Evolutionary analyses of non-genealogical bonds produced by introgressive descent

Eric Bapteste[a,1], Philippe Lopez[a], Frédéric Bouchard[b], Fernando Baquero[c], James O. McInerney[d], and Richard M. Burian[e]

[a]*Unité Mixte de Recherche 7138 Systématique, Adaptation, Evolution, Université Pierre et Marie Curie, 75005 Paris, France;* [b]*Département de Philosophie, Université de Montréal, Montréal, QC, Canada H3C 3J7;* [c]*Department of Microbiology, Ramón y Cajal University Hospital (IRYCIS, CIBERESP), 28034 Madrid, Spain;* [d]*Molecular Evolution and Bioinformatics Unit, Department of Biology, National University of Ireland Maynooth, County Kildare, Ireland; and* [e]*Department of Philosophy, Virginia Tech, Blacksburg, VA 24061*

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved September 24, 2012 (received for review April 20, 2012)

All evolutionary biologists are familiar with evolutionary units that evolve by vertical descent in a tree-like fashion in single lineages. However, many other kinds of processes contribute to evolutionary diversity. In vertical descent, the genetic material of a particular evolutionary unit is propagated by replication inside its own lineage. In what we call introgressive descent, the genetic material of a particular evolutionary unit propagates into different host structures and is replicated within these host structures. Thus, introgressive descent generates a variety of evolutionary units and leaves recognizable patterns in resemblance networks. We characterize six kinds of evolutionary units, of which five involve mosaic lineages generated by introgressive descent. To facilitate detection of these units in resemblance networks, we introduce terminology based on two notions, P3s (subgraphs of three nodes: A, B, and C) and mosaic P3s, and suggest an apparatus for systematic detection of introgressive descent. Mosaic P3s correspond to a distinct type of evolutionary bond that is orthogonal to the bonds of kinship and genealogy usually examined by evolutionary biologists. We argue that recognition of these evolutionary bonds stimulates radical rethinking of key questions in evolutionary biology (e.g., the relations among evolutionary players in very early phases of evolutionary history, the origin and emergence of novelties, and the production of new lineages). This line of research will expand the study of biological complexity beyond the usual genealogical bonds, revealing additional sources of biodiversity. It provides an important step to a more realistic pluralist treatment of evolutionary complexity.

biodiversity structure | evolutionary transitions | lateral gene transfer | network of life | symbiosis

Evolutionary biologists often study the origins of biodiversity through the identification of the units at which evolution operates. In agreement with the work by Lewontin (1), it is commonly assumed that such units present a few necessary conditions for evolution by natural selection, namely (*i*) phenotypic variation among members of an evolutionary unit, (*ii*) a link between phenotype, survival, and reproduction (i.e., differential fitness), and (*iii*) heritability of fitness differences (individuals resemble their relatives more than unrelated individuals). This view, however, raises at least two difficult questions. What can be selected? What evolves by selection?

This dual concern has prompted a distinction (2, 3) between units of selection and units of evolution, distinguishing between vehicles (or interactors) (4) on which selection can act (usually individuals or populations) and replicators (usually individual genes or small complexes of genes), the ultimate beneficiaries of evolution (2, 3). Replicators are consensually seen as central to evolutionary explanations (5). However, the consensus is more fluid regarding the definition of interactors. Debates about levels of selection and the multilevel selection theory (5–10) have led to investigations of whether interactors can be found at distinct levels of organization (cells, organisms, groups of organisms, and even for some, species) when survival of genes is affected by competition on various levels of organization in ways that may conflict across levels.

For instance, some considered that kin selection among related insects was sufficient to account for the seemingly higher level of organization in collectives of eusocial insects (2, 3, 11–13). For others, the colony existed as a selectable whole, irreducible to the simple addition of individual insects' fates (14–17). This multilevel perspective seems notably justified if some replicators (genes) are favored by their phenotype expressed in individual insects, whereas other genes are favored because selection acts on their extended phenotype expressed in the collective distributed behavior in groups of insects.

Although evolutionary biologists can agree that interactions of entities at different levels of organization influence which genes that are replicated across generations, they need to explain how a hierarchy of levels of organization had itself evolved. This question was tackled in the research program on evolutionary transitions (18–21). As many works have noted (18, 20–22), complex interactors corresponding to a special type of organization did not appear *ex nihilo*; they have evolved from simpler organizational levels, and evolution itself has shaped how each of these organizational levels is maintained.

Accordingly, studies of evolutionary units must address the order, constraints, and processes through which units from different levels emerged. Distinct cases were made to explain micro- and major evolutionary transitions. For instance, it was proposed that evolution of higher-level interactors results from the functional integration and suppression of competition between related lower-level interactors, like in scenarios for the "fraternal" transition from unicellularity to multicellularity (23), or from the "egalitarian" assortments of unrelated entities interacting in ways that lead to new entities (23), like in the symbiogenetic account of the eukaryotes in the work by Margulis (24).

Although evolutionary scenarios often focused on transitions affecting members within a single lineage, there is increasing evidence that processes using genetic material from multiple sources also had major effects on the evolution of a diversity of interactors. Recombination, lateral gene transfer (also called horizontal gene transfer) (*SI Text*, section 1), and symbiosis contribute to the structure of the biological world in ways that differ from vertical descent alone (25). Novelty-

Author contributions: E.B., F. Bouchard, F. Baquero, and R.M.B. designed research; P.L. performed research; J.O.M. contributed new reagents/analytic tools; E.B. and P.L. analyzed data; and E.B., P.L., F. Bouchard, F. Baquero, J.O.M., and R.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: eric.bapteste@snv.jussieu.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206541109/-/DCSupplemental.

generating genetic combinations have produced a variety of evolutionary outcomes at different hierarchical levels (26). Examples include domain sharing between gene families (27), transfer of adaptive genes in prokaryotic genomes (28–32), pangenomes (33), and sharing of transposases (34), integron gene cassettes (29), plasmids (35), and phages (28, 31) within genetic exchange communities (36); bacterial consortia, such as *Chlorochromatium aggregatum*, with partners undergoing synchronized separate cellular divisions (37); and endosymbiotic gene transfer (38, 39).

