsurface bacteria (Actinobacteria, Firmicutes, and Proteobacteria phyla present in metal- and radionuclide-contaminated soils of the FRC). It resulted in the amplification of 28 *zntA/cadA/pbrA*-like loci showing evidence of horizontal transfer among 10 Pbr *Arthrobacter spp.* and *Bacillus spp.* strains, illustrating the dissemination of PIB-type ATPases by LGT among their isolates. It was thus evidenced that, thanks to such an association of these unrelated, yet synergistic, physiological traits, *Arthrobacter sp.* and *Bacillus sp.* could be important in promoting the remediation of uranium through either intracellular sequestration or bioadsorption mechanisms [90]. Consequently, microbial phylogenetics, by teaching us more about such genomic flexibility and gene flow dynamics within biological systems, could play a new and significant role. Not only should this knowledge allow us to describe biodiversity and the mechanisms from which it emerges but it also should help us to design more efficient and sensible approaches to preserve extant life forms.

## References

- [1] C. Darwin, On the Origin of Species, John Murray, London, 1859.
- [2] E. Zuckerkandl, L. Pauling, Molecules as documents of evolutionary history, *J. Theor. Biol.* 8 (1965) 357–366.
- [3] C.R. Woese, G.E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms, *Proc. Natl. Acad. Sci. USA* 74 (1977) 5088–5090.
- [4] M.L. Sogin, H.J. Elwood, J.H. Gunderson, Evolutionary diversity of eukaryotic small-subunit rRNA genes, *Proc. Natl. Acad. Sci. USA* 83 (1986) 1383–1387.
- [5] N.R. Pace, G.J. Olsen, C.R. Woese, Ribosomal RNA phylogeny and the primary lines of evolutionary descent, *Cell* 45 (1986) 325–326.
- [6] R.P. Hirt, J.M. Logsdon Jr., B. Healy, M.W. Dorey, W.F. Doolittle, T.M. Embley, Microsporidia are related to fungi: evidence from the largest subunit of RNA polymerase II and other proteins, *Proc. Natl. Acad. Sci. USA* 96 (1999) 580–585.
- [7] T. Hashimoto, Y. Nakamura, T. Kamaishi, J. Adachi, F. Nakamura, K. Okamoto, M. Hasegawa, Phylogenetic place of kinetoplastid protozoa inferred from a protein phylogeny of elongation factor 1 alpha, *Mol. Biochem. Parasitol.* 70 (1995) 181–185.
- [8] H. Philippe, P. Lopez, H. Brinkmann, K. Budin, A. Germot, J. Laurent, D. Moreira, M. Müller, H. Le Guyader, Early branching or fast evolving eukaryotes? An answer based on slowly evolving positions, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 267 (2000) 1213–1221.
- [9] L. Shavit, D. Penny, M.D. Hendy, B.R. Holland, The problem of rooting rapid radiations, *Mol. Biol. Evol.* 24 (2007) 2400–2411.
- [10] W.J. Murphy, E. Eizirik, S.J. O'Brien, O. Madsen, M. Scally, C.J. Douady, E. Teeling, O.A. Ryder, M.J. Stanhope, W.W. de Jong, M.S. Springer, Resolution of the early placental mammal radiation using Bayesian phylogenetics, *Science* 294 (2001) 2348–2351.
- [11] P.J. Lockhart, D. Penny, The place of Amborella within the radiation of angiosperms, *Trends Plant Sci.* 10 (2005) 201–202.
- [12] J. Felsenstein, Cases in which parsimony or compatibility methods will be positively misleading, *Syst. Zool.* 27 (1978) 401–410.
- [13] C.R. Vossbrinck, J.V. Maddox, S. Friedman, B.A. Debrunner-Vossbrinck, C.R. Woese, Ribosomal RNA sequence suggests microsporidia are extremely ancient eukaryotes, *Nature* 326 (1987) 411–414.
- [14] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, *Syst. Biol.* 52 (2003) 696–704.
- [15] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics* 19 (2003) 1572–1574.
