

Too Much Eukaryote LGT

William F. Martin

The realization that prokaryotes naturally and frequently disperse genes across steep taxonomic boundaries via lateral gene transfer (LGT) gave wings to the idea that eukaryotes might do the same. Eukaryotes do acquire genes from mitochondria and plastids and they do transfer genes during the process of secondary endosymbiosis, the spread of plastids via eukaryotic algal endosymbionts. From those observations it, however, does not follow that eukaryotes transfer genes either in the same ways as prokaryotes do, or to a quantitatively similar degree. An important illustration of the difference is that eukaryotes do not exhibit pangenomes, though prokaryotes do. Eukaryotes reveal no detectable cumulative effects of LGT, though prokaryotes do. A critical analysis suggests that something is deeply amiss with eukaryote LGT theories.

1. Introduction

Few topics in evolutionary biology have caused as much stir in the last 20 years as lateral gene transfer (LGT). In prokaryotes, LGT is the normal means by which DNA is introduced into the cell for recombination. We knew about LGT in prokaryotes long before we had either genomes or phylogenetic trees.^[1] Claims for LGT among eukaryotes essentially did not exist before we had genomes because, in contrast to prokaryotes, there are no characters known among eukaryotes that require LGT in order to explain their distribution, except perhaps the spread of plastids via secondary symbiosis.^[2] Today, claims for eukaryote LGT are common in the literature,^[3] so common that students or nonspecialists might get the impression that there is no difference between prokaryotic and eukaryotic genetics. The time has come where we need to ask whether the many claims for eukaryote LGT – prokaryote to eukaryote LGT and eukaryote to eukaryote LGT – are true.

The reality checks are simple. If the claims are true, then we need to see evidence in eukaryotic genomes for the cumulative effects of LGT over time,^[4] as we see with pangenomes in prokaryotes,^[5] and as we see with sequence divergence. That is, the number of genes acquired by LGT needs to increase in eukaryotic lineages as a function of time. We also need to see evidence for genetic mechanisms that could spread genes across

eukaryote species (and order, and phylum) boundaries, as we see in prokaryotes.^[6] If we do not see the cumulative effects, and if there are no tangible genetic mechanisms, then we have to openly ask why, and entertain the possibility that the claims might not be true. Could it be that eukaryote LGT does not really exist to any significant extent in nature, but is an artefact produced by genome analysis pipelines?


Here, I explore the issue of eukaryote LGT from two angles. I inspect estimates for the prevalence of eukaryote LGT in genome sequence publications and I consider a specific example related to my own work – eukaryotic anaerobes – to illustrate the problem. I will argue that if we demand evidence for cumulative

effects, we will see that many of the claims for eukaryote LGT cannot be true, calling for more common sense, better analyses, and additional reality checks in the eukaryote LGT arena. We have benchmarks for comparison: we know that Darwin's principle of heritable variation produces lineage specific cumulative effects in morphology, and we know for sure that mutations produce lineage specific cumulative effects in the form of sequence divergence. If eukaryote LGT does not produce similar cumulative effects, then something is likely wrong with a fairly large segment of evolutionary literature that is more often the subject of topical reviews^[7–11] than it is the subject of critical inspection.

2. In the Beginning, There Was the Human Genome. . .

Back in the old days before genomes, claims for LGT from prokaryotes to eukaryotes were generally restricted to the literature on endosymbiotic theory and gene acquisitions at the origins of chloroplasts and mitochondria,^[12] something that made good biological sense, both then and now. But then came genomes, first from prokaryotes. By the millennium, the prevalence of LGT in prokaryotic genomes had become tangible in data for those who had always expected it and undeniable for those who had not. Many of the big headlines in early prokaryotic genome papers were the massive amounts of LGT in each genome.^[13–16] Then came the human genome sequence,^[17] and among the main storylines in that milestone paper we could read (many of us in utter disbelief): “Hundreds of human genes appear likely to have resulted from horizontal transfer from bacteria at some point in the vertebrate lineage.” That eyebrow-raising bioinformatics inference did not sit well with cognizant

Prof. W. F. Martin
 University of Düsseldorf
 Universitätsstr. 1, Düsseldorf 40225, Germany
 E-mail: bill@hhu.de

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/bies.201700115>.

DOI: 10.1002/bies.201700115

evolutionary biologists, who quickly showed that differential loss and analysis artifacts were behind the anomalous human genes, not LGT,^[18,19] as recently discussed by Salzberg.^[20]

Subsequent eukaryote genome papers regularly reported the amount or degree of LGT in each genome as a kind of standard parameter. Many genome sequence papers laudably did not report LGT values: they are too numerous to list here. But a number of them did focus on LGT to help bolster the novelty uncovered by the new eukaryote genome sequence, often reported as part of the abstract. Prominent examples include *Trypanosoma*,^[21] *Entamoeba*,^[22] *Dictyostelium*,^[23] *Giardia*,^[24] *Trichomonas*,^[25] *Meloidogyne*,^[26] *Hydra*,^[27] and two algae.^[28,29] In each case, the amount of LGT claimed was usually around 1–2% of all analyzed genes, sometimes more, the source of LGT was almost always bacteria. The more startling report that a tardigrade (an animal) had about 17% LGT in its genome^[3] prompted a second paper for the same tardigrade species^[30] in which several independent animals were sequenced to sort out gut bacteria, epiphyte contamination and the like, and that used different sequencing methods. The second tardigrade report uncovered 40 times less LGT, about 0.4%: most of the originally published 17% tardigrade LGT was bacterial contamination.^[30]

But even in the careful and conservative tardigrade assembly, there still remained 0.4% of the coding sequences in the genome that was scored as LGT.^[30] The tardigrade is not alone. Other animals and other eukaryotes are regularly reported to have 1–2% LGT in their genomes^[31,32] based on what appear to be widely accepted analytical tools. My point is this: If eukaryote LGT is occurring in nature at a rate that leads to an LGT content of 1–2% per genome,^[31,32] where are the cumulative effects? Why does it not accrue along lineages?

A recent state of the art study by Yoshida et al.^[32] helps illustrate the problem. In reporting two new tardigrade genomes, the authors clearly showed how crucial it is to use carefully curated data before publishing estimates for LGT, because the raw data can lead to 10-fold higher LGT estimates. They reported conservative estimates for LGT content in a number of invertebrate genomes, as well as the new tardigrades, using a method that compares BLAST values. I will not address the details of their method here. Salzberg has already done that.^[20] I am not criticizing their paper either: it is an outstanding paper with state of the art methods and results, and LGT is not the main message. But still, each genome investigated turned up about 0.5% LGT, usually more – LGTs that were apparently acquired from a donor residing phylogenetically outside the metazoa. If we plot the per genome LGT estimates from Yoshida et al.^[32] for *Drosophila* onto a phylogeny of the same 12 *Drosophila* genomes from Hahn et al.^[33] we obtain **Figure 1**. The phylogeny is uncontroversial. The timescale is uncontroversial. The amounts of LGT per eukaryote are apparently also uncontroversial, to almost everybody except me, it seems. I find claims of 0.5% LGT per fly genome very difficult to digest.

Why? If the individual per genome LGT estimates reported in^[32] and elsewhere^[21–30] are based on a process that really occurs in nature, common sense demands that there be observable cumulative effects of LGT in eukaryotes. If the LGTs are real – as opposed to being some kind of mass produced artefact in data, analysis, or both – the LGTs need to accrue over

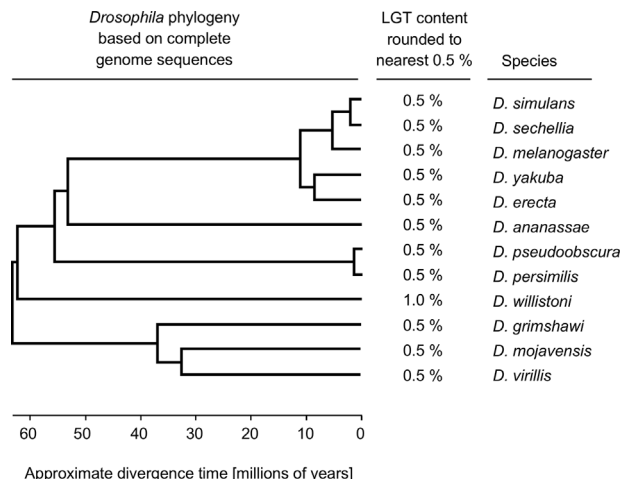


Figure 1. A phylogeny of *Drosophila* genomes with a rough geological timescale, redrawn from Figure 3 of Ref. ^[33] onto which the per genome estimates of LGT in each genome reported in Figure 2 of Ref. ^[32] rounded to the nearest 0.5%, are plotted.

time in each lineage, just like point mutations add up over time to produce sequence divergence among genomes.

This is not my idea: Darwin said it first. Darwin explained how natural variation acting on lineages over time generates the accumulation of lineage specific differences. That was a very important observation. Darwin was thinking about morphology – lineages accruing cumulative morphological differences over time. Today we know that Darwin's principle also works when it comes to mutations accumulating in genome lineages. That is the basis of both the modern synthesis^[34] and molecular evolution.^[35,36] But if LGT is a mechanism of *natural* variation in eukaryotes – that is, really occurring in nature, which its proponents (and the peer review system) are saying – then eukaryote LGT needs to show the same kinds of long-term Darwinian effects as morphology and mutation. Eukaryote LGT needs to accumulate over time, and it needs to show cumulative effects. Are they there? Let's look.

