Review in Advance first posted online on July 6, 2016. (Changes may still occur before final publication online and in print.)

# The Modern Synthesis in the Light of Microbial Genomics

Austin Booth,[1,2] Carlos Mariscal,[1,2,3] and W. Ford Doolittle[2]

[1]Department of Philosophy, Dalhousie University, Halifax B3H 4R2, Nova Scotia

[2]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax B3H 4R2, Nova Scotia; email: ford@dal.ca

[3]Department of Philosophy, University of Nevada, Reno, Nevada 89557



Annu. Rev. Microbiol. 2016. 70:279–97

The *Annual Review of Microbiology* is online at micro.annualreviews.org

This article's doi: 10.1146/annurev-micro-102215-095456



## Keywords

Modern Synthesis, microbiology, genomics, tree of life, lateral gene transfer, endosymbiosis

## Abstract

We review the theoretical implications microbial genomics poses for evolutionary biology since the Modern Synthesis. We examine the ways in which microbial genomics has influenced our understanding of the last universal common ancestor, the tree of life, species, lineages, and evolutionary transitions. We conclude by advocating a piecemeal toolkit approach to evolutionary biology, in lieu of a grand unified theory updated to include microbial genomics.

## Contents



## INTRODUCTION

> Even prior to recent developments stemming from the growth of genomic technology, philosophy of biology has been culpable in its failure to take serious account of the microbiological realm. Today this omission is inexcusable. (O'Malley & Dupré, 92, p. 179)

The evolutionary Modern Synthesis (MS, also known as the New Synthesis or neo-Darwinism) is a powerful conceptual edifice, erected in the early to middle part of the last century by evolutionary biology's leading theoreticians on a foundation of systematics, population genetics, organismal biology and paleontology, much aided in the latter half of that century by molecular biology. The MS as usually construed is committed to natural selection operating on successive undirected mutations of small effect as the dominant cause of adaptation, macroevolution (speciation and larger evolutionary trends) as a consequence of microevolution (anagenesis, or evolution within species), the reality of "species," and (more peripherally) the naturalness of their representation in a singly rooted tree-like pattern of phylogenetic relationships (17, 52, 77, 96).

The last decade or two have seen proposals for a more holistic extended evolutionary synthesis (EES), in which "developmental processes, operating through developmental bias, inclusive inheritance and niche construction, share responsibility for the direction and rate of evolution, the origin of character variation and organism-environment complementarity" (65; see also 53). Whatever the value of these proposals (23), both the MS and the EES as developed by biologists and philosophers focus theoretical attention on multicellular eukaryotes, to the neglect of the majority of Earth's contemporary living creatures, and its exclusive inhabitants for three-quarters of its history—microbes.

There are exceptions, of course (29, 31, 60, 70, 76, 90, 92), and indeed much of the molecular genetic articulation of the MS rests on microbial experimentation (55). But it is undoubtedly the exponentially growing databases of microbial genome sequences (from 1 in 1995 to more than 40,000 as of this writing) that calls into question certain tenets of the MS, as usually formulated. And it is largely the fluidity and mixing of genic, genomic, and organismal lineages occasioned by lateral gene transfer (LGT; sometimes denoted HGT, or horizontal gene transfer) and organism-level lineage merger processes that call the loudest.

However, the MS is itself a complex and contingent evolved historical entity, structured as much by relations of ancestry and descent among practitioners and their ideas as by propositions and proofs (50). Scholars will disagree about its tenets (21, 44)—often in the service of their own intellectual agendas—and the forms in which it is presented to the public can be oversimple, even rhetorical in purpose. That the MS has survived countless reformulations and caveats speaks as much to disciplinary allegiances of its advocates (and the desire not to give comfort to creationists) as it does to any deeper theoretical unity.

The toolkit approach we advocate in the final section of this article seems a reasonable alternative to a further general reformulation of the MS in order to accommodate microbial genomic fluidity. Such an approach begins with the recognition that we already understand and have documented a great many individual genetic, population, and ecological processes that currently operate in microbes and/or nonmicrobes (macrobes) and can infer others. The operation of these and similar mechanisms over four billion years are adequate in principle to explain the diversity and adaptiveness of contemporary living forms. There is no need for any more comprehensive or unifying evolutionary theory, no deeper metanarrative other than that the biological present is in principle explicable in terms of processes that also operated in the biological past.

## ORIGINS, ANCESTRY, AND ANCESTORS

> All living things have much in common, in their chemical composition, their germinal vesicles, their cellular structure, and their laws of growth and reproduction. We see this even in so trifling a circumstance as that the same poison often similarly affects plants and animals; or that the poison secreted by the gall-fly produces monstrous growths on the wild rose or oak-tree. Therefore I should infer from analogy that probably all organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed. (Darwin, 16, p. 484)

The hypothesis that there was a single ancestor for all of life was among the most revolutionary ideas in Darwin's *Origin*. Although there are still ways of making sense of this claim, extensive LGT problematizes the concept.

Contemporary discussion exploring the nature of the universal ancestor began in the late 1970s, when Carl Woese and George Fox discovered a division within prokaryotes so fundamental that they argued it was even deeper than distinctions at the kingdom level: domains Archaea and Bacteria (as they are now called) (125, 129). Woese's view of the early origins of these groups (as well as eukaryotes) was that each had annealed from a more interconnected network of primitive cells, which he called progenotes (126). The progenote era was characterized by rampant lateral transfer of molecules, and progenote lineages as such were not the primary beneficiaries of evolutionary processes. In Woese's model, early life-forms were metabolically interdependent. (There are interesting resonances here with recent findings of communities of interdependent, small-genome bacteria and archaea (11, 13). In the three domains, cells eventually evolved integrated multicomponent structures, enabling each lineage to resist the erosive effects of LGT (127). Woese viewed

the transition between an LGT-dominated community and lineages evolving primarily by vertical descent as the "Darwinian Threshold" (128), a key event in the history of life, marking a dramatic change in the tempo and mode of evolution.

In Woese's model, each lineage arose from a separate ancestor with a more rudimentary coupling between genotype and phenotype than that exhibited by modern prokaryotes (126). Indeed, he discounted Darwin's inference of a single ancestor, musing that,

> The universal phylogenetic tree, therefore, is not an organismal tree at its base but gradually becomes one as its peripheral branchings emerge. The universal ancestor is not a discrete entity. It is, rather, a diverse community of cells that survives and evolves as a biological unit. This communal ancestor has a physical history but not a genealogical one. Over time, this ancestor refined into a smaller number of increasingly complex cell types with the ancestors of the three primary groupings of organisms arising as a result. (Woese, 126, p. 6854)

However, others argue that there are so many similarities in sophisticated properties between all three lineages that a single common and much more prokaryote-like last universal common ancestor (LUCA) must be assumed, and they use comparative genomics and parsimony to deduce its likely gene content. Results are contradictory. Several studies support the notion of a generally sophisticated LUCA, possibly physiologically nearly indistinguishable from a contemporary prokaryote, with more or fewer genes (5, 20, 94, 99). Other studies using such methods to reconstruct a singular LUCA conclude that it had advanced metabolism but primitive ribosomes and no transcription (58), or other preprokaryotic, primitive features (22). Indeed even an RNA-genomed creature has been mooted (85).

