Horizontal Gene Transfer and the History of Life

Vincent Daubin, Gergely Szöllősi

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Horizontal Gene Transfer and the History of Life

Vincent Daubin[1,2] and Gergely J. Szöllősi[3]

[1]Laboratoire de Biométrie et Biologie Evolutive, Université de Lyon, 69000 Lyon, France

[2]Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5558, Université Lyon 1, 69622 Villeurbanne, France

[3]ELTE-MTA "Lendület" Biophysics Research Group, 1117 Budapest, Hungary

Correspondence: vincent.daubin@univ-lyon1.fr

Microbes acquire DNA from a variety of sources. The last decades, which have seen the development of genome sequencing, have revealed that horizontal gene transfer has been a major evolutionary force that has constantly reshaped genomes throughout evolution. However, because the history of life must ultimately be deduced from gene phylogenies, the lack of methods to account for horizontal gene transfer has thrown into confusion the very concept of the tree of life. As a result, many questions remain open, but emerging methodological developments promise to use information conveyed by horizontal gene transfer that remains unexploited today.

The discovery of the existence of prokaryotic microbes dates back more than 300 years. Since then, our picture of our distant microscopic relatives has undergone several revolutions: from being the living "proofs" of the existence of spontaneous generation, they became later the "archaic" representatives of our distant ancestors, to finally be legitimately recognized as exceptionally diverse organisms, keystone to any ecosystem, including the most familiar and the most hostile environments on Earth. Similarly, although they were first seen as elementary and unbreakable bricks of life, they are now seen as genetically composite bodies, heavyweight champions of "gene robbery." The most recent of these revolutions has indeed been the realization of their unparalleled ability to integrate genetic material coming from more or less evolutionarily distant organisms. This mechanism is called "horizontal gene transfer" as opposed to vertical transmission from mother to daughter cell.

MECHANISMS OF HORIZONTAL TRANSFER

Horizontal Gene Transfer and the Nature of Heredity

The first description of a horizontal gene transfer has been a major advance in molecular biology, and can even be seen as its founding experiment. By demonstrating, in 1928, that nonvirulent pneumococcus bacteria can become pathogenic simply by contact with virulent bacteria, even bacteria destroyed by heat, Griffith (1928) showed that there is a thermostable principle, capable of modifying heredity. This principle would be identified years later as DNA (Avery et al. 1944). This discovery, however, could only take place because of the re-

Editor: Howard Ochman
Additional Perspectives on Microbial Evolution available at www.cshperspectives.org

 doi: 10.1101/cshperspect.a018036
Cite this article as *Cold Spring Harb Perspect Biol* 2016;8:a018036

markable ability of pneumococci to acquire DNA horizontally. We now know that in this experiment, a gene responsible for the synthesis of the polysaccharide capsule of the bacterium is transferred, and incorporated in place of its deficient counterpart in nonvirulent strains.

The cytoplasm of the bacteria, wherein the genome is located, is effectively isolated from the external medium by one or more membranes, depending on the groups of bacteria. DNA cannot passively traverse these obstacles. There are specific mechanisms that facilitate foreign DNA's access to the genome. Three of these are well documented.

- Transformation is an active mechanism by which free DNA present in the medium, typically derived from dead organisms and is taken up into the cytoplasm. This could be mainly for nutritional purposes, but some bacteria are very selective on the type of DNA that they allow into the cell, suggesting that it also serves to favor recombination with close relatives (Redfield et al. 1997; Szöllősi et al. 2006; Mell and Redfield 2014).
- Conjugation is a one-way transmission mechanism of DNA from one cell to another via a "sexual pilus" by which DNA is transported. This mechanism has spuriously been compared with eukaryotic sex. The donor bacterium is described as male, whereas the recipient bacterium is called female. In fact, the genes responsible for conjugation are carried by plasmids or bacteriophages known as "conjugative," that use conjugation to insure their transmission (and thus transform the female into male). Sometimes, however, these conjugative elements accidentally carry with them the DNA of the host, in which case they can promote transfer of genes other than their own.
- Transduction is a type of transfer that occurs via a bacteriophage that transmits the DNA from one cell to another. At the end of its replication cycle, the host cell undergoes lysis, and fragmented DNA of the host genome is occasionally packaged inside infectious particles. This DNA can then be injected into another individual, in place of virus DNA. Some species of bacteria have hijacked this mechanism to their advantage and have recruited bacteriophage genes to facilitate genetic exchange. Such defective phage capsids, present, in particular, in many α-proteobacteria, are called "gene transfer agents" (GTAs) (Lang and Beatty 2007).

