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

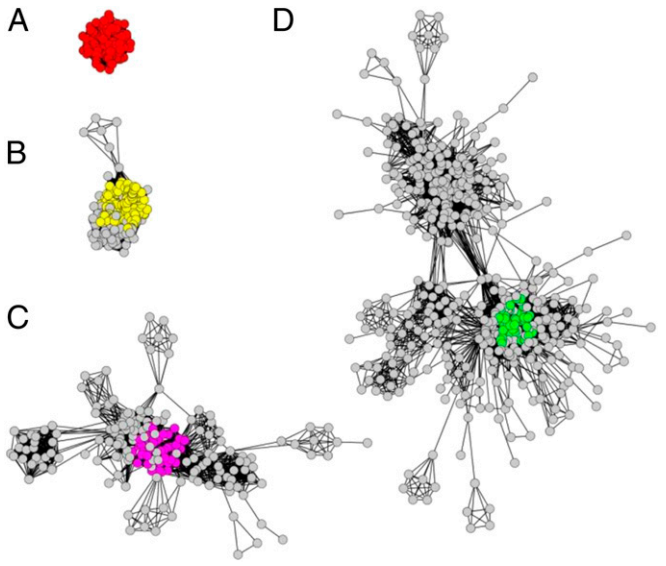

Fig. 1. Selection of gene network components displaying their largest maximal clique. Genes (nodes; in gray) aligning >80% of their sequences with their match in a BLAST analysis (showing >50% identification and a BLAST score < 1 e^{-20}) are directly connected. Sequences belonging to the largest maximal 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).

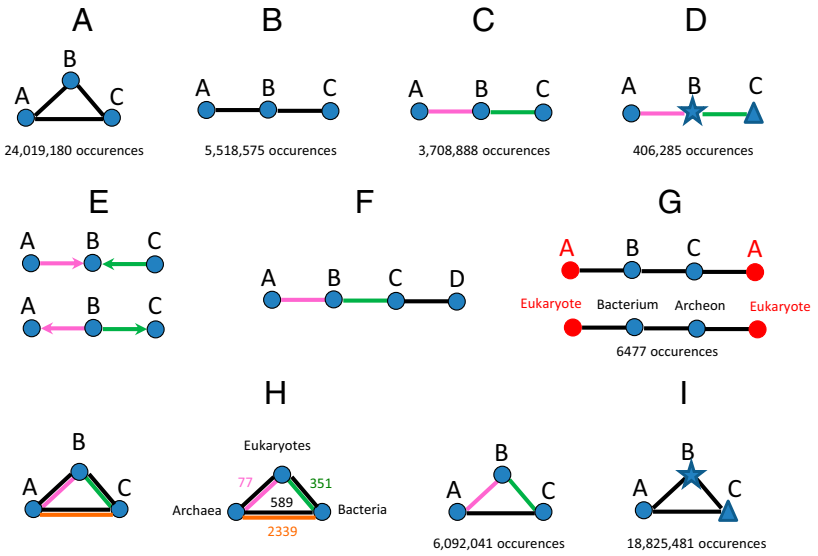

Fig. 2. Patterns with evolutionary significance in resemblance networks. Each symbol indicates an entity (node) from a distinct level of biological organization. Similarly colored edges indicate vertically inherited shared characters. Occurrences in our test dataset at >50% identification are quantified 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.

Supporting Information

Bapteste et al. 10.1073/pnas.1206541109

SI Text

1. Glossary. A club is an evolutionary unit composed of a group of entities that reproduces in separate events but shares specific features, most commonly genes. Clubs may originate from a single lineage (e.g., a bacterial colony), separate lineages (e.g., a multispecies biofilm or lichen), or different levels (e.g., a coalition of cyanophages and cyanobacteria).

Evolutionary novelties are evolutionary changes with important effects, such as the emergence of lineages of evolutionary units, a major reorganization of the structure of mosaic evolutionary units through the acquisition of genetic material, or the opening up of ecological niches to allow adaptive radiation.

Evolutionary processes are the diverse processes (such as mutation, replication, recombination, lateral gene transfer, genetic drift, etc.) that contribute to the diversity of evolutionary units, including higher-level processes such as those processes involved in recognizing and counteracting pathogens, cooperative interactions benefiting the cooperating parties, obligate endosymbioses, and natural selection.

Evolutionary units are individuals (understood broadly) that reproduce (understood broadly) and form lineages with individuals that vary, with positive correlations between the traits of parent(s) and offspring traits (i.e., with some degree of heredity). At least some of these traits may contribute to fitness. In this paper, we consider evolutionary units that possess genetic systems in their own right (e.g., viruses, plasmids, and single-celled organisms) or are composed of individuals with genetic systems, regardless of whether those component individuals are genetically uniform. Biofilms, multicellular eukaryotes, and beehives are examples of composite evolutionary units that are not genetically uniform. Such compound individuals are or at least, have the potential to be higher-level evolutionary units (i.e., are composed of lower-level units).

Horizontal gene transfer and lateral gene transfer are synonyms that designate the processes in which genetic material from a donor is passed on to a different host that is not its direct descendent.

Introgressive descent designates those evolutionary processes that propagate the genetic material of a particular evolutionary unit into a different host structure (or structures; e.g., a transposon into a series of different plasmids, a plasmid in different bacterial clones, or a nuclear gene into mitochondrial genomes) and then replicate these transformed host structures. It is, thus, a means by which lineages are created and lineages of features (e.g., genes and antibiotic resistance) may not coincide with the lineages of the wholes within which they were contained (e.g., plasmids, viruses, cells, and organisms).

A merger occurs when genetic material, not hitherto coexisting within the same evolutionary unit, is brought within a single evolutionary unit and subsequently replicated within a novel single corporate body. The result may be that the merger not only has novel combinations of traits but also, traits that arise from interactions not previously feasible when the genetic material belonged to separate units. An intralineage merger occurs when the sources of the genetic material all come from the same lineage, an interlineage merger occurs when at least two sources of genetic material belong to different lineages, and an interlevel merger occurs when the sources belong to different levels (e.g., when a viral genome containing material from a plasmid is inserted into a cellular genome).