A simple resolution of these contrasting views was suggested by one of us (24) and rearticulated recently by Peter Gogarten and colleagues (35). In this view, LUCA was a single cell or species not unlike contemporary prokaryotes, but many and possibly all of the genes in its genome have been replaced by LGT in its many descendant lineages, sometimes by orthologs and sometimes by genes encoding alternative metabolisms. The donors in these LGT events will have been (*a*) other still-extant or now-extinct lineages descended from LUCA, and (*b*) lineages contemporary with LUCA, but now extinct as such, though some of their genes live on in LUCA's progeny.

What distinguishes this view from Woese's is the implication that, although cellularity no doubt arose multiple times, all extant cells descend from one cell, as indeed the pre-Darwinian cell theory would have it (3). If that cell were a member of a species that engaged in some mating process entailing cell fusions, then perhaps we should speak of the species as ancestral. But if a diverse community as described by Woese diverged into three quasiseparated diverse communities, ancestral to the three domains, perhaps we must speak only of a common ancestry, not a common ancestor. And without some agreed upon definition of life—of which philosophers currently despair (6)—we would be at a loss to say how many times it arose. Indeed, even with the above and conceptually simpler single (cellular) LUCA, genomic ancestry and cellular ancestry are radically uncoupled in ways that neither Darwin nor the architects of the MS could have anticipated.

## THE TREE OF LIFE

> This connection of the former and present buds by ramifying branches may well represent the classification of all extinct and living species in groups subordinate to groups. (Darwin, 16, p. 129)

## Darwin's Tree of Life Hypothesis

Darwin described the tree of life (TOL) as a simile that "largely speaks the truth" (16; p. 129). The tree-like pattern represented in popular classifications of groups subordinate to groups results, he hypothesized, from an underlying tree-like process, namely speciation and divergence driven by natural selection. Phenetic classifications had a phylogenetic rationale and were not, as creationists held, a reflection of the Creator's orderly mind.

In the century after 1859, most classificatory trees presented simply assumed such an evolutionary causal process (18, 93), having no direct access to such a process or way to prove it. As de Queiroz writes of this period, "the relationships expressed in existing taxonomies were merely reinterpreted as the result of evolution, and evolutionary concepts were developed to justify existing methods" (18, p. 238).

It was not until the 1960s that Emile Zuckerkandl and Linus Pauling proposed that molecular sequence data, known to be produced by an inherently tree-like process (replication and mutation) could offer independent support for Darwin's claim that phylogenetics explains phenetics. They wrote that,

> It will be determined to what extent the phylogenetic tree, as derived from molecular data in complete independence of organismal biology, coincides with the tree constructed on the basis of organismal biology. If the two phylogenetic trees are mostly in agreement with respect to the topology of branching, the best available single proof of the reality of macroevolution would be furnished. (Zuckerkandl & Pauling, 131, p. 101)

Three objections could have been raised to this notion of proof at the time and are still relevant. First, it should surprise no one (creationists included) that similar genotypes are required to encode similar phenotypes. Surely only an especially unintelligent creator would invent new genetic codes and means for their expression for each of billions of species, making rats from entirely different genes than mice. Second, for prokaryotes there was (and remains) no agreed upon phenetic classification, no "tree constructed on the basis of organismal biology" with which to compare molecular results. Third, such a proof requires that trees derived from molecular data for different genes will agree with each other, unless only some are assumed to speak the truth. It is the disagreement between trees for different genes due to LGT that creates problems for the TOL as a hypothesis relating phenetic classification and phylogeny. Most of these disagreeing data have come from prokaryotic genome sequencing projects completed in the last two decades.

## Is The Tree of Life Hypothesis False?

It is still unclear just how extensive disagreement between prokaryotic gene trees actually is, how it should be measured, and over what time scale. For many species, genomes of many strains have been sequenced and show astonishing variation in gene content. Among more than 2,000 *Escherichia coli* strains—with an average genome size of about 5,000—genes shared between all or almost all number little more than 3,000, but the number of gene families with a representative in at least one *E. coli* genome is approaching 90,000 (66). This can only mean extensive gene gain and loss within the species, as Eugene Koonin concludes:

> The wide spread and high rate of gene exchange and loss in the prokaryotic world translate into "network genomics." The rates of gene gain and loss are comparable with the rate of point mutations

> but are substantially greater than the duplication rate. Thus, evolution of prokaryotes is primarily shaped by gene gain and loss. (Koonin, 61, p. 244)

Many transient genes involved in strain-specific adaptations (virulence, antibiotic resistance, substrate utilization) are links in this genomic network. These clearly comprise a sort of genetic public goods resource—a pangenome of accessory genes—that can be drawn on as needed (82). Also inhabiting this space are the various selfish elements (phages, plasmids, and transposable elements) responsible for much LGT, and it may be that short-term additions to genomes are actually on average fitness-reducing at the cellular level (4, 121).

More phylogenetically stable would be genes for metabolic processes characterizing taxa more inclusive than prokaryotic species. Martiny and colleagues (75) have nicely correlated trait depth (the average 16S distance between members of a clade and the clade's root for clades in which at least 90% of members show a particular trait) with trait complexity (number of genes required). Oxygenic photosynthesis (a defining feature of cyanobacteria) runs deepest, followed by methane oxidation, methanogenesis, and sulfate reduction. Shallowest are sugar utilization traits, variable at the species pangenome level. This and other recent demonstrations of a negative correlation between transferability and the extent of protein-protein interactions are roughly consistent with Jim Lake's complexity hypothesis (54), although multiple causal forces are likely at play (43).

Thus the extent to which any phenetic classification of prokaryotes corresponds to their phylogeny depends on the phenotypic trait considered and the complexity of its determination. Recognizing this will go some way to cooling down the argument about whether LGT falsifies Darwin's TOL hypothesis: Whether or not this is so is gene- and system-dependent.

## The Statistical Tree of Life and What It Might Represent

The only suite of traits whose trait depth is as deep as the presumed universal TOL comprises those involved in translation and transcription. It is tempting to think that this tells us something fundamental about cell biology and the evolutionary process (27, 129), but of course some suite of traits has to have the greatest phylogenetic depth and it is not surprising that we have built our phylogenetic worldview around that suite.

In any case, we are left with relatively few universally retained genes with which to build a TOL. Puigbò et al. (97) call 102 universal (all-domain) trees, based mostly on this largely translational core, nearly universal trees. Nearly universal tree topologies are "far more congruent than expected by chance," and "appear to reflect a significant central trend, an attractor in the tree space that could be equated with the STOL [statistical tree of life]" (97, p. 46). However, if all but a small percentage of the genes in a typical genome show different topologies or a patchy distribution indicative of recurring gain and loss, what is the statistical tree meant to prove or represent the tree *of*, over the full evolutionary timescale? We think that most investigators would admit that, because of LGT, there is no unique tree of genomes and that network patterns would more accurately describe genome relationships or histories, at almost any scale.