Once inside the cytoplasm, DNA has several possible fates. It can be destroyed by DNA degradation systems that are present in the cytoplasm of the host (restriction enzymes, DNAses, etc.) or persist as autonomous replicative entities, such as plasmids. Finally, all or part of the DNA may be integrated into the host chromosome. This integration depends on several factors such as the degree of similarity with genomic DNA of the host, in the case of homologous recombination, or physical association with other sequences capable of integration such as transposable elements or bacteriophage genes. When homologous recombination occurs, the foreign DNA sequence replaces existing homologous sequences in the host genome—this is what happens in the pneumococcus example. On the contrary, when DNA is integrated into the genome by other means, it is often simply inserted as an entirely new gene.

Evolutionary Consequences of Horizontal Transfer

The true evolutionary role and impact that horizontal gene transfer has had on the evolution of life were only realized recently with the advent of genome sequencing. The above mechanisms have relatively low specificity, and thus allow movement of genetic information even between distant species, with correspondingly profound consequences on the modes of adaptation and the concept of bacterial species (Ochman et al. 2005).

In comparison to descent with modification, horizontal gene transfer offers the possibility for quite drastic adaptation. However, far from questioning the principle of Darwinian evolution, as has been suggested, this mode of evolution underscores the importance of taking into account different levels of selection (e.g., genes vs. genomes) for understanding the evo-

lution of genomes. Many genes present in bacterial genomes come from prophages and hence have evolved under different constraints than the rest of the genome. Nonetheless, bacterial genomes have repeatedly co-opted functions from their genomic parasites (Canchaya et al. 2004; Bobay et al. 2014). Also, it has been suggested that the organization in operons of bacterial genomes (i.e., in groups of functionally related genes and cotranscribed) could be the result of a need for coregulation as well as a selection pressure for genes that interact to perform a function to remain together during horizontal transfers. This model is known as the "selfish operon" (Lawrence and Roth 1996; Price et al. 2006). Horizontal gene transfer has also been seen as a barrier to defining species in prokaryotes (and, as we shall see later, also for the concept of prokaryotic phylogeny). Defined in animals on a criterion of interfertility, or on a more molecular level, by the limits of recombination, the biological species is a concept that is difficult to apply to bacteria.

INTRA- AND INTERSPECIFIC GENETIC EXCHANGE

Better than Sex

In sexual eukaryotes, the theoretical advantages of genetic exchange, through the elimination of deleterious mutations and the combination of favorable ones, are well established. Horizontal gene transfer provides the same advantages when the acquisition of DNA is from conspecifics, although recombination is not associated with reproduction. However, these movements of genetic information extend well beyond the species boundaries. Many cases of horizontal transfers have been described, particularly in connection with the acquisition of new functions and colonization of new ecological niches. In particular, the capacity shown by some bacterial strains of acquiring virulence genes or antibiotic resistance remains a major health problem. Yet, one can wonder whether these cases are anecdotal, or if, instead, the horizontal transfer is a mechanism that plays a key role in the evolution of prokaryotes. Answering this question involves detecting transfers between species, to quantify their importance.

The Many Facets of Horizontal Transfer

There are three major types of approaches to identifying horizontal transfers in completely sequenced genomes:

1. Methods for comparison of gene repertoires contrast genomes of related species or strains of the same species and often reveal very different gene content. These differences in gene repertoires are explained by gene losses (deletions) and/or gains. However, as in the case of *Escherichia coli*, bacterial strains of the same species frequently have hundreds of genes that are strain specific. In such cases, we must either assume an ancestral genome of astronomical size, to interpret these differences by genes losses, or accept that differences between closely related strains are mostly the result of recent integrations (Daubin et al. 2003a; Touchon et al. 2009).
2. Methods of gene composition analysis rely on the long-standing observation that there is a great diversity of G+C content among bacterial genomes (Sueoka 1962) and that each genome has a base composition and codon usage that can be diagnostic of the species. However, regions having contrasting nucleotide or codon composition are often found in bacterial genomes. This suggests that such regions originate from recent transfer from genomes having different compositions (Lawrence and Ochman 1998; Ochman et al. 2000). As relatively closely related organisms (such as enterobacteriaceae, for example) have comparable G+C content, genes with diverging G+C content are generally considered as originating from more distant organisms. Overall, these methods are expected to underestimate the number of transfers, for two reasons. First, they are not able to identify transfers from species having a composition similar to the host genome. Second, these methods cannot find ancient transfers, because these genes, once integrated, gradually acquire the char-

acteristics of the host genome. It is worth noting, although, that some mechanisms probably maintain an heterogeneity of G+C content of genes within bacterial genomes, which could be deceiving for these methods (Guindon and Perrière 2001; Daubin and Perrière 2003; Lassalle et al. 2015).