In cases of symbiosis, mutualistic, commensal, and even parasitic relationships, gene exchange is not a necessary condition for the formation of higher-order entities that are composed of separate units with their own genomes. The contributing entities can profit from the combined resources made possible by interactions between the products encoded by the genes of the partners and can also yield an entity that is subject to selection in its own right. For instance, biofilms; colonial organisms [*Volvox* (40), sponges, Portuguese Man-O'-War, and the aggregates and slugs of *Dictyostelium discoideum* (41)]; multicellular eukaryotes, insect hosts, and the *Wolbachia* that determine their sex or other traits (42, 43); lichens (44, 45), herds and packs of social animals, communally organized (quasisocial and eusocial) social insects; and commensal and symbiotic gut microbes of insects and vertebrates together with their hosts are all excellent candidates to count as higher-order entities (or collective reproducers) (18). The genome of such collective reproducers should be counted as including all of the genetic material of their components (18, 42, 46).

Although such (compound) multigenomic and mosaic beings are widely recognized, disagreements about the establishment of their boundaries and pervasiveness of the processes involved in their making affect thinking about evolutionary units. If these processes are frequent, it becomes necessary to track entities in non-tree patterns, because their component parts depend on genetic material originating by introgressive descent from more than one lineage. Consider, for example, the microevolution of humans. Whereas the metaphor of a human genealogical tree is often used to back up the tree metaphor in evolution, it is only when focusing exclusively on the paternal or maternal line of descent that a portion of human evolution (in fact, of any sexual species) can be put on a dichotomously branching tree. To do justice to the evolutionary processes at play in sexual species, the genealogy of all organisms with two parents (not just humans) would be better described by a model that accounts for these dual origins and the process of sexual reproduction between two partners at each generation. The same logic holds true, we believe, not only for sexual organisms but also, all cellular organisms and evolutionary entities (i.e., phages, plasmids, lichens, eusocial insect communities, etc.) resulting from assortment of genetic material from more than one source.

We are not the first to suggest that a different formalization of evolutionary processes is useful to investigate the diversity of evolutionary units. For instance, the work by Godfrey-Smith (18) recently used a multidimensional space as a heuristic device to handle entities that evolve under processes with a non-Darwinian character (*SI Text*, section 2). In particular, it models evolutionary transitions that proceed through the aggregations of different reproducers (e.g., individual cells) with independent evolutionary activities that are increasingly constrained as their collective (e.g., a colonial organism) engages in a form of reproduction in its own right, gains autonomy (e.g., through central control), and acquires differential fitness. Importantly, this formalization highlights that biological complexity and evolutionary transitions do not occur solely in paradigmatically Darwinian populations that are characterized by (*i*) a relatively high fidelity of heredity, (*ii*) dependence of their reproductive differences on intrinsic characters, and (*iii*) similar organisms, to a large extent, having similar fitness. Following this lead, the study of interactors evolving by non-paradigmatically Darwinian processes could benefit from a network-based formalization that explicitly models the provenance of their genes (replicators) (36, 47–49).

We elaborate below on the evolutionary transition research program to propose that interactors are much more varied than is often assumed, and we suggest how to apply network tools to genomic datasets to detect genetically mosaic interactors. We argue for the importance of selectable entities comprised of replicators or components from more than one ancestral source as the result of either evolutionary transitions or combinations of elements that might be on the way to such a transition. Some evolutionary structures produced by such an assortment between distantly related lineages and even unrelated lineages (e.g., viruses and cells or cooperating individuals from different phyla in a symbiotic relationship) can be detected through remarkable patterns in genetic and genomic resemblance networks (36, 47–49) that differ from the transitive relationships of homology between objects evolving from a last common ancestor produced by vertical descent. We introduce network-based notions to facilitate recognition of these patterns in gene and genome networks and the patterns of additional classes of evolutionary units. Finally, we discuss how identifying these additional evolutionary patterns, orthogonal to the patterns produced by homology relationships, could stimulate radical rethinking of key questions in evolutionary biology.

Mergers and Clubs as Relevant Evolutionary Units. Members of monophyletic groups, evolving by clonal division and allowing for continuing mutational diversification in members of clonal complexes, characteristically share genes that trace back to a single locus in a single individual (in fact, the same locus in a single genome of a last common ancestor). We call such genes coalescent orthologs to distinguish them from shared genes originating from different processes. Indeed, many genetic similarities between biological objects are not caused by vertical descent, where the genetic material of a particular entity is propagated by replication inside its own lineage. For instance, adaptive lateral genetic transfer between genomes of entities from different lineages that share the same environment or lifestyle (29, 32, 46) indicates additional (non-vertical) mechanisms for the integration of genetic material into one host. Hence, another type of descent is fundamental to the reconstruction of an accurate evolutionary picture of the evolutionary units.

What we call introgressive descent occurs precisely when the genetic material of a particular evolutionary unit first propagates into different host structure(s) and then is propagated within or by the resulting unit(s). Examples include a transposon inserted into a series of different plasmids, a plasmid in different bacterial clones, a clone in different microbiomes, the mitochondrial genes present in a eukaryotic cell (regardless of whether those genes have been transferred into the nuclear genome), and the commensal combination of an alga and a fungus in a lichen that is propagated by vegetative reproduction or diaspores (44, 45, 50). The typical biological outcomes of these interlineage and interlevel assortments, namely the mosaic objects, and the multilineage coalitions of genetic partners involved in these processes can be stabilized and selected, becoming important evolutionary players in their own right (46). Therefore, introgressive descent generates non-genealogical bonds between biological objects, producing a reticulate evolutionary framework.