- [16] N. Lartillot, H. Philippe, A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process, *Mol. Biol. Evol.* 21 (2004) 1095–1109.
- [17] M. Steel, Should phylogenetic models be trying to “fit an elephant”? *Trends Genet.* 21 (2005) 307–309.
- [18] F. Delsuc, H. Brinkmann, H. Philippe, Phylogenomics and the reconstruction of the tree of life, *Nat. Rev. Genet.* 6 (2005) 361–375.
- [19] B. Snel, M.A. Huynen, B.E. Dutilh, Genome trees and the nature of genome evolution, *Annu. Rev. Microbiol.* 59 (2005) 191–209.
- [20] V. Daubin, M. Gouy, G. Perriere, A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history, *Genome Res.* 12 (2002) 1080–1090.
- [21] N. Rodriguez-Ezpeleta, H. Brinkmann, S.C. Burey, B. Roure, G. Burger, W. Löffelhardt, H.J. Bohnert, H. Philippe, B.F. Lang, Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes, *Curr. Biol.* 15 (2005) 1325–1330.
- [22] S.L. Baldauf, The deep roots of eukaryotes, *Science* 300 (2003) 1703–1706.
- [23] A.G. Simpson, A.J. Roger, The real ‘kingdoms’ of eukaryotes, *Curr. Biol.* 14 (2004) R693–R696.
- [24] S. Gribaldo, C. Brochier-Armanet, The origin and evolution of Archaea: a state of the art, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361 (2006) 1007–1022.
- [25] L.W. Parfrey, E. Barbero, E. Lasser, M. Dunthorn, D. Bhattacharya, D.J. Patterson, L.A. Katz, Evaluating support for the current classification of eukaryotic diversity, *PLoS Genet.* 2 (2006) e220.
- [26] W.J. Murphy, T.H. Pringle, T.A. Crider, M.S. Springer, W. Miller, Using genomic data to unravel the root of the placental mammal phylogeny, *Genome Res.* 17 (2007) 413–421.
- [27] E. Sober, M. Steel, Testing the hypothesis of common ancestry, *J. Theor. Biol.* 218 (2002) 395–408.
- [28] O. Zhaxybayeva, P. Lapierre, J.P. Gogarten, Ancient gene duplications and the root(s) of the tree of life, *Protoplasma* 227 (2005) 53–64.

- [29] A. Stechmann, T. Cavalier-Smith, The root of the eukaryote tree pinpointed, *Curr. Biol.* 13 (2003) R665–R666.
- [30] W.F. Doolittle, E. Baptiste, Pattern pluralism and the tree of life hypothesis, *Proc. Natl. Acad. Sci. USA* 104 (2007) 2043–2049.
- [31] J.P. Gogarten, W.F. Doolittle, J.G. Lawrence, Prokaryotic evolution in light of gene transfer, *Mol. Biol. Evol.* 19 (2002) 2226–2238.
- [32] W.F. Doolittle, Y. Boucher, C.L. Nesbo, C.J. Douady, J.O. Andersson, A.J. Roger, How big is the iceberg of which organellar genes in nuclear genomes are but the tip?, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358 (2003) 39–57; discussion 57–58.
- [33] L.S. Frost, R. Leplae, A.O. Summers, A. Toussaint, Mobile genetic elements: the agents of open source evolution, *Nat. Rev. Microbiol.* 3 (2005) 722–732.
- [34] Y. Boucher, M. Labbate, J.E. Koenig, H.W. Stokes, Integrons: mobilizable platforms that promote genetic diversity in bacteria, *Trends Microbiol.* 15 (2007) 301–309.
- [35] J. Filee, P. Forterre, J. Laurent, The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies, *Res. Microbiol.* 154 (2003) 237–243.
- [36] A.O. Richardson, J.D. Palmer, Horizontal gene transfer in plants, *J. Exp. Bot.* 58 (2007) 1–9.