3. Phylogenetic methods are the most general, and potentially most sensitive, methods for inferring horizontal transfers. These methods reconstruct the history of a family of homologous genes (a gene tree), and compare it to a putative history of the species in which the genes are found (the species tree). Phylogenetically well-supported disagreements between the two trees can be interpreted in terms of transfers. Such methods have the advantage of providing information on the donor species (not just the recipient) and can, in theory, identify ancient transfers. The difficulty of this approach is that it requires a reference phylogeny, and the ability to differentiate between different types of events that can change the history of a gene such as duplications, losses, and transfers (see Fig. 1). Most studies aimed at evaluating the role of gene transfer using phylogenetic approaches have tried to circumvent the problem of duplications and loss of genes by focusing on genes that are present in at most one copy in each genome (Beiko et al. 2005; Than et al. 2008; Abby et al. 2010, 2012; Puigbò et al. 2010). Only recently, new methods have been developed that can sort out the role of duplication, transfer, and loss in gene histories (Bansal et al. 2012; Szöllősi et al. 2012, 2013b; Sjöstrand et al. 2014). A crucial ingredient of any phylogenetic method that aims at detecting gene transfer is the ability to account for phylogenetic uncertainty. Taking into consideration phylogenetic uncertainty is important, because limited signal leads to reconstruction errors that subsequently result in a gross overestimate of the amount of horizontal gene transfer. To overcome this problem, it is possible to exploit the fact that, although each homologous gene family has its own unique story, they are all related by a shared species history, and this history can be helpful for gene tree inference. In a study that attempted to take into account shared species history for gene tree reconstruction in cyanobacteria (Szöllősi et al. 2013a), the majority of phylogenetic discord was found to result from reconstruction errors. This can be corrected

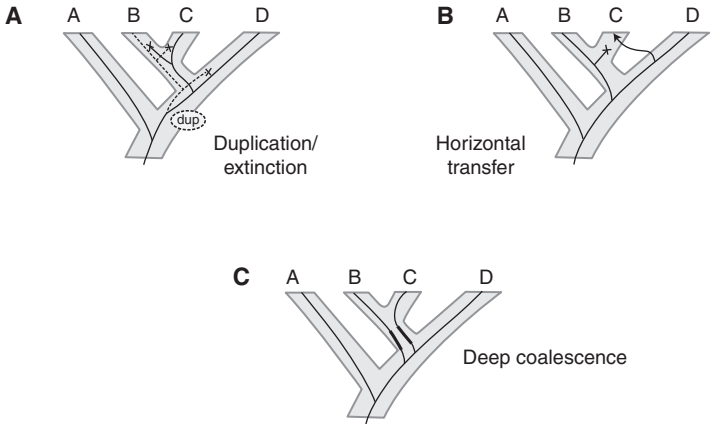

Figure 1. The processes of discord. Three biological processes can generate gene trees that differ from the species tree. (*A*) The combined action of gene duplication and loss, (*B*) horizontal gene transfer, and (*C*) deep coalescence, where polymorphic alleles can remain present in a population for a time than spans two speciations (black squares show the alleles that coexist for this period). In each of these examples, the genes from species *C* and *D* are closest relatives, although species *C* is more closely related to species *B* (adapted from Maddison 1997).

by combining information from sequence alignment with information from a putative species tree and probabilities of duplication, transfer, and loss optimized across gene families. The result is a striking reduction in apparent phylogenetic discord, with 24%, 59%, and 46% reductions, respectively, in the mean numbers of duplications, transfers, and losses per gene family.

More generally, the three approaches described above give seemingly very different ideas about the impact of horizontal transfer on genomes (Ragan 2001; Lawrence and Ochman 2002). The first two show that genomes can contain high proportions of "foreign" genes, up to 20%, acquired recently. The third approach, in contrast, has limited power to detect very recent transfers, but phylogenetic incongruities are clearly evident when studying the genomes of distant species (Daubin et al. 2003b). This apparent contradiction between the different approaches can, in fact, be interpreted as a fundamental underlying difference in the time scale considered. An individual genome sequence corresponds to the shortest time scale. It is a snapshot of the genetic information of a species (or strain) at a particular instance in time. Such a genomic snapshot can contain hundreds of recently acquired genes, the overwhelming majority of which are destined to disappear in the short term, leaving little trace in gene phylogenies (Daubin et al. 2003b; Lerat et al. 2005). Analysis of recently acquired genes shows that the vast majority are orphans, or "ORFans" (i.e., genes that have no homologs in the other known genomes) (Siew and Fischer 2003). One possible explanation is that these genes originate from bacteriophages that have integrated into the genome. To be retained in the longer term, they would have to turn out to be useful for their host (Daubin et al. 2003a; Daubin and Ochman 2004a,b; Cortez et al. 2009; Bobay et al. 2014).

Horizontal Transfer and Adaptation

The adaptive role of horizontal transfer in bacteria is well established. In particular, there are long stretches of genomes called "islands" that are only present sporadically in a given species, and are associated with pathogenicity, symbiosis with another organism (e.g., a plant), or other ecological characteristics (Dobrindt et al. 2004). The grouping of these genes in islands is probably a result of the fact that these genes are associated with mobile elements (transposons, bacteriophages) that tend to recombine with each other and thus are inserted in close proximity to each other, which then promotes their simultaneous transfer from one genome to another. Interestingly, these islands usually contain numerous "ORFans" whose function, if they have one, has not yet been discovered (Siew and Fischer 2003).

