Frédéric Bouchard
Département de philosophie
Université de Montréal
f.bouchard@umontreal.ca

Philosophy and evolution: minding the gap between Evolutionary Patterns and Tree-like Patterns

Final version published in
Eric Bapteste, Frederic Bouchard, and Richard M. Burian (2012),"Philosophy and Evolution: Minding the Gap Between Evolutionary Patterns and Tree-Like Patterns", in Evolutionary Genomics: Statistical and Computational Methods (Vol 2), Series: Methods in Molecular Biology (Vol. 856), ed by Maria Anisimova, New York: Humana Press (Springer)
http://www.springer.com/biomed/human+genetics/book/978-1-61779-584-8

Eric Bapteste(1), Frédéric Bouchard(2), and Richard M. Burian (3)

(1)UMR CNRS 7138, UPMC, 7, Quai Saint Bernard, 75005 Paris, France; e-mail: eric.bapteste@snv.jussieu.fr

(2)Département de Philosophie, Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Québec, Canada H3C 3J7; email: f.bouchard@umontreal.ca; www.fredericbouchard.org

(3) Department of Philosophy – 0126, Virginia Tech, Blacksburg, VA 24061, USA; e-mail: rmburian@vt.edu : www.phil.vt.edu/Burian

ABSTRACT: Ever since Darwin, the familiar genealogical pattern known as the Tree of Life has been prominent in evolutionary thinking and has dominated not only systematics, but also the analysis of the units of evolution. However, recent findings indicate that the evolution of DNA, especially in prokaryotes and in such DNA vehicles as viruses and plasmids, does not follow a unique tree-like pattern. Because evolutionary patterns track a greater range of processes than those captured in genealogies, genealogical patterns are in fact only a subset of a broader set of evolutionary patterns. This fact suggests that evolutionists who focus exclusively on genealogical patterns are blocked from providing a significant range of genuine evolutionary explanations. Consequently, we highlight challenges to tree-based approaches, and point the way toward more appropriate methods to study evolution (although we do not present them in technical detail). We argue that there is significant benefit in adopting wider range of models, evolutionary representations and evolutionary explanations, based on an analysis of the full range of evolutionary processes. We introduce an ecosystem –orientation into evolutionary thinking that highlights the importance of "type 1 coalitions" (functionally related units with genetic exchanges, aka 'friends with genetic benefits'), "type 2 coalitions" (functionally related units without genetic exchanges), "communal interactions", and "emergent evolutionary properties". On this basis, we seek to promote the study of (especially prokaryotic) evolution with dynamic evolutionary networks, which are less constrained than a TOL, and to provide new ways to analyze an expanded range of evolutionary units (genetic modules, recombined genes, plasmids, phages and prokaryotic genomes, pangenomes, and microbial communities), and of evolutionary processes. Finally, we discuss some of the conceptual and practical questions raised by such network-based representation.

Keywords: Network, Lateral Gene Transfer, Horizontal Gene Transfer, Evolution, Prokaryotes, Philosophy of Biology, units of evolution

"*Is the phylogenetic or a definitely nonphylogenetic system (e.g.., an idealistic-morphological system) better suited to serve as a general reference system, or does one of these systems for intrinsic reasons demand this precedence over all others?*' [1]

1. Genealogical Patterns and Evolutionary Patterns are two different things

Decades of phylogenetic research and practice provided Hennig's followers with a firm answer to his question: they held that the phylogenetic system should be preferred for the study of evolution and that such work allows the reconstruction of a Tree of Life (TOL). For his supporters, a TOL provides a universal, natural, practical and heuristic framework for evolutionary research [2-5]. One of the key arguments in favour of this position is that nonphylogenetic systems (i.e. evolutionary studies that do not give the priority to the reconstruction of a common genealogical tree) cannot provide adequate heuristics for adaptive explanations. In this chapter, we argue that this claim is wrong because units not recognized in the TOL are required in many adaptive explanations, and because the assumption that the units of evolution are supplied by phylogenetic genealogies forecloses the understanding of key evolutionary processes. The appeal of genealogical modeling depends on the uniformity and relative simplicity of its explanatory structure, based on a single tree of life, in which, in the absence of extinction, diversity increases over time and there is no reticulation between 'branches'. Although tree-based practices and the virtue of the uniformity and structural simplicity of the Tree of Life have been explicit in evolutionary thinking since Darwin, many recent findings show that no single phylogenetic tree can represent the evolutionary history of many 'microbes' or of such DNA vehicles as viruses and plasmids [6-18]. Furthermore, to make a more theoretical point, even when a tree obtains in some parts of the 'macrobial' world it does so for purely contingent reasons. Thus, although we grant that tree shaped patterns correctly characterize some sections of evolutionary history, we argue that this genealogical canalization is contingent. Tree-like modes of evolution result from some, but not of all the evolutionary processes at play (e.g., cell division, preferential mating); other evolutionary processes are also relevant to model evolution [19]. Familiar examples of process that do not respect phylogenetic boundaries are introgression across genera in plants, resulting in reticulated evolution, and incorporation of viral DNA, often with additional exogenous DNA, into both prokaryotic and eukaryotic genomes[17, 20, 21]. Other such processes will be discussed below.

By privileging mostly (or exclusively) nicely contained genealogical patterns and the constraints fashioning them, the phylogenetic system is a priori blind to the other patterns and constraints that are an integral part of evolution. 'Purely' genealogical explanations of the patterns of life do not include many microbial adaptations. To cite one example in passing, adaptation to high temperature (> 50 Celsius degrees) in Archaea and Bacteria involves multiple and important exchanges of genetic material between these distantly related organisms [22]. Thus the adaptive hyperthermophile and thermophile phenotypes cannot be tracked solely by their genealogy. Yet, no evolutionist studying microbes would assert that this adaptation is an epiphenomenon. On the basis of theoretical considerations and by use of several examples along these lines, we argue that comprehensive evolutionary analyses should take a variety of evolutionary processes that are not captured by conventional genealogical thinking into account. Genealogical patterns (GP)

and evolutionary patterns (EP) can be two different things, two distinct outcomes of evolution, that can be summarized by distinct drawings (see figure 1). In this figure, the trees are temporally oriented: the vertical axis in the top left (GP) and top right (EP) diagrams is time with earlier at the bottom, later at the top. GP (as in the left-hand diagram) consider only splitting lineages and no interactions across lineages, while EP (as in the right-hand diagram) considers both. Therefore EP can be broader, because not only reticulations of various kinds (symbioses, genetic partnerships, etc.) are important, but also because these interactions are crucial to evolutionary fates and contents of lineages.

Proponents of the Tree of Life hold that (some) monophyletic groups on the TOL provide a fruitful representation of (all of the) natural groups and thus provide a fruitful representation of (all important) evolutionary scenarios (see, e.g. [23, 24]. But when GP and EP differ, this approach suffers from two significant limitations whose importance is becoming widely recognized. By definition, the TOL can only represent branching processes and it focuses solely and explicitly on subsets of evolutionary processes, namely the evolution of (monophyletic) species understood as reproductively isolated populations. The proponents of the notion of species, defined as the least relevant monophyletic groups on an appropriately-scaled or constructed tree, have identified some of the limitations of that notion. As Mayr and Ghiselin separately note many plants do not fit this account, and therefore would deserve to be distinguished from the other species, and called instead paraspecies or pseudospecies. For the former, "only sexually reproducing organisms qualify as species. Some other terminology, for instance paraspecies, will have to be found for uniparentally reproducing forms" [25]. For the latter, "asexual lineages do not form reproductive populations, and have to be considered 'pseudospecies'" [26]. We are not claiming that clonal plants or bacteria cannot be accommodated in the TOL, but that an understanding of evolutionary patterns focused on clearly demarcated, fully encapsulated, monophyletic group leads to counterintuitive claims about **how** lineages are formed and maintained, e.g., clonally, sexually, etc. [27]. These limits of tree-based approaches are the basis for insisting on the importance of providing a less constrained way of modeling and interpreting more (and ideally, all) of the fundamental evolutionary processes. In consequence, we will point the way toward more appropriate methods, although we cannot present them in technical detail. We will, no doubt, fall short of persuading all readers of our approach, but we will at least show that greater inclusivity yields a considerable improvement in modeling of evolutionary patterns and processes.

Our primary motivation, the idea that evolutionary patterns encompass genealogical patterns but not the reverse, is illustrated in Figure 1. For phylogeneticists, GP are the bedrock of evolutionary thinking [23], but many evolutionary biologists have come to accept that at least some adaptations do not translate into one clean genealogical pattern [12, 27]. Here we restrict the argument to adaptations, with concentrating to some extent on adaptations of prokaryotes, but in fact we think it holds for a much broader class of phenomena (e.g. typically traits that emerge from multi-level selection, carried on mobile elements). Consider the spread of antibiotic drug resistance in prokaryotes: drug resistant phenotypes result from the action of a wide diversity of mechanisms that move DNA between distantly related organisms: plasmids, phages, integrons, transformation, cell-cell fusion, activation of the SOS system, and successful gene expression after a lateral gene transfer etc. [28-32]. Most of these mechanisms yield

regularities discordant with phylogeny, therefore GPs certainly do not explain the EPs that result from the acquisition and loss of antibiotic resistance in microbes.

Given the broad acceptance of adaptive traits emerging from multi-level selection in the prokaryotic world [33-40], the historical reduction of the process of evolution of natural groups to tree like patterns is no longer fully satisfactory. Recent findings force evolutionists to entertain a richer set of patterns [19, 41, 42]. Because EP are broader in scope than GP, it may not be the best explanatory strategy to go from a limited pattern (the evolution of monophyletic groups in genealogical relationships) to a universal characterization of evolutionary processes. This concern motivates some evolutionary studies that explicitly attempt to accommodate heterogeneous evolutionary models for evolving natural groups, instead of trying to constrain evolutionary patterns to match the branching genealogical patterns of the Tree of Life [9, 11, 13, 43-47].

The studies we have cited focus on microbial evolution, using alternative approaches to classic tree-based approaches. Importantly, they are not only justified by the question of which patterns are broader and more encompassing (EP > GP or EP < GP). Indeed, the deeper problem is that genealogical patterns and evolutionary patterns do not track the same processes; rather, they aim to capture distinct phenomena. The fact that there is a gap between those patterns suggests that we are missing out on a lot of genuine evolutionary explanations when exclusively adopting GP. Minding the gap could have profound consequences.

2. What does the gap between Genealogical Patterns and Evolutionary Patterns imply?

In genealogical patterns, the basic explanatory unit has been species or monophyletic groupings. Since isomorphy of evolutionary and genealogical patterns (or convergence of EP on GP) was assumed, it has also been assumed that the basic explanatory unit for evolutionary patterns are species or monophyletic groupings (Figure 1) [1, 23]. The assumed superiority of genealogical thinking is in part a function of this perceived isomorphy between monophyletic groups as the sole unit of evolution and monophyletic groups as the sole unit of evolutionary explanation. By contrast, the gap between GP and EP shows us that monophyletic groupings may not be the only (or best) explanatory unit in evolutionary patterns (Figure 1). Evolutionists may need other units.

There is a connection here with important methodological issues recently discussed by philosophers of biology (Franklin 2005; Burian 2007; Elliott 2007; O'Malley 2007; Strasser 2008; Strasser 2010). The studies that seek potential units of evolution are exploratory in character, deploying some of the methods of traditional natural history together with the laboratory-intense methods of molecular biology and bioinformatics. This combination requires exploratory use of sequence databases such as those used in recent '-omic' sciences in combination with the molecular tools (e.g. these that allow replacement of one gene by another), and new computer methods designed to sample and analyze protein and gene sequences from various natural and experimental contexts. Thus exploratory experimentation does not follow the standard methods of hypothesis testing; instead it deploys a variety of means for varying parameters to examine what follows from, e.g., the incorporation of a novel plasmid into a population of microbes or by changing the timing of a developmental switch, and to extract

'surprising' patterns from an 'hypothesis-neutral' data set (which, of course cannot have been gathered in the absence of hypotheses). The patterns unraveled in these exploratory approaches are important because they capture certain (molecular) sequelae of some event or process. The spirit of such exploratory experiments, characteristic of much new work in the -omic sciences and in systems biology, could be embraced to improve evolutionary studies by identifying additional evolutionary units and the processes that generated them, without depending on the central hypothesis of a TOL.