To account for the origins and features of these objects, we propose that, in

addition to single lineages resulting from the splitting processes of vertical descent, evolutionists should formally recognize a range of mosaic evolutionary units produced by introgressive descent. This range has two extremes. First, there are mergers. Mergers arise when two or more components, not hitherto coexisting within the same unit, are brought together, and these components are subsequently replicated or propagated within or by a new single corporate body (9). Often, component parts of mergers do not trace back to a single locus (or set of loci) in a single last common ancestor. Mergers exist at multiple levels of biological organization [molecular (27, 51), genomic (25, 52–54), and organismal (39, 55, 56)] and do not all subtend the same genetic consequences. Fused genes conferring drug resistance (35), new viral genomes (49), lineages created from symbioses (39, 56), and Russian dolls of mobile genetic elements (52, 53) are among the best known examples of mergers. The offspring of sexual reproduction are also obligate mergers, because their parts come from distinct—although closely related—sources (two parents instead of one last common ancestor). Many mergers bring together elements that were capable of independent replication before and can replicate only as part of a larger whole after their union (19); in such cases, they present typical signs of evolutionary transitions.

Second, there are multilineage clubs. Members of these clubs form coalitions of entities that replicate in separate events and exploit some common genetic material that does not trace back to a single locus in a single last common ancestor of all of the members (26, 29, 31, 32, 57, 58). Multispecies biofilms (59), environmental coalitions of cells and mobile genetic elements like those elements of marine cyanophages and cyanobacteria (28), and genetic exchange communities in gut microbiomes (31, 60, 61) provide examples of such multilineage clubs. These assortments may result in evolutionary transitions if the club exhibits some form of reproduction in their own right.

Some independently reproducing components of a larger whole will also fall between these two extreme poles that are produced by introgressive descent. Thus, the mycobionts and photobionts of most lichens may reproduce independently (although in such cases, the offspring of the mycobiont must find and incorporate an appropriate photobiont to be lichenized again), but they may also reproduce by vegetative reproduction or diaspores; therefore, they may be treated as facultative mergers (44, 45, 50). In contrast, the mycobionts of some populations of lichens seem to have lost the power of independent reproduction; such lichens are (obligate) mergers for their components that cannot reproduce independently (62). Consequently, empirical evidence regarding reproduction, maintenance mechanisms, integration, and fitness of each proposed merger (or club) is required for a detailed evaluation of why particular genetic assortments (or coalitions) based on the sharing of genetic material count (or not) as bona fide evolutionary units or are on a path to an evolutionary transition (*SI Text*, section 2 and Fig. S1).

In fact, when embracing the common definition of lineages (where groups of closely related entities belong to the same lineage by contrast to different lineages, which refer to groups of more distantly related entities) and the common definition of levels of biological organization (with cells and mobile genetic elements belonging to different levels), we propose to distinguish no fewer than five main classes of candidate evolutionary units. These units are (*i*) intralineage mergers, (*ii*) interlineage mergers, (*iii*) interlevel mergers, (*iv*) multilineage clubs, and (*v*) multilevel clubs, depending on whether the genetic material shared by introgressive descent comes from a single lineage and level of biological organization or more (*SI Text*, sections 1 and 2).

Examination of the importance of such units should broaden (and may challenge) traditional descriptions of evolutionary history, which are still largely focused on single lineages with evolution that can be modeled by a tree. We must, therefore, think about methodological innovations to deal with these additional interactors, which can include the use of directed or undirected cyclical graphs known as networks and the use of a simple graph-based terminology.

Tracking Non-genealogical Bonds in Evolutionary Networks. Networks, consisting of nodes connected by edges, are a natural way to capture specific patterns resulting from the distribution of genetic material from more than one source (36, 47–49). These graphs can represent genetic diversity at different levels of biological organization. For instance, gene networks represent sequences by nodes, and these nodes are connected by edges when they manifest significant similarity (63). Genome networks represent genomes as nodes, and these nodes are connected by edges when they share common features (e.g., the same sequence or the same gene family) (47–49).

In genome networks, monophyletic groups will generally produce cliques (Figs. 1 and 2*A* and Table 1) (i.e., subgraphs in which all nodes are directly connected to one another), because all entities under study share some coalescent orthologs. However, when the similarity of characters decreases under a given threshold through evolution, a different

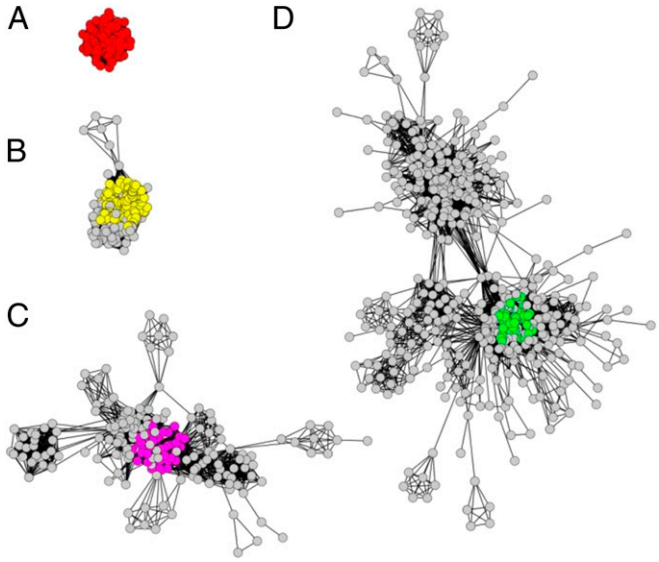

Fig. 1. Selection of gene network components displaying their largest maximal clique. Genes (nodes; in gray) aligning >80% of their sequences with their match in a BLAST analysis (showing >50% identification and a BLAST score < 1 e^{-20}) are directly connected. Sequences belonging to the largest maximal clique, defining the largest set easily amenable for a single phylogenetic analysis, are highlighted in bright colors. The largest maximal clique only covers a portion of each component, meaning that numerous similarity relationships and evolutionary relationships cannot be investigated using a single tree. *A* corresponds to the Holin BlyA family (only plasmids; 87 nodes in the clique). *B* contains AC3 and replication enhancer proteins (only viruses; 140 nodes in the clique). *C* corresponds to transposases OrfB family (mostly plasmids and a few prokaryotes; 50 nodes in the clique). *D* corresponds to oligopeptide ABC transporter ATP-binding proteins (mostly prokaryotes and some plasmids; 38 nodes in the clique).