- [37] U. Bergthorsson, A.O. Richardson, G.J. Young, L.R. Goertzen, J.D. Palmer, Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*, *Proc. Natl. Acad. Sci. USA* 101 (2004) 17747–17752.
- [38] C.C. Davis, K.J. Wurdack, Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales, *Science* 305 (2004) 676–678.
- [39] W. Martin, Gene transfer from organelles to the nucleus: Frequent and in big chunks, *Proc. Natl. Acad. Sci. USA* 14 (2003) 14.
- [40] J.M. Archibald, M.B. Rogers, M. Toop, K. Ishida, P.J. Keeling, Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*, *Proc. Natl. Acad. Sci. USA* 100 (2003) 7678–7683.
- [41] W.F. Doolittle, You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes, *Trends Genet.* 14 (1998) 307–311.
- [42] K.E. Nelson, R.A. Clayton, S.R. Gill, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, W.C. Nelson, K.A. Ketchum, L. McDonald, T.R. Utterback, J.A. Malek, K.D. Linher, M.M. Garrett, A.M. Stewart, M.D. Cotton, M.S. Pratt, C.A. Phillips, D. Richardson, J. Heidelberg, G.G. Sutton, R.D. Fleischmann, J.A. Eisen, C.M. Fraser, Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*, *Nature* 399 (1999) 323–329.
- [43] J.G. Lawrence, Gene transfer in bacteria: speciation without species?, *Theor. Popul. Biol.* 61 (2002) 449–460.
- [44] R.A. Welch, V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E.L. Buckles, S.R. Liou, A. Boutin, J. Hackett, D. Stroud, G.F. Mayhew, D.J. Rose, S. Zhou, D.C. Schwartz, N.T. Perna, H.L. Mobley, M.S. Donnenberg, F.R. Blattner, Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* 99 (2002) 17020–17024.
- [45] R. Sorek, Y. Zhu, C.J. Creevey, M.P. Francino, P. Bork, E.M. Rubin, Genome-wide experimental determination of barriers to horizontal gene transfer, *Science* 318 (2007) 1449–1452.
- [46] J.O. McInerney, D. Pisani, Genetics. Paradigm for life, *Science* 318 (2007) 1390–1391.
- [47] L. Olendzenski, L. Liu, O. Zhaxybayeva, R. Murphey, D.G. Shin, J.P. Gogarten, Horizontal transfer of archaeal genes into the deinococcaceae: detection by molecular and computer-based approaches, *J. Mol. Evol.* 51 (2000) 587–599.
- [48] Y. Boucher, C.J. Douady, R.T. Papke, D.A. Walsh, M.E. Boudreau, C.L. Nesbo, R.J. Case, W.F. Doolittle, Lateral gene transfer and the origins of prokaryotic groups, *Annu. Rev. Genet.* 37 (2003) 283–328.
- [49] Y. Boucher, H. Huber, S. L'Haridon, K.O. Stetter, W.F. Doolittle, Bacterial origin for the isoprenoid biosynthesis enzyme HMG-CoA reductase of the archaeal orders thermoplasmatales and archaeoglobales, *Mol. Biol. Evol.* 18 (2001) 1378–1388.
- [50] Y. Boucher, M. Kamekura, W.F. Doolittle, Origins and evolution of isoprenoid lipid biosynthesis in archaea, *Mol. Microbiol.* 52 (2004) 515–527.
- [51] M. Ibba, S. Morgan, A.W. Curnow, D.R. Pridmore, U.C. Vothknecht, W. Gardner, W. Lin, C.R. Woese, D. Soll, A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases, *Science* 278 (1997) 1119–1122.
- [52] Y.I. Wolf, L. Aravind, N.V. Grishin, E.V. Koonin, 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 (1999) 689–710.
- [53] C.R. Woese, G.J. Olsen, M. Ibba, D. Soll, Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process, *Microbiol. Mol. Biol. Rev.* 64 (2000) 202–236.
- [54] C. Brochier, H. Philippe, D. Moreira, The evolutionary history of ribosomal protein RpS14: horizontal gene transfer at the heart of the ribosome, *Trends Genet.* 16 (2000) 529–533.