These examples are evidence for the existence of recent horizontal transfer in conjunction with immediate environmental benefits, but there are also many indications suggesting that ancient transfers have been able to touch even the most fundamental cellular function. One example is the case of reverse gyrase, a protein that changes the conformation of the chromosome, is found in all hyperthermophilic and some thermophilic organisms (Bacteria and Archaea) and is thought to have been acquired repeatedly to adapt to this unique environment (Brochier-Armanet and Forterre 2006). Also, many examples of transfers of genes encoding tRNA synthetases, whose function is to load the amino acids onto their tRNAs before translation, have been described between phylogenetically distant organisms (Fournier et al. 2015). The selection pressures that promote such transfers are still poorly understood, but it is possible that these enzymes, which generally operate without interacting with other proteins, and whose substrates (an amino acid and tRNA) are highly conserved in evolution and can hence adapt relatively easily to a new cellular environment. This type of reasoning is consistent with the "hypothesis of complexity" (Jain et al. 1999), which maintains that the number of molecular interactions of the protein encoded by a gene is a barrier to transfer. For example, genes involved in complex molecular structures, such as ribosomes or DNA replication machinery, are considered less likely to be replaced by remote homologs, because they have a low

probability of having preserved their numerous sites of interactions intact. However, there are exceptions to this rule, and, specifically, certain ribosomal genes show clear traces of horizontal transfer (Brochier et al. 2000).

EVOLUTION OF GENE REPERTOIRE

The Genomic Diversity of Life

Our appreciation of the magnitude of extant diversity has broached new frontiers with the advent of comparative genomics and the recognition of the differences in gene repertoires between genomes. These differences are surprising at all evolutionary scales, from the oldest to most recent. As we have seen above, the genomes of different strains of the same bacterial species can differ by several hundred genes. The importance of this phenomenon is such in bacteria that it has led to the creation of the concept of the "pan-genome," the union of genes represented in all individuals of the species. In *E. coli*, the pan-genome comprises >20,000 genes (and continues to gradually expand as new genomes are sequenced), whereas the "core genome," the set of genes shared by all strains, comprises <2000 genes (Touchon et al. 2009). At the other extreme, although all cellular organisms have a DNA genome that they replicate, transcribe into messenger RNA, and translate using a quasi-universal genetic code, only about 60 genes have been found to be common to all living genomes (Koonin 2003; Charlebois and Doolittle 2004), a number that is clearly insufficient to perform all, or even any one, of these universal functions. Several phenomena have been proposed to account for the marked differences in the gene repertoire:

- Gene duplication;
- Gene loss;
- Horizontal transfer of genes from cellular organisms or viruses;
- Saturation of the signal of homology. After diverging for a very long time or with very high rates of evolution, genes can accumulate so many substitutions that they will no longer be recognizable as homolog;
- The origin of new genes from nonfunctional sequences, or combinations of pieces of preexisting genes.

The relative contribution of the different phenomena has been difficult to assess, and each probably accounts for disparities at different evolutionary scales. Differences in gene repertoires between strains of the same species are most probably explained by recent horizontal transfer. Although the diversity of genes present in the pan-genome of certain bacterial and archaeal species contains a reservoir of genes allowing to adapt to a variety of conditions (Szöllősi et al. 2006, cf. above; Touchon et al. 2009), it seems more plausible to see these variations of gene content essentially as the result of a highly dynamic process in which DNA integration from selfish elements is quickly counterbalanced by deletion of even slightly harmful DNA (Daubin et al. 2003a; Collins et al. 2011; Szöllősi and Daubin 2012). In a way, the pan-genome observed in bacteria could be an equivalent to the junk DNA found in many eukaryotic genomes, with the difference that eukaryotic species with small population sizes are usually inapt to eliminate this DNA (Lynch and Conery 2003). In Bacteria and Archaea, some lineages show strong tendencies to reduction of their genome, a feature that is generally related to lifestyle such as obligatory parasitism or endosymbiosis (Ochman and Moran 2001). It has been proposed that there is a deletion bias in all bacterial genomes, normally offset by horizontal transfer, and that endosymbionts and other parasites living in very confined spaces have lost this source of new genes (Mira et al. 2001). On the other hand, if the gene duplications seem to occur spontaneously in bacterial genomes, multigene families are relatively rare in these organisms, compared with eukaryotes, and a substantial portion of the genes of these multigene families could have expanded by horizontal gene transfer rather than duplication (Lerat et al. 2005; Treangen and Rocha 2011). At deeper evolutionary time scales, the differences between gene content among domains of life could be explained by the saturation of the signal of homology. Basic approaches searching for

nearly universal genes usually identify only a few genes (<60) that can be safely traced to the last universal cellular ancestor (LUCA). This is obviously not sufficient to accomplish any of the functions that are believed to have been present in this organism (e.g., functions that are universal to cellular organisms), such as DNA replication, transcription, translation, or membrane synthesis. It is of course possible that some of these functions were acquired after LUCA independently by its descendants (Koonin and Martin 2005; Forterre 2013), but the loss of the signal for recognizing homologs is very likely to play a significant role in the failure to assign genes to ancestral genomes (Daubin and Ochman 2004a; Elhaik et al. 2006).