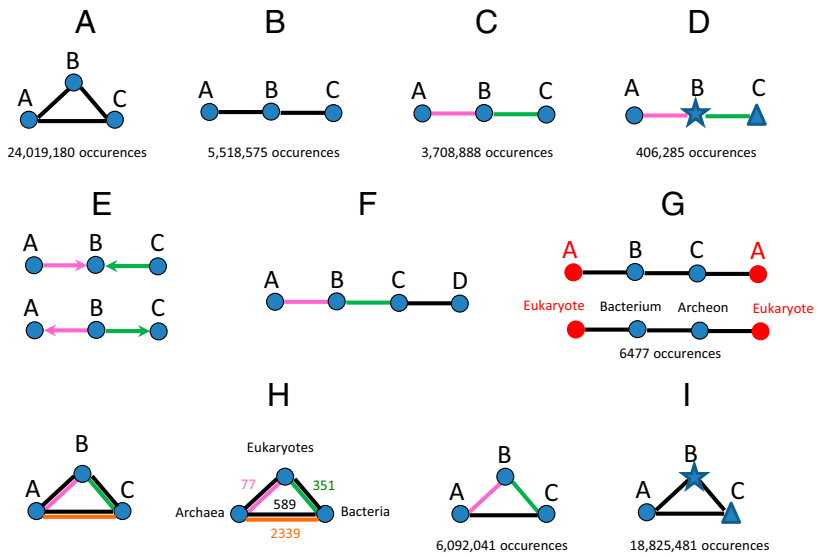

Fig. 2. Patterns with evolutionary significance in resemblance networks. Each symbol indicates an entity (node) from a distinct level of biological organization. Similarly colored edges indicate vertically inherited shared characters. Occurrences in our test dataset at >50% identification are quantified when available. (*A*) Clique (here, a triangle) capturing a homology relationship between A, B, and C. (*B*) P3 occurring when a homologous character evolved beyond recognition between A and C. (*C*) M-P3 indirectly connecting two entities through a third one by different (pink and green) shared characters. (*D*) Multilevel M-P3 indicating multilevel evolutionary units. (*E*) Polarized M-P3 showing B as a merger or as a fissioning unit. (*F*) P*n* (here, four). (*G*) P*n* with the distantly related parts from a merger entity A. (*H*) Hardly detectable M-P3s. (*Left*) Ancient core characters mask a recent combination of characters in B. (*Center*) Real numbers of shared gene families between domains of life. (*Right*) Aggregation of three M-P3s looking like a clique. (*I*) Multilevel cliques.

pattern is produced: some edges disappear, and cliques are replaced by intransitive chains, with adjacent objects of the chain presenting similarity up to a certain threshold (Fig. 2*B*). In agreement with the terminology of graph theory, we call such a subgraph of three nodes (A, B, and C) a P3 (64), where A is linked to B, B is linked to C, and A is not linked to C. This concept can be easily extended to the case where A, B, and C are not nodes but instead, cliques; in graph theory, B is called a minimal clique separator (65).

By contrast, we call mosaic-P3 (M-P3) a P3, in which two entities, A and C, are indirectly connected through a third entity, B, by one or more characters that are not coalescent orthologs (Fig. 2*C*). Such an M-P3 unites at least two distantly related and/or unrelated lineages through a third entity acting as an intermediate binder. By definition, this structure is beyond the reach of a single-tree analysis; A and C cannot be compared directly, because they lack homology for the traits under study. The relationship between A and C is not an intrinsic property of these two objects, and it is distinct from homology. Consequently, such M-P3s offer non-genealogical bonds to detect multilineage clubs (when all nodes of the M-P3 represent entities from different lineages but at the same level of biological organization) or multilevel clubs (when some of its nodes represent entities from different levels of biological organization; e.g., cellular chromosomes, phages, and plasmids) (Fig. 2*D*). Moreover, when polarized, M-P3s can be used to detect mergers (Fig. 2*E*) when the binder receives genetic contributions from two sources (*ex pluribus unum*), or M-P3s can be used to detect that a fissioning entity has contributed materially distinct objects (*ex unibus plurum*) (66). In both mergers and contributions to separate entities, the involved entities may belong to the same level or to different levels of biological organization.

We define P*n*, when *n* entities can be arranged, as a chain of n-2 P3s (Fig. 2*F*). Importantly, a P*n* can also detect mosaic units, when entities at its extremities are distinct parts of the same entity (Fig. 2*G*) (e.g., when the terminal nodes in a gene network are two genes present in the same organism but acquired from distinct sources).

Such simple patterns of the connections can facilitate the study of introgressive descent in networks. As a quick proof of concept, we assembled and BLASTed all-against-all, a dataset of 336,402 cellular protein sequences, from the complete genomes of 54 Archaebacteria, 70 Eubacteria, and 7 Eukaryotes sampled all over the web of life (the taxa are listed in *SI Text*, section 3) and 228,042 mobile genetic element protein sequences, comprising all viral and plasmid sequences available from the National Center for Biotechnology Information as of May of 2011. These sequences are available in the download section at www.evol-net.fr. We built gene networks (www.evol-net.fr) by connecting two sequences if they shared a BLAST hit displaying more than a given percentage identity (e.g., 50%, 70%, 90%, and 99%) and considered edges corresponding to a BLAST hit covering more than 80% of both sequences as sequence-homologous. In this case, we observed 6,477 P*n* patterns in our gene network, with distantly connected genes from the same homologous family in eukaryotes: one acquired from an archaebacterial ancestor, and the other acquired from a bacterial endosymbiont (mitochondria or chloroplast). Many of these P*n* were tracking the same ancient event of endosymbiotic transfer.