3. No Long-Term Effect of Eukaryote LGT

Figure 1 shows us that the amount of inferred LGT per genome in *Drosophila* is independent of both time and phylogeny. Critics will interject: “Dr. Martin, the method used to infer LGT in those genomes will not exclude the possibility that all the LGTs are present in the *Drosophila* common ancestor.” Yes, I counter, but then there is no evidence for accrual of LGTs in 60 million years of *Drosophila* evolution (the phylogeny actually samples about 300 million years of eukaryote genome evolution), is there? The possible reasons why *D. willistoni* shows 1% LGT versus 0.5% in the rest – rounded to the nearest 0.5% – are immaterial here. Other critics will interject: “Dr. Martin, the method used to infer LGT in those genomes will not exclude the possibility that all those LGTs were present in the eukaryote common ancestor.” Yes, I counter, but then they are more likely the result of differential loss than LGT, aren't they, and therefore have no business being tallied as LGTs.

But let's give those genes the full benefit of doubt as LGTs, and call the 0.5% values in Figure 1 real. A *Drosophila* genome has about 15 000 coding sequences. At 0.5%, that means 75 LGTs per genome. Now let's assume that they entered the fly lineage about 70 million years ago, just before divergence. Rounded, that is a convenient one-new-gene-via-LGT-per-million years, as a rough and conservative estimate on the rate. Recall that if we assume any of those acquisitions occurred during *Drosophila* lineage divergence, the rate just gets higher. For comparison, an old conservative estimate for the *Escherichia coli* rate was 16 kb per million years.^[37,38] "Aha" say LGT proponents, the eukaryote rate is about 20 times lower than that of the prokaryote: everything is fine. No, nothing is fine.

At a rate of one gene via LGT per million years, different lineages of animals that trace to the Cambrian explosion^[39] should have acquired 700 different prokaryotic genes each. Furthermore, different major lineages (supergroups) of eukaryotes are about 1.6 billion years old,^[39] so each supergroup should have accumulated 1600 different prokaryotic genes each because the LGT mechanism should produce lineage specific cumulative effects, just like sequence divergence does. Do we see eukaryotes acquiring new prokaryotic genes in a lineage-specific manner? Do we see a cumulative effect? No.^[40]

The more accurate answer is "No with one big exception." We found no evidence for lineage specific acquisitions in eukaryotes^[40] except at the origin of the plant lineage, where we do see a big influx of cyanobacterial genes that corresponds to the origins of plastids. How does one even look for cumulative effects? First one has to know whether a gene in one lineage is shared with a gene in another, perhaps closely related, lineage; and one has to know that information for all genes and lineages. For that, one has to cluster the genes, and the standard clustering algorithm^[41] does that quite efficiently. If we cluster 956 053 eukaryotic genes from 55 sequenced genomes and then cluster them with their homologues among 6 103 025 genes from 1847 prokaryotic genomes, how many prokaryotic genes are shared by two eukaryotes and prokaryotes? The estimate from that genome sample is 2585. Not 2585 per lineage, or 2585 per supergroup, but rather 2585 in total for 55 eukaryotes. Furthermore, they harbor no evidence for cumulative effects of lineage specific acquisitions, except at plastid origin.^[40] Moreover, the distributions and the phylogenies of those genes, which is reprinted in **Figure 2** with permission, shows upon detailed inspection that they are not lineage specific acquisitions.^[40]

Because no cumulative effects of eukaryote LGT are observed,^[4,40,42] values of the order of 0.5–2% per genome that people have grown accustomed to from genome sequence papers^[21–32] are suspect in my view. At one LGT per million years or even one every 10 million years, eukaryotic supergroups should have accumulated fundamentally different collections of prokaryotic genes because the LGT mechanism should produce lineage specific cumulative effects, just like sequence divergence does and just like morphological change does – changes that undeniably do exist.

How can it be that many reports indicate eukaryote LGT but that there are no cumulative effects? One, perhaps the, crucial factor is that the kinds of eukaryote LGT analyses that we see in the genomics literature are concentric, based on all genes in one genome. We need evolutionary information about all genes in all

genomes. In order to obtain that, one has to cluster all genes and make trees for all clusters from all genomes, not just BLAST, align and make trees for all the proteins from one genome. The clusters also need to be unique so as to avoid counting the same prokaryotic homologues redundantly. That can amount to a lot of work.^[40,43] If the clusters are constructed or analyzed haphazardly or if they contain redundancies, the inferences drawn from them will be erroneous.^[44]

Why should I care about eukaryote LGT anyway? Is not the practical solution to just believe what everyone else does and "get with the programme" as a prominent eukaryote LGT proponent recently recommended that I do (Dan Graur is my witness). At eukaryote genome meetings, where folks pride themselves on the amounts and kinds of LGT they are finding in a particular eukaryote genome (not in all genomes), I feel like Winston Smith in Orwell's novel 1984, listening to an invented truth recited by members of the Inner Party. My mentors taught me that students of the natural sciences are not obliged to get with anyone's program, instead we are supposed to think independently and always to critically inspect, and re-inspect, current premises. Doing "get with the program" science in herds can produce curious effects. For example, the well-managed ENCODE project that ascribed a function to 80% of the human genome was a textbook case of everyone "getting with the program," and everyone, however, also missing the point, obvious to evolutionary biologists, that the headline result of 80% function cannot be true.^[45] Get with the program? Get cumulative effects. Because cumulative effects are not there, per-genome eukaryote LGT values approaching roughly 1% cannot be true. It is too much eukaryote LGT.

4. Case Study: Eukaryote Anaerobes and LGT in Endosymbiotic Theory

There is also another kind of eukaryote LGT out there to discuss, inferred from "unexpected" branching patterns in trees, the topic of this section. Endosymbiotic theory did a good job of explaining the origin of oxygen respiration and oxidative phosphorylation in eukaryotic cells. Mitochondria were once free living bacteria that brought, as endosymbionts, the whole respiratory electron transfer chain, Krebs cycle and biosynthetic pathways for the cofactors involved (quinones, heme and the like) into the eukaryotic lineage. The theory also did a good job accounting for plastids, and the same reasoning applied: the photosynthetic electron transport chain, oxygen synthesis, and chlorophyll entered the eukaryotic lineage via the cyanobacterial antecedent of plastids at the origin of the plant lineage.^[46] Today's organelle genomes are highly reduced in terms of gene content, such that the vast majority of proteins germane to the functions of mitochondria are encoded in the nucleus, having been transferred during the course of evolution from the genomes of the endosymbionts to the chromosomes of their host.^[40] When it comes to core carbon and energy metabolism in eukaryotes, endosymbiotic theory almost covered it all. Almost? Why "almost"?

Eukaryote anaerobes never fit into classical formulations of endosymbiotic theory: Margulis (under her married name at the time, Sagan) initially contended that "all eukaryotic organisms