HORIZONTAL GENE TRANSFER AND THE TREE OF LIFE

After a century of stasis, distinguished only by the promotion of "Monera" and Fungi from the rank of phyla to kingdoms (Haeckel 1866; Whittaker 1969), the tree of life underwent radical reorganization with the rise of molecular phylogeny. It was in the 1970s, with the work of Fox and Woese (Fox et al. 1977; Woese and Fox 1977; Woese et al. 1990) on the RNA of the small subunit of the ribosome (termed 16S in prokaryotes and 18S in eukaryotes), that molecular phylogeny was systematically implemented to establish a phylogenetic classification of the prokaryotic world, and more generally a tree of life. Since then, the diversity of life has been seen as comprised of three major domains, Archaea, Bacteria, and Eukarya. Most of the diversity recognized before Woese (e.g., by Haeckel or Whittacker) belonged to the last domain, Eukarya, but because we have the tools to perceive genome diversity, the first two domains have come to be appreciated to be at least as diverse.

However, as popular as the tree of life based on 16/18S RNA has been, it quickly proved to be limited in its ability to resolve important parts of the history of life as other gene markers contradicted many of the relationships it suggested. Most notably, other phylogenetic markers supported different relationships in the early branches of the tree, especially among bacterial (or archaeal) phyla and between Archaea and Eukarya (Brown and Doolittle 1997). With the development of genome sequencing, the combination of genetic markers to resolve deep relationships in the tree of life became popular (Delsuc et al. 2005), but again led only to ephemeral conclusions: different so-called "phylogenomic" studies identified very different numbers of phylogenetic markers potentially informative to infer the tree of life (with variations among studies from 14 to >50 genes), each combination of these markers pleading for different trees, and different phylogenetic methods yielded conflicting signals (see review in Gribaldo et al. 2010). Metagenomic studies are now bringing new incentive to these approaches by providing genomic sequences for previously unknown organisms, which turn out to be key to resolve, for instance, the relationships between Eukarya and Archaea (Spang et al. 2015). These new results indicate that Eukarya are in fact nested within the archaeal domain rather than being its sister group, a relationship that was proposed earlier based on individual markers (Rivera and Lake 1992; Tourasse and Gouy 1999). However, none of these approaches account for the fact that individual genes have complex and, most certainly, different histories, but rather simply combine phylogenetic markers, constraining them to a shared story, and as a result are bound to produce indecisive representations of the history of life. Reconstructing the history of life based on molecular data will require the development of more principled methodologies that explicitly deal with the complex processes of gene evolution (Boussau and Daubin 2010; Szöllősi et al. 2015).

Gene Tree/Species Tree Models Are Key to Reconstructing the History of Life

Although phylogeny seeks to reconstruct the relationships among species (etymologically, the genesis of species), molecular phylogenetics has long focused on a different objective. For contingent and practical reasons, molecular phylogenetic analysis has mainly developed in the context of the analysis of the histories of "genes." The evolution of biological sequences

is subject to many factors and complex processes, but one of these processes has been the subject of most studies: the replacement of a letter of the nucleotide (or protein) alphabet by another. Models attempting to describe these substitutions are numerous, and are becoming more and more complex, gradually relaxing simplifying assumptions made by previous models (Felsenstein 2004). We now know that the probability of replacement of a nucleotide or amino acid by another varies according to its biochemical properties; that all sites of a molecule are not subject to the same constraints and that their evolutionary rate varies accordingly; that these constraints also vary over time and that changes within a molecule can facilitate or prevent others; and that the conditions in which an organism lives can also affect these processes. Increasingly sophisticated attempts have been made to incorporate these into models of sequence evolution, but much remains to be accounted for (Boussau and Gouy 2006; Dutheil and Boussau 2008; Lartillot et al. 2009; Lartillot and Poujol 2010; Matsumoto et al. 2015).