It is one thing to show the incompleteness of existing evolutionary explanations based on the TOL [12] and quite another to show that one could step outside the TOL to recognize additional units of evolution of diverse sorts. Defenders of the TOL might argue that existing explanations, although incomplete, are powerful enough to encompass the majority of additional evolutionary patterns as outliers, as acceptable noise. We disagree because the inclusion of evolutionary processes and units in evolutionary representations and explanations beyond those envisaged in the TOL entails an inescapable pluralism. Yet, as we argue, the additional units are required to recognize the importance of interactions among hierarchical processes at several levels in bringing about evolutionary change. For us, **the gap between EP and GP encourages conceptual and practical developments aimed at capturing all the adaptations in which the phylogeneticist is interested, as well as other adaptations, objects and process beyond those revealed by studies restricted to the usual monophyletic groups relevant to phylogenetic studies** [48].

What are these additional evolutionary objects? Consider, for instance, the impact of lateral gene transfer (LGT) and recombination, which produce evolutionary **modules** (genes, groups of genes, operons) with their own individual fates. One example based on LGT is the suite of coevolving genes coding for gas vesicles in cyanobacteria and Haloarchaea; this suite of genes defines a functional and evolutionary unit [49]. This genetic module codes for a clear adaptive phenotype, conferring buoyancy to its hosts, and can be inherited by LGT and by vertical descent from ancestors to descendants. These (adaptive) genes and groups of genes are distributed across prokaryotes and mobile genetic elements in ways that do not match species genealogies. LGT and recombination also create phylogenetically mosaic entities (e.g. recombined genes [50], recombined plasmids [10], viral [16] or prokaryotic genomes [22, 51]. Quite generally, microbial genomes harbor genes with multiple distinct phylogenetic affinities and from distantly related sources. These processes thus impact the size of bacterial **pangenomes** (e.g. the overall gene pool of a set of organisms considered as belonging to a single species) [52]. Consequently, pangenomes of various sizes, composition and origins are also remarkable evolving entities that are outcomes of evolution. Finally, LGT and recombination are also greatly involved in the evolution of **microbial communities** [53, 54]. These ecologically-shuffled evolutionary units are often phylogenetically composite: they associate distinct DNA donors and hosts (also referred to as 'genetic partners' [41]) in a genetic network [9], mixing both mobile elements and cell lineages. Many examples beyond that of antibiotic resistance, mentioned above are known – for example, communities of cyanobacteria, cyanophages and plasmids in the ocean [55-58], natural communities in acid mine drainage [50], or in gut microbiomes of various metazoans [59, 60]. All include many ecologically-shared genetic partners that do not occupy a single branch in a tree of life. Evolution of microbes and their mobile elements is greatly affected by such a communal lifestyle.

By focusing anew on the evolutionary processes in these and other cases, we may be able to model additional evolutionary patterns that *cannot* appear within genealogical patterns. Species and monophyletic groups as the sole units of evolution are not as explanatorily exhaustive as many evolutionary biologists would like to believe, a fact that should be reflected in our explanatory models. For many this has led to efforts to redefine species in order to make the concept refer to something that is simultaneously an evolutionary, a classificatory, a functional, *and* an explanatory unit [61]. In our view, this effort cannot succeed. In fact, to reduce the gap between model and phenomena, i.e. to improve explanations of evolutionary processes when EP and GP are not isomorphic, evolutionists may wish to re-examine the 'units of explanation' they employ and ask whether additional 'units of evolution' are involved in the processes underlying the patterns they have found.

3. Richer conceptualization and representation of evolution

The biological world is not easily carved up at its joints. The use of species/ monophyletic groups as the primary unit of evolutionary change assumes a strong form of uniformity and continuity in what evolves. LGT is but one of many processes that transgresses these frontiers; it serves us as one indicator that this assumption does not always obtain. Speciation patterns are of course patterns of increased discontinuity. But various indicators suggest that many processes distinct from lineage splitting yield clumping patterns [7-11, 13, 16, 43, 62]; such patterns are found at many levels (from infra-cellular to supra-specific) in evolution. Thus, **evolutionists need to study the dynamics of the many sorts of clumping and splitting that occur in evolution**, far beyond those provided in standard genealogical studies (Figure 1).

A first step toward a broader conceptualization and representation of evolution consists in recognizing that evolution by natural selection is not necessarily a linear transformation within a lineage; it often involves the **intersection of many processes across many different types of entities**. Thus, LGT and recombination cause differential rates of recombination in various regions of prokaryotic and eukaryotic genomes. For example, in prokaryotes, gene evolution varies between genomic islands and the rest of the chromosome. Recent data indicate that environmental *Vibrio* differentiate rapidly into endemic subpopulations by tapping into a local gene pool as they acquire and express local newly acquired gene cassettes by LGT in their integrons (Boucher et al., in prep). However, most of their gene content outside the integron remains unchanged. Thus, a gene's occurrence in the chromosome of a *Vibrio* is not a sufficient indicator of whether it will be conserved or recombined; another process, such as occurs when the mechanistic processes that yield a higher rate of recombination between integron gene cassettes than between bacterial chromosomes and a local environmental pool of integrons, intersects with the canalization that stabilizes *Vibrio* chromosomes. Processes affecting organisms at a higher level of organization also intersect with the genealogical canalization. Bacteria living in dynamic and genetically diverse environments, with many partners, typically have larger pangenomes than obligate intracellular pathogens [52].

In such contexts the concept of a **coalition** may be more useful than that of a species or monophyletic group. This concept enables us to focus on functionally related units that swap functions and sometimes parts (e.g., segments of DNA) within or across communities and populations. Metazoan species are coalitions, for the functional relations that count for building

a coalition include reproductive relations, but for many biological systems, a more fluid category than **species** is needed to reflect how evolutionary change occurs. We distinguish two kinds of coalitions, depending on the type of material that is swapped. In type 1 coalitions, some of the swapped material is DNA; therefore members of a given coalition can be seen as "**friends with genetic benefits**". For example, cyanobacteria and cyanophages sometimes form such a coalition. The genes encoding the photosystem-II (PSII) or the Photosystem-I (PSI) reaction center have been found in many cyanophage genomes, and some phages, like plants and cyanobacteria, even contain both PSII and PSI genes and NADH dehydrogenase genes. As these viruses infect their cyanobacterial host, they can use different options to maximize their survival and that of their host by enhancing either cyanobacterial photosynthesis or ATP production [63]. Similarly, phylogenetically heterogeneous communities known as gut microbiomes, comprised of archaea and bacteria, converged in their repertoires of carbohydrate-active enzymes to adapt to shared challenges, in large part thanks to lateral gene transfer mediated by mobile elements rather than gene family expansion [64]. Gut microbiomes of metazoans are full of friends with genetic benefits. Last but not least, although the chimeric nature of many eukaryotic genomes is often under-appreciated in deep eukaryotic phylogenetics, type 1 coalitions can also be observed in eukaryotes. Using the diatoms as an example, Moustafa et al.[65] found that 16% of the *P. tricornutum* nuclear genes may have green algal origins[66]. Ignoring the probability that additional genes have been contributed to the genome over time in a non-vertical manner, this means at least one in five of this diatom's genes could be expected to produce a phylogenetic signal at odds with vertically inherited genes due to endosymbioses followed by gene transfer to the host nucleus.

On another hand, **tight functional interactions between phylogenetically unrelated partners in symbioses, consortia, etc. can also occur with few if any gene exchanges.** We will refer to functionally related units with a shared evolutionary fate in which no genetic material is swapped between communities and populations as type 2 coalitions. Many biologists might find that evolutionary studies of type 2 coalitions do not require new models of evolution that go beyond the tree of life. However the consideration of these type 2 coalitions argue for the **dependence of the change in the evolutionary fate of various subgroups on what others** (often members of other species or other types of partner) **in the community do, a phenomenon that cannot be represented with a genealogical tree alone**. Consider the oft-studied *Vibrio* fisheri-Hawaiian Bobtail squid interaction where bio-luminescence of the squid allows it to avoid predators. Bio-luminescence is generated by quorum-sensing of the bacteria in the constrained environment (i.e. high density conditions) of the squid's mantle that they colonize. The fitness-gain from bio-luminescence is not obvious for the *Vibrio sans* symbiosis and the squid alone cannot generate light, but as a coalition they allow for novel adaptations for both the squid and the *Vibrio*. To put things a bit simply: *Vibrio* don't need to glow, and squids can't glow, but they have co-evolved the adaptations of bio-luminescence and those required for their cooperative behaviors. This illustrates our claim that we should not expect EP to match GP, since it is the ecological interaction that allows for these adaptations to occur, not the genealogical confinement alone [67]. Many cases of genuine co-evolution[68] , e.g., between pollinators and plants, or hosts and parasites, support this same conclusion. Cases of type 2 coalitions are also well-known in prokaryotes. An example is the interspecific associations of anaerobic methane oxidizing archaea (ANME) and sulfate-reducing bacteria (Desulfosarcina,

Desulfobulbaceae, Desulfobacteriaceae, Betaproteobacteria and Alphaproteobacteria) [69]. These consortia, in which the archaeal member oxidizes methane and shuttles reduced compounds to the sulfate-reducing bacteria, are globally distributed. This metabolic cooperation enables the partners to thrive on low energy carbon sources, which neither partner could utilize on its own [40]. Together, ANME-sulfate reducer coalitions are estimated to be responsible for more than 80% of the consumption of methane in the oceans. Another obvious microbial coalition, '*Chlorochromatium aggregatum',* an interspecific phototrophic consortium with worldwide distribution, may constitute as much as 2/3 of bacterial biomass at the oxic/anoxic interface in stratified lakes [54]. These are tight associations of green sulfur bacterial epibionts which surround a central, motile, chemotrophic bacterium. The epibionts act as light sensors and control the carbon uptake of the central bacterium, which confers motility to the consortium, assuring that the coalition occupies a niche in which it will grow [70]. The cell division of these bacterial partners is highly coordinated and it was estimated by proteomics and transcriptomics that 352 genes are likely to be involved in sustaining the coalition [71]. Many intricate cases of mutualism and commensalism display similar emergent adaptations in type 2 coalitions. Importantly such emergent adaptation have more than one genealogical origin, hence require other model to be thoroughly analyzed.

Precisely, a second step in proposing new models of evolution rests on the recognition that the **interactions between many processes and entities are structured**, and that their frequent intersections should be modeled carefully. After all this is exactly why the populational approach was adopted in preference to a typological approach: pre-Darwinian concepts treated species as fixed types with fixed characteristics. Transformist theories forced biologists to think about species as malleable. Mayr devised the "non-dimensional" Biological Species concept (BSC) as part of his effort to reconcile an established biological category, species, which had implied stable properties from Aristotle to Linnaeus, with a view of evolution hinted by Darwin and developed in population genetics, that species are metapopulations of populations of genealogically-related individuals with diverse traits. Because of the shuffling of individuals and the impact of selection, the frequency of traits within populations changed through time; the BSC picks out the supra-populational entity composed of all potentially interbreeding individuals as of a given time or short stretch of time. Although it has no essential properties, it has a separate evolutionary fate because of the limitations on interbreeding with members of other species. The sub-population trajectories determine the distribution of attributes within populations and therefore within the species, thus ultimately affecting its fate. But, moving beyond Mayr's development of the BSC, one needs to realize that such intersections go beyond the ebb and flow of populational mixings. Populational approaches implicitly adopt a network approach in that individuals and subpopulations exchange genes in ways that are spatially determined. Take a population of deer. Their spatial distribution will determine which ones can reproductively interact with which others. Ecological constraints (mountain range, rivers, etc.) will determine the placement of nodes, i.e., of bottlenecks delimiting sub-populations within which gene change occurs. Real populations have a clustered topology. This is often abstracted away in population models, but it is a fact that should remain in the forefront of our understanding of the processes involved (see for instance Sewall Wright's shifting balance theory). To fully account for this **natural clustered topology**, evolutionists should provide better accounts of the motley crew of types of partners and the very diverse class of types of

interactions between partners [41].