A P3 is a pattern in a resemblance network: an intransitive chain with three adjacent objects of the chain presenting similarity up to a certain threshold caused by the sharing of features between these objects. When these features are genetic, the patterns can be analyzed by gene or genome networks. A mosaic-P3 (M-P3) is of particular interest, because the P3 chain connects at least two distantly related and/or unrelated lineages through a third entity acting as an intermediate binder. This binder shares a particular feature with one of the objects (e.g., a given gene family) and a different feature (e.g., another gene family) with the other objects in the chain. A polarized P3 is one in which the donor can be distinguished from the recipient; it is represented by an arrow running from donor to recipient on the edges of the resemblance network.

A resemblance network is a network of objects represented by nodes that are linked when they share a character (represented by connecting them by an edge) at some threshold of resemblance.

2. Empirical Evidence That Introgressive Descent Produces Bona Fide Evolutionary Units. From Lewontin's perspective, for a merger or a club to be considered an evolutionary unit, it must be able to show heritable variation that leads to variation in the fitness of its descendants. From a gene's view, for a merger or a club to be a unit of selection, it must constitute an extended phenotype that contributes to the evolutionary success of the genes (replicators) present in these associations. From the perspective in the work by Godfrey-Smith (1), distinct sorts of more-or-less Darwinian units (e.g., paradigmatic and marginal) can be distinguished depending on the extent to which various criteria are fulfilled. In the 2009 book by Godfrey-Smith (1), the distinction between paradigmatic Darwinian evolutionary units (and populations) and marginal ones is based on their fidelity of heredity (H), abundance of variation (V), competitive interaction with respect to reproduction (α), the extent to which similar organisms in the population have similar fitness (C), and the dependence of their reproductive differences on intrinsic characters (S). With some qualifications and amplifications, paradigmatic Darwinian processes involve units with high values of H, S, and C.

Another general key condition for mergers and clubs to be considered paradigmatic Darwinian units is that they reproduce as a unit (i.e., that they are reproducers). It is sometimes difficult to determine when this condition is fulfilled where there is a gradient rather than a clear-cut boundary between the phenomenon of growth (through which entities persist based on maintenance mechanisms) and the phenomenon of reproduction [after which one or more new individuals of the same sort (the progeny) have been produced; good cases allow for the succession of generations to be clearly delineated]. Paradigm Darwinian units involve genuine reproduction, whereas growth is generally associated with less significant evolutionary change; therefore, it results at best in marginal Darwinian units. Moreover, entities must be autonomous or collective reproducers to be considered paradigm Darwinian units; they will be considered marginal ones if they are not reproduced by their own replication machinery but for example, as parts of a larger unit. The work by Godfrey-Smith (1) labels items such as viruses and genes that reproduce in this way as "scaffolded reproducers" (1). Importantly, the claim that some units are paradigmatically Darwinian should not be understood as meaning that they are the most prevalent. It simply means that they would be among the most characteristic examples of processes yielding evolution by means of natural selection, processes that serve as good models from which less ideal examples of Darwinian evolutionary processes depart.

The biological literature shows that clubs and mergers produced by introgressive descent fall on a gradient between paradigmatic and marginal according to a complex analysis of several parameters that can be used to describe them (see below). Before we consider some particular cases discussed in the text [specifically, fused genes, eukaryotic cells, and recombined viral genomes (candidate multilineage mergers), Russian genetic dolls (candidate multilevel mergers), bacterial clones (intralineage clubs), multispecies biofilms with mobile genetic elements, gut microbiomes, and coalitions of cyanophages and cyanobacteria (candidate multilevel clubs)], we need to underscore that there are various ways for mergers (or clubs) to achieve the properties mentioned above.

Building on the approach in the work by Godfrey-Smith (1), a merger (or club) can be said to have progeny when some of its components produce a new merger (or club) by a repeatable process that depends in some crucial way on the properties of the merger (or club) over and above the properties of its components considered separately. For instance, a merger (or club) will reproduce as a unit when there is reproductive dependence between the elements of the merger (or the members of a club; i.e., when they share replication machinery or when their reproduction is mutually constrained and a variant of the club or merger is reconstituted subsequent to the reproduction of its components). Heritability of a merger (or club) can then be defined by means of correlation between parent and offspring mergers (or clubs).

A merger (or club) then has fitness on its own, because intrinsic features of the unit (over and above the independent features of its components and elements) are responsible for its evolutionary success (as well as those features of their descendants). These features do not need to be emergent properties of the merger (or club) in a strong sense, because they can result from combinations of functions associated at the right frequencies that were not previously deployable.

Finally, a merger (or club) will most paradigmatically act as a unit when its elements or components are integrated. For instance, this integration can involve the development of some central control and/or some degree of differentiation of a germ line from a somatic line. In the latter case, the loss of reproductive competence for elements in the somatic line belongs to a process of de-Darwinization of some lower-level units integrated in a larger (compound) unit. De-Darwinizing typically involves decrease in the high heritability and/or dependence of fitness on intrinsic features and/or high continuity over the fitness landscape of the formerly autonomous lower-level units (1). As a result of integration, the best—or only—way of the lower-level units to serve their reproductive interests is to serve the reproductive interest of the higher-level unit. This result is true for the somatic cells of a eukaryote and the sterile workers of a bee hive. A less extreme case of integration can be observed when elements of mergers (or members of clubs) align their reproductive interests rather than compete with one another. In more general terms, some form of coordination is observed, because elements of a merger (or members of a club) are constrained to subordinate what might be best for them individually (e.g., to be free riders) that would interfere with the evolutionary fate of the larger unit to which they belong. This coordination/subordination can be diagnosed by identifying costs for (at least some) of the elements of a merger (or members of a club) when there is a good of the larger unit.