That said, molecular phylogeneticists might insist that they never sought—let alone required—a unique tree of genomes, and that proving Darwin's obviously true TOL hypothesis was never a goal. This position is not unlike that of premolecular systematists, as described above by de Queiroz (18). That there is an organismal tree (a tree of cells or species) is accepted as fact, and the task all along has been to use gene and genome sequence data in order to determine its topology. Patrick Foreterre perhaps spoke for the majority when he recently wrote, "The universal tree should depict evolutionary relationships between domains defined according to the translation apparatus reflecting the history of cells (and their envelope…) and not according to the global

genome composition that is influenced by LGT, virus integration and endosymbiosis, the history of which is incredibly complex" (34, p. 3).

## Prokaryotes Versus Eukaryotes

To be sure, incongruence between gene and species trees is not a new problem or unique to prokaryotes. That it is an inevitable consequence of incomplete lineage sorting in speciating populations of sexually reproducing eukaryotes (or introgression between recently speciated populations) is the subject of a well-developed body of theory (71). It is nevertheless widely believed that LGT in prokaryotes is more frequent and can span greater phylogenetic distances than can LGT in eukaryotes and is thus more profoundly disruptive of tree building and the MS (to the extent that it embraces a TOL).

This may well be so, and two recent influential reviews claim to have shown that most prokaryote-to-eukaryote LGT events are recent (of limited phylogenetic distribution among eukaryotes) and that the major prokaryotic contributions to eukaryotes are of organellar (mitochondrial or plastid) origin (56, 63). Moreover, the well-known trio of LGT-mediating prokaryotic processes (transduction, conjugation, and transformation) have no obvious counterparts in many eukaryotes, and the separation of germline and soma in many might seem to mitigate against evolutionarily significant transfers (113).

These things said, we arguably do not have fair metrics to determine the relative importance of prokaryotic and eukaryotic LGT. Comparing eukaryotes to prokaryotes is not like comparing apples to oranges (8). Rather it is like comparing apples to all fruit, since prokaryotes are more ancient and thus unsurprisingly more diverse. It is often noted for instance that eukaryotes are biochemically more monotonous than prokaryotes, but are they in fact less biochemically diverse than cyanobacteria or methanogens? And do they enjoy fewer LGT events? A recent genome-scale analysis in fact found that rates of LGT in cyanobacteria and fungi are roughly comparable (116), and there are many well-documented cases of major contributions of LGT to eukaryotic adaptations for specific niches (105, 106). Indeed, the term pangenome is now being used to describe gene content variation in eukaryotes, perhaps surprising in its extent (42, 100).

Moreover, surveys of LGT into eukaryotes are often blind methodologically to eukaryote-eukaryote transfers, and yet experience with prokaryotes suggests that intradomain exchanges are more frequent than interdomain exchanges. We submit that whether or not LGT is as disruptive a force in eukaryote evolution as it is for prokaryotes remains an open question, especially when apples-versus-fruit biases and methodological difficulties are factored in.

## Lateral Gene Transfer, Lamarck, and Punctuated Equilibria

Prokaryotic genome evolution is different in tempo and mode from genome evolution as modeled in population genetics of multicellular eukaryotes, embedded in the MS. In such models, innovations arise largely through reshuffling of standing variation and occasionally by gene duplication and neofunctionalization (117). Fixation of favorable mutations and rearrangements occurring within individuals within a species are taken to be the predominant modes of anagenesis. For many and maybe most prokaryotes, this is not so (see above quotation from Koonin). Indeed, it is now thought likely that, within prokaryotes, what appear to be duplications are more often the product of LGT events introducing orthologs from other strains or species (119).

Both philosophers (29) and biologists (62) have considered and even endorsed the idea that prokaryotic evolution is rendered Lamarckian (an epithet often contrasted with the MS) by such processes. But LGT is very far from anything that either Darwin or Lamarck could have dreamt of, whereas elements of the EES that have been called Lamarckian aim at dethroning DNA, not

making it a central focus, as LGT inevitably must (48). Our view is that reviving the terminology of Lamarckism when challenging the MS through microbial genomics makes for more heat than light and does little to foster theoretical clarity or agreement about how microbial evolution actually occurs.

A more fruitful avenue, we think, might be the importation into microbiology of punctuated equilibrium, to the extent that this is a challenge to the MS (45). It is increasingly popular to see microbial genome evolution as comprising short episodes of simultaneous (or nearly simultaneous) acquisition of multiple genes by LGT or endosymbiotic gene transfer (EGT) followed by prolonged periods of differential loss along different lineages (33, 130). Gradual cumulative evolution, so often viewed as essential to the MS, is arguably violated in such models.

## SPECIES

> Some would so emphasize variation in the characteristics of bacteria as to deny the value of the species concept as applied to this group . . .
>
> (Breed, 9, p. 144)

### Definitions and Concepts

In the 88 years since the above words were written, the grounds for bacterial species denial have changed radically, though the conclusion remains. The problems of bacterial species concepts are not different in kind from those plaguing eukaryotic systematics (49, 124), but at first blush they do seem different in degree. For eukaryotes there is general agreement that something resembling George Gaylord Simpson's lineage-based "evolutionary species concept" (108), as tightened up by de Queiroz, underlies many variant formulations. De Queiroz (19), writes, "Alternative species concepts agree in treating existence as a separately evolving metapopulation lineage as the primary defining property of the species category, but they disagree in adopting different properties acquired by lineages during the course of divergence (e.g., intrinsic reproductive isolation, diagnosability, monophyly) as secondary defining properties (secondary species criteria)" (19, p. 879).

Intrinsic reproductive isolation or exclusive interbreeding is of course the "secondary defining property" that informs Ernst Mayr's Biological Species Concept (BSC, 79), the most widely accepted by biologists and easily understood by laypersons. By insisting that species are maximally inclusive interbreeding groups, the BSC can tell us where to draw the line between varieties, species, and genera. The BSC leads to *cohesive divergence*, simultaneously promoting within-species genetic homogenization and between-species differentiation. But Mayr himself admitted (and did not seem to particularly care) that nonbreeding (asexual) organisms do not conform to the BSC, and he understood most bacteria to be asexual (78).

Similarly persuaded, bacteriologists (and those studying archaea) had largely contented themselves with an operational or practice-oriented "polyphasic" species definition agreed upon in the 1980s (114) and based on DNA-DNA hybridization. Indeed, many speak of operational taxonomic units rather than species, to avoid any ontological commitment. Since the mid-1990s, a 16S rRNA sequence similarity of 97% has been taken as an easier way of delimiting species (114), and there is increasing support for measures based on Average Nucleotide Identity (ANI) of shared single-copy genes, with ANI >96% corresponding to more traditional species circumscription in a well-studied case (15). In accepting such explicitly operational definitions, bacteriologists admitted that what they were missing was a species concept, a single model of

genetic, population genetic and ecological processes as these might together ensure that bacteria form genotypic and/or phenotypic clusters roughly corresponding to the species definition.

## A BSC for Bacteria

That something rather more like the BSC might after all drive cohesive divergence, at least in some bacteria, was apparent from gene sequence analyses beginning in the 1990s (30, 47, 110). These showed that in spite of clonal (asexual) reproduction at the cellular level, homologous recombination (HR) can be frequent (after acquisition of homologous foreign DNA *via* conjugation, transduction, transformation, or processes unknown). Dykhuizen & Green noted that insofar as interbreeding entails recombination, the BSC, "implies that the phylogenes [*sic*] of different genes from individuals of the same species should be significantly different, whereas the phylogeny of genes from individuals of different species should not be significantly different. Thus we have an operational criterion for the defining of bacterial species" (30. p. 7266).