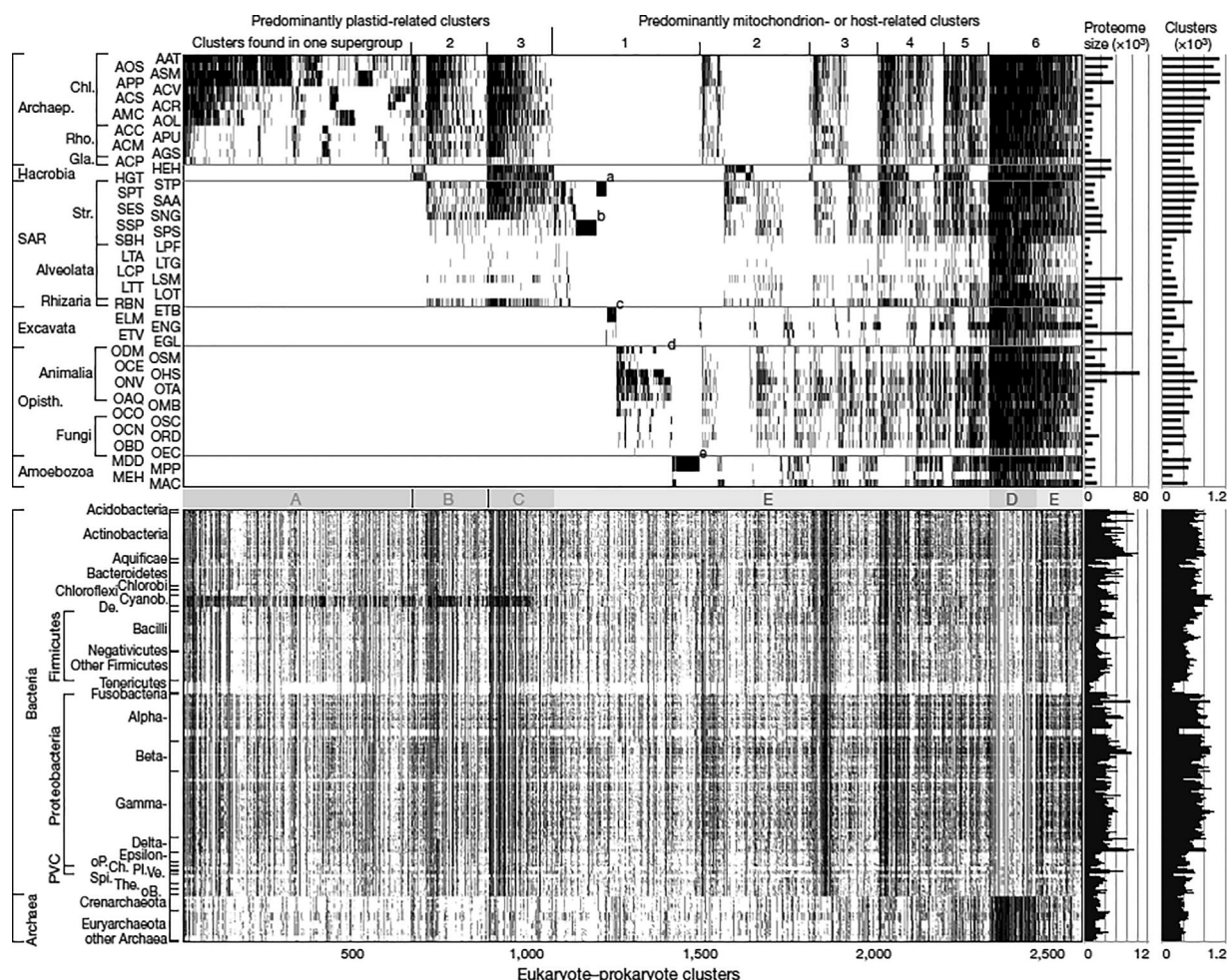


Figure 2. Distribution of 2585 gene families (clusters) occurring in at least two eukaryotic (top panel) and at least five prokaryotic (lower panel) genomes, reprinted from Figure 1 in Ref. [40]. The requirement that the gene needs to be present in at least two eukaryotic genomes was introduced in order to avoid artefacts stemming from bacterial contaminations in sequenced genomes, [40] contaminations being a very common problem in eukaryotic genome sequence data. [30,42] Each black tick in the figure indicates gene presence, white indicates gene absence. Eukaryote supergroup assignments and prokaryotic phyla are given at left. For full details of the figure, see Ku et al. [40] Note the presence of five gene distribution patterns that look very much as if they should represent tip-specific acquisitions from prokaryotes, labeled with lower case a–e in the top panel. However, if the genes in those five “blocks” are lineage specific acquisitions, they should have been acquired more recently in evolution than genes labeled in the block labeled with a capital D (right portion of the figure), which was tested and found not to be the case. [40] Those patterns were generated by differential loss.

contain mitochondria and are fundamentally aerobic” (p. 228), [47] while Gray and Doolittle’s more mainstream formulation of the endosymbiont hypothesis [48] did not mention eukaryote anaerobes at all. There was no place for eukaryotic anaerobes in the endosymbiont hypothesis, which is why the discovery of hydrogenosomes in trichomonads [49] and their subsequent characterization in ciliates [50] as well as other eukaryotes had virtually no impact whatsoever on endosymbiotic theory, at least initially. In the meantime, three proposals have accrued to explain how different lineages of eukaryotes came to possess the ability to survive without oxygen.

Early views had it that eukaryotic anaerobes represented basal branching lineages that diverged in eukaryotic phylogeny before the acquisition of mitochondria. [51,52] The idea that eukaryotic anaerobes are basal was extensively tested using phylogenetic methods, but it failed all tests. [53] The closely allied

idea that eukaryotic anaerobes were phagocytosing heterotrophs before the origin of mitochondria has also been tested and rejected. [54] A second proposal, the hydrogen hypothesis, had it that the common ancestor of mitochondria and hydrogenosomes was a facultative anaerobe that brought not only the respiratory chain but also the enzymes germane to anaerobic energy metabolism and redox balance into the eukaryotic lineage. [55]

In brief, the hydrogen hypothesis proposed that the host for the origin of mitochondria was a H₂-dependent archaeon, that the mitochondrial endosymbiont was a facultative anaerobic proteobacterium that was able to respire like a mitochondrion but was also able to perform H₂-producing fermentations under anaerobic conditions, like *Rhodobacter* or *E. coli* can. (As recently pointed out in these pages, most contemporary proteobacteria and a number of eukaryotes are facultative

anaerobes.^[56] It was the host's dependence upon H_2 – anaerobic syntrophy – that brought host and endosymbiont together at the outset of the symbiotic association, a symbiosis of prokaryotes from which eukaryotic cell complexity emerged.^[54,55,57] Its main predictions have fared well through 20 years of data: eukaryote aerobes and anaerobes should interleave in eukaryote phylogeny,^[53] eukaryotes lacking typical mitochondria should be secondarily amitochondriate,^[58,59] eukaryotic enzymes of anaerobic energy metabolism should trace to the eukaryote common ancestor and to a single bacterial origin,^[53,60] organelle forms intermediate between hydrogenosomes and mitochondria should be found^[61] and the host for the origin of mitochondria should turn out to be an archaeon^[62,63]; even some of the enzymes of anaerobic energy metabolism in hydrogenosomes are turning out to branch with alpha proteobacterial homologues.^[64]

More recently though, a third proposal based on eukaryote LGT (we can call it “lateral late”) has become quite popular.^[65–68] Common to its various formulations is the idea that eukaryotes were ancestrally unable to survive in anaerobic habitats and that the ability to survive anaerobiosis entered the eukaryotic lineage late in evolution (after diversification of the major eukaryotic lineages) via LGT from anaerobic prokaryotes. Because the genes for anaerobic redox balance in eukaryotes tend to reflect a single origin, recent formulations of lateral late entail the idea that the corresponding genes entered the eukaryotic lineage via LGT into one member of a well-diversified eukaryotic domain, and that the genes acquired by that eukaryote were then subsequently passed around to other eukaryotic lineages via a process described as eukaryote-to-eukaryote LGT, enabling gene recipients to “rapidly adapt to anaerobiosis.”^[68] Is laterally late really supported by the data upon which it rests? Closer inspection uncovers problems with the recourse to eukaryote LGT as a tool to explain unexpected branching patterns in phylogenetic trees.

4.1. Problems With Eukaryote Anaerobe LGT

Taking a specific case, Figure 4 of Leger et al.^[68] which is redrawn here in **Figure 3** solely for illustrative purposes, shows a phylogeny for an crucial enzyme of anaerobic energy metabolism in eukaryotes called pyruvate:ferredoxin oxidoreductase (PFO),^[55,69] and a fusion variant of PFO called pyruvate:NAD⁺ oxidoreductase (PNO),^[70] from various lineages including archaeplastida, alveolates, and excavates, in addition to the very interesting newly characterized anaerobes *Pygsuia biforma* and *Stygliella incarcerata*. PFO and PNO are oxygen sensitive enzymes that represent an alternative to pyruvate dehydrogenase for the oxidative decarboxylation of pyruvate in mitochondria, plastids, and the cytosol.^[60] Let us assume that their interpretation is correct, namely that PFO entered the eukaryotic lineage via lateral acquisition long after mitochondria arose and was then distributed among diverse eukaryotic lineages via LGT.^[68] Were that true, the trees for PFO and PNO would show nested phylogenies. How so?

Eukaryote to eukaryote LGT creates phyletic patterns in which the eukaryotic lineages branch in a nested manner, recipients branching within donors. This is not just a claim, it is an observation from established and uncontroversial cases of phylogenies involving eukaryote-to-eukaryote gene transfer: secondary endosymbiotic spread of plastids.^[71,72] In secondary endosymbiosis, there exist clear and well-known examples for such nested phylogenies that are furthermore independently corroborated by the presence of novel organelles (secondary plastids) in the recipient lineages. For example, the symbiotic origin of red secondary plastids generates trees in which plastid-derived genes of diatoms branch nested within red algae,^[73] and the symbiotic origin of green secondary plastids generates trees in which plastid-derived genes of chlorarachniophytes branch nested within chlorophytes.^[74] There are many such examples in the literature.^[71,72] Because secondary endosymbiosis also

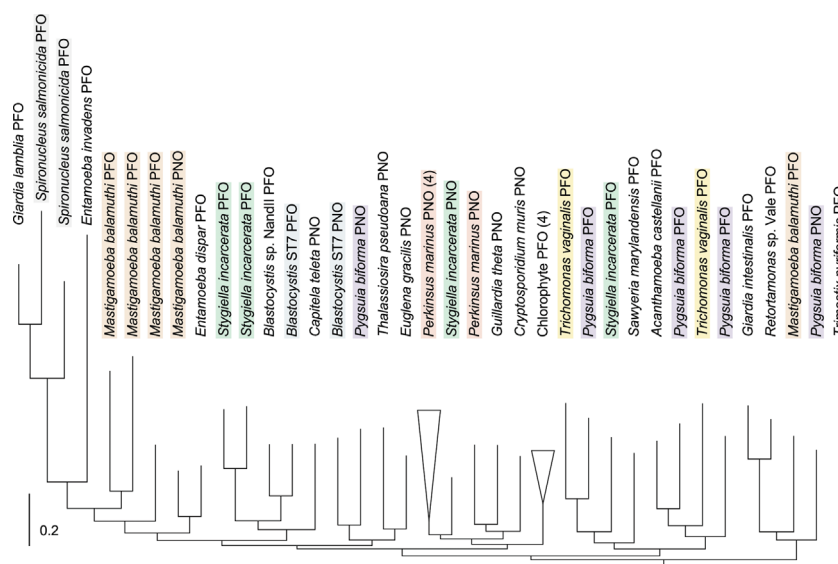


Figure 3. The phylogeny of PFO and PNO genes published as Figure 4 of Leger et al.^[68] redrawn here and included at the request of the editor and referees (see text).

entails LGT from eukaryotes to eukaryotes in order to spread an organelle with photosynthetic physiology across lineages, it generates nested phylogenies in which recipient eukaryotic lineages branch within the donor eukaryotic lineage. Do the trees in Leger et al.^[68] reveal nested phylogenies? No, their PFO phylogeny has, inter alia, *Trimastix* in a basal position, followed by a series of branches that harbor five highly divergent copies of *Pygsuia* PFO and *Pygsuia* PNO (a fusion protein), four highly divergent copies of *Stygliella* PFO and *Stygliella* PNO, five highly divergent copies of *Mastigamoeba* PFO and PNO, two highly divergent copies of *Giardia* PFO, two highly divergent copies of *Trichomonas* PFO, all branching with homologues from diverse clades.