For convenience, the evolutionarily significant interactions can be classified as genetic, structural and functional. The first type of interaction will be most prevalent in monophyletic groups of metazoa (which has led many to assume that EP and GP are the same thing). Nonetheless, one should not be surprised to find genuine functional interactions among non-related groups that lead to adaptive change, as observed in microbial evolutionary studies. Such findings force us to broaden our understanding of what to count as an efficacious partner in a coalition. The two prokaryotic coalitions (ANME-sulfate reducer and *Chlorochromatium aggregatum*) described above clearly associate organisms that are phylogenetically distant but nonetheless bona fide functional partners. And they are not exceptional. There are many cases of **communal evolution** with traits that GPs cannot properly describe, **because they involve both distinct phylogenetic microbial lineages and mobile elements**. These are reported with increasing frequency in the metagenomic literature, and strongly supported by molecular data (see Figure 2). For such communities, evolution is often coevolution, and functional, structural and genetic interactions matter. Such coalitions cannot be neglected. For instance, type 1 coalitions of cyanobacteria and cyanophages play a central role in marine photosynthesis, in the global carbon cycle and in the world oxygen supply. Type 2 coalitions such as the one observed between Glomerales and 60-80 % of the land plants for at least 460 million years [72-75], positively affected plant performance, nutrient mobilization from soil minerals, fixation of atmospheric nitrogen and protection of plants against root pathogens, and thus determined many aspects of community and ecosystem functioning. Overall, the impact of coalitions (be they genetic or not) should make communal interactions (and their resulting 'emergent evolutionary properties') essential features of evolutionary models, narratives and explanations, beside monophyletic groups.

Finally, a third step to improve our model of evolution is to acknowledge that these **coalitions evolved in ecosystems**. Odenbaugh [76] offers a detailed analysis of the concepts of community and ecosystem, most helpful to understand the latter. A *community* corresponds to the assemblage of most or all interacting species (populations) in a given area, ecological niche or environment. Communities are defined solely by the biotic entities that they include. Some [77],) think communities need to be functionally integrated but this view is arguably the minority view in contemporary ecology. An *ecosystem* corresponds to the functional assemblage of all communities as well as their abiotic (physical, chemical, geological, climatic) environment. Tansley [78] offered an early defense of such a view, according to which "community" is best considered a populational term focusing on the demographic distribution of the biotic individuals in a given context (e.g. predator-prey population interactions), whereas "ecosystem" is a functional term **focusing on the functional integration between biotic and abiotic subsystems in a given context**. The possibility that whole ecosystems can be said to evolve has recently been gaining some traction [79, 80]. But even if one rejects that possibility, **the ecosystem perspective improves on the evolutionary models of a purely populational-community perspective by highlighting functional integration and natural clustered topology over shared genealogical history**.

To sum up, that many sorts of processes and types of entities that intersect during evolution should have at least three consequences for evolutionary models and methods. First, understanding evolution should often mean understanding coalitions. Second, understanding

coalitions requires understanding the functional, genetic, and material interchanges that structure communal interactions among partners. Third, the interchanges underlying communal interactions in coalitions will be better understood by considering the ecosystems in which evolution occurs. According to this point of view, a more complete representation of (prokaryotic) evolution corresponds to a dynamic topology (Figure 3) rather than a TOL, tracking only the genealogical relationships. The various -omics are very good ways to define additional edges in dynamic evolutionary networks, as they capture aspects of these diverse relationships between evolving entities. Phylogenomics provides a phylogenetic distance between genes, genomes, and other operational taxonomic units (OTUs) of interest (e.g. these units may correspond to terminal taxa of a phylogenetic tree, such as species, genera, individuals, etc., and to any biotic nodes in the network). Comparative genomics produces estimates (e.g. percentages of identity, average nucleotide identity distances [81], etc.) based on the DNA shared between genomes and OTUs. It also provides physical distances between genes (e.g. by measuring their physical distance on chromosomes and organelles). Transcriptomics proposes co-expression matrices for genes, which can serve as bases for distances of genetic co-regulation, within cells and within environments; similarly proteomics provides measures of the physical and functional interactions of proteins within cells and within environments. Last but not least, metagenomics leads to identification of genetic partnerships (and incompatibilities) between and within environmental genes, populations, etc. The important claim here is that **if evolutionists intend to do so, they can represent coalitions, functional integration, and natural topologies along with genealogy in evolutionary studies**.

4. Exploiting dynamic evolutionary networks

When evolutionists reconstruct the dynamic evolutionary networks described above, they face a plethora of relations between biotic entities rather than a 'simple' unitary TOL. The patterns of evolution also reflect the impact of a wide range of disparate processes that link together the fates of entities at different levels, with varying degrees and kinds of connection to each other. Note that even though the examples described above mainly concern the evolution of organisms, the biotic entities entering coalitions, partnerships and ecosystems can be of many types, e.g. genes, operons, plasmids, genomes, organisms, coalitions, communities, etc. Whereas multi-level selection is usually focused on the very different levels at each of which entities of the same type interact (i.e. genes with genes, cells with cells, organisms with organisms, etc.), a coalition approach is open to the possibility that entities at different levels of organization can and do interact. The *Vibrio*-Squid symbiosis is such an example where a single organism interacts not with one individual organism but with a group of individuals (i.e. a bacterial colony). Gut flora in many metazoa have a similar profile: in those cases, an individual organism interacts with a community of different microbial species. However, a network-based representation of this complexity raises serious conceptual and practical questions. How could evolutionists make sense of such dynamic evolutionary networks (except by reconstructing a TOL) [13, 17, 82]? It is one thing to claim that whole ecosystems qua ecosystems, can evolve; it is another to try to model interactions where the monophyletic groups that are functional parts of those ecosystems are not the only relevant units that one needs to model to track evolutionary change. In the dynamic evolutionary networks approach, it is an open question, which units of

evolution deserve tracking and which explanatory units should be used in models.

To answer such questions, we need to think about relation between units of evolution (i.e. what actually evolves in response to natural selection) and units of explanation (i.e. the conceptual 'objects' should be used to model this change). In the GP approach, it was largely assumed that representations of the changes in the evolutionary units of the TOL were sufficient to provide the explanatory units of evolutionary explanations. Monophyletic genealogical relationships served both as evolutionary and explanatory units. We, like many others, have argued that while this representation *may* be appropriate for the evolution of some monophyletic groups (especially monophyletic groups of eukaryotes), it is woefully inadequate for many 'microbes' and is ruled out by definition in the evolution of more complex biological arrangements that we called coalitions [19, 41, 67]. Let us now see how other additional units of evolution and units of explanation play out in this coalition world.

4.1. Searching clusters in networks

Since we do not wish to rule out any type of organization as possibly being a coalition or a member of a coalition, we suggest adopting investigating clusters in our topologies as a first way to identify coalitions [9, 11, 83]. See Box 1 for a description of how such genetic networks are reconstructed with sequence data and the ways by which they are dynamically maintained. Our working hypothesis is that we will be able to identify and track coalitions. We have shown that clusters in networks, for instance in genome networks, are areas where nodes show a greater number of connections among themselves than with the other nodes of the graph. We expect to demonstrate that such patterns might be the result of evolution, as we will explain below.

But first, let us stress that looking for such clusters is consistent with the natural inclination of biologists to favor significant groupings of phenomena. In tree pattern analysis the search for clusters is also central, and it has translated in the classic problems of ranking and grouping [84]. The problem of grouping has been 'solved' by privileging a single unified type of relation, namely the genealogical relation exhibited by nodes. This allowed 'objective' pairs of nodes shown to share a last common ancestor in a data set to be grouped together and shown to be distally related. Ranking (e.g., the decision to classify a genealogical group as a species instead of genus, an order, etc.) was never truly solved and remains largely arbitrary [85]. This point was explicitly made by Darwin himself in chapter 1 of the *Origin*. It is therefore somewhat ironic that evolutionary explanations have reified clusters as 'real' encapsulated (bounded) evolutionary units by privileging genealogical relations. That is, evolutionary explanations have treated evolutionary clusters as if they were stable unitary units impervious to interference from other clusters, apart from the change in the selective environment caused by changes in the abiotic environment and the changes that any one group causes in the other groups with which it interacts. Genealogical explanations have given absolute ontological priority to genealogical change of a certain type and been blind to other natural processes that have deep consequences in the process of adaptation. It behooves us to look at the neglected branches created by LGT, hybridization and other means of genetic exchange, coevolution, and reticulation between branches, in order to reexamine the adequacy of models that focus exclusively on well compartmentalized (i.e. modular) monophyletic groups. By looking at these usual 'outliers' in shared gene networks for instance, we will identify new clusters, some of which, we argue, are

created and maintained by selective pressures and evolutionary processes. Figure 4 illustrates how clusters of partners of different types (e.g., clusters of bacteria and plasmids, bacteria and phages, plasmids and phages) can unravel the presence of groups of entities affected by processes of conjugation, transduction and /or recombination, respectively. These entities are candidate 'friends with genetic benefits'.

Importantly, as the ecosystems approach to microbial evolution has taught us, the networks representing evolutionary dynamics should not be purely genealogical; they should also be structural and functional. Ecosystems involve both biotic and abiotic processes. Abiotic processes do not have genealogies (after all they are not genetic systems) and the arrangements of species in communities can be initiated or reorganized in ways that do not reflect or require deep evolutionary histories. Increasingly comprehensive pattern analyses of ecosystems will then require an increasing number of types of edges and types of nodes as compared to the genome network of figure 4. Some of the edges (those involved in abiotic processes) will be of a physico-chemical nature [86], while others may (but will not necessarily) track more traditional biological relationships. Given the seemingly incommensurable nature of the possible types of relationships, it may appear that clustering in salient units becomes incredibly arduous.

Yet, the fact that analyses of comprehensive evolutionary networks are difficult doesn't mean they are impossible or useless. It merely relativizes the import of the conclusions that evolutionists may draw from their attempts at clustering vastly heterogeneous networks. If nature is not neatly cut at the joints, we should be suspicious of any overly simple model (e.g. a TOL) that assumes such simplicity. A pluralistic approach to clustering seems necessary to track the complex, messy and sometimes transient nature of evolutionary dynamics. The work of an evolutionary modeler goes from tracking 'simple' monophyletic groups (which we now know do not yield the universal history that they was expected to for most of the 20th Century) to analyzing the possible ways in which structural constraints and functional possibilities interact with hereditary systems in selective environments. It is not that genealogy is insignificant, but rather that it becomes one tool (among others) to track evolutionary change.

But how are evolutionists to identify the relevant interesting explanatory clusters? This chapter is an initial salvo in a broad project to reconceptualize evolution by natural selection. To describe the dynamics of the changes in both units and relationships, evolutionists will need to think about how the evolution of the processes translates into changes in the topology of dynamic evolutionary networks. Figure 4 is but the tip of the iceberg of interesting EPs that demand to be accommodated in our models. We know for instance how processes of conjugation and transduction translate into a topology of shared genes networks, as they generate remarkable clusters of bacteria and plasmids on the one hand, and of bacteria and phages on the other hand along lines suggested schematically in Figure 4 [9, 11, 13, 16]. Evolutionists need to learn how these and other processes translate into even more comprehensive dynamic evolutionary networks that include biotic and abiotic components.

4.2. Searching for 'correlations' in networks

Our second suggestion for identifying units that could play a significant role in evolutionary explanations is to display and to compare multiple networks including the same objects but connected according to different rules (e.g. functional similarity, genetic similarity, physical

interactions, etc.) to look for their common features. This approach is also consistent with scientific practice (see for instance the ongoing *National Geographic*-sponsored Genographic project that studies human evolution by searching for correlations between molecular analyses and non-molecular analyses of diverse traits that can be fairly well tracked (such as similarities of single nucleotide polymorphism (SNPs) in genomes, disease susceptibilities, gut flora, linguistic patterns, and ecological neighbors).