Although there are no prokaryotes in which sex and HR are essential for reproduction, a BSC-like model seems appropriate in some cases. A 2009 tabulation of recombination frequencies—as rates of nucleotide change from HR relative to changes from point mutation (r/m)—ranged from a minimum of 0.02 for the commensal spirochaete *Leptospira interrogans* to 63.6 for *Flavobacterium psychrophilum*, a pathogenic Bacteroidetes (120). Archaea can also recombine avidly, and some haloarchaea can "mate" by cell fusion even between quite distinct species as defined by ANI, exchanging surprisingly long (>500 kbp) contiguous stretches of DNA (87).

HR is expected to fall precipitously in frequency as homologs diverge in sequence through mutation (72), and one can model situations in which cohesive divergence results solely from the interplay of these two processes, producing stable genotypic clusters (species), sympatrically. An early modeling study by Fraser et al. concluded, however, that dependence of HR rate on divergence is too weak for such speciation to occur without some degree or physical separation, or barriers to exchange between members of diverging population, as might be afforded by differential resistance to phages or transformation (37).

## Ecotypes

An alternative species concept, also dating to before accumulation of significant genomic data, was developed by Fred Cohan, who elaborated a variety of "ecotype" models based on periodic selection in asexual clones (14). In species with scant HR, favorable mutations (which can include acquisitions by LGT) will sweep to fixation in a population, along with the chromosome on which they occur. Diversity at all loci is thus purged in the sweep, only to accumulate anew until the next purge. The result is clusters of related sequences (at the 16S rRNA or other marker loci) defining ecotypes. Diversity within ecotype clusters is a measure of time since the last sweep. Diversity between clusters is an indication of ecological separation, effected physically (particle-associated versus free in the water column, for instance) or biochemically (metabolism of different substrates, resistance to different phages). As with asexual eukaryotes, if there were to be species-creating "secondary defining properties" (19), they would be ecological. Cohesion within ecotypes and divergence between them is thus the result of the same process: periodic selection.

Shapiro & Polz recently reviewed empirical and simulation studies that emphasize the relative importance of recombination (*r*, favoring HR models) and selection (*s*, favoring ecotypes) for the cohesive divergence of prokaryotic genomes. When $r/s \gg 1$, selection succeeds only in fixing the positively selected locus (107). Two ecologically differentiated subpopulations (suppose a locus under selection determines which of two plentiful substrates will be used) will look like a single

species by Dykhuizen & Green's criteria, sporting different trees at all but that locus. If $r/s \ll 1$, then something like Cohan's ecotype model will obtain: diversity at all loci will be purged and all gene trees within a cluster will be the same. As Shapiro & Polz argue, however, "...even with $r/s \ll 1$, given enough time before any further selective events, and assuming that the two [substrate-defined] niches remain sympatric, neutral alleles will eventually become randomly distributed across genotypes, with only adaptive alleles being selectively maintained" 107, p. 9). Same result, more slowly achieved.

### Speciation Without Species

Similar to Fraser et al. (37), Shapiro & Polz conclude that, "some kind of microgeographic separation between niches, akin to the 'mosaic sympatry' described by Mallet (73) might be required to reduce gene flow between niche-adapted genotypes before clusters of selectively neutral genome-wide diversity may develop" (107, p. 9). Importantly, though, these authors write that, "We wish to make a strong distinction between this process of speciation – which we define as any stage of the dynamic process of ecological and genetic differentiation – and the concept of species, which we are not attempting to address" (107, p. 5). Indeed, the excitement in contemporary microbial ecology lies in mapping genomic diversity to microgeographic ecological differentiation or other possibly ephemeral barriers to gene flow, not in developing any unified species concept.

Both the BSC and ecotype models are lineage concepts in de Queiroz's sense, but as secondary defining properties they are antithetical. Although interplay between the two speciation processes does corrfelate ecological and genetic differentiation, the result will be what Shapiro & Polz call a, "'speciation spectrum,' which microbial populations traverse in different ways depending on their balance of gene flow and natural selection" (107, p. 1). Some bacteria or archaea will form clusters tight enough that no one would be reluctant to call them species, but others would not. That is, any enumeration of all the prokaryotic "species" in the world will fail to take in all the individuals.

Moreover, most microbial speciation theory addresses itself to shared orthologs and recombination and selection between and on them, treating LGT as a sort of macromutation. Formal population genetic treatment of pan-genomic turnover per se, as it affects species-like clustering, has scarcely begun (95, 121). Undoubtedly, LGT in prokaryotes is more promiscuous in terms of frequency and genomic contribution, and less demanding in terms of donor-recipient relatedness than is HR. Its impact on formal species conceptualization will be even more disruptive. Microbial genomics has brought home the impossibility of applying a single species concept to all organisms and stimulated a healthy focus on the interplay of genetic and ecological processes in both limiting and promoting the diversification of genomes – "speciation without species" (28, 36, 68). Elimination of "species" as a vestige of pre-Darwinian essentialism was in any case long overdue (25).

## LINEAGES AND TRANSITIONS

> Biological individuals are probably all situated at the confluence of several lineages of traditional organismal entities, and the adaptive benefits and biological constraints conferred by such arrangements underpin whether these meta-entities or their contributing elements constitute evolutionary individuals. (O'Malley, 90, p. 117)

### Endosymbiosis

LGT between ephemerally interacting cells is not the only source of revisionism about the overall "shape" of life, from a tree-like representation to a more reticulated network-like representation.

Not long after Lynn Margulis's (re)articulation of the symbiotic theory of organelles in 1967 (104) molecular phylogenetic techniques were brought definitively to bear as an alternative source of evidence (1, 46). The bacterial origin of mitochondria and plastids is now uncontroversially accepted. Moreover, as microbial genomic research has progressed, other endosymbiotic relationships, involving bacteria (51), protists and bacteria (2, 123), and multicellular eukaryotes and bacteria (123) are also now extensively documented and studied. While microbial genomics has played no small part in establishing the truth of the endosymbiotic theory of organelles and continues to shed light on other endosymbiotic relationships, it is possible to think of these phenomena primarily as examples of cell lineage fusion, and so as a reticulating evolutionary process distinct from LGT.

That is not say that mechanisms of genomic fluidity and integration have not affected the evolution of endosymbiotic systems. EGT, from endosymbiont to host nuclear DNA, does occur, the best understood cases being those involving transfer from the bacterial precursors of what are now considered organelles (mitochondria and plastids) (118). Some of the products of those genes are targeted to the organelle. Moreover, EGT is not limited to cells containing the classical organelles, as has been observed in *Paulinella chromatophora*. *P. chromatophora* contains one or two blue-green chromatophores per cell, derived from *Synechococcus* species (cyanobacteria unrelated to plastids). The chromatophores have the hallmarks of traditional organelles, including EGT to the host nucleus and posttranslational targeting of nuclear gene products back to chromatophores (88).