Thus, a sufficient condition for showing that mergers and clubs are bona fide evolutionary units is that they are the focus of (multilevel) selection and that their elements or members have been de-Darwinized or paid a cost for the good of the larger unit. This case is likely for numerous mergers characterized by the possession of some collective-level function(s); under the appropriate selective regimens, they would then have a positive influence on their own fitness. Examples of mergers fitting this description include bifunctional fused genes and mosaic genomes in eukaryotes. It is also very likely for clubs, such as biofilms, in which the sharing of genetic material often seems to generate selectable benefits for the club, whereas it has some cost for some of its individual members.

2.1. Multilevel clubs of cyanophages–cyanobacteria. Various coalitions of cyanophages (myoviruses, siphoviruses, and podoviruses) and cyanobacteria (either *Synechococcus* or *Prochlorococcus*), adapted to particular sets of conditions in aquatic environments, have been documented (1). It seems plausible that these coalitions qualify as clubs and marginally Darwinian evolutionary units.

Some marginal reproduction at the club-level is very likely, although only a weak case can be made for the Darwinian reproduction of these coalitions, because it is difficult to define clear-cut generations of parent and offspring clubs of cyanophages and cyanobacteria. Except when an infected cyanobacterium or a group of cyanobacteria migrates and seeds a new club, these coalitions probably grow (thus, are collective growers) more than they reproduce and are collective reproducers. Indeed, coalitions of cyanophages and cyanobacteria lead to coalitions of cyanophages and cyanobacteria when cyanophages, cyanobacteria, or both leave direct offspring within a coalition. The first situation happens when cyanobacteria are lysed; hence, they do not leave descendants, but a novel generation of cyanophages is produced that will interact with surviving cyanobacterial cells. The second situation happens when reproducing cyanobacteria comprise temperate phages in stable relationship; the third situation happens in case of pseudolysogeny, when a phage-infected cell grows and divides, although its virus is pursuing a lytic infection (1).

There are strong signs that cyanophages (and possibly, cyanobacteria) (2) pay a cost for the good of their club. Although there is a cost for a virus to maintain a gene in a size-limited genome, there is a collective benefit of encoding this additional metabolic function in photoautotroph hosts living in low-nutrient waters (3). In clubs cemented by the sharing of photosynthetic genes, cyanomyoviruses notably provide the *psbA* genes (frequently combined with the psbD gene), which are expressed to help to repair photodamage in light-harvesting antenna complexes (2, 4). As a result, there is no loss in photosynthetic efficiency during the infection cycle (2); in addition, phages with a broader host range are more likely to carry both *psbA* and *psbD* (2). Thus, the sharing of genes benefits the club. [Other cyanophage–cyanobacterial clubs rely on the sharing of other genes involved in carbon metabolism, phosphate acquisition, and ppGpp metabolism (2, 3), ensuring the swapping of metabolic components critical to phage and host reproduction (5)]. Cases have also been reported where lysogenic infection seems to protect *Synechococcus* against viral infections (2). For instance, the LPS genes in the cyanomyovirus S-PM2, with a supposed protective function against infection or grazing, are among the earliest and most abundantly transcribed genes expressed in infected cyanobacteria (2). Such intertwining of benefits suggests that many different phage and/or bacterial cells profit from what they cannot individually produce.

Likewise, cyanophages do not always behave as free riders: they seem to subordinate what might be best for them individually to the good of the club. This finding is suggested by the fact that the probabilities of lysogenization and induction of the lytic cycle are affected by environmental and host genetic factors, with the consequence that lysogeny maintains the phage population when host abundance is too low to support maintenance of a population of lytic phages (1). The length of the latent period is under strong selection pressure determined by the concentration of sensitive host cells (1). Furthermore, some traces of coordination for the use of the shared genetic material can also be suggested. Although cyanophages genomes have an average GC content of 40% and *Synechococcus* genomes have an average GC content of

60%, the GC content of *psbA* genes in phage has drifted to a value of 50%, further underlying the functional integration at play in these members of the club (2). In addition, the level of phosphorus starvation in cyanobacterial hosts selectively influences the degree of up-regulation of the phage-encoded phosphate binding protein gene *pstS*, which suggests a coevolution of regulatory systems between host and phage (6): indeed, the *pstS* (and *phoA*) phage genes seem to be regulated by the host PhoR/PhoB two-component system. Finally, cyanophages obviously depend on cells for their reproduction, but aspects of this dependence seem stronger than the general dependence of viruses on their host for reproduction, suggesting that some weak form of de-Darwinization might have affected cyanomyoviruses coevolving with cyanobacteria. These phages all lack homologs to genes essential to moderate the specificity of the host RNA polymerase by recognizing the early promoters of these phage genes and genes responsible for the production of a transcription factor that replaces the σ-70 factor of their host (2). Therefore, the early (and possibly middle) expression of these phage genes depends on mechanisms of their host (2).

Consequently, the maintenance of clubs with favorable distribution of virus types and lytic vs. lysogenic viruses seems more likely than the maintenance and flourishing of less promising distribution of cellular and viral members. A mechanism that reduces the effective host population size for infection by a given virus and eases long-term coexistence between viruses and their hosts has recently been shown (2). Accordingly, in a stratified water column, maximum *Synechococcus* myovirus diversity correlates with maximum *Synechococcus* population density (1). Moreover, there are specific cyanophage–host relationships, because most abundant cyanophages show a parallel pattern of abundance with the most abundant and second most abundant *Synechococcus* clones in summer and autumn (1). These situations reflect that, to some extent, cyanophages and cyanobacteria have aligned their reproductive interests if not demonstrably de-Darwinized.

This finding is important, because a greater likelihood of club reproduction can then lock in favorable combinations of functions associated at the right frequencies. For instance, the sharing of genes, such as *hli*, that encode high light-inducible proteins seems to be under selection: myoviruses isolated on *Prochlorococcus* have two times as many hli genes as myoviruses isolated on *Synechococcus*. Likewise, photosynthetic shared genes are under strong purifying selection and continue to be exchanged through homologous recombination among phages and possibly between phages and their hosts (7). In some cases, these combinations led to heritable novel properties over time, such as the evolution and spread of an original unique phage-encoded gene, *PebS*, that performs reactions requiring two consecutives enzymes in its cyanobacterial host and can replace the canonical pathway to maintain bilin biosynthesis (5).