- [55] W.H. Yap, Z. Zhang, Y. Wang, 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 (1999) 5201–5209.
- [56] P.J. Keeling, J.D. Palmer, Lateral transfer at the gene and sub-genic levels in the evolution of eukaryotic enolase, *Proc. Natl. Acad. Sci.* 98 (2001) 10745–10750.
- [57] J.O. Andersson, R.P. Hirt, P.G. Foster, A.J. Roger, Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes, *BMC Evol. Biol.* 6 (2006) 27.
- [58] J.O. Andersson, A.J. Roger, Evolution of glutamate dehydrogenase genes: evidence for lateral gene transfer within and between prokaryotes and eukaryotes, *BMC Evol. Biol.* 3 (2003) 14.
- [59] J.O. Andersson, A.M. Sjogren, L.A. Davis, T.M. Embley, A.J. Roger, Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes, *Curr. Biol.* 13 (2003) 94–104.
- [60] J.O. Andersson, A.M. Sjogren, D.S. Horner, C.A. Murphy, P.L. Dyal, S.G. Svard, J.M. Logsdon Jr., M.A. Ragan, R.P. Hirt, A.J. Roger, A genomic survey of the fish parasite *Spirionucleus salmonicida* indicates genomic plasticity among diplomonads and significant lateral gene transfer in eukaryote genome evolution, *BMC Genomics* 8 (2007) 51.
- [61] B. Loftus, I. Anderson, R. Davies, U.C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R.P. Hirt, B.J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoef, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M.A. Quail, E. Rabinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S.E. Stroup, S. Bhattacharya, A. Lohia, P.G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N.M. El-Sayed, W.A. Petri Jr., C.G. Clark, T.M. Embley, B. Barrell, C.M. Fraser, N. Hall, The genome of the protist parasite *Entamoeba histolytica*, *Nature* 433 (2005) 865–868.

- [62] A. Stechmann, M. Baumgartner, J.D. Silberman, A.J. Roger, The glycolytic pathway of *Trimastix pyriformis* is an evolutionary mosaic, *BMC Evol. Biol.* 6 (2006) 101.
- [63] B. Striepen, A.J. Pruijssers, J. Huang, C. Li, M.J. Gubbels, N.N. Umejiego, L. Hedstrom, J.C. Kissinger, Gene transfer in the evolution of parasite nucleotide biosynthesis, *Proc. Natl. Acad. Sci. USA* 101 (2004) 3154–3159.
- [64] J. Raymond, R.E. Blankenship, Horizontal gene transfer in eukaryotic algal evolution, *Proc. Natl. Acad. Sci. USA* 100 (2003) 7419–7420.
- [65] K. Suzuki, I. Yamashita, N. Tanaka, Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution, *The Plant Journal* 32 (2002) 775–787.
- [66] R. Zardoya, X. Ding, Y. Kitagawa, M.J. Chrispeels, Origin of plant glycerol transporters by horizontal gene transfer and functional recruitment, *Proc. Natl. Acad. Sci. USA* 99 (2002) 14893–14896.
- [67] S.D. Copley, J.K. Dhillon, Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes, *Genome Biol.* 3 (2002), research0025.
- [68] H. Won, S.S. Renner, Horizontal gene transfer from flowering plants to *Gnetum*, *Proc. Natl. Acad. Sci. USA* 100 (2003) 10824–10829.
- [69] M. Mitreva, M.L. Blaxter, D.M. Bird, J.P. McCarter, Comparative genomics of nematodes, *Trends Genet.* 21 (2005) 573–581.
- [70] J.C. Hotopp, M.E. Clark, D.C. Oliveira, J.M. Foster, P. Fischer, M.C. Torres, J.D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R.V. Nene, J. Shepard, J. Tomkins, S. Richards, D.J. Spiro, E. Ghedin, B.E. Slatko, H. Tettelin, J.H. Werren, Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes, *Science* 317 (2007) 1753–1756.