Other cases of endosymbiosis show complex mixing of cell lineage mergers and LGT from diverse bacterial sources. The sap-feeding insect *Pachypsylla venusta*, for example, contains an endosymbiont (*Carsonella ruddii*) with an extremely reduced genome. The endosymbiont expresses genes that synthesize amino acids necessary for host nutrition, whereas several genes in host nuclear DNA, derived from various bacterial origins (including one likely from *C. ruddii* itself), are involved in maintaining metabolic pathways interrupted by gene loss in the reduced endosymbiont (109). *C. ruddii* thus has some organelle-like properties (genome reduction and EGT to the host) but likely no others (in particular, host-derived proteins specifically targeted to the endosymbiont) (80). Conversely, Nakabachi et al. demonstrated that a gene of alphaproteobacterial origin in the pea aphid genome targets the aphid's gammaproteobacterial endosymbiont (*Buchnera aphidicola*) (86). Here we see another similarity to the biology of organelles: A gene in the host, of bacterial origin, produces a protein that is targeted at an endosymbiont, itself derived from a different bacterial phylum.

We think it clear that endosymbioses provide an extremely interesting assemblage of case studies and potential challenges for the evolutionary theoretical commitments of the MS. To begin with, endosymbiotic systems are implicated not only in cell-level reticulating processes, but also in diverse and apparently iterated gene-level reticulating processes. Minimally, this gives rise to the possibility that it is not just rampant LGT among prokaryotes that might give theorists reason to rethink an overly strong commitment to a universal, branching, tree-like representation of ancestry and descent among all life. Endosymbiotic systems show that reticulation sometimes occurs even among cells and genes in derived animal lineages, and research continues to erode any supposedly robust distinction between mere endosymbionts and organelles. We find ourselves compelled to agree with McCutcheon & Keeling: "As the prevalence of intimate and stable endosymbiotic associations has become more clear, the degree to which host and endosymbiont are integrated has been revealed to be far less discontinuous than previously believed" (80, p. R655).

Another conceptual issue relevant to potential revision of the MS is the role of saltational evolution occurring as a direct result of the sudden acquisition of endosymbionts (59). Endosymbiosis makes it clear that the evolution of new metabolic capacities can occur via mechanisms other than gradualist point mutations in endogenous host DNA, and this challenges to some extent tenets

of the MS holding that evolution always occurs via small, cumulative steps. Some theorists have gone further, suggesting that explaining endosymbiotic evolution requires that a robust metabolic explanatory component be added to a more traditional, gene-based evolutionary framework (91). Proposals like this strike us as interesting and provocative, but we do not wholeheartedly endorse them.

## Metagenomics and Community Ontology

Endosymbiosis raises questions about the extent to which the resulting system is most accurately conceived as a community, or as an individual in its own right (41, 74). In a recent review, Jennifer Wernegreen distinguishes between different varieties of integration that can occur among endosymbiotic systems, remarking, "Such profound functional and genetic interconnectedness raises the question of whether such partnerships represent two distinct organisms or a unified amalgamation" (123, p. R560-R561). We agree that questions about the emergence of new forms of cohesive biological individuality are especially relevant in cases of endosymbiosis. In addition, we suggest, echoing other recent commentators (7, 26, 38, 90), that multilineage microbial communities might in general exhibit varying degrees and kinds of integration that influence our conceptualization of them as emergent biological individuals.

Recent metagenomic analyses have revealed the extent to which microbes from domains Bacteria (11) and Archaea (13) exhibit extremely small, streamlined genomes. Extreme genome streamlining has also been documented among intracellular symbiotic bacteria (81), though the forces responsible for their streamlining may well differ from those affecting free-living prokaryotes (39). Such findings have led to speculation that, at least among environmental samples of Archaea and Bacteria, there is "strong interdependence in both domains" (13, p. 690). A recent theoretical innovation, the black queen hypothesis (BQH), makes the case for the ubiquity of cooperative multispecies microbial communities even more plausible. The BQH purports to explain extreme genome streamlining in free-living organisms, and concomitant loss of function, by proposing that individual microbes can lose essential function if situated in an environment in which some members of the community retain the function, and in which those resources are available to other members of the community (84). This explanatory strategy gives special status to the idea of the pangenome of a microbial community, understood as the set of core genes (possessed by all members of the community) and accessory genes (possessed by just one or a few members) (38). The combined products of the pangenome are a full complement of metabolically essential components that are in effect a shared resource among a microbial community, regardless of the phylogenetic heritage (shared or not) of the members of the community.

The BQH thus provides not only a coherent mechanism for the evolution of extremely streamlined genomes, but also a mechanism for how the evolutionary fates of (potentially unrelated) microbes can become tied together. Sachs & Hallowell (103) suggest that such cooperation might open the door to further integration and stabilization of particular communities by fostering the evolution of costly cooperative traits that maximize the essential functions provided by members of the community. The idea that shared evolutionary fates of individuals within a group might result in higher levels of evolutionarily relevant organization is not new (112), but the idea arguably has a special relevance for microbial systems given their large populations, capacity for interdomain gene and resource sharing, extremely fast evolutionary rates, coupling of gene loss with loss of function, and propensity for deletion bias (64).

Enthusiasm for the idea that black queen phenomena might result in the evolution of important higher levels of organization is articulated by Fullmer et al., who argue that, under a strong BQH, "individual cells would be integrated into a meta-organism, whose genome is the pan-genome of

the population. . .” (38, p. 3). We are intrigued by this and similar ideas, but our enthusiasm is perhaps more tempered. What exactly constitutes organismality is a matter of some debate (98), and there is arguably more to it than being a beneficiary of a set of shared genomic resources, including substantial structural (cellular, morphological) integration (123), spatial boundedness, and the extent to which there are mechanisms for autonomous reproduction of the whole community (40). One of us (Doolittle) previously advocated an approach to thinking about the emergence of microbial metaorganismal communities that envisions a continuous range,

> in the extent to which multi-lineage consortia or communities can be seen as reproductive individuals subject to natural selection. Most paradigmatic would be integrated endosymbiosis, such as eukaryotic cells in which organellar and nuclear genomes function in obligatory cooperation, despite their separate origins and distinct means of replication and segregation. . . Less tightly but still necessarily integrated would be many prokaryotic consortia whose members, because they exchange essential nutrients, do not grow well (or at all) alone. (Doolittle, 26, p. 370)

Conceiving organismality and, more generally, individuality as a continuous condition is, today, the received wisdom (40, 90, 98). Nevertheless, we consider this set of issues theoretically important (given the requirement in many evolutionary explanatory contexts of delimiting individuals and populations), conceptually wide open for further elaboration, and also likely intrinsically linked to the interests of researchers and the particular systems and processes they study.

## AN EVOLUTIONARY EXPLANATORY TOOLKIT

> Beyond the astonishing, unexpected diversity of genome organization and modes of evolution revealed by comparative genomics, is there any chance to discover underlying general principles? (Koonin, 59, p. 1027)

### Metatheory

The title initially assigned us for this review, “Revising the Neo-Darwinian Synthesis with Whole-Genome Mapping of Microbes,” presupposes a need for some grand unifying approach, tying together all evolutionary phenomena, updated to include newly understood processes affecting microbial genomes. The metatheoretical tradition in evolutionary biology is strong, of course. Darwin himself meant the *Origin* as “one long argument” (16, p. 459), intended to persuade readers that speciation, diversity, and adaptation can be explained in terms of uniform processes of change, natural selection primary among them, operating over long periods of time. The need for a comprehensive grand theory to replace the then default natural theology—and the rhetorical strategies Darwin used to promote it—are nicely reviewed by Moore (83).