Although M-P3s can be characterized in terms of graph theory, their detection can be complex. For instance, M-P3 patterns can be masked by additional bona fide homology bonds between the entities caused by other characters (Fig. 2*H*). Other M-P3s can be missed when two characters assumed to be coalescent orthologs are not. This situation can occur for gene families with significant amounts of in and out paralogy (67) or in the extreme case of nearly identical replacement of genetic material by sequence-homologous copies. Finally, cliques with unrelated entities (Fig. 2*I*) also deserve particular consideration, because they are not united by vertical descent. Their topology suggests the sharing of genetic material in multilevel clubs.

Formally naming these P3s (and cliques) is a first step for implementing their systematic detection to better track evolutionary transitions and evolutionary units using both genealogical and non-genealogical bonds. Typically, in our real dataset, no single tree can analyze all of the connected sequences in the gene network, because no single clique with more than four sequences entirely covers a connected component uniting sequences with significant similarities (Fig. 1 and Table 1). Only a fraction of the sequences in a gene network included in such cliques (counted using maximal clique enumerator) (68) are amenable to classic phylogenetic analysis; 11.5–36.3% of the sequences are present in P3, meaning that their relationships of homology are also too distant to be accounted for by a single tree. In addition, a fair proportion of sequences (from 3.8% to 28.9%) belongs to M-P3 and multilevel P3 (up to 11.4%) subgraphs, further hinting at phenomena of introgressive descent (Table 1). Likewise, although numerous sequences belong to triangles connected by homology edges, suggesting that their similarity

Table 1. Counts of maximal cliques, P3, and M-P3 in a real test dataset of over 330,000 sequences

Identity threshold (%)	Nodes	Average number of cliques by CC	Percent nodes in cliques (in MLvl cliques)	Percent nodes in H triangles (in MLvl H triangles)	Percent nodes in Syn triangles (in MLvl Syn triangles)	Percent nodes in P3 (in MLvl P3)	Percent nodes in M-P3 (in MLvl M-P3)
50	295,606	35.1	46.8 (10.1)	66.3 (40.4)	27.4 (18.3)	36.3 (11.4)	28.9 (8.5)
70	178,558	60.8	35.8 (2.7)	59.4 (44.8)	17.9 (14.9)	16.7 (2)	15.2 (1.1)
90	104,851	0.2	36.9 (0.8)	57.6 (52.4)	14.1 (12.3)	12.1 (0.3)	10.9 (0.4)
99	44,592	0.2	31.8 (0.7)	55.3 (50.9)	4.7 (4.2)	11.5 (0.1)	3.8 (0.1)
Examples	NA	NA	Fig. 1	*	†	‡	§

Maximal cliques of four nodes and more that were amenable to phylogenetic studies were referred to as cliques. Triangles, based on homology edges only (called H triangles) or sharing of distinct genetic material (called Syn triangles), and P3s were enumerated using in-house scripts, which are available from Philippe Lopez on request. P3s for which at least one of two edges was not homologous were labeled M-P3s. Triangles, P3s, and cliques harboring both cellular and mobile genetic elements sequences were labeled multilevel (MLvl). The percentage of sequences involved in each pattern was estimated. It does not sum to 100%, because a given sequence can simultaneously be part of distinct patterns, in which they are involved through different sets of neighbors. A few real examples corresponding to these patterns are provided for the network at 50% identification threshold (genInfo identifier numbers are indicated). CC, connected component.

*Sharing of cyanophycin synthetase by A (*Cyanothece sp*. ATCC 51142_172037152), B (*Nostoc punctiforme* PCC 73102_186685868), and C (*Gloeobacter violaceus* PCC 7421_37523895). Sharing of fosfomycin resistance protein by A (a plasmid of *Staphylococcus aureus*_170780437), B (a chromosome of *Bacillus cereus* Q1_222095687), and C (a virus, *Bacillus* phage Cherry_77020211).

†The bifunctional protein HldE, glycerol-3-phosphate cytidylyltransferase, and ADP-heptose synthase of *Thermodesulfovibrio yellowstonii* DSM 11347_206890027, *Fusobacterium nucleatum* subsp. nucleatum ATCC 25586_19704265, and *Bdellovibrio bacteriovorus* HD100_42524647 follow this pattern. Late competence protein, S-layer protein, and β-lactamase domain protein of a virus, *Geobacillus* phage GBSV1_115334647, a chromosome of *Bacillus cereus* Q1_222096303, and a plasmid of *Geobacillus sp*. WCH70_239828744, respectively, follow this pattern.

‡Sharing of ammonium transporter Amt by A (*Methanobrevibacter ruminantium* M1_288560581), B (*T. yellowstonii* DSM 11347_206890102), and C (*Leptospira interrogans serovar Lai str*. 56601_294828399). Sharing of 6-phosphogluconate dehydrogenase-like protein by A (a virus, Synechococcus phage syn9_162290189), B (a plasmid of *Anabaena variabilis* ATCC 29413_75812812), and C (a chromosome of *Chloroflexus aurantiacus* J-10-fl_163846093).

§B (*Bacteroides fragilis* YCH46_53714858) shares parts of its bifunctional methionine sulfoxide reductase A/B with the methionine sulfoxide reductase of A (*Clostridium acetobutylicum* ATCC 824_15893384) and other parts with the methionine sulfoxide reductase B of C (*Bordetella pertussis* Tohama I_33594433). B (the chromosome of *Rickettsia rickettsii str*. Iowa_165933859) shares parts of its lysozyme with A (the lysozyme of a virus, Bacteriophage APSE-2_212499717) and other parts with the lysozyme of C (a plasmid of *Azospirillum sp*. B510_2_288961413).

results from vertical descent, in a vast majority, these triangles contain sequences from genomes from distinct levels of organization, indicating important amounts of genetic sharing between unrelated entities. Moreover, depending on the threshold retained to construct the gene network, an additional 4.7–27.4% of triangles present in the network would rather be explained by the introgressive sharing of unrelated (or extremely divergent) fragments of DNA between the three connected elements. Thus, the detection and recognition of such non-genealogical bonds possibly yield deep consequences for evolutionary knowledge.

Evolutionary Thinking Beyond Genealogical Bonds. The systematic analysis of M-P3 patterns in networks suggests that one should assign comparable ontological importance to evolutionary transitions in both single lineages and phylogenetically mosaic units to broaden the analysis of four types of evolutionary questions.