Therefore, it seems plausible that, within clubs of cyanobacteria and cyanophages, some selection acts at the level of the club, and it is powerful enough to alter what can be gotten from individual selection acting phage by phage and cyanobacteria by cyanobacteria. Such coalitions may, thus, be seen as consistent with the marginal evolutionary units in the work by Godfrey-Smith (1) (Fig. S1).

2.2. Multilevel clubs of multispecies biofilms. Cells growing as part of a biofilm are usually embedded within a matrix of extracellular polymeric substances, which can include environmental DNA (eDNA), central to its formation. Multispecies biofilms host a rich genetic diversity (4). They grow and reproduce when cells from the biofilm or fragments of the biofilms detach from the parental coalition, drift, disperse, and seed a novel biofilm (7). Therefore, fidelity of biofilm reproduction varies greatly, making them rather marginal Darwinian units.

However, some forms of coordination are remarkable in these clubs. Whereas in some multispecies biofilms, microbial populations are in competition and do not align their reproductive interests (3), in other multispecies biofilms (8), such as acid mine drainage biofilms (4), there is division of metabolic roles between bacterial and archaeal populations indicative of integration at the collective level. The collective integration of some multispecies biofilm is notably manifested by the fact that individual biofilms migrate (5). It is also manifested by the fact that the eDNA involved in the development and reproduction of biofilms is not only derived from dead cells: some of it is also actively transported from intact cells, which is the case for *Streptococci* (6). DNA from the club is then transferred between biofilm members by conjugative plasmids (9, 10) and conjugative transposons (8, 11, 12). In fact, the ability of oral streptococci to integrate eDNA by transformation (a state known as competence) is linked to biofilm formation through the production of the quorum-sensing molecule competence stimulating peptide and subsequent cell death, lysis, and release of eDNA in a subpopulation of cells (13, 14).

Some of the members of these multispecies biofilms also tolerate a cost to produce genetic material that will be shared within the club and benefit to the club. For example, a conjugative transposon from the Tn916 family, capable of broad host-range conjugative transfer between bacterial cells (often of different genera), has been shown to spread antibiotic resistance between members of a multispecies oral consortium from *Veillonella dispar* to four different *Streptococcus spp.* (13). Likewise, different bacteria (*V. harveyi* and *V. parahaemolyticus*) produce autoinducer molecules that induce lateral gene transfer to *V. cholerae* (15). Such mechanisms of stabilization and development of the biofilm result in greater sharing of genetic material between different members of the biofilm.

The integration described above is certainly selected. Bacteria in mixed-species biofilms have been shown to gain a fitness advantage over single-species biofilms, which is illustrated by greater productivity in cell numbers (16, 17) and greater stability produced, in particular, by greater resistance to antimicrobial treatment (18, 19).

Finally, a stronger argument based on intrinsic properties of multispecies biofilms suggests that such biofilms are evolutionary players in their own right: they can achieve emergent collective properties. For instance, in contrast to when it grows in a multispecies oral biofilm, *V. dispar* is unable to transform any of the members to tetracycline resistance when they are grown as a monoculture. Biofilms, as wholes, show some phenotypes of their own.

All these properties qualify them as Darwinian units (Fig. S1).

2.3. Intralineage and multilineage mergers from recombining viral populations. The case is clear that viruses, such as T4 bacteriophages, lambdoid phages, and mycobacteriophages, have highly mosaic genomes and high heritability because of their intrinsic properties. These entities recombine a lot over a large geographical span (20), generating offspring with genetic variability that can be selected based on their unique genetic combinations (review in ref. 21). This recombination occurs more readily among closely related phages. The resulting mergers are the focus of multilevel selection, which is strongly suggested both by theories that allow illegitimate recombination to take place almost randomly in the recombining phage genomes, generating many mergers that will be defective for phage growth and eliminated by natural selection (22), and by theories for which modular evolution of phage requires homologous recombination to take place at special intergenic sites (23). Selection can act on entire phages. For instance, the RB49 virion seems to have recently acquired the ability to infect *Escherichia coli* by acquiring the *g38* tail fiber adhesin sequence from a T-even phage (24). It can also be argued that selection for recombination in viral populations acts at a higher level than the gene level. Indeed, in mycobacteriophages recombining genomes, morons that consist of a protein-coding region flanked by a putative s70 promoter and

a putative factor-independent transcription terminator located between two genes otherwise adjacent in phages (25) have been reported. These morons take advantage of the perpetuation of highly recombinogenic viral phages by inserting between functional units involving several genes that will remain together in selective contexts, where they remain selectively valuable.

Selection can also act on some of the recombinant genes or fractions of those genes in these mosaic phages; for instance, the T4-type tail fiber loci have a mosaic design caused by recombination events in this region of the phages genomes (24). Mosaic phages have evolved specific intrinsic mechanisms that ensure the stabilization of the functional units that profit the communal entity and provide it with selectable traits. These features range from the evolution of recombinational hotspots (glycine islands in T2-types fibers and His-boxes in T4 phage) (24) to the evolution of very efficient recombinases (λ-red genes) (26); it is, in some cases, essential for the phage lifecycle, like in P22 bacteriophages (27), where reproduction depends on recombination.

The unparalleled abundance of such mosaic phages testifies to their lasting evolutionary success: they are among the most numerous entities on Earth, and they occupy a wide range of ecological niches from animal gut to open ocean (24). This success indicates that the intralineage mergers and multilineage mergers built by recombination between distantly related phages, entities that demonstrably have high heritability, are the bearers of fitness and form paradigmatic Darwinian populations when they are in competition for the same resources (Fig. S1).