The eukaryotic PFO sequences are monophyletic, which would suggest a single origin. But there is no nesting of eukaryotic lineages, which eukaryote-to-eukaryote LGT (lateral late) would generate. Rather the PFO phylogeny indicates at face value, as Stairs et al.^[65] state, that “early gene duplications within eukaryotes followed by differential loss is conceivable and would explain the odd phylogenetic patterns observed” because several “of the eukaryotes examined so far have retained multiple putative ancient paralogs expected under such a scenario” (quoted passage from^[65] p. 2095, in the context of interpreting a different gene phylogeny). The *Pygsuia* and *Stygliella* sequences^[68] present the patterns expected for ancient gene duplications and differential loss that Stairs et al.^[65] describe, but when such patterns are actually observed, the same team of authors interprets them as evidence for eukaryote-to-eukaryote LGT.^[68]

4.2. Donating to the Rich

The central conceptual pillar of the LGT theory for eukaryotic anaerobes is that lateral gene acquisition confers access to a new ecological niche. For access to anaerobic environments, all other genes required for anaerobic redox balance would also have to be simultaneously acquired, because a gene for a single subunit in a multienzyme pathway of redox balance is a useless acquisition.^[75] That is why Leger et al.^[68] assume the existence of a transferred “module.” That is cumbersome, but it is not the main problem.

The main problem concerns the five different and highly divergent PFO copies of *Pygsuia* and *Mastigamoeba* (and four in *Stygliella*, etc.). Under the interpretations of Leger et al.,^[68] the deeply divergent multiple copies of PFO in *Pygsuia* and *Stygliella* (and *Mastigamoeba* and *Giardia*, etc.) indicate that each of those recipient lineages underwent multiple independent LGT acquisition from different eukaryotic donors while already in possession of preexisting active PFO genes, because the multiple *Pygsuia* (and other eukaryote) PFO copies do not branch together as recent duplicates within the same genome.

Yet multiple independent acquisitions of the same gene directly contradict the central pillar of the LGT theory,^[65–68] namely that eukaryotes acquired genes such as PFO in order to survive in and/or colonize anaerobic environments. How so? LGT for access to anaerobic environments can only be invoked to account for the origin of the first copy. A lineage in possession of one PFO gene can deal with anaerobiosis at the reaction catalyzed by PFO. Adaptation or novel niche access cannot be

invoked to account for the three to four additional putative LGTs per lineage, however, because the adaptation exists and the niche is already colonized. This simple observation in a simple gene family tree deflates LGT theories for eukaryote anaerobe origin that invoke fixation of acquired genes based on adaptive value.

Finally, if enzymes for anaerobic energy metabolism were being distributed among eukaryotic lineages as a “module,” then the question of genetic mechanisms emerges. LGT proponents are quick to change the subject when it comes to mechanisms and eukaryote genetics. (For gene transfers from organelles to the nucleus, the mechanism is known: it is non-homologous end joining.^[76,77]) What kind of genetics would spread the module – which would have to consist of about a dozen or so genes at least^[60] – across eukaryotic supergroups? In prokaryotes the situation is simple: the corresponding genes could be organized as an operon, then copied onto a plasmid and passed around via conjugation, like photosynthesis (ca. 100 genes) is among members of the *Roseobacter* group of α -proteobacteria.^[78] But the eukaryotes studied so far (the ones from which LGT claims stem) do not use sex pili or plasmids to distribute operons or single genes via LGT.

In the absence of plasmids, the only other putative genetics-based mechanism that remains is trans-supergroup hybridization: The eukaryotes that Leger et al.^[68] posit to have been involved in gene transfers (all of the eukaryotic lineages in their trees) could have undergone some kind of interkingdom cell fusion or interkingdom gamete fusion followed by nuclear fusion in order to get the chromosomes into contact so that the module of genes for anaerobic redox balance in question could enter the recipient lineage. Then, in order to incorporate the genes into the genome, two possibilities can be imagined. Perhaps illegitimate recombination via double crossover targeting only anaerobiosis genes could occur specifically with the recipient genome so that only the donor anaerobiosis genes are incorporated, the remaining 10 000 genes of the donor being specifically degraded by a mechanism yet to be discovered. Alternatively, perhaps the fusion was wholesale (as in interspecific hybrids) such that the ensuing hybrids were facultative anaerobes with two kinds of genomes in the nucleus, and two very different kinds of mitochondria in the cytosol, as it occurs in cybrids.^[79,80] The “cybrid” would however have to sort itself out each time such that, mysteriously, only the module of anaerobiosis genes remained from the fusion, the remainder of the genome adhering to the paradigm of vertical eukaryotic lineage evolution as multigene studies of eukaryote evolution indicate. Breeders would rejoice to have such tools for targeted multi-trait transkingdom genetics at their disposal.

If transkingdom genetics between plants, animals, fungi, and protists is really going on in nature as Leger et al.^[68] suggest, why have 100 years of plant, animal, and yeast breeding never discovered evidence for its existence? It is either hidden somehow from our realm of observation or it does not exist in nature. Looking at the matter openly, the claims that eukaryotes acquire genes from anaerobes and then inherit them so that progeny may survive in anaerobic habitats are completely Lamarckian in the terms of what biologists today associate with Lamarck, regardless of whether it was Lamarck’s intent or not.^[81] To be fair, I too am saying that there are gene acquisitions in eukaryote evolution, but I take the minimalist (in terms of

frequency) stance on acquisition by saying that gene acquisitions coincide with two undeniable events of acquisition – the origin of mitochondria and chloroplasts^[40] – and that Darwinian selection operated on the natural variation introduced by those events. My position thus contrasts sharply to the “acquire as needed” theory of the eukaryote LGT camp in this debate.

5. Too Much Eukaryote LGT

As with the per-genome LGT estimates discussed at the outset of this paper, the recent interpretations of Leger et al.^[68] and Stairs et al.^[65–67] exemplify a popular current trend in phylogenetic reasoning and data interpretation^[3,9,82] that is far too implicit and needs to be spelled out more clearly by its proponents. Eukaryote LGT is never penalized: that is, LGT proponents freely assume that it occurs wherever convenient to fit a given data pattern. In some investigations, it might even be the first line of explanation for unexpected branching patterns in trees rather than the last resort. It can be assumed to occur among eukaryotes even though no characterized genetic or molecular mechanisms as vehicles for such peculiar inheritance are known, and – uniquely among evolutionary mechanisms – it need not produce cumulative effects over time.

It is apparent neither to me nor to others^[20,34,83,84] why eukaryote LGT should be preferred over both vertical inheritance and gene loss as the first line of interpretation for eukaryote gene trees, while less spectacular mechanisms that conform to the rules of eukaryote genetics such as vertical inheritance within species, gene duplication, and differential loss are implicitly penalized as unlikely processes. Why should gene duplication and gene loss in addition to the ever-present existence of random phylogenetic error be viewed as unlikely causes for unexpected database search results or unexpected branches in trees?

Next to mutation,^[35] gene duplication is probably the most normal process known in eukaryotic genome evolution biology,^[85] and is exacerbated by the pervasive prevalence of whole genome duplication in eukaryotic genome evolution.^[86] Differential loss, also known as reductive evolution, is one of the most prominent underlying themes of genome evolution, whether prokaryotic^[87] or eukaryotic.^[88,89]

LGT admittedly provides a much more colorful story for unexpected branches in phylogenetic trees of eukaryotic genes than duplication and loss.^[7,82] LGT interpretations generate interesting and unusual narratives for high visibility papers about otherwise soberingly black and white genome data. Has a segment of the field studying genome evolution subordinated accurate depictions of evolutionary history to LGT sensationalism? Current studies among fungi examine many trees and find many unexpected branches, which are interpreted as evidence for widespread eukaryote LGT,^[90,91] whereby, those studies take the existence of eukaryote LGT as a given and employ parameters to estimate its magnitude. Nearly every phylogenetic tree ever constructed contains unexpected branches. But is every unexpected branch in eukaryote phylogeny evidence for LGT? That is what LGT proponents are having us believe.