First, the origin of evolutionary novelties is generally considered through the impact of (selective/selected) mutations and recombination in nucleotide sequences within a genome (69) or random drift in populations. Although a number of mutations in key regulatory nodes might produce quite complex phenotypes, this focus must be expanded to solve the problem of how big novelties are acquired (e.g., how assembly of original combinations of preexisting, often unrelated biological entities increases diversity at every level of biological organization) (70, 71). A compelling example can be found in the recent expansion of a bacterial gene $bla_{\text{CTX-M-15}}$, which inactivates most modern cephalosporin antibiotics in *Escherichia coli*. The ancestral gene of this detoxifying enzyme was a housekeeping gene in an organism ecologically accessible by *E. coli* and its plasmids, captured by an insertion sequence, and then moved into plasmids that were captured by particular cosmopolitan *E. coli* clones, including the widespread high-risk clone ST131-O25:H4-B2, which contributed to its worldwide spread. The $bla_{\text{CTX-M-15}}$ gene was then captured by new plasmids, which were captured in their turn by other *E. coli* clones. Because some of these clones are particularly suited to be integrated in the intestinal microbiome of different types of animals, the $bla_{\text{CTX-M-15}}$ gene expanded multidimensionally, finally reaching even the hemolytic–uremic *E. coli* O104 responsible for food poisoning in Germany in 2011 (72–74). Consideration of M-P3s, the true binding of unlike to unlike at the origin of original evolutionary units, explicitly includes such evolutionary quantum leaps in studies of evolutionary novelties.

Second, introgressive and vertical descent can enrich models pertaining to the Darwinian threshold (75) (i.e., the time at which cellular lineages acquired sufficient autonomy, as lineages, to diverge from each other). After this threshold was crossed, the bonds of homology became more striking than the structures produced by M-P3s, but homology is not the only guideline to explain this early transition in the history of life. Considerations of vertical descent alone suggest that the more recent common ancestor of life would be more ancient than the Earth (76, 77), which seems impossible. Introgressive descent can, therefore, also contribute to understanding of early evolution. Interlevel mergers and multilevel clubs were likely key elements in the pre-Darwinian world (78). Investigations of ancient evolution should benefit from research to define the pool of shared genes of early multilineage and multilevel clubs rather than hinge on the definition of the single minimal cellular genome inferred from genealogical bonds between extant cellular beings. Unless introgressive descent is acknowledged, there will be Lost Common Ancestors: the contemporary mosaic evolutionary units of the hypothetical last common ancestor.

Third, the origin of lineages is often considered as a problem of branching order on a tree. However, genetic assortments crossing lineages and levels also yield lineages of major evolutionary players. The entry of eukaryotes on the scene, whether as the product of some sort of fusion (56) or successive endosymbioses

(24, 79), provides an obvious example. Any selective pressure favoring the stabilization of a merger (e.g., when the merged entity acquires better resistance to parasites or pathogens) can produce the non-tree like evolution of ecologically successful novel lineages. For example, a selective sweep might occur in the descendants of an individual bacterium that harbored a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) that acquired new spacers conferring greater resistance to phages in a given ecological niche (80, 81). Thus, considerations of M-P3 patterns could explain the origin of what we recognize as lineages (e.g., some microbial lineages corresponding to ecotype) (82, 83) without a tree. Likewise, members of sexual species can be studied as the result of the stabilization of the obligate mergers produced during sex (84). Hence, non-genealogical bonds could replace the series of dichotomies often used to model some intermediate stages of lineage evolution in prokaryotes and eukaryotes.

Fourth, evolutionary explanations generally rely on the comparison of variations of vertically inherited features. However, the systematic detection of mergers and clubs, defined by non-genealogical bonds, can increase the number of evolutionarily relevant comparisons. This enlarged comparative scope accommodates more complex questions regarding "egalitarian" evolutionary transitions. The origin of (compound) multilineage units is possibly no less fundamental than the origin of multicellularity. Both phenomena require explaining how distantly related entities (e.g., cells or mobile elements) reach their current level of integration and the mechanisms deployed for passing on traits that belong to the complex rather than particular individuals or lineages. Similar questions can be raised for multilevel organizations. Thus, comparative analyses of multiple multilineage/multilevel clubs could identify convergent mechanisms, features, genomic properties, ecological affinities, or functional capacities of the members of such clubs. Analyses of M-P3s can set up an analytical framework to define the possible rules (85, 86) (the grammar of associations between the different entities), even in the absence of genealogical continuity.

Conclusions

Richard Owen proposed that instances of the same organ under every variety of form and function should be considered homologs. Darwin proposed a genealogical cause for that homology. He, thus, established a hidden bond particularly suited to diagnose and explain evolution of single lineages. Ever since that time, biologists have preferentially investigated evolutionary changes through relationships of homology and tree-like genealogical patterns. However, increasingly many evolutionary units and transitions seemed to depend on and arise from non-tree like processes. In particular, the analysis of the evolution of mergers and clubs requires us to uncover other bonds, reaching beyond strict kinship and beyond one biological level. Because introgressive descent structures biodiversity in ways that vertical descent does not, it seems essential to study the patterns caused by intersections and genetic exchanges between lineages (and not just within lineages). By starting with patterns as simple as M-P3s, it should be possible to improve our understanding of past, present, and future biological evolution significantly and encourage the inclusion of additional evolutionary units in our description of biological evolution. This line of research can expand the study of biological complexity beyond the usual genealogical bonds, revealing additional sources of biodiversity, and promote additional developments of the analytical apparatus required for network analysis to handle even more complex patterns generated by introgressive descent. We commend it to our readers.

ACKNOWLEDGMENTS. The work of F. Bouchard is funded by the Social Sciences and Humanities Research Council of Canada. The research of F. Baquero is sponsored by the Seventh Framework Programme of the European Union (PAR-241476 and EvoTAR-282004) and the Carlos III Institute Research Fund (FIS-PI10-02588). J.O.M. is funded by the Science Foundation Ireland Research Frontiers Programme (09/RFP/EOB2510).