2.4. Multilevel mergers (Russian genetic dolls) of nested mobile elements collectively selected: The case of extended-spectrum β-lactamases. The elements involved in the worldwide spread of genes determining resistance to the newest β-lactam antibiotics [extended-spectrum β-lactamases (ESBLs)] are

(*i*) the resistance genes themselves;
(*ii*) integrons (genetic platforms capable of capturing and mobilizing genes);
(*iii*) transposons (larger segments of DNA frequently harboring integrons capable of independent replication and insertion of the copy within other transposons, plasmids, or chromosomes);
(*iv*) plasmids (autonomously replicating extrachromosomal circular DNA molecules able to be transferred from cell to cell, even among different species, and frequently harboring transposons);
(*v*) clones (subspecific groups of bacteria frequently specialized in particular habitats and frequently carrying plasmids);
(*vi*) species (ensembles of clones with the same core genome);
(*vii*) genetic exchange communities (ensembles of species able to exchange genetic material, commonly by sharing plasmids); and
(*vii*) specific microbiomes (ensembles of species symbiotically associated with particular animal or human hosts, which contain genetic exchange communities).

Note that the epidemiology and evolution of antibiotic resistance, primarily determined by a particular resistance gene, is dependent on the interactions of a diversity of evolutionary individuals at different hierarchical levels, with each of these individuals hosted by an evolutionary unit superior in the hierarchy (Russian dolls model) (28).

As an example, consider β-lactam ESBL-mediated resistance contained in the transposon Tn*21*. This transposon is known as the flagship of the floating genome (29). This transposon frequently contains another transposon, Tn*402,* which might contain class 1 integrons harboring ESBL resistance genes. Tn*21* is harbored, in turn, by different groups of plasmids, such as IncFI and IncHI1. The wide use of antibiotics and probably, industrial pollutants (such as mercury, because Tn*21* also has genes for mercury resistance) collaborates in selecting plasmids carrying this transposon, the clones harboring these plasmids, and subsequently, the species and genetic exchange communities carrying Tn*21* (30, 31). Of course, at each one of these levels, each evolutionary individual (such as the transposon Tn*21*, the plasmid IncFI, or the *E. coli* clone ST131) is in competition with other individuals (other genes, other transposons, or other clones) at the same hierarchical level, and therefore, differential fitness exerts its selective effects at each level. Indeed, this excellent example shows the levels of selection debate (32) centered on two questions: (*i*) How does natural selection acting on lower-level biological units create higher-level units? (*ii*) Given that multiple levels exist, how does natural selection at one biological level affect selection at lower or higher levels?

Therefore, the units in the hierarchy of Russian dolls can be considered as Darwinian units (Fig. S1).

2.5. Multilineage clubs: The human gut microbiome. In humans, bacterial and archaeal cells, plus occasional eukaryotic commensals, belonging to more than 1,000 taxa are condensed in a particular functional collective domain, the intestinal microbiome. This microbiome is composed in part of a core microbiome (i.e., an assemblage of microbial species and consortia with fairly constant taxic composition) and in part of a more fluid assemblage, probably dispensable for gut physiology but with possible local adaptive value in particular environments. Despite this complexity, the microbiome can be considered as an evolutionary unit, applying the criteria of reproducibility, heritability, genetic variation, fitness, and integration.

First, the criteria of reproducibility and heritability are considered. The human gut microbiota turns over fairly rapidly, but the composition of its core is maintained in a highly reproducible way (with considerable circumstance-dependent variation in the proportions of key species within the biome cycling according to nutrition, physiological condition, presence of pathogens, and the like), thus giving rise to a consistent heritage of a common core microbiome with interpersonal variations maintained over generations within a kinship (33). Each human reproduction gives rise to a reproduction event of the microbiome. Note that this finding does not mean that the newborn acquires the complex mother's microbiome immediately after birth, but it is known that, after 1 or 2 y of age, the full microbiome has been reproduced almost in its full integrity. In humans, a number of starting bacteria, such as *Lactobacillus*, *Prevotella*, and *Sneathia*, might be acquired during vaginal delivery (34, 35), whereas other pioneering taxa may be acquired by breastfeeding (36). It might be suggested that these early colonizers serve as sinks or attractors for other microbial partners, and those for others thereafter; eventually, pairs or higher consortia of organisms create novel niches for other organisms. The corresponding law of attraction remains one of the most important items to be investigated in microbiome biology (37). These laws might be related to genomic functional complementarity after genetic reductions using a model proposed for coevolution of bacterial and eukaryotic cells (37). Additionally, we can obtain events of reproduction by techniques of microbiome transplantation. The possibility of establishing new microbiota from a donor source (38) has been shown; 14 d posttransplantation, the recipient microbiota was shown to be highly similar to the microbiota of the donor (39).

With respect to genetic variation, this finding should be understood as the ability to modify the composition of the microbiota in a heritable way. Indeed, the microbiota is exposed during life to environmental challenges (such as invasion of environmental microbes, undernutrition, and exposure to drugs) and even behavioral influences (such as vegetarianism). Adaptive changes in the microbiome follow these environmental challenges, and these changes are heritable. Such adaptive changes are transmitted just as antibiotic resistance is transmitted in kin

communities (see above), Stably inherited changes in microbiome composition within an individual (e.g., after immune response to a pathogen) often provide fitness benefits for the host, eventually in competition with other hosts. Indeed, the microbiota composition and its relation with the gut has resulted from the dynamics of selection and competition (40). [See the work by Pradeu (41) for a philosophically oriented review of the interactions of the microbiome and the immune system.] It has even been suggested that the composition of the microbiota might influence the behavior of the host (33). Finally, the developmental processes that build up microbiotic bacterial communities and the long-term persistence of core microbiota over human lifetimes suggest that the microbiotic system is highly integrated, acting, in fact, as a biological individual or something approximating an organ of its host, which would then have to count as a more genetically complex individual than it is typically thought to be (41).

2.6. Many other candidates. In addition to these rather well-documented cases, there is scattered evidence from many sources that numerous other mergers and clubs qualify as evolutionary units.