As sophisticated as these models become, they cannot solve the problem of phylogeny in its primary sense. For the history of the species is not a simple transposition of the history of a gene or a genomic sequence (Maddison 1997). As we have seen before, the history of a gene is marked by a series of speciation, duplication, losses, and horizontal transfer (Fig. 1). In addition, genes evolve in populations, and different allelic forms of a gene can coexist for periods that may span several speciation events. All of these events prevent us from interpreting the phylogeny of a gene directly as a phylogeny of species. Hence, to reconstruct the history of species based on molecular data it is necessary, in addition to modeling sequence evolution, to be able to model the relationships between a gene tree and a species tree. This is currently an active avenue of research, although a complete model accounting for all mechanisms is still lacking (Szöllősi et al. 2015). Most relevant to the inference of the tree of life, probabilistic models combining the likelihood of gene duplication, transfer, and loss scenarios and sequence evolution have been recently developed (Szöllősi et al. 2013a; Sjöstrand et al. 2014). The previously mentioned application to cyanobacteria has shown that these models allow the reconstruction of a species tree based on the full set of genes present in genomes, and not only those widely represented among species (Szöllősi et al. 2012, 2013a,b). Importantly, under a model accounting for lateral gene transfer, the reconstructed species tree contains information on the relative timing of speciation of different groups, because gene transfer can only occur between species that have been contemporaries (Fig. 2) (Szöllősi et al. 2012). This information is completely independent from any other information about the timing of events, such as fossils or molecular clocks, and can therefore be used to complete these data. It has been shown that, using the complete set of gene trees that can be reconstructed from a set of genomes, one can reconstruct a fully resolved species tree, with events of speciation ordered in time (Szöllősi et al. 2012). This approach has been applied only to parts of the tree of life because it is computationally intensive and cannot yet be applied to hundreds or thousands of species. However, its advantages are many because it also provides an explicit reconstruction of the events of duplication, loss, and transfer along the phylogeny, and hence also reconstruct ancestral genomes at each node of the species tree. Future developments will hopefully allow the generalization of such methods to larger data sets and an application to a well-sampled tree of life.

CONCLUDING REMARKS

Lateral gene transfer is a powerful evolutionary force, allowing the combination of molecular functions well beyond the species boundaries. However, its success as a process has long been seen by evolutionary biologists as a barrier to reconstructing the patterns of evolution (i.e., the tree of life). This difficulty is not so much conceptual as methodological and recent developments allow to use lateral gene transfer as information for reconstructing the history of life. Under this view, gene transfer even unveils unforeseen opportunities to address long-