- [71] T.L. Friesen, E.H. Stukenbrock, Z. Liu, S. Meinhardt, H. Ling, J.D. Faris, J.B. Rasmussen, P.S. Solomon, B.A. McDonald, R.P. Oliver, Emergence of a new disease as a result of interspecific virulence gene transfer, *Nat. Genet.* 38 (2006) 953–956.
- [72] T.A. Richards, J.B. Dacks, J.M. Jenkinson, C.R. Thornton, N.J. Talbot, Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms, *Curr. Biol.* 16 (2006) 1857–1864.
- [73] Y.-L. Qiu, Y. Cho, J.C. Cox, J.D. Palmer, The gain of three mitochondrial introns identifies liverworts as the earliest land plants, *Nature* 394 (1998) 671–674.
- [74] W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, D. Penny, Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus, *Proc. Natl. Acad. Sci. USA* 99 (2002) 12246–12251.
- [75] R.L. Charlebois, R.G. Beiko, M.A. Ragan, Microbial phylogenomics: Branching out, *Nature* 421 (2003) 217.
- [76] M.A. O'Malley, Y. Boucher, Paradigm change in evolutionary microbiology, *Studies in History and Philosophy of Biological and Biomedical Science* 36 (2005) 183–208.
- [77] T. Dagan, W. Martin, The tree of one percent, *Genome Biology* 7 (2006) 118.
- [78] E. Bapteste, Y. Boucher, Lateral gene transfer challenges: principles of microbial systematics, *Trends Microbiol.* 16 (2008) 200–207.
- [79] A.C. Retchless, J.G. Lawrence, Temporal fragmentation of speciation in bacteria, *Science* 317 (2007) 1093–1096.
- [80] D. Medini, C. Donati, H. Tettelin, V. Masignani, R. Rappuoli, The microbial pan-genome, *Curr. Opin. Genet. Dev.* 15 (2005) 589–594.
- [81] D. Gevers, F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, E.J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F.L. Thompson, J. Swings, Opinion: Re-evaluating prokaryotic species, *Nat. Rev. Microbiol.* 3 (2005) 733–739.
- [82] F. Doolittle, R.T. Papke, Genomics and the bacterial “species problem”, *Curr. Biol.* 7 (2007) 116.
- [83] M. Hijri, H. Niculita, I.R. Sanders, Molecular characterization of chromosome termini of the arbuscular mycorrhizal fungus *Glomus intraradices* (Glomeromycota), *Fungal Genet. Biol.* 44 (2007) 1380–1386.
- [84] J.T. Staley, A. Konopka, Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats, *Annu. Rev. Microbiol.* 39 (1985) 321–346.
- [85] K. Nauhaus, M. Albrecht, M. Elvert, A. Boetius, F. Widdel, In vitro cell growth of marine archaeal-bacterial consortia during anaerobic oxidation of methane with sulfate, *Environ. Microbiol.* 9 (2007) 187–196.
- [86] J. Huang, J.P. Gogarten, Ancient horizontal gene transfer can benefit phylogenetic reconstruction, *Trends Genet.* 22 (2006) 361–366.
- [87] E. Bapteste, Y. Boucher, Epistemological impacts of HGT on classification in microbiology, in: J.P. Gogarten (Ed.), *Genomes in Flux*, Humana Press, 2008, in press.
- [88] O. Zhaxybayeva, J.P. Gogarten, R.L. Charlebois, W.F. Doolittle, R.T. Papke, Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events, *Genome Res.* 16 (2006) 1099–1108.
- [89] E. Lerat, V. Daubin, H. Ochman, N.A. Moran, Evolutionary origins of genomic repertoires in bacteria, *PLoS Biol.* 3 (2005) e130.
- [90] R.J. Martinez, Y. Wang, M.A. Raimondo, J.M. Coombs, T. Barkay, P.A. Sobecky, Horizontal gene transfer of PIB-type ATPases among bacteria isolated from radionuclide- and metal-contaminated subsurface soils, *Appl. Environ. Microbiol.* 72 (2006) 3111–3118.