For instance, Apicomplexans with their apicoplast (a modified descendant of chloroplasts), most especially *Plasmodium falciparum*, are multilineage mergers that would probably be considered as an extended phenotype of the endosymbiont genes under the gene perspective. An apicomplexan constitutes a unit of selection on its own, in which the bringing together of formally independent DNA-based entities is, in large part, explained by contemporary functions that are subject to natural selection rather than some genealogical trend (42, 43). The entity composed of the endosymbiosed organelle and its host is actually a composite of distinct, unrelated genealogical units. However, selection acting on the composite, perhaps also acting at several levels (44), is key to the retention of the organelle. Although many of its genes have migrated to the nuclear genome and the endosymbiont has given up essential parts of its reproductive machinery (as in a typical evolutionary transition, during which component parts de-Darwinize), this endosymbiont has retained around 500 genes, some of which are connected to its obligate function of coding for isopentenyl pyrophosphate (45, 46). This example is a clear example in which the alignment of reproductive interests does not require kinship but depends on emergent collective adaptations (a shared body).

A second example occurs in chickens: selection favors individuals with particular heterozygous combinations of histocompatibility haplotypes, depending on their exposures to pathogens, thus yielding new lineages with particular combinations of histocompatibility complex haplotypes (47). Indeed, the maintenance of high heterozygosity is selected mainly by the increased disease susceptibility of chickens homozygous for various particular haplotypes, a fact that breeders had to learn by hard experience and the sleuthing done by chicken geneticists. Because these intralineage mergers obtained from the swapping of genetic material (outbreeding within a species) are subject to selection, they are also good candidates to qualify as evolutionary units.

Overall, the properties of the mergers and clubs discussed here can be summarized on a multidimensional space (axis D represents the extent to which they correspond to paradigmatic/marginal evolutionary units, and axes M and L represent the number of lineages and levels that contributed the genetic material, respectively).

Readers who remain unconvinced that our mergers and clubs are evolutionary units should find in our approach a formal analytical apparatus to diagnose cases requiring coevolutionary explanations. However, we maintain that the surprising prevalence in every domain that we have examined of novelties and evolutionary transitions that depend on the deployment of material and functions from distinct sources indicates that there will be increasing recognition of the importance of introgressive descent.

3. List of Chromosomes of Cellular Organisms Used in the Dataset. ***3.1. Archaea.*** *Aeropyrum pernix K1_1, Archaeoglobus fulgidus DSM 4304_1, Archaeoglobus profundus DSM 5631_1, Caldivirga maquilingensis IC-167_1, Candidatus Korarchaeum cryptofilum OPF8_1, Candidatus Methanoregula boonei 6A8_1, Candidatus Methanosphaerula palustris E1-9c_1, Cenarchaeum symbiosum A_1, Desulfurococcus kamchatkensis 1221n_1, Haloarcula marismortui ATCC 43049_1, Haloarcula marismortui ATCC 43049_2, Halobacterium sp. NRC-1_1, Halomicrobium mukohataei DSM 12286_1, Haloquadratum walsbyi DSM 16790_1, Halorhabdus utahensis DSM 12940_1, Halorubrum lacusprofundi ATCC 49239_1, Halorubrum lacusprofundi ATCC 49239_2, Haloterrigena turkmenica DSM 5511_1, Hyperthermus butylicus DSM 5456_1, Ignicoccus hospitalis KIN4/I_1, Metallosphaera sedula DSM 5348_1, Methanobrevibacter ruminantium M1_1, Methanocaldococcus fervens AG86_1, Methanocella paludicola SANAE_1, Methanococcoides burtonii DSM 6242_1, Methanococcus aeolicus Nankai-3_1, Methanococcus maripaludis C6_1, Methanococcus vannielii SB_1, Methanocorpusculum labreanum Z_1, Methanoculleus marisnigri JR1_1, Methanopyrus kandleri AV19_1, Methanosaeta thermophila PT_1, Methanosarcina acetivorans C2A_1, Methanosarcina barkeri str. Fusaro_1, Methanosarcina mazei Go1_1, Methanosphaera stadtmanae DSM 3091_1, Methanospirillum hungatei JF-1_1, Nanoarchaeum equitans Kin4-M_1, Natronomonas pharaonis DSM 2160_1, Nitrosopumilus maritimus SCM1_1, Picrophilus torridus DSM 9790_1, Pyrobaculum aerophilum str. IM2_1, Pyrobaculum arsenaticum DSM 13514_1, Pyrococcus abyssi GE5_1, Pyrococcus furiosus DSM 3638_1, Pyrococcus horikoshii OT3_1, Staphylothermus marinus F1_1, Sulfolobus acidocaldarius DSM 639_1, Sulfolobus solfataricus P2_1, Thermococcus gammatolerans EJ3_1, Thermofilum pendens Hrk 5_1, Thermoplasma acidophilum DSM 1728_1, Thermoplasma volcanium GSS1_1, Thermoproteus neutrophilus V24Sta_1.*

3.2. Bacteria. *Acholeplasma laidlawii PG-8A_1, Acidobacterium capsulatum ATCC 51196_1, Akkermansia muciniphila ATCC BAA-835_1, Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446_1, Aquifex aeolicus VF5_1, Bacillus cereus Q1_1, Bacillus pseudofirmus OF4_1, Bacteroides fragilis YCH46_1, Bdellovibrio bacteriovorus HD100_1, Bordetella pertussis Tohama I_1, Borrelia afzelii PKo_1, Borrelia burgdorferi B31_1, Borrelia burgdorferi ZS7_1, Borrelia duttonii Ly_1, Borrelia garinii PBi_1, Borrelia hermsii DAH_1, Borrelia recurrentis A1_1, Borrelia turicatae 91E135_1, Campylobacter jejuni subsp. jejuni 81–176_1, Candidatus Amoebophilus asiaticus 5a2_1, Candidatus Cloacamonas acidaminovorans_1, Candidatus Endomicrobium sp. RsD17_1, Carboxydothermus hydrogenoformans Z-2901_1, Chlamydia trachomatis 434/Bu_1, Chlorobium chlorochromatii CaD3_1, Chloroflexus aurantiacus J-10-fl_1, Clostridium acetobutylicum ATCC 824_1, Corynebacterium glutamicum ATCC 13032_1, Coxiella burnetii RSA 493_1, Cupriavidus taiwanensis_1, Cupriavidus taiwanensis_2, Cyanothece sp. ATCC 51142_1, Cyanothece sp. ATCC 51142_2, Dehalococcoides ethenogenes 195_1, Deinococcus radiodurans R1_1, Deinococcus radiodurans R1_2, Dictyoglomus thermophilum H-6–12_1, Elusimicrobium minutum Pei191_1, Fibrobacter succinogenes subsp. succinogenes S85_1, Flavobacterium psychrophilum JIP02/86_1, Fusobacterium nucleatum subsp. nucleatum ATCC 25586_1, Gemmata obscuriglobus UQM 2246_1, Gemmatimonas aurantiaca T-27_1, Gloeobacter violaceus PCC 7421_1, Leptospira interrogans serovar Lai str. 56601_1, Leptospira interrogans serovar Lai str. 56601_2, Magnetococcus sp. MC-1_1, Methylacidiphilum infernorum V4_1, Mycoplasma genitalium G37_1, Nostoc punctiforme PCC 73102_1, Opitutus terrae PB90-1_1, Pedobacter heparinus DSM 2366_1, Pirellula staleyi DSM 6068_1, Prochlorococcus marinus str.*