Indeed, comprehensiveness (both theoretical and rhetorical) seems to be part and parcel of evolutionary thinking. Even before the *Origin* appeared, Herbert Spencer had invoked evolution (though not via selection) as a natural law applicable to everything in the universe (32), and attempts to create or sustain a grand unified theory of evolution continue even now. Authors have attempted to derive evolution via natural selection from thermodynamic principles (10), to characterize its underlying logic (69), and to give an account based on statistical laws (57). Others have thought to bring under the traditional Darwinian umbrella new or hitherto unconsidered biological phenomena, such as multiple levels of selection (89), major transitions in evolution (76, 115), or principles from developmental evolution (evo-devo, niche construction, developmental plasticity) (65). Some approaches seek even to incorporate previously arguably abiotic

phenomena into a biological framework, as in theories of cultural evolution (101) or digital evolution (67). These efforts clearly inform specific practices and presentation within biology, but most who do not see evolution as a primary concern are content to work within more narrowly understood disciplinary frameworks.

As to the broader evolutionary biological mainstream, Rose & Oakley describe biology as having been "reintegrated twice already, first by Darwin in 1859 and then during the 'Modern Synthesis' in the 1920s and 30s" (102, p. 1) and consider that we now have a transdisciplinary "new biology" that "knits together genomics, bioinformatics, evolutionary genetics, and other such general-purpose tools to supply novel explanations for the paradoxes that undermined Modernist biology [the MS]" (102, p. 1). These paradoxes included (*a*) too much genetic variation in DNA (and too much DNA) to be adaptive in terms of any one gene–one (functionally optimized) protein model, (*b*) gene homology detectable across all of life, (*c*) LGT, symbiogenesis, lineage mergers (as in major transitions), and incomplete lineage sorting as disruptors of tree-like phylogeny, and (we would add) (*d*) the saltational consequences of LGT and lineage merger. Recognition of all of these as conceptually problematic depends heavily on microbial genomic data.

## Does the New Biology Require a New Synthesis?

Rose and Oakley see genomics more generally as integrating previously independent biological disciplines, in the way that the periodic table integrated physics and chemistry a century before. We too see genomics as integrative, but mostly methodologically, not as a conceptual framework. We endorse Rose & Oakley's emphasis on "general-purpose tools" (102, p. 1), including but not limited to many neo-Darwinian concepts. The new biology is method driven and pluralistic, embracing selection, drift, and ratchet-like neutral processes at several levels as process, and trees seamlessly melting into networks at several levels as pattern. Tempo and mode as well are liberalized: There is no reason to assume all fixed mutations are of small effect, and speciation is a messy process not necessarily entailing the formation of discrete species. Further, microbial genomic data have led to foundational discussion and potential revisionism of several traditional biological categories, even such synthesis stalwarts as gene, organism, and lineage.

Rose and Oakley seem to us neutral on the question of whether the new biology comprises a new synthesis. Koonin, on the other hand, believes that "although at present only isolated elements of a new 'postmodern' synthesis of evolutionary biology are starting to be formulated, such a synthesis is indeed feasible" (59, p. 1028). In a similar vein, Depew & Weber express confidence that "a new and more general theory of evolution is evolving that will explain the different strategies by which unicellular organisms and complex metazoa have acquired their various forms of 'evolvability'" (21, p. 100). We doubt that any such general theory is possible or will advance understanding in any significant way, including understanding of evolvability. There is no difficulty in explaining the strategies by which this phenomenon has arisen in specific instances. For example, the role of mutators in adaptive evolution in *Escherichia coli* is understood both experimentally and through simulations (111), as is the way in which variable surface antigens in trypanosomes allow evasion of host immune responses, as are the roles of individual transposable elements in specific evolutionary innovations in humans (12). All can be understood without reference to any more "general theory of evolution" (21). Given our wealth of what Rose & Oakley (102) call general purpose tools (often genomic), it is seldom the understanding of specifics in biology that is problematic. Rather it is the attempt to cast specifics as exemplars of reified entities or processes playing roles in some more comprehensive and foundational general theory. We agree with O'Malley's argument that "making universalized claims about evolution but then restricting them to particular organisms is a self-defeating strategy," but are tempted to balk at her further injunction to "find your generalizations

in the most common and pervasive life forms and then work out where to put your occasional anomalies such as animals" (90, p. 131). By contrast, we advocate an entirely piecemeal approach.

Moreover, although generalized biological theory is perhaps essential for pedagogy, it makes evolutionary biology more vulnerable to attack from creationists, the heirs of the natural theological traditions Darwinism eventually supplanted. It is often apparent disunity within the edifice of evolutionary biology that creationists attack, as if to weaken what they see as a necessarily cohesive, monolithic theoretical construct (122). If evolutionary biological theory were recast instead as a historically and loosely connected toolkit of concepts, methods, models, and mechanisms, concatenations of which can explain how individual changes might have been effected in individual molecules, organisms, or lineages—as the majority of publications in the discipline actually reflect—this vulnerability would disappear, despite the apparent disunity of the subject matter as viewed from the outside. We suggest that if the lessons of microbial biology help liberate us from the perceived need for unifying metatheory or all-embracing syntheses, that might very well be a good thing for the future of biological theorizing.



## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## LITERATURE CITED

1. Archibald JM. 2014. *One Plus One Equals One: Symbiosis and the Evolution of Complex Life*. Oxford, UK: Oxford Univ. Press
2. Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* 25:R911–21
3. Baker JR. 1953. The cell-theory: a restatement, history and critique; Part IV, the multiplication of cells. *J. Cell Sci.* S3-94:407–40
4. Baltrus DA. 2013. Exploring the costs of horizontal gene transfer. *Trends Ecol. Evol.* 28:489–95
5. Beccera A, Delaye L, Islas S, Lazcano A. 2007. *Annu. Rev. Ecol. Evol. Syst.* 38:361–79
6. Bedau MA, Cleland C, eds. 2010. *The Nature of Life: Classical and Contemporary Perspectives from Philosophy and Science*. Cambridge, UK: Cambridge Univ. Press
7. Boon E, Meehan C, Whidden C, Wong D, Langille M, Beiko R. 2013. Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol. Rev.* 38:90–118
8. Booth A, Doolittle WF. 2015. Eukaryogenesis, how special really? *PNAS* 112:10278–85
9. Breed RS. 1928. The present status of systematic bacteriology. *J. Bacteriol.* 15:143–63
10. Brooks DR, Wiley EO. 1988. *Evolution as Entropy*. Chicago: Univ. Chicago Press
11. Brown C, Hug L, Thomas B, Sharon I, Castelle C, et al. 2015. Unusual biology across a croup comprising more than 15% of domain Bacteria. *Nature* 523(9):208–11
12. Brunet TD, Doolittle WF. 2015. Multilevel selection theory and the evolutionary functions of transposable elements. *Genome Biol. Evol.* 7(8):2445–57
13. Castelle C, Wrighton K, Thomas B, Hug L, Brown C, et al. 2015. Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25:690–701
14. Cohan FM. 2001. Bacterial species and speciation. *Syst. Biol.* 50:513–24
15. Colston S, Fullmer MS, Beja L, Lamy B, Gogarten JP, Gral J. 2014. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *mBio* 5:e20136-14
16. Darwin C. 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray. 1st ed.
17. Dawkins R. 2004. *The Ancestors Tale: A Pilgrimage to the Dawn of Life*. London: Orion
18. de Queiroz K. 1988. Systematics and the darwinian revolution. *Philos. Sci.* 55:238–59
19. de Querioz K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–86
20. Delaye L, Becerra A. 2012. *Cenancestor*, the last universal common ancestor. *Evol. Educ. Outreach* 5:382–88