1. Lewontin RC (1970) The units of selection. *Annu Rev Ecol Syst* 1:1–18.
2. Dawkins R (1976) *The Selfish Gene* (Oxford Univ Press, Oxford).
3. Dawkins R (1982) *The Extended Phenotype: The Gene as the Unit of Selection* (Oxford Univ Press, Oxford).
4. Hull DL (1980) Individuality and selection. *Annu Rev Ecol Syst* 11:311–332.
5. Lloyd EA (2012) *The Stanford Encyclopedia of Philosophy*, ed Zalta EN (Metaphysics Research Lab, Stanford, CA), pp 1–40.
6. Brandon RN (1982) The levels of selection. *PSA 1982*, eds Asquith P, Nickles T (Philosophy of Science Association, East Lansing, MI), vol. 1, pp 315–323.
7. Gould SJ, Lloyd EA (1999) Individuality and adaptation across levels of selection: How shall we name and generalize the unit of Darwinism? *Proc Natl Acad Sci USA* 96(21):11904–11909.
8. Keller L (1999) *Levels of Selection in Evolution* (Princeton Univ Press, Princeton).
9. Okasha S (2006) *Evolution and the Levels of Selection* (Oxford Univ Press, Oxford).
10. Wilson DS (1997) Altruism and organism: Disentangling the themes of multilevel selection theory. *Am Nat* 150(Suppl 1):S122–S134.
11. Hamilton WD (1964) The genetical evolution of social behaviour. I. *J Theor Biol* 7(1):1–16.
12. Hamilton WD (1964) The genetical evolution of social behaviour. II. *J Theor Biol* 7(1):17–52.
13. Williams GC (1966) *Adaptation and Natural Selection* (Princeton Univ Press, Princeton).
14. Nowak MA, Tarnita CE, Wilson EO (2010) The evolution of eusociality. *Nature* 466(7310):1057–1062.
15. Strassmann JE, Queller DC (2007) Insect societies as divided organisms: The complexities of purpose and cross-purpose. *Proc Natl Acad Sci USA* 104(Suppl 1):8619–8626.
16. Turner JS (2000) *The Extended Organism: The Physiology of Animal-Built Structures* (Harvard Univ Press, Cambridge, MA).
17. Wilson DS, Sober E (1989) Reviving the superorganism. *J Theor Biol* 136(3):337–356.
18. Godfrey-Smith P (2009) *Darwinian Populations and Natural Selection* (Oxford Univ Press, Oxford).
19. Maynard Smith J, Szathmáry E (1995) *The Major Transitions in Evolution* (Freeman, New York).
20. Michod RE (1999) *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality* (Princeton Univ Press, Princeton).
21. Okasha S (2008) *A Companion to the Philosophy of Biology*, eds Sarkar S, Plutynski A (Wiley-Blackwell, Oxford), pp 138–157.
22. Buss LW (1987) *The Evolution of Individuality* (Princeton Univ Press, Princeton).
23. Queller DC (1997) Cooperators since life began. *Q Rev Biol* 72:184–188.
24. Margulis L (1981) *Symbiosis in Cell Evolution* (Freeman, San Francisco).
25. Bapteste E, et al. (2009) Prokaryotic evolution and the tree of life are two different things. *Biol Direct* 4:34.
26. Baquero F (2004) From pieces to patterns: Evolutionary engineering in bacterial pathogens. *Nat Rev Microbiol* 2(6):510–518.
27. Levitt M (2009) Nature of the protein universe. *Proc Natl Acad Sci USA* 106(27):11079–11084.
28. Alperovitch-Lavy A, et al. (2011) Reconstructing a puzzle: Existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environ Microbiol* 13 (1):24–32.
29. Boucher Y, et al. (2011) Local mobile gene pools rapidly cross species boundaries to create endemicity within global Vibrio cholerae populations. *MBio* 2(2): pii: e00335-10.
30. Brazelton WJ, Baross JA (2010) Metagenomic comparison of two Thiomicrospira lineages inhabiting contrasting deep-sea hydrothermal environments. *PLoS One* 5(10):e13530.
31. Lozupone CA, et al. (2008) The convergence of carbohydrate active gene repertoires in human gut microbes. *Proc Natl Acad Sci USA* 105(39):15076–15081.
32. Smillie CS, et al. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480(7376):241–244.
33. Lukjancenko O, Wassenaar TM, Ussery DW (2010) Comparison of 61 sequenced Escherichia coli genomes. *Microb Ecol* 60(4):708–720.
34. Xie W, et al. (2011) Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J* 5(3):414–426.
35. Zhang T, Zhang XX, Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS One* 6(10):e26041.
36. Skippington E, Ragan MA (2011) Lateral genetic transfer and the construction of genetic exchange communities. *FEMS Microbiol Rev* 35(5):707–735.
37. Overmann J (2010) The phototrophic consortium "Chlorochromatium aggregatum"—a model for bacterial heterologous multicellularity. *Adv Exp Med Biol* 675:15–29.
38. Marin B, Nowack EC, Melkonian M (2005) A plastid in the making: Evidence for a second primary endosymbiosis. *Protist* 156(4):425–432.
39. Moustafa A, et al. (2009) Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324 (5935):1724–1726.