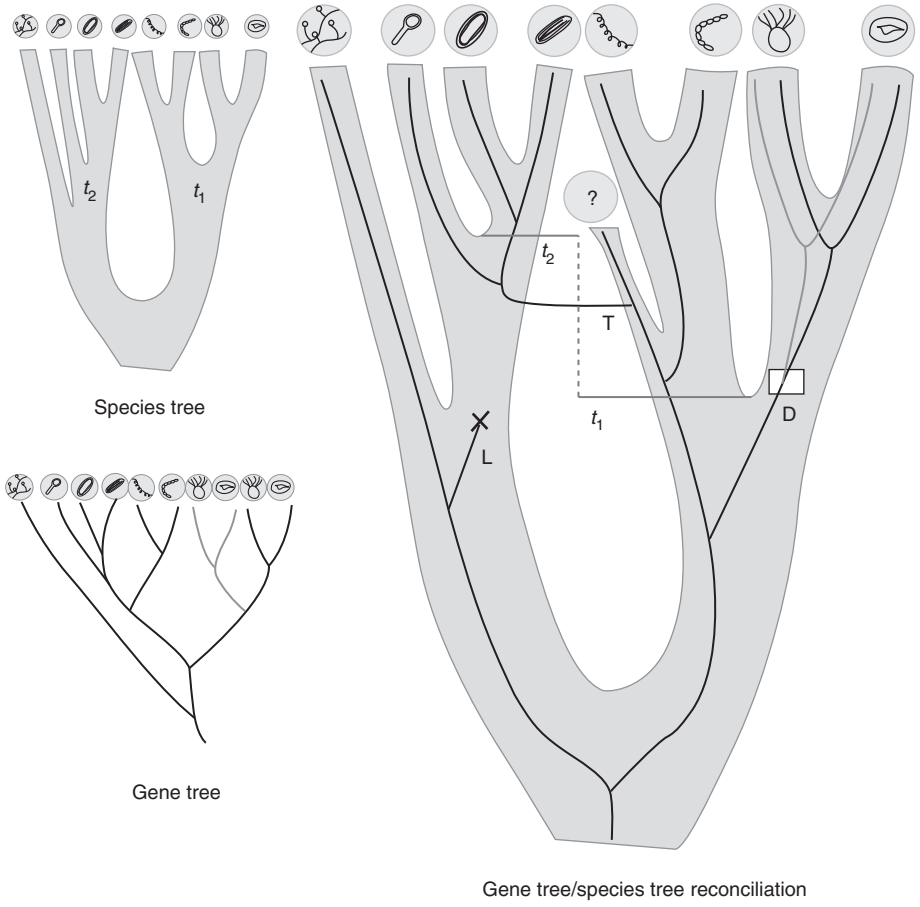

Figure 2. Gene tree/species tree reconciliation and the timing of events. Models of reconciliation invoking horizontal gene transfer (T, in addition to duplications, D, and losses, L) implicitly or explicitly imply a partial order of evolutionary events in a tree. Here, the scenario of reconciliation of the gene tree and the species contains a transfer that implies that the speciation at time t_1 occurred before the speciation at t_2. The reconciliation of a large number of gene trees (typically, from all the homologous genes represented in the genomes under study) with a species tree can yield a fully resolved time order of evolutionary events (Szöllősi et al. 2012).

standing questions such as the reconstruction of the timing of life evolution or the root of the tree of life. The recent realization that horizontal gene transfer has been extensive throughout the evolution of the eukaryotic domain (Keeling and Palmer 2008; Andersson 2009) will allow to compare the information derived from genome histories to the fossil record.

ACKNOWLEDGMENTS

We thank Sophie Abby for allowing us to use her drawings in Figure 2. V.D. is supported by the French Agence Nationale de la Recherche (ANR) through Grant (ANR-10-BINF-01-01) "Ancestrome." G.J.S. is supported by Marie Curie Grant CIG 618438 "Genestory."

REFERENCES

Abby SS, Tannier E, Gouy M, Daubin V. 2010. Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests. *BMC Bioinformatics* **11:** 324.

Abby SS, Tannier E, Gouy M, Daubin V. 2012. Lateral gene transfer as a support for the tree of life. *Proc Natl Acad Sci* **109:** 4962–4967.

Andersson JO. 2009. Horizontal gene transfer between microbial eukaryotes. *Methods Mol Biol* **532:** 473–487.

Avery OT, Macleod CM, McCarty M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J Exp Med* **79:** 137–158.

Bansal MS, Alm EJ, Kellis M. 2012. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. *Bioinformatics* **28:** 283–i291.

Beiko RG, Harlow TJ, Ragan MA. 2005. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci* **102:** 14332–14337.

Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. *Proc Natl Acad Sci* **111:** 12127–12132.

Boussau B, Daubin V. 2010. Genomes as documents of evolutionary history. *Trends Ecol Evol* **25:** 224–232.

Boussau B, Gouy M. 2006. Efficient likelihood computations with nonreversible models of evolution. *Syst Biol* **55:** 756–768.

Brochier-Armanet C, Forterre P. 2006. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea* **2:** 83–93.

Brochier C, Philippe H, Moreira D. 2000. The evolutionary history of ribosomal protein RpS14: Horizontal gene transfer at the heart of the ribosome. *Trends Genet* **16:** 529–533.

Brown JR, Doolittle WF. 1997. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* **61:** 456–502.

Canchaya C, Fournous G, Brüssow H. 2004. The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53:** 9–18.

Charlebois RL, Doolittle WF. 2004. Computing prokaryotic gene ubiquity: Rescuing the core from extinction. *Genome Res* **14:** 2469–2477.

Collins RE, Merz H, Higgs PG. 2011. Origin and evolution of gene families in bacteria and archaea. *BMC Bioinformatics* **12:** S14.

Cortez D, Forterre P, Gribaldo S. 2009. A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol* **10:** R65.

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

Daubin V, Ochman H. 2004b. Start-up entities in the origin of new genes. *Curr Opin Genet Dev* **14:** 616–619.

Daubin V, Perrière G. 2003. G+C3 structuring along the genome: A common feature in prokaryotes. *Mol Biol Evol* **20:** 471–483.

Daubin V, Lerat E, Perrière G. 2003a. The source of laterally transferred genes in bacterial genomes. *Genome Biol* **4:** R57.

Daubin V, Moran NA, Ochman H. 2003b. Phylogenetics and the cohesion of bacterial genomes. *Science* **301:** 829–832.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* **6:** 361–375.

Dobrindt U, Hochhut B, Hentschel U, Hacker J. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2:** 414–424.

Dutheil J, Boussau B. 2008. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. *BMC Evol Biol* **8:** 255.

Elhaik E, Sabath N, Graur D. 2006. The "inverse relationship between evolutionary rate and age of mammalian genes" is an artifact of increased genetic distance with rate of evolution and time of divergence. *Mol Biol Evol* **23:** 1–3.

Felsenstein J. 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.

Forterre P. 2013. Why are there so many diverse replication machineries? *J Mol Biol* **425:** 4714–4726.

Fournier GP, Andam CP, Gogarten JP. 2015. Ancient horizontal gene transfer and the last common ancestors. *BMC Evol Biol* **15:** 70.

Fox GE, Magrum LJ, Balch WE, Wolfe RS, Woese CR. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci* **74:** 4537–4541.

Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C. 2010. The origin of eukaryotes and their relationship with the Archaea: Are we at a phylogenomic impasse? *Nat Rev Microbiol* **8:** 743–752.

Griffith F. 1928. The significance of pneumococcal types. *J Hyg (Lond)* **27:** 113–159.

Guindon S, Perrière G. 2001. Intragenomic base content variation is a potential source of biases when searching for horizontally transferred genes. *Mol Biol Evol* **18:** 1838–1840.

Haeckel EH. 1866. *Generelle morphologie der organismen*. Georg Reimer, Berlin.

Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proc Natl Acad Sci* **96:** 3801–3806.

Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9:** 605–618.

Koonin EV. 2003. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* **1:** 127–136.

Koonin EV, Martin W. 2005. On the origin of genomes and cells within inorganic compartments. *Trends Genet* **21:** 647–654.

Lang AS, Beatty JT. 2007. Importance of widespread gene transfer agent genes in α-proteobacteria. *Trends Microbiol* **15:** 54–62.

Lartillot N, Poujol R. 2010. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol Biol Evol* **28:** 729–744.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25:** 2286–2288.

Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, Daubin V. 2015. GC-content evolution in bacterial genomes: The biased gene conversion hypothesis expands. *PLoS Genet* **11:** e1004941.

Lawrence JG, Ochman H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci* **95:** 9413–9417.

Lawrence JG, Ochman H. 2002. Reconciling the many faces of lateral gene transfer. *Trends Microbiol* **10:** 1–4.

Lawrence JG, Roth JR. 1996. Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics* **143:** 1843–1860.

Lerat E, Daubin V, Ochman H, Moran NA. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3:** e130.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302:** 1401–1404.

Maddison WP. 1997. Gene trees in species trees. *Syst Biol* **46:** 523–536.

Matsumoto T, Akashi H, Yang Z. 2015. Evaluation of ancestral sequence reconstruction methods to infer nonstationary patterns of nucleotide substitution. *Genetics* **200:** 873–890.

Mell JC, Redfield RJ. 2014. Natural competence and the evolution of DNA uptake specificity. *J Bacteriol* **196:** 1471–1483.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17:** 589–596.

Ochman H, Moran NA. 2001. Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. *Science* **292:** 1096–1099.

Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405:** 299–304.

Ochman H, Lerat E, Daubin V. 2005. Examining bacterial species under the specter of gene transfer and exchange. *Proc Natl Acad Sci* **102:** 6595–6599.

Price MN, Arkin AP, Alm EJ. 2006. The life-cycle of operons. *PLoS Genet* **2:** e96.

Puigbò P, Wolf YI, Koonin EV. 2010. The tree and net components of prokaryote evolution. *Genome Biol Evol* **2:** 745–756.

Ragan MA. 2001. On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol Lett* **201:** 187–191.

Redfield RJ, Schrag MR, Dean AM. 1997. The evolution of bacterial transformation: Sex with poor relations. *Genetics* **146:** 27–38.

Rivera MC, Lake JA. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257:** 74–76.

Siew N, Fischer D. 2003. Twenty thousand ORFan microbial protein families for the biologist? *Structure* **11:** 7–9.

Sjöstrand J, Tofigh A, Daubin V, Arvestad L, Sennblad B, Lagergren J. 2014. A Bayesian method for analyzing lateral gene transfer. *Syst Biol* **63:** 409–420.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521:** 173–179.

Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci* **48:** 582–592.

Szöllősi GJ, Daubin V. 2012. Modeling gene family evolution and reconciling phylogenetic discord. *Methods Mol Biol* **856:** 29–51.

Szöllősi GJ, Derényi I, Vellai T. 2006. The maintenance of sex in bacteria is ensured by its potential to reload genes. *Genetics* **174:** 2173–2180.

Szöllősi GJ, Boussau B, Abby SS, Tannier E, Daubin V. 2012. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc Natl Acad Sci* **109:** 17513–17518.

Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013a. Efficient exploration of the space of reconciled gene trees. *Syst Biol* **62:** 601–912.

Szöllősi GJ, Tannier E, Lartillot N, Daubin V. 2013b. Lateral gene transfer from the dead. *Syst Biol* **62:** 386–397.

Szöllősi GJ, Tannier E, Daubin V, Boussau B. 2015. The inference of gene trees with species trees. *Syst Biol* **64:** e42–e62.

Than C, Ruths D, Nakhleh L. 2008. PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9:** 322.

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, et al. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5:** e1000344.

Tourasse NJ, Gouy M. 1999. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol Phylogenet Evol* **13:** 159–168.

Treangen TJ, Rocha EPC. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7:** e1001284.

Whittaker RH. 1969. New concepts of kingdoms of organisms. *Science* **163:** 150–160.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci* **74:** 5088–5090.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci* **87:** 4576–4579.