I am not questioning the obvious and long-known existence of LGT among prokaryotes,^[1] nor am I questioning gene acquisitions by eukaryotes during endosymbiosis.^[40] Nor am I

saying that LGT in prokaryotes is Lamarckian: LGT in prokaryotes is natural variation upon which natural selection may act. I am questioning the claims for LGT among eukaryotes based solely in BLAST searches or single gene phylogenies, because at some point the numbers need to add up. If there is as much LGT between eukaryotes going on in evolution as proponents claim in their reviews,^[9–11] then we need to see cumulative effects – not cumulative effects in the literature, cumulative effects in nature.

5.1. Looking for Cumulative Effects of Eukaryote LGT

Is anybody even looking for cumulative effects? My group has recently looked, in two ways. First, we looked to see if there is evidence in data from 55 sequenced eukaryotic genomes (Figure 2) to support claims for the widespread occurrence of eukaryote-to-eukaryote LGT.^[40] We asked using straightforward statistical methods whether eukaryote genes that have readily detectable prokaryotic homologues and that generate trees recovering eukaryote monophyly produce sets of eukaryotic topologies that significantly differ from the topologies generated by eukaryote specific genes.^[40] The answer was “no”: by the measure of phylogenetic trees, eukaryotic genes with prokaryotic homologues are inherited just as vertically among eukaryotes as eukaryote specific genes, patchy distributions being attributable to differential loss, not to eukaryote-eukaryote LGT.^[40]

Second, we looked to see if genome wide phylogenies provide any evidence for recent LGT from prokaryotes to eukaryotes.^[42] Our test there was also straightforward. If eukaryotes are acquiring genes from prokaryotes via LGT in the same manner, at a similar rate, or to a comparable degree, as prokaryotes are acquiring genes from other (distantly related) prokaryotes, then eukaryote genomes should harbor evidence for recent prokaryote-to-eukaryote LGTs, just like prokaryotes genomes harbor evidence for recent LGTs from different prokaryotic phyla.^[42] That is, eukaryote genomes should harbor genes that are nearly identical in sequence to genes in prokaryotic genomes, just like prokaryote genomes harbor genes that are nearly identical in sequence to genes from other phyla. Do they? Again the answer was “no”.^[42] While prokaryote genomes are replete with genes recently acquired from distant prokaryote phyla, eukaryote genomes are devoid of such genes.^[42] This indicates 1) that many or most claims for prokaryote to eukaryote LGT (outside the context of organelle origins) are just genome annotation contaminations – or over-interpreted phylogenetic trees^[92] – and 2) that there exists a biological barrier in nature to LGT from prokaryotes to eukaryotes.^[42]

5.2. Jumping to Eukaryote LGT Conclusions

Gene duplication, genome duplication, differential loss, alignment errors,^[36] contamination, and annotation problems^[30] as well as simple phylogenetic errors based in the non-uniformity of the evolutionary process of sequence change across lineages^[93] explain unexpected branching patterns in trees that are currently interpreted as eukaryote LGT. Such normal mechanisms fit well with the view that eukaryotes generate

and inherit new combinations of genes via standard eukaryote genetics: meiotic recombination, gamete formation, gamete fusion, karyogamy and gene dynamics in populations.^[34] The once-every-billion-years exception is endosymbiosis, when cellular lineages and whole genomes merge during eukaryotic evolution to generate organelles bounded by two or more membranes. Endosymbiosis is rare in eukaryote evolution,^[54] but when it occurs, major transitions, gene transfer from organelles and the origin of novel taxa at the highest levels are the result.^[4,40,72,94]

Should LGT be the new default explanation for “unexpected branches” in eukaryotic phylogenetic trees? We can easily remedy the “unexpected branch problem” with measures less dire than LGT, for example, by assuming that occasional phylogeny artefacts are a normal and unavoidable component of phylogenetic reconstruction. Part of the problem is that trees with unexpected branches can readily be published as evidence for eukaryote LGT in many journals, whereas, almost nobody has an interest in publishing a phylogeny artefact declared as such, unless it is in the context of debunking some prominently published LGT claims.^[92] We should also recall that the ancestors of plastids and mitochondria were just normal prokaryotes with normal pangenomes and that gene loss is prevalent in eukaryote evolution.^[4,38,40,42] Those two remedies alone would reclassify almost all of the odd branches that underlie phylogeny based claims for eukaryote LGT^[7–10,31] into normal gene acquisitions from organelles,^[4,40] a known and ongoing process^[94] with known and observable mechanisms.^[76,77]

The phycologist Robert E. Lee^[95] summarized it well 50 years ago when he wrote (concerning the distribution of plastids among eukaryotes): “Any evolutionary scheme should adhere to the following three principles. 1) A monophyletic origin of any organism, chemical compound or cytoplasmic structure has the greatest statistical probability of being correct. 2) The loss of a non-essential structure can require just the mutation of a single gene but the acquisition of a structure generally requires many mutations and a considerable amount of time. 3) Most organisms in evolutionary sequences would have been lost, yet in postulating phylogenetic events the plausibility of the theory can be enhanced by the existence of organisms similar to those in the proposed scheme.”^[95] (p. 44). Lee’s rules seem more important today than ever before.

6. How Did We Get Here Anyway?

Literature dealing with the possibility of eukaryote LGT traces back more than 30 years^[12,38,96], into the days before LGT was popularized by Doolittle.^[97] Many of the eukaryote LGT reports emerged from phylogenetic pipeline analyses of genome data. I confess that genome-scale phylogenetics pipelines came out of my laboratory.^[73] When we made hundreds and thousands of trees to investigate the question of how many genes in plants come from cyanobacteria, we found that about 18% of *Arabidopsis* genes have readily detectable homologues in sequenced prokaryote genomes and yeast entered the plant lineage via the plastid.^[73] But we also found many odd phylogenies that “would suggest at face value that the *Arabidopsis* lineage acquired genes from all organisms sampled in this

study,” whereby, we continued in the same breath that “such interpretations can hardly be true”, inter alia, because “lateral gene transfer between free-living prokaryotes occurs to a great extent.”^[73]

Other groups followed the mass phylogenetics approach to analyzing genomes,^[21–31] but heeded no warnings, taking the odd trees where genes branch “unexpectedly” at face value: that is, as evidence suited to detecting and measuring eukaryote LGT.^[21–31,65–68] There are many reasons not to interpret trees that way.^[34–36,93] I will only list three: 1) phylogenetics is much better at testing hypotheses than at generating hypotheses out of sequences from scratch, 2) it has always been obvious (to me anyway) that LGT among prokaryotes figures into the inference of eukaryote gene origin,^[4,38,40,42,73,94] and 3) some small proportion of trees will always contain odd branches but not because of LGT. We know that single gene trees for proteins that share a common history can differ widely,^[98] not because of LGT, but because phylogeny is an imperfect art.^[36,99] If we look at phylogenetic trees for thousands of different proteins, each tree having hundreds of branches each, we will find many unexpected branches. For a tree with 60 sequences there are more than 10^{80} possible trees, 10^{80} being the number of protons in the universe, roughly. In addition to the 57 “expected” branches among those 10^{80} possible trees there are $2^{59}–61–57$ ($=5.8 \times 10^{17}$) unexpected branches; in the search for a tree of 60 sequences, unexpected branches outnumber expected ones by a factor of 10^{16} . A ratio of 1 in 10^{16} exceeds the accuracy of today’s best clocks, which miss one second every 30 million years. Even if phylogeny were error-free, computers (and scientists) still have only finite time. We should expect to find unexpected branches. Phylogenetic pipelines were never intended for mass production of eukaryote LGT papers, which explains why the use of single genome phylogeny pipelines,^[21–31] like single genome BLAST comparisons,^[30,32] leads to estimates of LGT that do not add up.

And what about tree-independent evidence, “loner” sequences that are present only in prokaryotes and in one eukaryote or one tip eukaryote lineage, and that are not annotation artifacts? Gene loss can – and will – generate such patterns. Gene loss is very widespread in eukaryotes.^[40,42,89] If a gene can be lost in one lineage it can be lost in others, and when gene loss is at work, the last lineage to lose the gene appears unexpectedly alone, because “the last one out looks like an LGT,”^[42] as in the case of gene blocks a–e labeled in Figure 2.