21. Depew DJ, Weber BH. 2011. The fate of Darwinism: evolution after the Modern Synthesis. *Biol. Theor.* 6:89–102
22. di Giulio M. 2011. The last universal common ancestor and the ancestors of archaea and bacteria were progenotes. *J. Mol. Evol.* 72(1):119–26
23. Dickins TE, Rahman Q. 2012. The extended evolutionary synthesis and the role of soft inheritance in evolution. *Proc. R. Soc. B* 279:2913–21
24. Doolittle WF. 2000. The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* 10(3):355–58
25. Doolittle WF. 2009. Eradicating typological thinking in prokaryotic systematics and evolution. *Cold Spring Harb. Symp. Quant. Biol.* 74:197–204
26. Doolittle WF. 2013. Microbial neopleomorphism. *Biol. Philos.* 28:351–78
27. Doolittle WF, Brown JR. 1994. Tempo, mode, the progenote, and the universal root. *PNAS* 91(15):6721–28
28. Doolittle WF, Zhayxybayeva. 2009. On the origin of prokaryotic species. *Genome Res.* 19:744–56
29. Dupré J. 2010. Postgenomic Darwinism. In *Darwin*, ed. W Brown, A Fabian, pp. 150–71. Cambridge, UK: Cambridge Univ. Press
30. Dykhuizen DE, Green L. 1991. Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* 173:7257–68
31. Elena SF, Lenski RE. 2003. Microbial genetics: evolution experiments with microorganisms: the dynamics and genetic basis of adaptation. *Nat. Rev. Genet.* 4:457–69
32. Elliott P. 2003. Erasmus Darwin, Herbert Spencer, and the origins of the evolutionary worldview in British provincial scientific culture, 1770–1850. *Isis* 94(1):1–29
33. Forterre P. 2013. The common ancestor of Archaea and Eukarya was not an archaeon. *Archaea* 2013:372396
34. Forterre P. 2015. The universal tree of life: an update. *Front. Microbiol.* 6:717
35. Fournier GP, Andam CP, Gogarten JP. 2015. Ancient horizontal gene transfer and the last common ancestors. *BMC Evol. Biol.* 15:70
36. Franklin LR. 2007. Bacteria, sex and systematics. *Phil. Sci.* 74:69–95
37. Fraser C, Hanage W, Spratt BG. 2007. Recombination and the nature of bacterial speciation. *Science* 315:476–80
38. Fullmer M, Soucy S, Gogarten JP. 2015. The pan-genome as a shared genomic resource: mutual cheating, cooperation and the black queen hypothesis. *Fron. MicroBio.* 6:1–5
39. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J.* 8(8):1553–64
40. Godfrey-Smith P. 2009. *Darwinian Populations and Natural Selection*. Oxford: Oxford Univ. Press
41. Godfrey-Smith P. 2015. Reproduction, symbiosis, and the eukaryotic cell. *PNAS* 112(33):10120–25
42. Golicz AA, Batley J, Edwards D. 2106. Towards plant pangenomics. *Plant Biotechnol. J.* 14:1099–1105. doi: 10.1111/pbi.12499
43. Gophna U, Ofran Y. 2011. Lateral acquisition of genes is affected by the friendliness of their products. *PNAS* 108:343–48
44. Gould SJ. 1982. Darwinism and the expansion of evolutionary theory. *Science* 216(4544):380–87
45. Gould SJ, Eldredge N. 1986. Punctuated equilibrium at the third stage. *Syst. Zool.* 35:143–48
46. Gray MW, Doolittle WF. 1982. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* 46:1–42
47. Guttman DS, Dykhuizen DE. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266:1380–83
48. Haig D. 2007. Weismann rules! OK? Epigenetics and the Lamarckian temptation. *Biol. Phil.* 22:415–28
49. Hey J. 2006. On the failure of modern species concepts. *Trends Ecol. Evol.* 21:447–50
50. Hull D. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: Univ. Chicago Press
51. Husnik F, Nikoh N, Koga R, Ross L, Cuncan RP, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567–78
52. Huxley J. 1942. *Evolution: The Modern Synthesis*. London: George Allen Unwin

53. Jablonka E, Lamb MJ. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral and Symbolic Variation in the History of Life*. Cambridge, MA: MIT Press
54. Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *PNAS* 96:3801–6
55. Judson HF. 1996. *The Eighth Day of Creation: The Makers of the Revolution in Biology (Commemorative Edition)*. Cold Spring Harbor, NY: Cold Spring Harb. Press
56. Katz LA. 2015. Recent events dominate interdomain lateral gene transfers between prokaryotes and eukaryotes and, with the exception of endosymbiotic gene transfers, few ancient events persist. *Phil. Trans. R. Soc. B* 370(1678):20140324
57. Kauffman SA. 2000. *Investigations*. Oxford, UK: Oxford Univ. Press
58. Kim KM, Caetano-Anollés G. 2011. The proteomic complexity and rise of the primordial ancestor of diversified life. *Evol. Biol.* 11:140
59. Koonin E. 2009. Darwinian evolution in the light of genomics. *Nucleic Acids Res.* 37(4):1011–34
60. Koonin EV. 2009. The *Origin* at 150: Is a new evolutionary synthesis in sight? *Trends Genet.* 11:473–75
61. Koonin EV. 2015. The turbulent network dynamics of microbial evolution and the statistical tree of life. *J. Mol. Evol.* 80:244–50
62. Koonin EV, Wolf YI. 2009. Is evolution Darwinian and/or Lamarckian? *Biol. Direct.* 4:42
63. Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, et al. 2015. Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524:427–32
64. Kuo CH, Ochman H. 2009. Deletional bias across the three domains of life. *Genome Biol. Evol.* 1:145–52
65. Laland KN, Uller T, Feldman MW, Sterelny K, Müller GB, et al. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc. R. Soc. B* 282:20151019
66. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, et al. 2015. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15:141–61
67. Langton CG. *Artificial Life*. Redwood City, CA: Addison-Wesley
68. Lawrence JG. 2002. Gene transfer in bacteria: speciation without species. *Theor. Pop. Biol.* 61:449–60
69. Lewontin RC. 1970. The units of selection. *Annu. Rev. Ecol. Syst.* 1:1–18
70. Luria SE, Delbrück M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511
71. Maddison WP, Knowles LL. 2006. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* 55:21–30
72. Majewski J, Cohan FM. 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* 153:1525–33
73. Mallet J. 2008. Hybridization, ecological races, and the nature of species: Empirical evidence for the ease of speciation. *Phil. Trans. R. Soc. Lond. B.* 363:2971–86
74. Margulis L. 1991. Symbiogenesis and symbionticism. In *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis*, ed. L Margulis, R Fester, pp. 17–36. Cambridge, MA: MIT Press
75. Martiny AC, Treseder K, Pusch G. 2012. Phylogenetic conservation of functional traits in microorganisms. *ISME J.* 7:830–38
76. Maynard Smith J, Szathmary E. 1995. *The Major Transitions in Evolution*. Oxford, UK: Oxford Univ. Press
77. Mayr E. 1942. *Systematics and the Origin of Species*. New York: Columbia Univ. Press
78. Mayr E. 1996. What is a species and what is not? *Phil. Sci.* 63:262–77
79. Mayr E. 2000. The biological species concept. In *Species Concepts and Phylogenetic Theory*, ed. Q Wheeler, R Meier, pp. 17–29. New York: Columbia Univ. Press
80. McCutcheon JP, Keeling PJ. 2014. Endosymbiosis: protein targeting further erodes the organelle/symbiont distinction. *Curr. Biol.* 24:R654–55
81. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.* 10:13–26
82. McInerney JO, Pisani D, Bapteste E, O'Connell MJ. 2011. The public goods hypothesis for the evolution of life on Earth. *Biol. Direct* 6:41
83. Moore R. 1997. The persuasive Mr. Darwin. *BioScience* 47(2):107–14