Importantly, the richness and great diversity of the biological world has always been perceived as a significant methodological research opportunity as a well as genuine problem. As Hennig has rightly pointed out, 'each organism may be conceived as a member of the totality of all organisms in a great variety of ways, depending on whether this totality is investigated as a living community, a community of descent, as the bearer of the physiological characters of life, as a chorologically differentiated unit, or in still other ways. The classification of organisms or specific groups of organisms as parasites, saprophytes, blood suckers, predators, carnivores, phytophages, etc.; into lung-, trachea-, or gill-breathers, etc.; into diggers of the digging wasp type, mole type, and earthworm type; into homoiothermous or poikilothermous; into inhabitants of the paleartic, neotropical, and ethiopian regions, etc., are partial pieces of such systematic presentations that have been carried out for different dimensions of the multidimensional multiplicity' [1]. However, for Hennig and the many evolutionists that his thinking influenced, this multiplicity was in part reducible, since one dimension (the genealogical) provided the best proxy for all the others. As Hennig put it: 'making the phylogenetic system the general reference system for special systematics has the inestimable advantage that the relations to all other conceivable biological systems can be most easily represented through it. This is because the historical development of organisms must necessarily be reflected in some way in all relationships between organisms. Consequently, direct relations extend from the phylogenetic system to all other possible systems, whereas they are often no such direct relations between these other systems' [1]. However, the –omic disciplines reveal that the number of processes, interactions, systems, and relationships affecting evolutionary – and the various entities that are, in fact, units of evolution – are more astonishingly diverse than Hennig (and for that matter, Darwin) recognized. Phylogenomics also provides a strong case that the TOL is a poor proxy for all the features of biodiversity [87], as it would explain only the history of 1 % of the genes in a complete tree for prokaryotes [12] or of about 10-15 % at the level of bacterial phyla [88, 89], and, by definition, none of the emergent and communal microbial properties. Likewise some functional analyses of metagenomic data show that the functional signal is, in some cases, stronger than the genealogical signal in portions of the genome, showing that the presence of genetic material with a given function matters more than the presence of a given genealogical lineage in some ecosystems[90]. Thus the claim that one system has precedence over the others deserves empirically reassessment. We maintain that such reassessment has potential to unravel important hidden correlations in the relationships between evolving entities, overlooked thus far when they were not consistent with the genealogy.

Network approaches (in contrast to branching genealogical representations) are precisely the right tool to use for this purpose; they are better suited to the evolutionary modeling needed here in that they are agnostic about the structure of the relevant topologies. Network-based studies can easily represent the multiplicity of relationships discovered by –omics approaches, and test whether, indeed, one system (i.e., one of the networks) is a better proxy than the others.

In fact, all sorts of relationships between evolving entities can be represented on these graphs. Proteomics allows one to draw connections based on protein-protein interaction and functional associations. Metagenomics proposes environmental and functional connections. And so on. Correlation studies between multiple networks reconstructed for the same objects (e.g. thousands of genes) by using different rules with respect to connections should expose, without preconceptions, which networks (e.g. functional, regulatory, genetic) and parts of networks can be placed in direct relation to each other. Evolutionary studies can then examine the shared connections (paths, edges, modules) present in these networks (Figures 5 & 6), e.g. to identify units that are worthy of note for their shared functional, structural and genetic features and for the possibility that these are the result of evolutionary significant interactions.

Correlation analyses of this sort have in fact already been initiated for organisms for which metabolic networks, protein-protein interaction networks, and phylogenetic information are available. For instance, Cotton and McInerney [45] recently showed that the phylogenetic origin of eukaryotic genes (e.g., from archaea or from bacteria) is correlated with the centrality of these genes in metabolic network (e.g., genes of archaeal origin occupy less terminal positions in yeast metabolic network). This result suggests that eukaryotes evolved as bits of bacterial metabolisms were added to a backbone of archaeal pathways. Also, Dilthey and Lercher characterized spatially and metabolically coherent clusters of genes in gamma-proteobacteria. Though these genes share connections in spatial and metabolic networks, they present multiple inconsistent phylogenetic origins with the rest of the genes of the genomes hosting them. This lack of correlation between the genealogical affinities of genes otherwise displaying remarkable shared connections in their spatial and functional interactions suggests that analyses of correlations in these particular networks could be used to predict LGT of groups of tightly associated genes (Dilthey and Lercher, in prep.). Here, additional evolutionary units (gene coalitions), consistent with the selfish operon theory, could be identified (Lawrence 1999).

Our more general point is that, if – at some level of evolutionary analysis – no network is an objectively better proxy for all the others, local parts of different networks could still show significant correlations, useful to elaborate evolutionary scenarios (e.g., involving genetic modules, pathway evolution, etc.). Just as Dilthey and Lercher suggested for clusters of metabolic genes, locally common paths between physical and functional networks reconstructed for many organisms could define clusters of genes with physical and functional interactions that are found in multiple taxa. If the genes making these clusters are distantly related in terms of phylogeny, such findings suggest that these genes may have been laterally transferred, possibly between distantly related members of a type 1 coalition. With further investigation, the physical and functional associations observed between these genes, in multiple taxa, could be interpreted as emerging phenotypes owing to LGT.

Correlations between networks based on transcriptomics, proteomics, and metagenomics could also inform evolutionists about the robustness of coalitions (e.g. the presence of resilient and recurring edges in various OTUs/ coalitions/ environments/ over time). Think of a trophic cycle in a given ecosystem. Various species can play the same functional role, but the cycle remains. A species can be replaced (via competition, migration, etc.) within a trophic cycle. Representing this in networks, we would observe that some clusters have changed (a network focused on genealogical relationships) while others are stable (those focused on functional properties). The fact that some functional relationships persist longer than some genealogical

ones may be an indication of an evolutionary cluster that cannot be tracked by GP alone [91], i.e. when the functional composition of a community remains stable over longer times than the taxonomic composition. Again, this is typically observed in gut flora: the functional network and the phylogenetic network are not always well correlated, since the composition and diversity of microbial populations changes within the gut, even if the microbes keep thriving on a shared gene pool [90]. It would also be observed in natural geochemical cycles [86], which has the potential to introduce functional, genetic and environmental signatures in evolution, that might outlive genealogical ones.

Since this search for correlation between networks does not impose an a priori dominant pattern on biodiversity, it could offer an improved and finer-grained representation of some aspects of evolution. In particular, this approach would facilitate the recognition of evolutionary units not revealed in analyses based solely on monophyletic groupings. The evaluation of the evolutionary importance of such units cannot properly begin until they are made into explicit objects of evolutionary study. If significant correlations reveal a pattern worth naming and deserving evolutionary explanation, they will thus have opened up pathways in the study of evolutionary origins not accessible in a strictly phylogenetic evolutionary system (Figure 6).

5. Conclusion

We suggest that in nature coalitions (both friends with genetic benefits and type 2 coalitions) are an important category of evolving entities. Developing the tools (e.g., of network analysis) to analyze the evolutionary impact of the processes into which coalitions enter and the various roles that coalitions (and their evolutionarily interesting components) play will provide an improved basis for the study of evolution, one that can include, but also go beyond what can be achieved with TOL-based modeling. We also suggest that modeling of evolutionary adaptive processes can be significantly improved by examining the evolutionary dynamics of coalitions, in particular by including parameters informative about the topology and structure of the components of the networks classified in various ways, including their evolutionary roots. Such modeling is open to various types of assortments of partners (whereas GP focus on same types of associations), various durations of association (whereas GP focus on the long term relative to organismal scale), all the degrees of functional integration (whereas GP focus almost exclusively on the maximally integrated associations such as mitochondria or on the shallow associations of co-evolution). Because genealogical patterns and evolutionary patterns are not isomorphic, evolutionists should not be too strict in maintaining the ontological superiority of genealogical patterns. In genealogical patterns evolutionists had (rightly or not) an intuition about what persisted through time: species and monophyletic groups. This allowed for the changing of parts while maintaining continuity of some entity (which was assumed to be what evolution was about). In the broader (and a priori less constrained) perspective for which we argued, i.e., in ecosystem-oriented evolutionary thinking, what persists through evolution needs be pinned down more carefully since monophyletic groups are not the exclusive units and do not provide all of the ways of carving out the patterns. In particular, studies of the correlations and clusters in evolutionary dynamic networks could offer a possible future alternative approach to complete the TOL perspective.

Acknowledgments
This paper was made possible through a series of meetings funded by the Leverhulme Trust (Perspectives on the Tree of Life), organized by Maureen O'Malley, whom we want to thank dearly. We also thank P. Lopez, S. Halary & K. Schliep for help with some analyses and figures, and P. Lopez & L. Bittner for critical discussions.

Exercise

1. What are the computational steps required to reconstruct a genome network?
2. Cite four examples of 'communal evolution'?
3. Cite three examples of 'coalitions'?
4. In your opinion, is the genealogical pattern the best proxy for all evolutionary patterns? What aspects of evolution in particular cannot be described by a Tree of Life only? Are there aspects of evolution that can be described by the Tree of Life that cannot be captured in a network-based approach?
5. Are genes from all functional categories found in the genomes of mobile elements?

Box 1: Reconstructing genome and gene networks
The various networks described in this chapter can easily be reconstructed, for instance using genetic similarities.

For genome networks, a set of protein and/or nucleic sequences from complete genomes must be retrieved from a relevant database (e.g. the NCBI (http://www.ncbi.nlm.nih.gov/Entrez). All these sequences are then BLASTed against one another. To define homologous DNA families, sequences are clustered when they shared a reciprocal best-BLAST hit relationship with at least one of the sequences of the cluster, and a minimum sequence identity. For each pair of sequences, all best BLAST hits with a score of 1e-20 are stored in a mySQL database. To define homologous DNA families, sequences must be clustered, for instance using a single-linkage algorithm or MCL. With the former approach, a sequence is added to a cluster if it shares a reciprocal best-BLAST hit (RBBH) relationship with at least one of the sequences of the cluster. We call CHDs (for cluster of homologous DNA families) the DNA families so defined. Requirement that RBBH pairs share a minimum sequence identity, in addition to a BLAST homology, can also be taken into account to define the CHDs. Thus, distinct sets of CHDs can be produced, e.g. for various identity thresholds (from 100% – to study recent events – to 20% to study events of all evolutionary ages). Based on these sets of CHDs and their distribution in the genomes, genome networks can be built to summarize the DNA-sharing relationships between the genomes under study, as summarized by Figure 7. A network layout can be produced by Cytoscape software, using an edge-weighted spring-embedded model.

Several different evolutionary gene networks (EGN) can be reconstructed to be contrasted with protein-protein interaction networks, or networks of metabolic pathways. For instance, EGN based on sequence similarity can be reconstructed when each node in the graph corresponds to a sequence. Two nodes are connected by edges if their sequences show significant similarity, as assessed by BLAST. Hundreds of thousands of DNA (or protein) sequences can thus be all BLASTed against each other. The results of these BLASTs (the best

BLAST scores between two sequences, their percent of identity, the length over which they align, etc.) are stored in databases. Groups of homologous sequences are then inferred using clustering algorithms (such as the simple linkage algorithm). The BLAST score or the percentage of identity between each pair of sequences, or in fact any evolutionary distance inferred from the comparison of the two sequences, can then be used to weight the corresponding edges. Most similar sequences can then be displayed closer on the EGN. The lower the BLAST score cut-off (e.g. 1e-5), the more inclusive the EGNs. Since not all gene forms resemble one another however, discontinuous variations will structure the graph.

Finally, Clusters in genome and gene networks can be found by computing modules, using packages for graph analysis such as MCODE 1.3 Cytoscape plugin (default parameters), Igraph[92], or by modularity maximization (as described in [11], and [93]).