7. What About Selection?

Harsh critics of the views expressed here (there will doubtless be many) will complain that I am underestimating the power of natural selection to fix rare variants created by LGT. Beyond the issue of lacking evidence for any cumulative effects,^[40,42] beyond the issue of Lamarckian evolution of eukaryotes acquiring traits from the environment, and couching the issue in the more general terms of the modern synthesis,^[34] where is the adaptive value of eukaryote LGT in the bigger picture of the evolutionary process? Where would we even look? Our focus would obviously be on physiological traits, clearly, because prokaryotes have nothing to offer eukaryotes in terms of genes that govern morphology, development, or behavior.^[54,100] Cytochrome *bd*

terminal oxidases would be a prime target for prokaryote to eukaryote LGT because they are common and readily transferred among prokaryotes,^[101] and because they are sulfide resistant,^[102] allowing use of O₂ as a terminal acceptor in sulfidic environments, which abound today and were even more abundant in the distant eukaryotic past.^[103] But eukaryotes show no acquisitions of such selectively useful *cyt bd* oxidase genes. Why not? No interest in a spectacularly useful adaptation? Or is there really a natural barrier to LGT from prokaryotes to eukaryotes.^[42]

The endosymbiotic bacteria of insects also offer opportunity for adaptive acquisition of traits from the environment.^[104] *Buchnera aphidicola* lives within a special organ of aphids, the bacteriome, and has been vertically transmitted within the aphid lineage for perhaps 250 million years.^[105] *Buchnera* supplies the insect with essential amino acids and receives non-essential amino acids from the animal in return.^[104] It turns out that many insects have such endosymbionts, which typically provide amino acids, though they sometimes provide vitamins.^[106] There have even been incorporations of whole *Wolbachia* genomes to insect nuclear DNA,^[107] but without corresponding acquisition of bacterial traits. The aphid genome curiously revealed almost no transfers, and none for amino acid biosynthesis.^[108] The insects have had every opportunity over 250 million years to become prototrophic for amino acids (capable of amino acid biosynthesis) by merely acquiring the bacterial genes (and vice versa). What a huge adaptive advantage amino acid prototrophy would confer. It would allow the insects to colonize a plethora of new amino acid poor niches – immense opportunity for speciation. The insect endosymbionts offer a splendid opportunity for eukaryotes to acquire hugely beneficial genes, a smorgasbord of opportunity for adaptive evolution. The crucial point is that despite all that opportunity no such adaptively advantageous transfer is observed.

8. Conclusion

Before the age of genomes, there were no traits in eukaryotes that required eukaryote-to-eukaryote LGT to account for their evolution or distribution to begin with, except secondary plastids.^[95] Nor were there traits that required prokaryote-to-eukaryote LGT, except the origins of chloroplasts and mitochondria.^[46] That raises the issue of what eukaryote LGT actually explained in the first place, beyond curious patterns of sequence similarity from which it was inferred. I am not saying that LGT to eukaryotes never, ever, ever occurs outside the context of symbiosis. Biology is a science of exceptions. Transposable elements are such an exception because if they enter the genome – by whatever mechanism, known or not – they can multiply across chromosomes and become fixed rapidly. I am biased regarding transposons because my PhD mentor discovered transposons in prokaryotes, the IS elements of *E. coli*,^[109] hence I “grew up” knowing that mobile DNA can multiply, spread and become fixed in genomes not because it is selected, useful or neutral, but because it is genomically infectious at rates that dwarf the standard dynamics of mutations in populations. The promiscuity of transposable elements was evident to conservative classical geneticists, and it was evident early on that P-elements in *Drosophila* had some ability to spread.^[110] Recent

reports of massive transposable element spread during insect genome evolution are hence not only completely credible,^[111] they make excellent biological sense.

The reports whose truth I am doubting here do not concern eukaryotic LGT of transposable elements, they concern eukaryotic LGT of normal protein coding genes. If eukaryote LGT is real as opposed to being an artefact, the LGTs need to accrue, just like point mutations add up over time to generate sequence divergence among genomes. When we look for cumulative effects of LGT in prokaryotes, where LGT is real, we see them.^[112] When we look for pangenomes in prokaryotes, where LGT is real, we see them.^[4,5,113] When we look for pangenomes in eukaryotes, cumulative effects of LGT from prokaryotes or LGT from other eukaryotes into eukaryotic genomes, they are not there. Transposable elements are an exception. The idea of LGT in eukaryotes has led to suggestions that genes are transferred between eukaryotes via meteorites,^[114] that LGT occurs via organisms touching one another,^[8] or that tardigrade genomes consist to 17% out of genes acquired via recent LGT.^[3] Far less spectacular results reporting about 0.5% LGT for the same tardigrade^[30] tend more accurately to reflect nature's workings, but 0.5% is still way too much eukaryote LGT. Even the highly touted observations suggesting LGT in rotifers have found more sobering explanations,^[115] while LGT claims in schistosomes fail under critical inspection.^[116] In the face of pressure to publish evolutionary insights from genomic investigations – symptomatically when no other storyline is readily found in a given set of genome data – have evolutionary biologists become uncritical? We need to remain critical, perhaps more attentive than ever before. The LGT numbers in eukaryotes do not add up. There is something wrong with eukaryote LGT theories.

Acknowledgements

I thank Dan Graur, Steven Salzberg, Olivia Judson, Murray Cox, Giddy Landan, Mike Steel, Lilli Martin, Sriram Garg, Verena Zimorski, Howard Ochman, and especially Sven Gould for discussions and comments. I thank the ERC (666053), the GIF (I-1321-203.12/2015), and the Volkswagen Stiftung (Life) for financial support.

Conflict of Interest

The author has declared no conflict of interest.

Keywords

lateral gene transfer, horizontal gene transfer, phylogenetic artefact, genome analysis, Lamarckian evolution

Received: June 30, 2017
Revised: September 26, 2017
Published online:

- [1] D. Jones, P. H. A. Sneath, *Bacteriol. Rev.* **1970**, *34*, 40.
- [2] R. E. Lee, *S. Afr. J. Sci.* **1977**, *73*, 179.
- [3] T. C. Boothby, J. R. Tenlen, F. W. Smith, J. R. Wang, K. A. Patanella, E. O. Nishimura, S. C. Tintori, Q. Li, C. D. Jones, M. Yandell,

- D. N. Messina, J. Glasscock, B. Goldstein, *Proc. Natl. Acad. Sci. USA* **2016**, *112*, 15976.
- [4] C. Ku, S. Nelson-Sathi, M. Roettger, S. Garg, E. Hazkani-Covo, W. F. Martin, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10139.
 - [5] D. Medini, C. Donati, H. Tettelin, V. Massignani, R. Rappuoli, *Curr. Opin. Genet. Dev.* **2005**, *15*, 589.
 - [6] O. Popa, T. Dagan, *Curr. Opin. Microbiol.* **2011**, *14*, 615.
 - [7] J. O. Andersson, *Cell. Mol. Life Sci.* **2005**, *62*, 1182.
 - [8] P. J. Keeling, J. D. Palmer, *Nat. Rev. Genet.* **2008**, *9*, 605.
 - [9] J. Huang, *Bioessays* **2013**, *35*, 868.
 - [10] L. Boto, *Proc. Biol. Sci.* **2014**, *281*, 20132450.
 - [11] R. Bock, *Annu. Rev. Genet.* **2017**, *51*, [In press]. <https://doi.org/10.1146/annurev-genet-120215-03532>
 - [12] W. Martin, R. Cerff, *Eur. J. Biochem.* **1986**, *159*, 323.
 - [13] F. R. Blattner, G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, Y. Shao, *Science* **1997**, *277*, 1453.
 - [14] N. T. Perna, G. Plunkett, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Pósfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamouisis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, F. R. Blattner, *Nature* **2001**, *409*, 529.
 - [15] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Cordani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denzot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppe, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Togoni, V. Tosato, S. Uchiyama, M. Vandebol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H. F. Yoshikawa, E. Zumstein, H. Yoshikawa, A. Danchin, *Nature* **1997**, *390*, 249.
 - [16] K. E. Nelson, R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, C. M. Fraser, *Nature* **1999**, *399*, 323.
 - [17] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczyk, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendt, K. D. Delehaunty, T. L. Miner, R. S. Kucherlapati, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowski, International Human Genome Sequencing Consortium, *Nature* **2001**, *409*, 860.
 - [18] S. L. Salzberg, O. White, J. Peterson, J. A. Eisen, *Science* **2001**, *292*, 1903.
 - [19] M. J. Stanhope, A. Lupas, M. J. Italia, K. K. Koretke, C. Volker, J. R. Brown, *Nature* **2001**, *411*, 940.
 - [20] S. L. Salzberg, *Genome Biol.* **2017**, *18*, 85.
 - [21] M. Berriman, E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Böhme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. M. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivins, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. MacLeod, P. J. Mooney, S. Moule, D. M.