AS9601_1, Psychrobacter arcticus 273–4_1, Rhizobium leguminosarum bv. trifolii WSM1325_1, Rhodopirellula baltica SH 1_1, Rhodospirillum rubrum ATCC 11170_1, Rickettsia rickettsii str. Iowa_1, Shewanella putrefaciens CN-32_1, Solibacter usitatus Ellin6076_1, Synechococcus elongatus PCC 6301_1, Thermanaerovibrio acidaminovorans DSM 6589_1, Thermoanaerobacter tengcongensis MB4_1, Thermobaculum terrenum ATCC BAA-798_1, Thermobaculum terrenum ATCC BAA-798_2, Thermodesulfovibrio yellowstonii DSM 11347_1, Thermomicrobium roseum DSM 5159_1, Thermotoga maritima MSB8_1, Thermus thermophilus HB8_1.

3.3. Eukaryotes. *Entamoeba histolytica HM-1:IMSS_1, Oryza sativa (japonica cultivar-group)_1, Oryza sativa (japonica cultivar-group)_2, Oryza sativa (japonica cultivar-group)_3, Oryza sativa (japonica cultivar-group)_4, Oryza sativa (japonica cultivar-group)_5, Oryza sativa (japonica cultivar-group)_6, Oryza sativa (japonica cultivar-group)_7, Oryza sativa (japonica cultivar-group)_8, Oryza sativa (japonica cultivar-group)_9, Oryza sativa (japonica cultivar-group)_10, Oryza sativa (japonica cultivar-group)_11, Oryza sativa (japonica cultivar-group)_12,Paramecium tetraurelia strain d4-2_1, Saccharomyces cerevisiae_1,Bigelowiella natans_nucleormorph 1,Guillardia theta_nucleormorph 1,Hemiselmis andersenii_nucleormorph1, Saccharomyces cerevisiae_1, Saccharomyces cerevisiae_2, Saccharomyces cerevisiae_3, Saccharomyces cerevisiae_4, Saccharomyces cerevisiae_5, Saccharomyces cerevisiae_6, Saccharomyces cerevisiae_7, Saccharomyces cerevisiae_8, Saccharomyces cerevisiae_9, Saccharomyces cerevisiae_10, Saccharomyces cerevisiae_11, Saccharomyces cerevisiae_12, Saccharomyces cerevisiae_13, Saccharomyces cerevisiae_14, Saccharomyces cerevisiae_15, Saccharomyces cerevisiae_16.*