Figure legends

Figure 1: Relationships between GP (black) and EP (grey).
Evolutionary Patterns (EP) encompass Genealogical Patterns (GP) but not the reverse

Figure 2: Distribution of genes of various functional categories in genomes of mobile elements.
All functional categories genes except genes of nuclear structure can be found in mobile elements, many of which should benefit communal evolution since expression of genes with cellular functions will increase the fitness of cells containing the mobile elements, which, in turn, will increase the likelihood of the mobile elements being carried forward to the next cellular generation. Bars for plasmids are in black; bars for phages are in white. The X-axis corresponds to the functional categories defined by clusters of orthologous groups (COGs) [94]. The Y-axis indicates the percentage of occurrences of these categories in an unpublished dataset of 148864 plasmids and 79413 phage sequences, annotated using RAMMCAP [95]. Functional categories are sorted as follow: 1) ***Information storage and processing***; A: RNA processing and modification; B: Chromatin structure and dynamics; J: Translation; K : Transcription; L: Replication and repair; 2) ***Cellular processes***; D: Cell cycle control and mitosis; Y: Nuclear structure; V: Defense mechanisms; T: Signal Transduction; M: Cell wall/membrane/envelop biogenesis; N: Cell motility; Z: Cytoskeleton; W: extracellular structures; U: Intracellular trafficking, secretion and vesicular transport; O: Post-translational modification, protein turnover, chaperone functions; 3) ***Metabolism***; C: Energy production and conversion; E: Amino Acid metabolism and transport; F: Nucleotide metabolism and transport; G: Carbohydrate metabolism and transport; H: Coenzyme metabolism and transport; I: Lipid metabolism and transport; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; 4) ***Poorly characterized***; R: General functional prediction only; S: Function Unknown.

Figure 3: Theoretical scheme of a dynamic evolutionary network & real polarized network of genetic partnerships between Archaea and Bacteria
A. Nodes are apparent entities that can be selected during evolution. Various –omics help determine the various edges in such network, in order to describe covariation of fitness between nodes. Note that nodes can contain other nodes (nodes are multi-level). Smaller grey nodes are

genes. Some of these genes have phylogenetic affinities indicated by long dashed black edges, other connected by plain thin edges are co-expressed. Collectively, some of these genes associations define larger units (here the two *Vibrio* genomes, or ecological organisms, like the *Vibrio*-Squid emergent ecological individual). Some of these genes and genomes interact functionally with the products of other genes and other genomes defining coalitions (dashed grey lines). In many coalitions the interaction between partners may be transient, ephemeral and not the result of a long co-evolution, yet the adaptations they display still deserve evolutionary analysis. Thus, edge length corresponds to the temporal stability of the association (closer nodes are in a more stable relationships over time). B. Network adapted from [47] computed from gene trees including only archaea and a single bacterial OTU in a phylogenetic forest of 6901 gene trees with 59 species of Archaea and 41 species of Bacteria. The isolated bacterial OTU (that can differ in different trees) is odd, since the rest of the tree comprises only archaeal lineages. For this reason, the single odd taxon is called an intruder [47]. Archaea are represented by squares, bacteria are represented by circles. Edges are colored based on the lifestyle distance between the pairs of partners, from 0 (darkest edges, same lifestyle) to 4 (clearest edges, 50 % similar lifestyle). The largest lifestyle distance in that analysis was 8, so the organisms with the greater number of LGT had all a close to moderately distant lifestyle. Edge length is inversely proportional to the number of transferred genes: the greater the number of shared genes between distantly related organisms, the shorter the edge on the graph. The networks are polarized by arrows pointing from donors to hosts, here showing LGT from Archaea to Bacteria.

Figure 4: Remarkable patterns and processes in shared genome networks
A. Schematic diagram of a connected component, showing a candidate coalition of friends with genetic benefits, where each node represents a genome, either cellular (white for bacterial chromosome), plasmidic (grey) or phage (black). Data are real and were kindly provided by [9]. Two nodes are connected by an edge if they share homologous DNA (reciprocal best BLAST hit with a minimum of 1e-20 score, and 100% minimum identity). Edges are weighted by the number of shared DNA families. The layout was produced by Cytoscape, using an edge-weighted spring-embedded model, meaning that genomes sharing more DNA families are closer on the display[96]. Clusters of bacteria and plasmids suggest events of conjugation; clusters of bacteria and phages suggest events of transduction; clusters of phages and plasmids suggest exchange of DNA between classes of mobile elements, etc. B. Three connected components corresponding to three genetic worlds, defined by displaying connections between genomes (same color code) for a reciprocal best BLAST hit with a minimum of 1e-20 score, and a minimum of 20% identity. Their three gene pools are absolutely distinct, which suggest that some mechanisms and barriers structure the genetic diversity, and the genetic evolution outside the TOL. These real data were also kindly provided by [9].

Figure 5: Schematic correlations between -omics network.
Each node corresponds to one individual gene. 4 networks illustrate the relationships inferred by –omics for these genes: black edges between nodes indicate the shortest distances in terms of phylogenetics, functional interaction, physical distance and regulatory distances for these genes. The question whether one of these networks is a better proxy for all the others (within an organism or an environment or between organisms or environments) is an open (empirical)

question. Shaded edges indicate paths that are identical between more than 2 networks of a single organism, bold edges indicate paths that are identical between comparable networks of distinct organisms. For instance, in this graph, a cluster of 3 interconnected genes showed functional, physical and regulatory coherence both in organisms / environments *i* and *j*. However, this pattern was not captured by their phylogenetic affinities in gene trees.

Figure 6: Functional networks of shared genes for plasmids, phages and prokaryotes.
Four functional genome networks including 2209 genomes of plasmids, 3477 genomes of phages and 116 prokaryotic chromosomes (from the same dataset as figure 2) were reconstructed by displaying only edges that correspond to the sharing of genetic material involved in each of these functions on a separated graph. Here, we only showed the giant connected components of four functional genomes network: A) for J: Translation, B for C: Energy production, C) for T: Signal transduction and D) for U: Intracellular trafficking. Bacterial genomes are in black, archaeal genomes in white, plasmids in light grey and phages in dark grey. It is clear that these functional networks are quite different, because the histories of the genes coding for these functions were distinct. However, some local correspondence can be found between the GCC of these functional graphs, suggesting that some functional categories underwent the same evolutionary history in some groups of genomes, sometimes consistently with the taxonomy (e.g. translation and energy production in bacteria and archaea), sometimes not. The layout was produced by Cytoscape[96].

Figure 7: Illustration for Box 1
Genes found in each type of DNA vehicle and belonging to the same homologous DNA family are represented by a similar dash. The distribution of DNA families in mobile elements and cellular chromosomes can be summarized by a presence/absence matrix, which can be used to reconstruct a network. With real data, the network of genetic diversity is disconnected yet highly structured. It presents multiple connected components.

References

1. Hennig, W. (1966) Phylogenetic systematics. Urbana.
2. Daubin, V., Moran, N.A., and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. Science 301, 829-832.
3. Galtier, N., and Daubin, V. (2008) Dealing with incongruence in phylogenomic analyses. Philos Trans R Soc Lond B Biol Sci 363, 4023-4029.
4. Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283-1287.
5. Kurland, C.G., Canback, B., and Berg, O.G. (2003) Horizontal gene transfer: a critical view. Proc Natl Acad Sci U S A 100, 9658-9662.
6. Lawrence, J.G., and Retchless, A.C. (2009) The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol 532, 29-53.
7. Retchless, A.C., and Lawrence, J.G. (2010) Phylogenetic incongruence arising

from fragmented speciation in enteric bacteria. Proc Natl Acad Sci U S A 107, 11453-11458.
8. Retchless, A.C., and Lawrence, J.G. (2007) Temporal fragmentation of speciation in bacteria. Science 317, 1093-1096.
9. Halary, S., Leigh, J.W., Cheaib, B., Lopez, P., and Bapteste, E. (2010) Network analyses structure genetic diversity in independent genetic worlds. Proc Natl Acad Sci U S A 107, 127-132.
10. Brilli, M., Mengoni, A., Fondi, M., Bazzicalupo, M., Lio, P., and Fani, R. (2008) Analysis of plasmid genes by phylogenetic profiling and visualization of homology relationships using Blast2Network. BMC Bioinformatics 9, 551.
11. Dagan, T., Artzy-Randrup, Y., and Martin, W. (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc Natl Acad Sci U S A 105, 10039-10044.
12. Dagan, T., and Martin, W. (2006) The tree of one percent. Genome Biology 7, 118.
13. Dagan, T., and Martin, W. (2009) Getting a better picture of microbial evolution en route to a network of genomes. Philos Trans R Soc Lond B Biol Sci 364, 2187-2196.
14. Doolittle, W.F., Nesbo, C.L., Bapteste, E., and Zhaxybayeva, O. (2007) Lateral Gene Transfer. In: Evolutionary Genomics and Proteomics: Sinauer.
15. Doolittle, W.F., and Bapteste, E. (2007) Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci U S A 104, 2043-2049.
16. Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008) Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol 25, 762-777.
17. Ragan, M.A., McInerney, J.O., and Lake, J.A. (2009) The network of life: genome beginnings and evolution. Introduction. Philos Trans R Soc Lond B Biol Sci 364, 2169-2175.
18. Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E., Nesbo, C.L., Case, R.J., and Doolittle, W.F. (2003) Lateral gene transfer and the origins of prokaryotic groups. Annu Rev Genet 37, 283-328.
19. Bapteste, E., O'Malley, M., Beiko, R.G., Ereshefsky, M., Gogarten, J.P., Franklin-Hall, L., Lapointe, F.J., Dupré, J., Dagan, T., Boucher, Y., and Martin, W. (2009) Prokaryotic evolution and the tree of life are two different things. Biology Direct 4, 34.
20. Lopez, P., and Bapteste, E. (2009) Molecular phylogeny: reconstructing the forest. C R Biol 332, 171-182.
21. Brussow, H. (2009) The not so universal tree of life or the place of viruses in the living world. Philos Trans R Soc Lond B Biol Sci 364, 2263-2274.
22. Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc Natl Acad Sci U S A 106, 5865-5870.
23. O'Hara, R.J. (1997) Population thinking and tree thinking in systematics.

Zoologica Scripta 26, 323-329.
24. Kuntner, M., and Agnarsson, I. (2006) Are the linnean and phylogenetic nomenclatural systems combinable? Recommendations for biological nomenclature. Syst Biol 55, 774-784.
25. Mayr, E. (1987) The ontological status of species. Biology and Philosophy 2, 145-166.
26. Ghiselin, M.T. (1987) Species concepts, Individuality, and Objectivity. Biology and Philosophy 4, 127-143.
27. Doolittle, W.F., and Zhaxybayeva, O. (2009) On the origin of prokaryotic species. Genome Res 19, 744-756.
28. Tsvetkova, K., Marvaud, J.C., and Lambert, T. (2010) Analysis of the mobilization functions of the vancomycin resistance transposon Tn1549, a member of a new family of conjugative elements. J Bacteriol 192, 702-713.
29. D'Auria, G., Jimenez-Hernandez, N., Peris-Bondia, F., Moya, A., and Latorre, A. (2010) Legionella pneumophila pangenome reveals strain-specific virulence factors. BMC Genomics 11, 181.
30. Barlow, M. (2009) What antimicrobial resistance has taught us about horizontal gene transfer. Methods Mol Biol 532, 397-411.
31. Manson, J.M., Hancock, L.E., and Gilmore, M.S. (2010) Mechanism of chromosomal transfer of Enterococcus faecalis pathogenicity island, capsule, antimicrobial resistance, and other traits. Proc Natl Acad Sci U S A 107, 12269-12274.
32. Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74, 417-433.
33. Krakauer, D.C., and Komarova, N.L. (2003) Levels of selection in positive-strand virus dynamics. J Evol Biol 16, 64-73.
34. Lee, H.H., Molla, M.N., Cantor, C.R., and Collins, J.J. (2010) Bacterial charity work leads to population-wide resistance. Nature 467, 82-85.
35. Dupre, J., and O'Malley, M.A. (2007) Metagenomics and biological ontology. Stud Hist Philos Biol Biomed Sci 38, 834-846.
36. Shah, S.A., and Garrett, R.A. (2010) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res Microbiol.
37. Lyon, P. (2007) From quorum to cooperation: lessons from bacterial sociality for evolutionary theory. Stud Hist Philos Biol Biomed Sci 38, 820-833.
38. Koonin, E.V., and Wolf, Y.I. (2009) Is evolution Darwinian or/and Lamarckian? Biol Direct 4, 42.
39. Van Melderen, L., and Saavedra De Bast, M. (2009) Bacterial toxin-antitoxin systems: more than selfish entities? PLoS Genet 5, e1000437.
40. DeLong, E.F. (2007) Microbiology. Life on the thermodynamic edge. Science 317, 327-328.
41. Bapteste, E., and Burian, R.M. (2010) On the Need for Integrative Phylogenomics, and Some Steps Toward its Creation. Biology and Philosophy 25, 711-

