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.

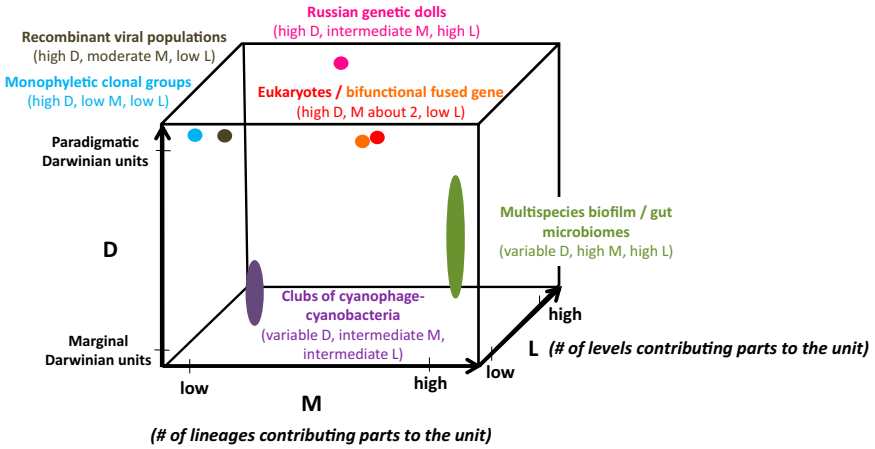

Fig. S1. Multidimensional representation of mergers/clubs as Darwinian evolutionary units. Location of our mergers and clubs in a multidimensional space in 3D: M corresponds to the number of lineages contributing parts to a unit, L corresponds to the number of levels of biological organization contributing 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).