- A. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M. A. Quail, E. Rabinowitsch, M.-A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. R. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. Suh, M. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, S. E. Melville, N. M. El-Sayed, *Science* **2005**, 309, 416.
- [22] B. Loftus, I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillén, C. Gilchrist, S.E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri Jr, C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser, N. Hall, *Nature* **2005**, 433, 865.
- [23] L. Eichinger, J. A. Pachebat, G. Glöckner, M. A. Rajandream, R. Sugang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Louseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, A. Kuspa, *Nature* **2005**, 435, 43.
- [24] H. G. Morrison, A. G. McArthur, F. D. Gillin, S. B. Aley, R. D. Adam, G. J. Olsen, A. A. Best, W.Z. Cande, F. Chen, M. J. Cipriano, B. J. Davids, S. C. Dawson, H. G. Elmendorf, A. B. Hehl, M. E. Holder, S. M. Huse, U. U. Kim, E. Lasek-Nesselquist, G. Manning, A. Nigam, J. E. Nixon, D. Palm, N.E. Passamaneck, A. Prabhu, C. I. Reich, D. S. Reiner, J. Samuelson, S. G. Svard, M. L. Sogin, *Science* **2007**, 317, 1921.
- [25] J. M. Carlton, R. P. Hirt, J. C. Silva, A. L. Delcher, M. Schatz, Q. Zhao, J. R. Wortman, S. L. Bidwell, U. C. Alsmark, S. Besteiro, T. Sicheritz-Ponten, C. J. Noel, J. B. Dacks, P. G. Foster, C. Simillion, Y. Van de Peer, D. Miranda-Saavedra, G. J. Barton, G. D. Westrop, S. Müller, D. Dessi, P. L. Fiori, Q. Ren, I. Paulsen, H. Zhang, F. D. Bastida-Corcuera, A. Simoes-Barbosa, M. T. Brown, R. D. Hayes, M. Mukherjee, C. Y. Okumura, R. Schneider, A. J. Smith, S. Vanacova, M. Villalvazo, B. J. Haas, M. Peretea, T. V. Feldblyum, T. R. Utterback, C.-L. Shu, K. Osoegawa, P. J. de Jong, I. Hrdy, L. Horvathova, Z. Zubacova, P. Dolezal, S.-B. Malik, J. M. Logsdon Jr, K. Henze, A. Gupta, C.C. Wang, R. L. Dunne, J. A. Upcroft, P. Upcroft, O. White, S. L. Salzberg, P. Tang, C.-H. Chiu, Y.-S. Lee, T. M. Embley, G. H. Coombs, J. C. Mottram, J. Tachezy, C. M. Fraser-Liggett, P. J. Johnson, *Science* **2007**, 315, 207.
- [26] P. Abad, J. Gouzy, P. Abad, J. Gouzy, J. M. Aury, P. Castagnone-Sereno, E. G. Danchin, E. Deleury, L. Perfus-Barbeoch, V. Anthouard, F. Artiguenave, V. C. Blok, M. C. Caillaud, P. M. Coutinho, C. Dasilva, F. De Luca, F. Deau, M. Esquibet, T. Flutre, J. V. Goldstone, N. Hamamouch, T. Hewezi, O. Jaillon, C. Jubin, P. Leonetti, M. Magliano, T. R. Maier, G. V. Markov, P. McVeigh, G. Pesole, J. Poulain, M. Robinson-Rechavi, E. Sallet, B. Séguens, D. Steinbach, T. Tytgat, E. Ugarte, C. van Ghelder, P. Veronico, T. J. Baum, M. Blaxter, T. Blevé-Zacheo, E. L. Davis, J. J. Ewbank, B. Favery, E. Grenier, B. Henrissat, J. T. Jones, V. Laudet, A. G. Maule, H. Quesneville, M.-N. Rosso, T. Schiex, G. Smant, J. Weissenbach, P. Wincker, *Nat. Biotechnol.* **2008**, 26, 909.
- [27] J. A. Chapman, E. F. Kirkness, O. Simakov, S. E. Hampson, T. Mitros, T. Weinmaier, T. Rattei, P. G. Balasubramanian, J. Borman, D. Busam, K. Disbennett, C. Pfannkoch, N. Sumin, G. G. Sutton, L. D. Viswanathan, B. Walenz, D. M. Goodstein, U. Hellsten, T. Kawashima, S. E. Prochnik, N. H. Putnam, S. Shu, B. Blumberg, C. E. Dana, L. Gee, D.F. Kibler, L. Law, D. Lindgens, D.E. Martinez, J. Peng, P. A. Wigge, B. Bertulat, C. Cuder, Y. Nakamura, S. Ozbek, H. Watanabe, K. Khalturin, G. Hemmrich, A. Franke, R. Augustin, S. Fraune, E. Hayakawa, S. Hayakawa, M. Hirose, J. Shan Hwang, K. Ikeo, C. Nishimiya-Fujisawa, A. Ogura, T. Takahashi, P. R. H. Steinmetz, X. Zhang, R. Aufschnaiter, M.-K. Eder, A.-K. Gorny, W. Salvenmoser, A. M. Heimberg, B. M. Wheeler, K.J. Peterson, A. Böttger, P. Tischler, A. Wolf, T. Gojobori, K. A. Remington, R. L. Strausberg, J.C. Venter, U. Technau, B. Hobmayer, T.C. Bosch, T.W. Holstein, T. Fujisawa, H. R. Bode, C. N. David, D. S. Rokhsar, R. E. Steele, *Nature* **2010**, 464, 592.
- [28] D. C. Price, C. X. Chan, H. S. Yoon, E. C. Yang, H. Qiu, A. P. Weber, R. Schwacke, J. Gross, N. A. Blouin, C. Lane, A. Reyes-Prieto, D. G. Durnford, J. A. Neilson, B. F. Lang, G. Burger, J. M. Steiner, W. Löffelhardt, J. E. Meuser, M. C. Posewitz, S. Ball, M. C. Arias, B. Henrissat, P. M. Coutinho, S. A. Rensing, A. Symeonidi, H. Doddapaneni, B. R. Green, V. D. Rajah, J. Boore, D. Bhattacharya, *Science* **2012**, 335, 843.
- [29] G. Schoenkecht, W. H. Chen, C. M. Ternes, C. G. Barbier, R. P. Shrestha, M. Stanke, A. Bräutigam, B. J. Baker, J. F. Banfield, R. M. Garavito, K. Carr, C. Wilkerson, S. A. Rensing, D. Gagneul, N. E. Dickinson, C. Oesterhelt, M. J. Lercher, A. P. Weber, *Science* **2013**, 339, 1207.
- [30] G. Koutsovoulos, S. Kumar, D. R. Laetsch, L. Stevens, L. Stevens, J. Daub, C. Conlon, H. Maroon, F. Thomas, A. A. Aboobaker, M. Blaxter, *Proc. Natl. Acad. Sci. USA* **2016**, 113, 5053.
- [31] R. P. Hirt, C. Alsmark, T. M. Embley, *Curr. Opin. Microbiol.* **2015**, 23, 155.
- [32] Y. Yoshida, G. Koutsovoulos, D. R. Laetsch, L. Stevens, S. Kumar, D. D. Horikawa, K. Ishino, S. Komine, T. Kunieda, M. Tomita, M. Blaxter, K. Arakawa, *PLoS Biol.* **2017**, 15, e2002266.
- [33] M. W. Hahn, M. V. Han, S. G. Han, *PLoS Genet.* **2007**, 3, e197.
- [34] D. Charlesworth, N. H. Barton, B. Charlesworth, *Proc. R. Soc. Lond. B* **2017**, 284, pii 20162864.
- [35] M. Nei, *Mutation-Driven Evolution*. Oxford University Press, Oxford, UK **2013**.
- [36] D. Graur, *Molecular and Genome Evolution*. Sinauer, Sunderland MA, USA **2016**.
- [37] J. G. Lawrence, H. Ochman, *Proc. Natl. Acad. Sci. USA* **1998**, 95, 9413.
- [38] Y. Martin, *Bioessays* **1999**, 21, 99.
- [39] L. W. Parfrey, D. J. G. Lahr, A. H. Knoll, L. A. Katz, *Proc. Natl. Acad. Sci. USA* **2011**, 108, 13624.
- [40] C. Ku, S. Nelson-Sathi, M. Roettger, F. L. Sousa, P. J. Lockhart, D. Bryant, E. Hazkani-Covo, J. O. McInerney, G. Landan, W. F. Martin, *Nature* **2015**, 524, 427.
- [41] A. J. Enright, S. Van Dongen, C. A. Ouzounis, *Nucleic Acids Res.* **2002**, 30, 1575.
- [42] C. Ku, W. F. Martin, *BMC Biol.* **2016**, 14, 89.
- [43] S. Nelson-Sathi, F. L. Sousa, M. Roettger, N. Lozada-Chávez, T. Thiergart, A. Janssen, D. Bryant, G. Landan, P. Schönheit, B. Siebers, J. O. McInerney, W. F. Martin, *Nature* **2015**, 517, 77.