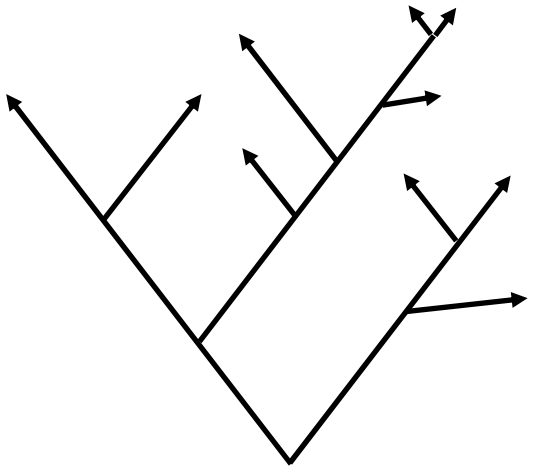
736.
42. Valas, R.E., and Bourne, P.E. (2010) Save the tree of life or get lost in the woods. Biol Direct 5, 44.
43. Dagan, T., and Martin, W. (2009) Microbiology. Seeing green and red in diatom genomes. Science 324, 1651-1652.
44. Dagan, T., Roettger, M., Bryant, D., and Martin, W. (2010) Genome networks root the tree of life between prokaryotic domains. Genome Biol Evol 2, 379-392.
45. Cotton, J.A., and McInerney, J.O. (2010) Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. Proc Natl Acad Sci U S A.
46. Lapointe, F.J., Lopez, P., Boucher, Y., Koenig, J., and Bapteste, E. (2010) Clanistics: a multi-level perspective for harvesting unrooted gene trees. Trends Microbiol 18, 341-347.
47. Schliep, K., Lopez, P., Lapointe, F.J., and Bapteste, E. (2010) Harvesting Evolutionary Signals in a Forest of Prokaryotic Gene Trees. Mol Biol Evol, ahead of print
48. Bapteste, E., and Boucher, Y. (2008) Lateral gene transfer challenges principles of microbial systematics. Trends Microbiol 16, 200-207.
49. Walsby, A.E. (1994) Gas vesicles. Microbiol Rev 58, 94-144.
50. Lo, I., Denef, V.J., Verberkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., Hettich, R.L., and Banfield, J.F. (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446, 537-541.
51. Nesbo, C.L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N.K., and Doolittle, W.F. (2009) The genome of Thermosipho africanus TCF52B: lateral genetic connections to the Firmicutes and Archaea. J Bacteriol 191, 1974-1978.
52. Wilmes, P., Simmons, S.L., Denef, V.J., and Banfield, J.F. (2009) The dynamic genetic repertoire of microbial communities. FEMS Microbiol Rev 33, 109-132.
53. Vogl, K., Wenter, R., Dressen, M., Schlickenrieder, M., Ploscher, M., Eichacker, L., and Overmann, J. (2008) Identification and analysis of four candidate symbiosis genes from 'Chlorochromatium aggregatum', a highly developed bacterial symbiosis. Environ Microbiol 10, 2842-2856.
54. Wanner, G., Vogl, K., and Overmann, J. (2008) Ultrastructural characterization of the prokaryotic symbiosis in "Chlorochromatium aggregatum". J Bacteriol 190, 3721-3730.
55. Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R., Church, G.M., and Chisholm, S.W. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. Nature 449, 83-86.
56. Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and

Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. Proc Natl Acad Sci U S A 101, 11013-11018.
57. Palenik, B., Ren, Q., Tai, V., and Paulsen, I.T. (2009) Coastal Synechococcus metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. Environ Microbiol 11, 349-359.
58. Zeidner, G., Bielawski, J.P., Shmoish, M., Scanlan, D.J., Sabehi, G., and Beja, O. (2005) Potential photosynthesis gene recombination between Prochlorococcus and Synechococcus via viral intermediates. Environ Microbiol 7, 1505-1513.
59. Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M., and Nelson, K.E. (2006) Metagenomic analysis of the human distal gut microbiome. Science 312, 1355-1359.
60. Qu, A., Brulc, J.M., Wilson, M.K., Law, B.F., Theoret, J.R., Joens, L.A., Konkel, M.E., Angly, F., Dinsdale, E.A., Edwards, R.A., Nelson, K.E., and White, B.A. (2008) Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. PLoS One 3, e2945.
61. Simpson, G.G. (1961) Principles of Animal Taxonomy. New York: Columbia Univ Press.
62. Lane, C.E., and Archibald, J.M. (2008) The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol Evol 23, 268-275.
63. Alperovitch-Lavy, A., Sharon, I., Rohwer, F., Aro, E.M., Glaser, F., Milo, R., Nelson, N., and Beja, O. (2010) Reconstructing a puzzle: existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. Environ Microbiol.
64. Lozupone, C.A., Hamady, M., Cantarel, B.L., Coutinho, P.M., Henrissat, B., Gordon, J.I., and Knight, R. (2008) The convergence of carbohydrate active gene repertoires in human gut microbes. Proc Natl Acad Sci U S A 105, 15076-15081.
65. Moustafa, A., Beszteri, B., Maier, U.G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324, 1724-1726.
66. Lane, C.E., and Durnford, D. (2010) Endosymbiosis and the evolution of plastids. In: Molecular Phylogeny of Microorganisms. Oren, A., and Papke, R.T. eds. Norwich: Horizon Press.
67. Bouchard, F. (2010) Symbiosis, Lateral Function Transfer and the (many) saplings of life. Biology and Philosophy 25, 623-641.
68. Janzen, D.H. (1980) When is it coevolution? Evolution 34, 611-612.
69. Pernthaler, A., Dekas, A.E., Brown, C.T., Goffredi, S.K., Embaye, T., and Orphan, V.J. (2008) Diverse syntrophic partnerships from deep-sea methane vents revealed by direct cell capture and metagenomics. Proc Natl Acad Sci U S A 105, 7052-7057.
70. Overmann, J. (2010) The phototrophic consortium "Chlorochromatium aggregatum" - a model for bacterial heterologous multicellularity. Adv Exp Med Biol 675, 15-29.

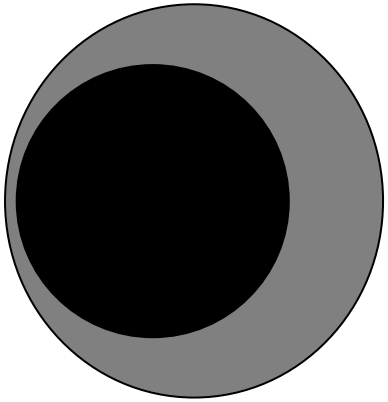
71. Wenter, R., Hutz, K., Dibbern, D., Li, T., Reisinger, V., Ploscher, M., Eichacker, L., Eddie, B., Hanson, T., Bryant, D.A., and Overmann, J. (2010) Expression-based identification of genetic determinants of the bacterial symbiosis 'Chlorochromatium aggregatum'. Environ Microbiol.
72. Ehinger, M., Koch, A.M., and Sanders, I.R. (2009) Changes in arbuscular mycorrhizal fungal phenotypes and genotypes in response to plant species identity and phosphorus concentration. New Phytol 184, 412-423.
73. Scheublin, T.R., Sanders, I.R., Keel, C., and van der Meer, J.R. (2010) Characterisation of microbial communities colonising the hyphal surfaces of arbuscular mycorrhizal fungi. ISME J 4, 752-763.
74. Hijri, I., Sykorova, Z., Oehl, F., Ineichen, K., Mader, P., Wiemken, A., and Redecker, D. (2006) Communities of arbuscular mycorrhizal fungi in arable soils are not necessarily low in diversity. Mol Ecol 15, 2277-2289.
75. Kuhn, G., Hijri, M., and Sanders, I.R. (2001) Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. Nature 414, 745-748.
76. Odenbaugh, J. (2007) Seeing the Forest and the Trees: Realism about Communities and Ecosystems. Philosophy of Science 74, 628-641.
77. Hutchinson, G.E. (1948) Circular Causal Systems in Ecology. Annals of the New York Academy of Sciences 50, 221-246.
78. Tansley, A.G. (1935) The Use and Abuse of Vegetational Terms and Concepts. Ecology 16, 284-307.
79. Swenson, W., Wilson, D.S., and Elias, R. (2000) Artificial Ecosystem Selection. Proceedings of the National Academy of Science 97, 9110-9114.
80. Bouchard, F. (2011) How ecosystem evolution strengthens the case for functional pluralism. In: Functions: selection and mechanisms. Huneman, P. ed.: Synthese Library, Springer.
81. Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102, 2567-2572.
82. Doolittle, W.F. (2009) Eradicating Typological Thinking in Prokaryotic Systematics and Evolution. Cold Spring Harb Symp Quant Biol.
83. Popa, O., Hazkani-Covo, E., Landan, G., Martin, W., and Dagan, T. (in prep.) Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes.
84. Broogard, B. (2004) Species as Individuals". Biology and Philosophy 19, 223-242.
85. Ereshefsky, M. (2010) Mystery of mysteries: Darwin and the species problem. Cladistics 26, 1-13.
86. Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. Science 320, 1034-1039.
87. Doolittle, W.F., and Zhaxybayeva, O. (2010) Metagenomics and the Units of Biological Organization. Bioscience 60, 102-112.
88. Lerat, E., Daubin, V., and Moran, N.A. (2003) From gene trees to organismal

phylogeny in prokaryotes: the case of the gamma-Proteobacteria. PLoS Biol 1, E19.
89. Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Medigue, C., Rocha, E.P., and Denamur, E. (2009) Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 5, e1000344.
90. Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M.A., Nelson, K.E., Nilsson, C., Olson, R., Paul, J., Brito, B.R., Ruan, Y., Swan, B.K., Stevens, R., Valentine, D.L., Thurber, R.V., Wegley, L., White, B.A., and Rohwer, F. (2008) Functional metagenomic profiling of nine biomes. Nature 452, 629-632.
91. Bouchard, F. (2008) Causal Processes, Fitness and the Differential Persistence of Lineages. Philosophy of Science 75, 560-570.
92. Csardi, G., and Nepusz, T. (2006) The igraph software package for complex network research. InterJournal Complex Systems, 1695.
93. Newman, M.E. (2006) Finding community structure in networks using the eigenvectors of matrices. Phys Rev E Stat Nonlin Soft Matter Phys 74, 36104.
94. Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. Science 278, 631-637.
95. Li, W. (2009) Analysis and comparison of very large metagenomes with fast clustering and functional annotation. BMC Bioinformatics 10, 359.
96. Killcoyne, S., Carter, G.W., Smith, J., and Boyle, J. (2009) Cytoscape: a community-based framework for network modeling. Methods Mol Biol 563, 219-239.

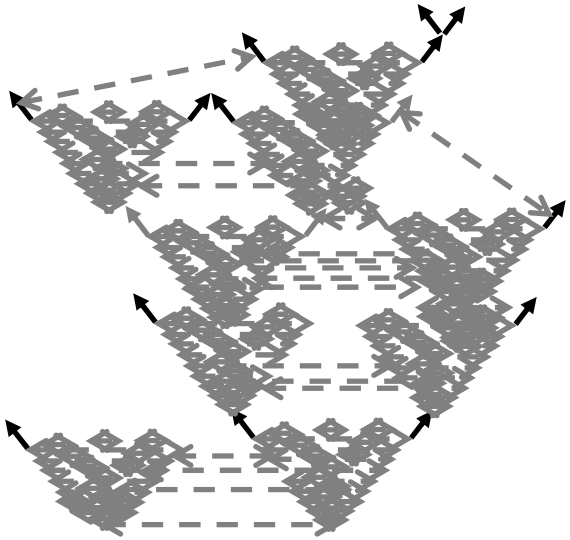
GP:
Evolutionary phenomena associated with the genealogy



Splitting events



EP:
GP + evolutionary phenomena that do not match the genealogy:



Splitting & clumping events

Evolutionary relationships
=
Genealogical relationships

Evolutionary relationships
=
Genealogical relationships
+
Other relationships

(ecological, functional, genetic partnerships)

Evolutionary units
=
Genealogical units

Evolutionary units
=
Genealogical units
+
Other evolving units

Figure 1.