1. Godfrey-Smith P (2009) *Darwinian Populations and Natural Selection* (Oxford Univ Press, Oxford).
2. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* 474(7353):604–608.
3. Denef VJ, Mueller RS, Banfield JF (2010) AMD biofilms: Using model communities to study microbial evolution and ecological complexity in nature. *ISME J* 4(5):599–610.
4. Moons P, Michiels CW, Aertsen A (2009) Bacterial interactions in biofilms. *Crit Rev Microbiol* 35(3):157–168.
5. Xavier JB, Foster KR (2007) Cooperation and conflict in microbial biofilms. *Proc Natl Acad Sci USA* 104(3):876–881.
6. Klayman BJ, Volden PA, Stewart PS, Camper AK (2009) Escherichia coli O157:H7 requires colonizing partner to adhere and persist in a capillary flow cell. *Environ Sci Technol* 43(6):2105–2111.
7. Kreth J, Vu H, Zhang Y, Herzberg MC (2009) Characterization of hydrogen peroxide-induced DNA release by Streptococcus sanguinis and Streptococcus gordonii. *J Bacteriol* 191(20):6281–6291.
8. Christensen BB, et al. (1998) Establishment of new genetic traits in a microbial biofilm community. *Appl Environ Microbiol* 64(6):2247–2255.
9. Molin S, Tolker-Nielsen T (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Curr Opin Biotechnol* 14(3):255–261.
10. Roberts AP, Pratten J, Wilson M, Mullany P (1999) Transfer of a conjugative transposon, Tn5397 in a model oral biofilm. *FEMS Microbiol Lett* 177(1):63–66.
11. Roberts AP, et al. (2001) Transfer of TN916-like elements in microcosm dental plaques. *Antimicrob Agents Chemother* 45(10):2943–2946.
12. Ready D, et al. (2006) Potential role of Veillonella spp. as a reservoir of transferable tetracycline resistance in the oral cavity. *Antimicrob Agents Chemother* 50(8): 2866–2868.
13. Hannan S, et al. (2010) Transfer of antibiotic resistance by transformation with eDNA within oral biofilms. *FEMS Immunol Med Microbiol* 59(3):345–349.
14. Perry JA, Cvitkovitch DG, Lévesque CM (2009) Cell death in Streptococcus mutans biofilms: A link between CSP and extracellular DNA. *FEMS Microbiol Lett* 299(2): 261–266.
15. 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.
16. Hansen SK, Rainey PB, Haagensen JA, Molin S (2007) Evolution of species interactions in a biofilm community. *Nature* 445(7127):533–536.
17. Filoche SK, Anderson SA, Sissons CH (2004) Biofilm growth of Lactobacillus species is promoted by Actinomyces species and Streptococcus mutans. *Oral Microbiol Immunol* 19(5):322–326.
18. Rickard AH, Gilbert P, High NJ, Kolenbrander PE, Handley PS (2003) Bacterial coaggregation: An integral process in the development of multi-species biofilms. *Trends Microbiol* 11(2):94–100.
19. Burmølle M, et al. (2006) Enhanced biofilm formation and increased resistance to antimicrobial agents and bacterial invasion are caused by synergistic interactions in multispecies biofilms. *Appl Environ Microbiol* 72(6):3916–3923.
20. Juhala RJ, et al. (2000) Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol* 299(1):27–51.
21. Marston MF, Amrich CG (2009) Recombination and microdiversity in coastal marine cyanophages. *Environ Microbiol* 11(11):2893–2903.
22. Pedulla ML, et al. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* 113(2):171–182.
23. Susskind MM, Botstein D (1978) Molecular genetics of bacteriophage P22. *Microbiol Rev* 42(2):385–413.
24. Desplats C, Krisch HM (2003) The diversity and evolution of the T4-type bacteriophages. *Res Microbiol* 154(4):259–267.
25. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S (2000) The origins and ongoing evolution of viruses. *Trends Microbiol* 8(11):504–508.
26. Martinsohn JT, Radman M, Petit MA (2008) The lambda red proteins promote efficient recombination between diverged sequences: Implications for bacteriophage genome mosaicism. *PLoS Genet* 4(5):e1000065.
27. Murphy KC (2012) Phage recombinases and their applications. *Adv Virus Res* 83: 367–414.
28. Baquero F (2009) Environmental stress and evolvability in microbial systems. *Clin Microbiol Infect* 15(Suppl 1):5–10.
29. Liebert CA, Hall RM, Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* 63(3):507–522.
30. Novais A, et al. (2010) International spread and persistence of TEM-24 is caused by the confluence of highly penetrating enterobacteriaceae clones and an IncA/C2 plasmid containing Tn1696:Tn1 and IS5075-Tn21. *Antimicrob Agents Chemother* 54(2): 825–834.
31. Novais A, et al. (2006) Dissemination and persistence of blaCTX-M-9 are linked to class 1 integrons containing CR1 associated with defective transposon derivatives from Tn402 located in early antibiotic resistance plasmids of IncHI2, IncP1-alpha, and IncFI groups. *Antimicrob Agents Chemother* 50(8):2741–2750.
32. Reeve HK, Keller L (1999) *Levels of Selection in Evolution*, ed Keller L (Princeton Univ Press, Princeton), pp 3–14.
33. Gonzalez A, et al. (2011) The mind-body-microbial continuum. *Dialogues Clin Neurosci* 13(1):55–62.
34. Dominguez-Bello MG, et al. (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 107(26):11971–11975.
35. Reid G, et al. (2011) Microbiota restoration: Natural and supplemented recovery of human microbial communities. *Nat Rev Microbiol* 9(1):27–38.
36. Robinson CJ, Bohannan BJM, Young VB (2010) From structure to function: The ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* 74(3): 453–476.
37. Moya A, Gil R, Latorre A (2009) The evolutionary history of symbiotic associations among bacteria and their animal hosts: A model. *Clin Microbiol Infect* 15(Suppl 1): 11–13.
38. Khoruts A, Sadowsky MJ (2011) Therapeutic transplantation of the distal gut microbiota. *Mucosal Immunol* 4(1):4–7.
39. Khoruts A, Dicksved J, Jansson JK, Sadowsky MJ (2010) Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol* 44(5):354–360.
40. Angelakis E, Armougom F, Million M, Raoult D (2012) The relationship between gut microbiota and weight gain in humans. *Future Microbiol* 7(1):91–109.
41. Pradeu T (2012) *The Limits of the Self. Immunology and Biological Identity* (Oxford Univ Press, Oxford).
42. Smillie CS, et al. (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480(7376):241–244.
43. 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.
44. Okasha S (2010) Levels of selection. *Curr Biol* 20(7):R306–R307.
45. Nair SC, Striepen B (2011) What do human parasites do with a chloroplast anyway? *PLoS Biol* 9(8):e1001137.
46. Yeh E, DeRisi JL (2011) Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage Plasmodium falciparum. *PLoS Biol* 9(8): e1001138.
47. 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.

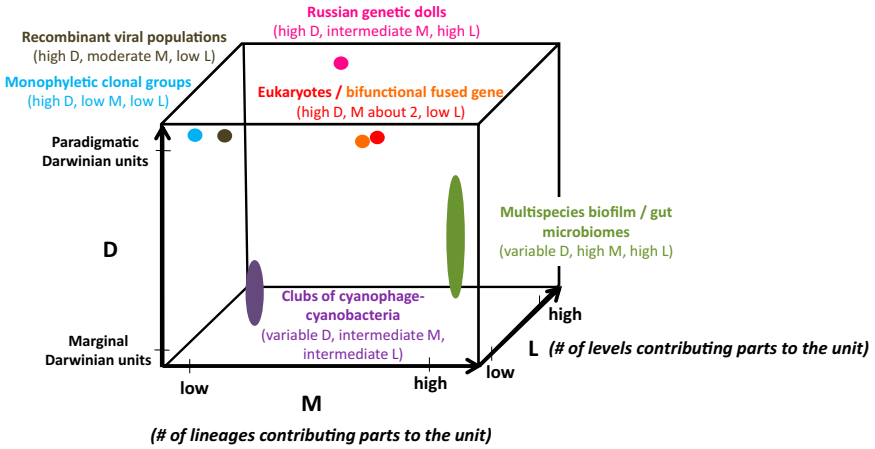

Fig. S1. Multidimensional representation of mergers/clubs as Darwinian evolutionary units. Location of our mergers and clubs in a multidimensional space in 3D: M corresponds to the number of lineages contributing parts to a unit, L corresponds to the number of levels of biological organization contributing parts to a unit, and D corresponds to the extent to which a unit is paradigmatically or marginally Darwinian based on the criteria selected by the work by Godfrey-Smith (1).