40. Kirk DL (1998) *Volvox: Molecular-Genetic Origins of Multicellularity and Cellular Differentiation* (Cambridge Univ Press, Cambridge, UK).
41. Kessin RH (2001) *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge Univ Press, Cambridge, UK).
42. Gilbert SF, Epel D (2009) *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution* (Sinauer, Sunderland, MA).
43. Werren JH, Baldo L, Clark ME (2008) Wolbachia: Master manipulators of invertebrate biology. *Nat Rev Microbiol* 6(10):741–751.
44. Büdel B, Scheidegger C (2008) *Lichen Biology*, ed Nash TH (Cambridge Univ Press, Cambridge, UK), 3rd Ed, pp 40–68.
45. Honegger R, Scherrer S (2008) *Lichen Biology*, ed Nash TH (Cambridge Univ Press, Cambridge, UK), 3rd Ed, pp 94–103.
46. Bouchard F (2010) Symbiosis, lateral function transfer and the (many) saplings of life. *Biol Philos* 25:623–641.
47. Dagan T, Artzy-Randrup Y, Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* 105(29):10039–10044.
48. Halary S, Leigh JW, Cheaib B, Lopez P, Bapteste E (2010) Network analyses structure genetic diversity in independent genetic worlds. *Proc Natl Acad Sci USA* 107(1):127–132.
49. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R (2008) Reticulate representation of evolutionary and functional relationships between phage genomes. *Mol Biol Evol* 25(4):762–777.
50. Lawrey JD (1984) *Biology of Lichenized Fungi* (Praeger, New York).
51. Zhang W, Fisher JF, Mobashery S (2009) The bifunctional enzymes of antibiotic resistance. *Curr Opin Microbiol* 12(5):505–511.
52. Böltner D, MacMahon C, Pembroke JT, Strike P, Osborn AM (2002) R391: A conjugative integrating mosaic comprised of phage, plasmid, and transposon elements. *J Bacteriol* 184(18):5158–5169.
53. Canchaya C, Giubellini V, Ventura M, de los Reyes-Gavilán CG, Margolles A (2010) Mosaic-like sequences containing transposon, phage, and plasmid elements among Listeria monocytogenes plasmids. *Appl Environ Microbiol* 76(14):4851–4857.
54. Zhaxybayeva O, et al. (2009) On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc Natl Acad Sci USA* 106(14): 5865–5870.
55. Cotton JA, McInerney JO (2010) Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc Natl Acad Sci USA* 107(40):17252–17255.
56. Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 430:623–630.
57. Andam CP, Fournier GP, Gogarten JP (2011) Multilevel populations and the evolution of antibiotic resistance through horizontal gene transfer. *FEMS Microbiol Rev* 35(5):756–767.
58. Colson P, Raoult D (2010) Gene repertoire of amoeba-associated giant viruses. *Intervirology* 53(5):330–343.
59. Antonova ES, Hammer BK (2011) Quorum-sensing autoinducer molecules produced by members of a multispecies biofilm promote horizontal gene transfer to Vibrio cholerae. *FEMS Microbiol Lett* 322(1):68–76.
60. Jones BV (2010) The human gut mobile metagenome: A metazoan perspective. *Gut Microbes* 1(6):415–431.
61. Qu A, et al. (2008) Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS One* 3(8):e2945.
62. Walser J-C (2004) Molecular evidence for limited dispersal of vegetative propagules in the epiphytic lichen Lobaria pulmonaria. *Am J Bot* 91(8):1273–1276.
63. Beauregard-Racine J, et al. (2011) Of woods and webs: Possible alternatives to the tree of life for studying genomic fluidity in E. coli. *Biol Direct* 6(1):39.
64. Brandstädt A, Le VB, Spinrad JP (1999) *Graph Classes: A Survey* (SIAM, Philadelphia).
65. Berry A, Pogorelcnik R, Simonet G (2010) An introduction to clique minimal separator decomposition. *Algorithms* 3:197–215.
66. Baquero F (2011) The 2010 Garrod Lecture: The dimensions of evolution in antibiotic resistance: Ex unibus plurum et ex pluribus unum. *J Antimicrob Chemother* 66(8):1659–1672.
67. Koonin EV (2011) *The Logic of Chance: The Nature and Origin of Biological Evolution* (FT Press, Upper Saddle River, NJ).
68. Makino K, Uno T (2004) New algorithms for enumerating all maximal cliques. *LNCS 3111* 260–272.
69. Feder ME (2007) Evolvability of physiological and biochemical traits: Evolutionary mechanisms including and beyond single-nucleotide mutation. *J Exp Biol* 210(Pt 9):1653–1660.
70. Paauw A, Leverstein-van Hall MA, Verhoef J, Fluit AC (2010) Evolution in quantum leaps: Multiple combinatorial transfers of HPI and other genetic modules in Enterobacteriaceae. *PLoS One* 5(1):e8662.
71. Sterelny K (2004) *Modularity in Evolution and Development*, eds Schlosser G, Wagner GP (Univ of Chicago Press, Chicago), pp 490–518.
72. Bezuidt O, Pierneef R, Mncube K, Lima-Mendez G, Reva ON (2011) Mainstreams of horizontal gene exchange in enterobacteria: Consideration of the outbreak of enterohemorrhagic E. coli O104:H4 in Germany in 2011. *PLoS One* 6(10):e25702.
73. Bielaszewska M, et al. (2011) Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: A microbiological study. *Lancet Infect Dis* 11(9):671–676.
74. Rogers BA, Sidjabat HE, Paterson DL (2011) Escherichia coli O25b-ST131: A pandemic, multiresistant, community-associated strain. *J Antimicrob Chemother* 66(1): 1–14.
75. Woese CR (2002) On the evolution of cells. *Proc Natl Acad Sci USA* 99(13):8742–8747.
76. Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3(9):679–687.
77. Zhaxybayeva O, Gogarten JP (2004) Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* 20(4):182–187.
78. Forterre P (2002) The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol* 5(5):525–532.
79. 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(10):743–752.
80. Levin BR (2010) Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genet* 6(10):e1001171.
81. Vale PF, Little TJ (2010) CRISPR-mediated phage resistance and the ghost of coevolution past. *Proc Biol Sci* 277(1691):2097–2103.
82. Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 17(10):R373–R386.
83. Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 35(5):957–976.
84. Martin A, Dunnington EA, Briles WE, Briles RW, Siegel PB (1989) Marek's disease and major histocompatibility complex haplotypes in chickens selected for high or low antibody response. *Anim Genet* 20(4): 407–414.
85. Bodnar JW, Killian J, Nagle M, Ramchandani S (1997) Deciphering the language of the genome. *J Theor Biol* 189(2):183–193.
86. Searls DB (2002) The language of genes. *Nature* 420 (6912):211–217.