Figure 2.

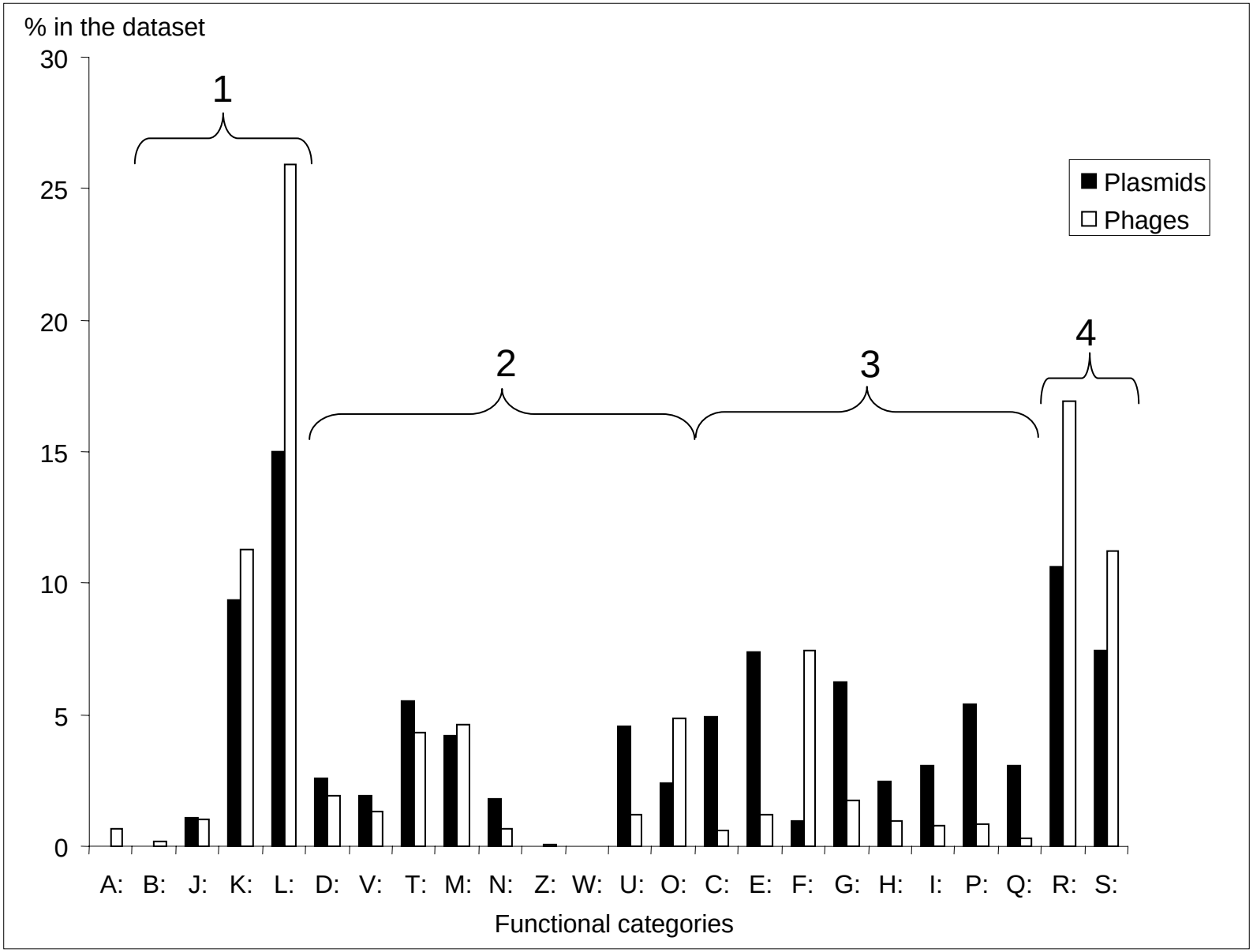
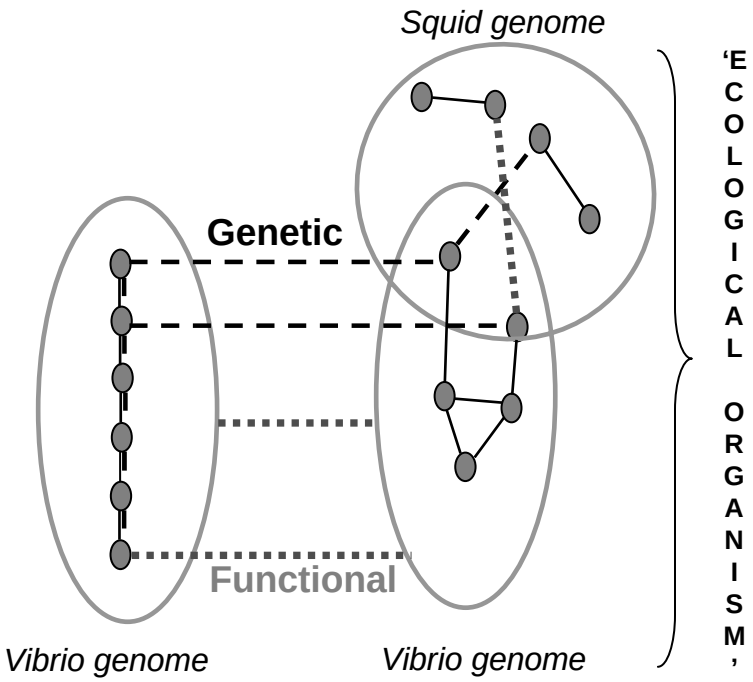
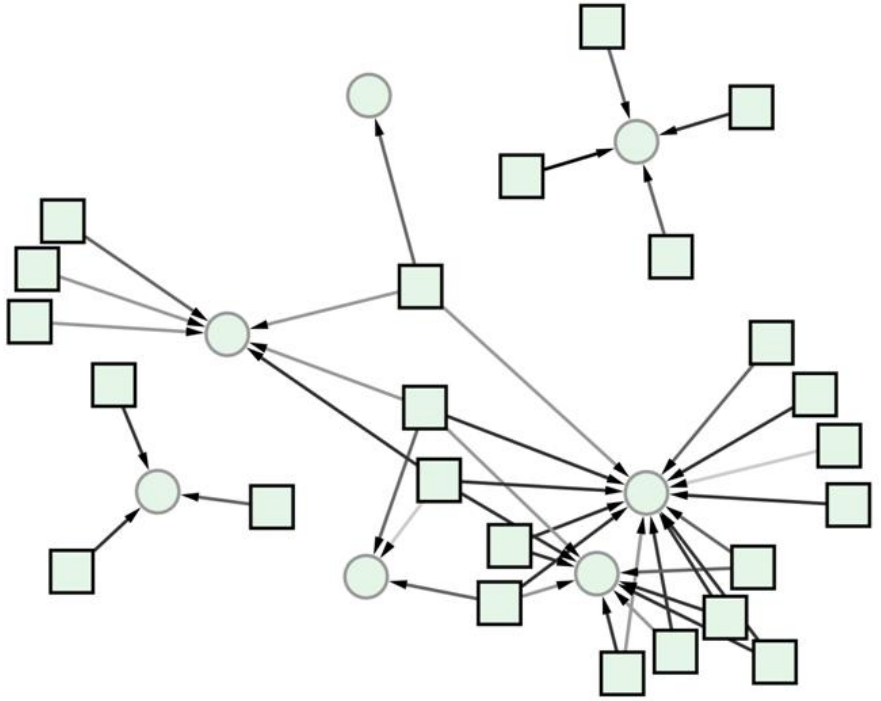

% in the dataset
30
25
20
15
10
5
0
1
2
3
4
Plasmids
Phages
A:
B:
J:
K:
L:
D:
V:
T:
M:
N:
Z:
W:
U:
O:
C:
E:
F:
G:
H:
I:
P:
Q:
R:
S:
Functional categories

Figure 3.

A.



B.



Highways of LGT exchanges deduced from a phylogenetic forest

Figure 4.

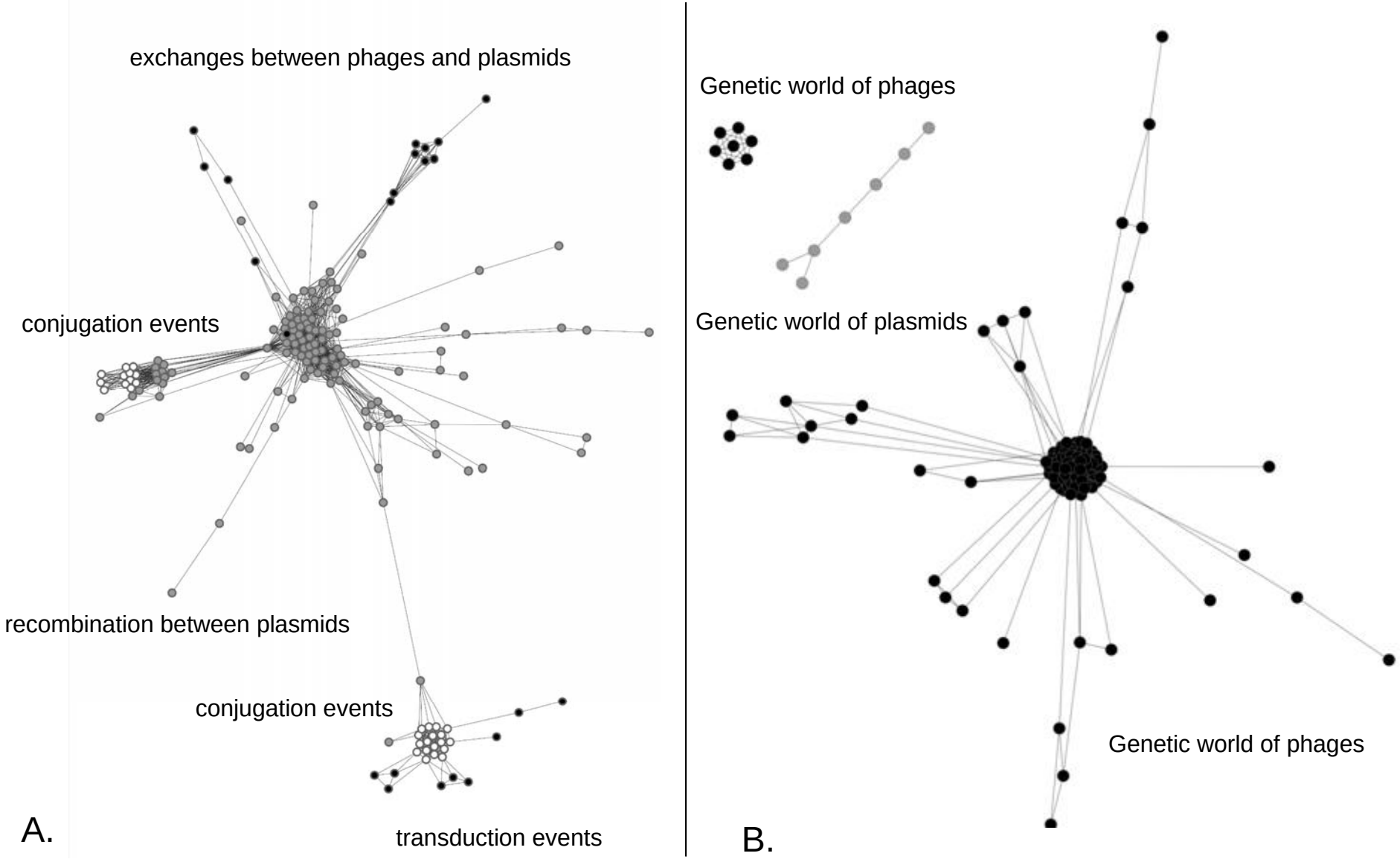

exchanges between phages and plasmids
conjugation events
recombination between plasmids
conjugation events
transduction events
A.
Genetic world of phages
Genetic world of plasmids
Genetic world of phages
B.

Figure 5.