- [44] W. F. Martin, M. Roettger, C. Ku, S. G. Garg, S. Nelson-Sathi, G. Landan, *Genome Biol. Evol.* **2017**, 9, 373.
- [45] D. Graur, Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall, E. Elhaik, *Genome Biol. Evol.* **2013**, 5, 578.
- [46] R. M. Schwartz, M. Dayhoff, *Science* **1978**, 199, 395.
- [47] L. Sagan, *J. Theoret. Biol.* **1967**, 14, 225.
- [48] M. W. Gray, W. F. Doolittle, *Microbiol. Rev.* **1982**, 46, 1.
- [49] D. G. Lindmark, M. Müller, *J. Biol. Chem.* **1973**, 248, 7724.
- [50] B. J. Finlay, T. Fenchel, *FEMS Microbiol. Lett.* **1989**, 65, 311.
- [51] L. M. van Valen, V. C. Maiorana, *Nature* **1980**, 287, 248.
- [52] C. R. Vossbrinck, J. V. Maddox, S. Friedman, B. A. Debrunner-Vossbrinck, C. R. Woese, *Nature* **1987**, 326, 411.
- [53] T. M. Embley, W. Martin, *Nature* **2006**, 440, 623.
- [54] W. F. Martin, A. G. M. Tielens, M. Mentel, S. G. Garg, S. B. Gould, *Microbiol. Mol. Biol. Rev.* **2017**, 81, e00008. 00008–17.
- [55] W. Martin, M. Müller, *Nature* **1998**, 392, 37.
- [56] W. F. Martin, *Bioessays* **2017**, 39, 1700041.
- [57] S. B. Gould, S. G. Garg, W. F. Martin, *Trends Microbiol.* **2016**, 24, 525.
- [58] B. A. Williams, R. P. Hirt, J. M. Lucocq, T. M. Embley, *Nature* **2002**, 418, 865.
- [59] A. Karnkowska, V. Vacek, Z. Zubáčová, S. C. Treitli, R. Petrželková, L. Eme, L. Novák, V. Žárský, L. D. Barlow, E. K. Herman, P. Soukal, M. Hroudová, P. Doležal, C. W. Stairs, A. J. Roger, M. Eliáš, J. B. Dacks, C. Vlček, V. Hampl, *Curr. Biol.* **2016**, 26, 1274.
- [60] M. Müller, M. Mentel, J. J. van Hellemond, K. Henze, C. Woehle, S. B. Gould, R.-Y. Yu, M. van der Giezen, A. G. Tielens, W. F. Martin, *Microbiol. Mol. Biol. Rev.* **2012**, 76, 444.
- [61] B. Boxma, R. M. de Graaf, G. W. van der Staay, T. A. van Alen, G. Ricard, T. Gabaldón, A. H. van Hoek, S. Y. Moon-van der Staay, W. J. Koopman, J. J. van Hellemond, A. G. Tielens, T. Friedrich, M. Veenhuis, M. A. Huynen, J. H. Hackstein, *Nature* **2005**, 434, 74.
- [62] T. A. Williams, P. G. Foster, C. J. Cox, T. M. Embley, *Nature* **2013**, 504, 231.
- [63] K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, M. B. Stott, T. Nunoura, J. F. Banfield, A. Schramm, B. J. Baker, A. Spang, T. J. Ettema, *Nature* **2017**, 541, 353.
- [64] M. Degli Esposti, D. Cortez, L. Lozano, S. Rasmussen, H. B. Nielsen, E. Martinez Romero, *Biol. Direct* **2016**, 11, 34.
- [65] C. W. Stairs, A. J. Roger, V. Hampl, *Mol. Biol. Evol.* **2011**, 28, 2087.
- [66] C. W. Stairs, L. Eme, M. W. Brown, C. Mutsaers, E. Susko, G. Dellaire, D. M. Soanes, M. van der Giezen, A. J. Roger, *Curr. Biol.* **2014**, 24, 1176.
- [67] C. W. Stairs, M. M. Leger, A. J. Roger, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, 370, 20140326.
- [68] M. M. Leger, L. Eme, L. A. Hug, A. J. Roger, *Mol. Biol. Evol.* **2016**, 33, 2318.
- [69] I. Hrdy, M. Müller, *J. Mol. Evol.* **1995**, 41, 388.
- [70] C. Rotte, F. Mjerskal, G. Zhu, J. S. Keithly, W. Martin, *Mol. Biol. Evol.* **2001**, 18, 710.
- [71] V. Zimorski, C. Ku, W. F. Martin, S. B. Gould, *Curr. Opin. Microbiol.* **2014**, 22, 38.
- [72] J. M. Archibald, *Curr. Biol.* **2015**, 25, R911.
- [73] W. Martin, T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, D. Penny, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 12246.
- [74] P. G. Hofstatter, A. K. Tice, S. Kang, M. W. Brown, D. J. Lahr, *Proc. Biol. Sci.* **2016**, 283, 20161453.
- [75] S. Nelson-Sathi, T. Dagan, G. Landan, A. Janssen, M. Steel, J. O. McInerney, U. Deppenmeier, W. F. Martin, *Proc. Natl. Acad. Sci. USA* **2012**, 109, 20537.
- [76] E. Hazkani-Covo, S. Covo, *PLoS Genet.* **2008**, 4, e1000237.
- [77] E. Hazkani-Covo, R. M. Zeller, W. Martin, *PLoS Genet.* **2010**, 6, e1000834.
- [78] J. Petersen, H. Brinkmann, B. Bunk, V. Michael, O. Pärker, S. Pradella, *Environ. Microbiol.* **2012**, 14, 2661.
- [79] H. M. Wilkins, S. M. Carl, R. H. Swerdlow, *Redox Biol.* **2014**, 2, 619.
- [80] G. Pelletier, F. Vedel, G. Belliard, *Hereditas Suppl.* **1985**, 3, 49.
- [81] R. W. Burkhardt, Jr., *Genetics* **2013**, 194, 793.
- [82] J. P. Gogarten, *Curr. Biol.* **2003**, 13, R53.
- [83] P. Y. Dupont, M. P. Cox, *G3 (Bethesda)* **2017**, 7, 1301.
- [84] E. G. J. Danchin, *BMC Biol.* **2016**, 14, 101.
- [85] S. Ohno, *Evolution by Gene Duplication*. Springer, Heidelberg, Germany **1970**.
- [86] Y. van de Peer, S. Maere, A. Meyer, *Nat. Rev. Genet.* **2009**, 10, 725.
- [87] N. Nikoh, T. Hosokawa, K. Oshima, M. Hattori, T. Fukatsu, *Genome Biol. Evol.* **2011**, 3, 702.
- [88] K. Hjort, A. V. Goldberg, A. D. Tsaousis, R. P. Hirt, T. M. Embley, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, 365, 713.
- [89] R. Albalat, C. Cañestro, *Nat. Rev. Genet.* **2016**, 17, 379.
- [90] M. A. Naranjo-Ortiz, M. Brock, S. Brunke, B. Hube, M. Marcet-Houben, T. Gabaldón, *Front. Microbiol.* **2016**, 7, 2001.
- [91] G. J. Szelloosi, A. A. Davin, E. Tannier, V. Daubin, *Phil. Trans. R. Soc. Lond. B* **2015**, 370, 20140335.
- [92] D. Domman, M. Horn, T. M. Embley, T. A. Williams, *Nat. Comm.* **2015**, 6, 6421.
- [93] P. J. Lockhart, M. Steel, M. D. Hendy, D. Penny, *Mol. Biol. Evol.* **1994**, 11, 605.
- [94] J. N. Timmis, M. A. Ayliffe, C. Y. Huang, W. Martin, *Nat. Rev. Genet.* **2004**, 5, 123.
- [95] R. E. Lee, *Nature* **1972**, 237, 44.
- [96] W. F. Martin, R. Cerff, *Protoplasma* **2017**, 254, 1823.
- [97] W. F. Doolittle, *Science* **1999**, 284, 2124.
- [98] W. Martin, B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, K. V. Kowalik, *Nature* **1998**, 393, 162.
- [99] C. Semple, M. Steel, *Phylogenetics*. Oxford University Press, Oxford, UK **2003**.
- [100] S. G. Garg, W. F. Martin, *Genome Biol. Evol.* **2016**, 8, 1950.
- [101] V. B. Borisov, R. B. Gennis, J. Hemp, M. I. Verkhovskiy, *Biochim. Biophys. Acta* **2011**, 1807, 1398.
- [102] E. Forte, V. B. Borisov, M. Falabella, H. G. Colaco, M. Tinajero-Trejo, R. K. Poole, J. B. Vicente, P. Sarti, A. Giuffrè, *Sci. Rep.* **2016**, 6, 23788.
- [103] K. R. Olson, K. D. Straub, *Physiology (Bethesda)* **2016**, 31, 60.
- [104] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, H. Ishikawa, *Nature* **2000**, 407, 81.
- [105] R. C. van Ham, J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernández, L. Jiménez, M. Postigo, F. J. Silva, J. Tamames, E. Viguera, A. Latorre, A. Valencia, F. Morán, A. Moya, *Proc. Natl. Acad. Sci. USA* **2003**, 100, 581.
- [106] C. Dale, N. A. Moran, *Cell* **2006**, 126, 453.
- [107] N. Kondo, N. Nikoh, N. Ijichi, M. Shimada, T. Fukatsu, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 14280.
- [108] International Aphid Genomics Consortium, *PLoS Biol.* **2010**, 8, e1000313.
- [109] E. Jordan, H. Saedler, *Mol. Gen. Genet.* **1967**, 100, 283.
- [110] M. Kidwell, *Ann. Rev. Genet.* **1993**, 27, 235.
- [111] J. Peccoud, V. Loiseau, R. Cordaux, C. Gilbert, *Proc. Natl. Acad. Sci. USA* **2017**, 114, 4721.
- [112] T. Dagan, Y. Artzy-Randrup, W. Martin, *Proc. Natl. Acad. Sci. USA* **2008**, 105, 10039.
- [113] J. O. McInerney, A. McNally, M. J. O'Connell, *Nat. Microbiol.* **2017**, 2, 17040.
- [114] U. Bergthorsson, K. L. Adams, B. Thomasson, J. D. Palmer, *Nature* **2003**, 424, 197.
- [115] C. G. Wilson, R. W. Nowell, T. G. Barraclough, *bioRxiv* **2017**, 150490. <https://doi.org/10.1101/150490>
- [116] B. K. Wijayawardena, D. J. Minchella, J. A. DeWoody, *Mol. Biochem. Parasitol.* **2015**, 201, 57.