84. Morris JJ, Lenski RE, Zinser ER. 2012. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3(2):e00036-12
85. Mushegian AR, Koonin EV. 1996. A minimal genes set for cellular life derived by comparison of complete bacterial genomes. *PNAS* 93(19):10268–73
86. Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr. Biol.* 24:R640–41
87. Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U. 2012. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22:R601–2
88. Nowack EC, Grossman AR. 2012. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *PNAS* 109:5340–45
89. Okasha S. 2006. *Evolution and the Levels of Selection*. Oxford, UK: Clarendon Press
90. O'Malley MA. 2014. *Philosophy of Microbiology*. Cambridge, UK: Cambridge Univ. Press
91. O'Malley MA. 2015. Endosymbiosis and its implications for evolutionary theory. *PNAS* 112:10270–77
92. O'Malley MA, Dupré J. 2007. Size doesn't matter: towards a more inclusive philosophy of biology. *Biol. Phil.* 22:155–91
93. O'Malley MA, Koonin EV. 2011. How stands the Tree of Life a century and a half after *The Origin*? *Biol. Direct.* 6:32
94. Ouzounis C, Kyrpides N. 1996. The emergence of major cellular processes in evolution. *FEBS Lett.* 390(2):119–23
95. Papke RT, Corral P, Ram-Mohan N, de la Haba RR, Sánchez-Porro C, et al. 2015. Horizontal gene transfer, dispersal and haloarchaeal speciation. *Life* 5:1405–26
96. Provine W. 1971. *The Origins of Theoretical Population Genetics*. Chicago: Univ. Chicago Press
97. Puigbò P, Wolf YI, Koonin EV. 2013. Seeing the Tree of Life behind the phylogenetic forest. *BMC Biol.* 11:46
98. Queller D, Strassmann J. 2009. Beyond society: the evolution of organismality. *Phil. Trans. R. Soc. B* 364(1533):3143–55
99. Ranea JA, Sillero A, Thornton JM, Orengo CA. 2006. Protein superfamily evolution and the last universal common ancestor (LUCA). *J. Mol. Evol.* 63(4):513–25
100. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, et al. 2013. Pan genome of the phytoplankton *Emiliana* underpins its global distribution. *Nature* 499:209–13
101. Richerson PJ, Boyd R. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: Univ. Chicago Press
102. Rose MR, Oakley TH. 2007. The new biology: beyond the Modern Synthesis. *Biol. Direct.* 2:30
103. Sachs JL, Hallowell AC. 2012. The origins of cooperative bacterial communities. *mBio* 3(3):e00099-12
104. Sagan L. 1967. On the origin of mitosing cells. *J. Theor. Biol.* 14:255–77
105. Savory F, Leonard G, Richards TA. 2015. The role of horizontal gene transfer in the evolution of oomycetes. *PLOS Pathog.* 11:e1004805
106. Schönknecht G, Weber APM, Lercher MJ. 2015. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays* 36:9–20
107. Shapiro BJ, Polz MF. 2015. Microbial speciation. *Cold Spring Harb. Perspect. Biol.* 7:a018143
108. Simpson GG. 1961. *Principles of Animal Taxonomy*. New York: Columbia Univ. Press
109. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, et al. 2014. Parallel histories of horizontal gene gransfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol. Biol. Evol.* 31:857–71
110. Smith JM, Smith NH, O'Rourke M, Spratt BG. 1993. How clonal are bacteria? *PNAS* 90:4384–88
111. Sniegowski PD, Gerrish PJ, Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387(6634):703–5
112. Sober E, Wilson DS. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard Univ. Press
113. Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16:472–82
114. Stackebrandt E, Goebel BM. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition of bacteriology. *Int. J. Syst. Evol. Microbiol.* 44:846–49

115. Szathmáry E. 2015. Toward major evolutionary transitions theory 2.0. *PNAS* 112(33):10104–11
116. Szöllosi GJ, Davin AA, Tannier E, Daubin V, Boussau B. 2015. Genome-scale phylogenetic analysis finds extensive gene transfer among Fungi. *Phil. Trans. R. Soc. B* 370:20140335
117. Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 38:615–43
118. Timmis J, Ayliffe M, Huang C, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5:123–35
119. Treangen TJ, Rocha EPC. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLOS Genet.* 7(1):e1001284
120. Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* 3:199–208
121. Vos M, Hesselman MC, te Beek TA, van Pssel MWJ, Eyre-Walker A. 2015. Rates of lateral gene transfer in prokaryotes: high but why? *Trends Microbiol.* 23:598–605
122. Wells J. *Icons of Evolution: Science or Myth? Why Much of What We Teach About Evolution Is Wrong*. Washington, DC: Regnery
123. Wernegreen J. 2012. Endosymbiosis. *Curr. Biol.* 22(14):R555–61
124. Wilkins JS. 2009. *Species: A History of the Idea*. Berkeley, CA: Univ. Calif. Press
125. Woese CR. 1982. Archaebacteria and cellular origins: an overview. *Zent. Bakteriol. Mikrobiol. Hyg.* 3(1): 1–7
126. Woese CR. 1998. The universal ancestor. *PNAS* 95(12):6854–59
127. Woese CR. 2000. Interpreting the universal phylogenetic tree. *PNAS* 97(15):8392–96
128. Woese CR. 2002. On the evolution of cells. *PNAS* 99(13):8742–47
129. Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *PNAS* 74(11):5088–90
130. Wolf YI, Koonin EV. 2013. Genome reduction as the dominant mode of evolution. *BioEssays* 35:829–37
131. Zuckerkandl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 97–166. New York: Academic