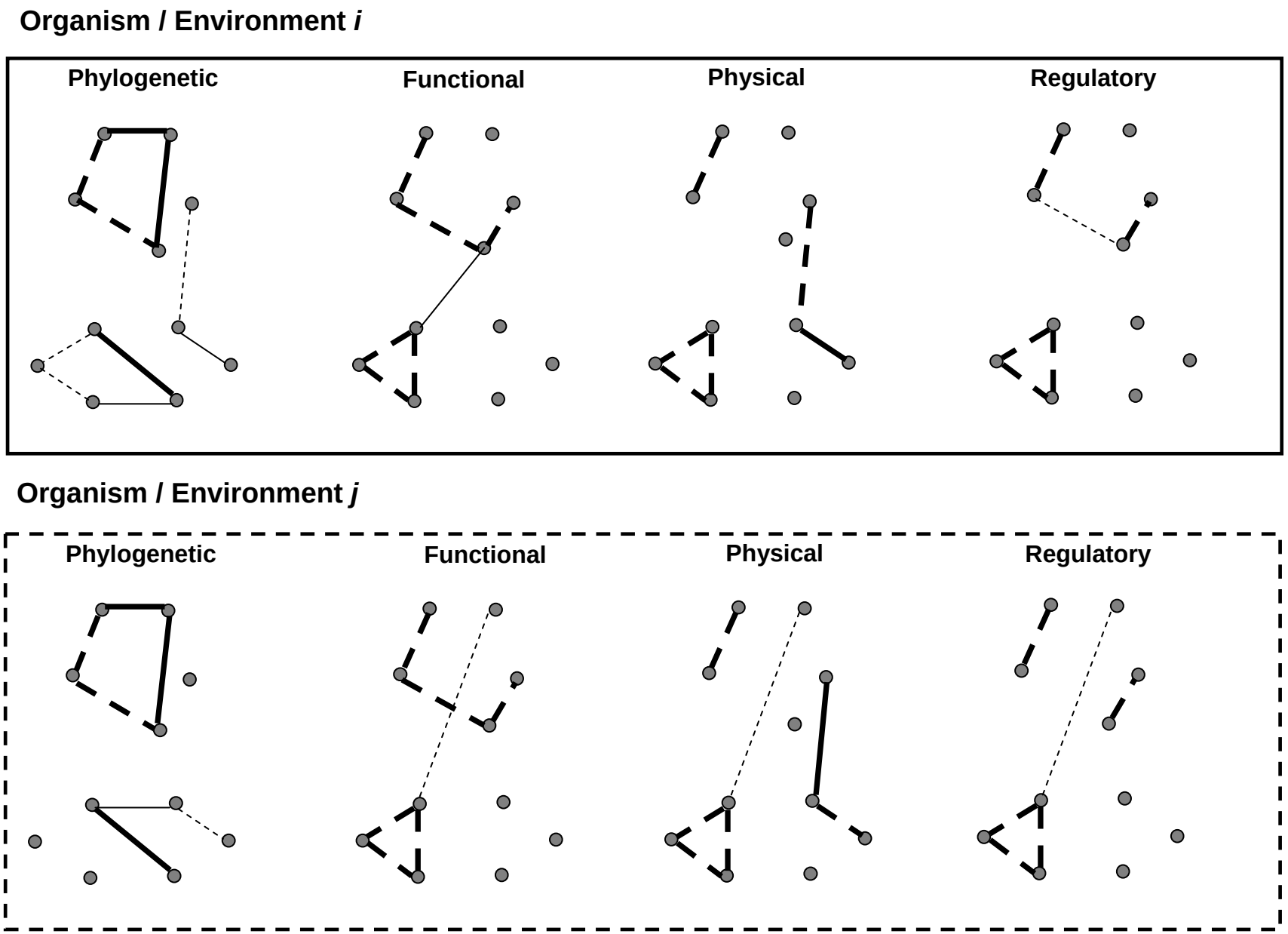

Organism / Environment *i*
Phylogenetic
Functional
Physical
Regulatory
Organism / Environment *j*
Phylogenetic
Functional
Physical
Regulatory

Figure 6.

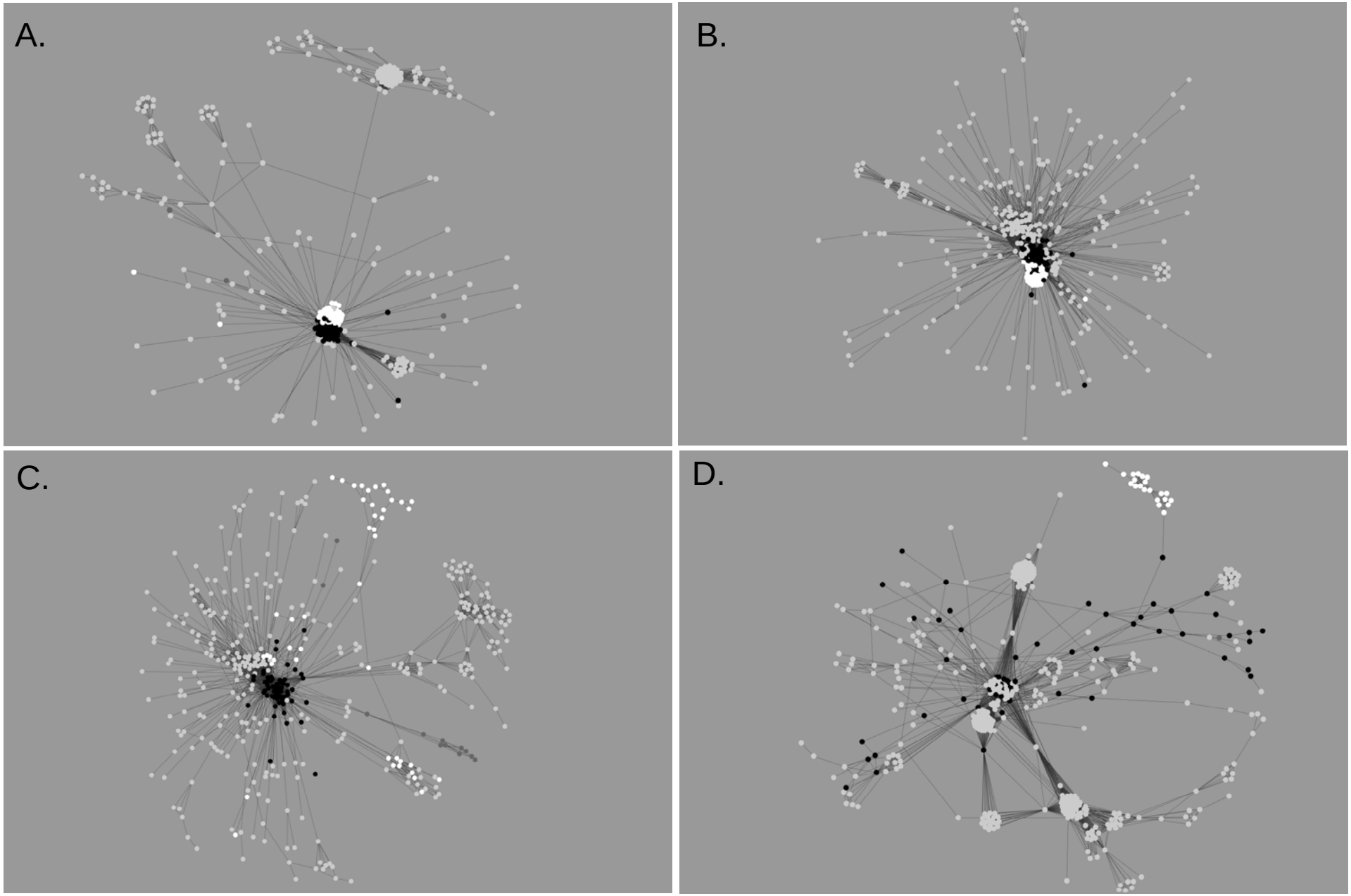
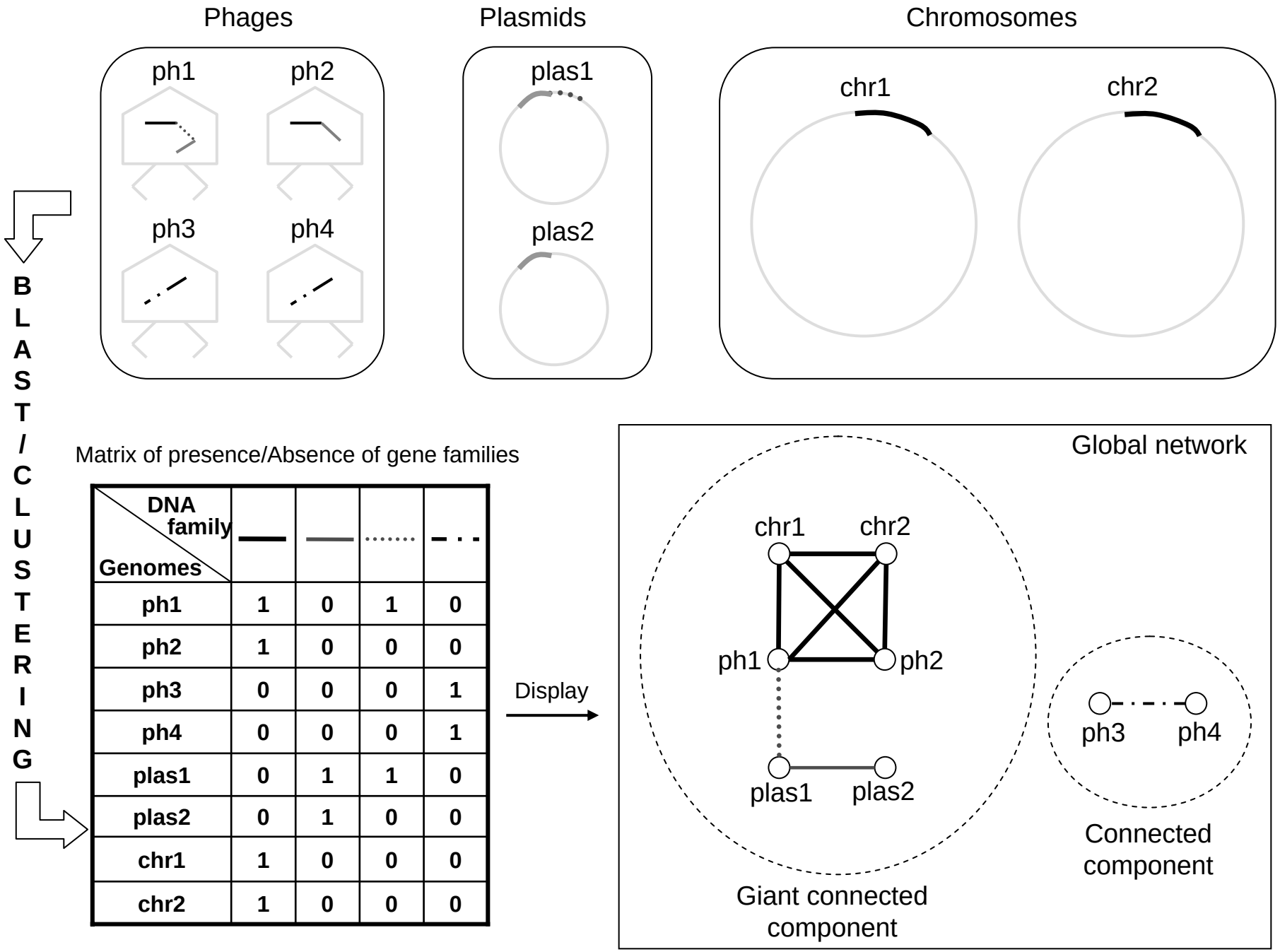


Figure 7.


Phages
Plasmids
Chromosomes
ph1
ph2
ph3
ph4
plas1
plas2
chr1
chr2
BLASTiCLUSTERING
Matrix of presence/Absence of gene families
DNA family
Genomes
ph1 1 0 1 0
ph2 1 0 0 0
ph3 0 0 0 1
ph4 0 0 0 1
plas1 0 1 1 0
plas2 0 1 0 0
chr1 1 0 0 0
chr2 1 0 0 0
Display
Global network
Giant connected component
Connected component