BioEssays
www.bioessays-journal.com

Demystifying Eukaryote Lateral Gene Transfer

(Response to Martin 2017 DOI: 10.1002/bies.201700115)

*Michelle M. Leger, Laura Eme, Courtney W. Stairs, and Andrew J. Roger**

In a recent BioEssays paper [W. F. Martin, *BioEssays* 2017, 39, 1700115], William Martin sharply criticizes evolutionary interpretations that involve lateral gene transfer (LGT) into eukaryotic genomes. Most published examples of LGTs in eukaryotes, he suggests, are in fact contaminants, ancestral genes that have been lost from other extant lineages, or the result of artefactual phylogenetic inferences. Martin argues that, except for transfers that occurred from endosymbiotic organelles, eukaryote LGT is insignificant. Here, in reviewing this field, we seek to correct some of the misconceptions presented therein with regard to the evidence for LGT in eukaryotes.

1. Mechanisms and Evidence of Lateral Gene Transfer in Eukaryotes

In a recent BioEssays article,[1] Martin suggests that there are no known genetic mechanisms that could explain LGT in eukaryotes except for hybridization. But that is not true. We know of several mechanisms for the introduction of foreign DNA into eukaryotic genomes, and we have direct evidence for some of them occurring on a recent time scale. Eukaryotic genomes are littered with remnants of viruses[2] and transposons that promiscuously hop between diverse eukaryotic host genomes, sometimes mobilizing host DNA in the process.[3–6] Conjugation between bacteria and eukaryotes in nature is well documented and can be reproduced under laboratory conditions (reviewed in ref. [7]). Furthermore, many eukaryotes harbour eukaryotic and/or prokaryotic endosymbionts (not just plastids and mitochondria); gene transfer from these endobionts into host chromosomes has been extensively reported in arthropods, nematodes, and *Paulinella* (e.g.,[8,9]).

Examples of large chunks of DNA being recently transferred into eukaryote genomes are not restricted to endosymbiont-containing organisms. Many metabolic gene clusters have been transferred into and between Fungi—conferring new functions such as lactose metabolism,[10] secondary metabolite biosynthesis,[11] or virulence.[12] Similarly, a 27 gene-encoding fragment of DNA from a *Peptinophilus*-related firmicute was confirmed to be encoded in the genomes of several strains of the protistan parasite *Trichomonas vaginalis*.[13] The foregoing cases are detectable not only by phylogenetic analyses of the encoded genes, but also by a high degree of sequence similarity to, and synteny with, the chromosomal segment of the donor lineage.

Careful analyses have built a strong body of evidence for older events of lineage-specific LGTs in eukaryotes as well. These inferences necessarily rely on phylogenetic analyses in which genes from recipient lineages robustly nest within subgroups of distantly related donor lineages in the phylogeny. There are dozens of such well-verified cases of older and younger events of LGT into various lineages of the Fungi, some as old as the fungal kingdom itself.[14] Furthermore, careful analyses of oomycete genomes revealed well-supported LGT events that occurred on sequential ancestral branches of the phylogeny going back to the earliest splits within the group.[15,16] Analyses of LGTs in other eukaryotic groups, including kinetoplastid parasites[17,18] and metazoans,[19,20] have similarly revealed clear evidence for ancient[21] and recent lineage-specific acquisitions of genes from bacteria and other eukaryotes. For many of these cases, contamination can be ruled out, as the presence of the genes has been verified in multiple closely related species, or the eukaryotic genomic contiguous segments containing the genes have been verified with long-read sequencing data, the presence of introns, or follow-up experiments (reviewed in ref. [22]).

2. Eukaryote LGT is Consistent with Accepted Evolutionary Mechanisms

Regardless of the specific mechanisms, the evolutionary fates of laterally transferred genes are not fundamentally different from other kinds of mutations. There is likely a low-level barrage of external genetic material that, once accidentally incorporated into the nuclear genome (by processes such as non-homologous

Dr. M. M. Leger
Institute of Evolutionary Biology (CSIC-UPF)
Pg. Marítim de la Barceloneta, Barcelona ES 08003, Spain
Dr. L. Eme, Dr. C. W. Stairs
Department of Cell and Molecular Biology
Science for Life Laboratory
Uppsala University
Box 596, Uppsala SE 751 25, Sweden
Prof. A. J. Roger
Centre for Comparative Genomics and Evolutionary Bioinformatics
Department of Biochemistry and Molecular Biology
Dalhousie University
P.O. Box 15000, Halifax CAN B3H 4R2 Nova Scotia, Canada
E-mail: andrew.roger@dal.ca

This article is in response to the article by W. F. Martin, DOI: https://doi.org/10.1002/bies.201700115

DOI: 10.1002/bies.201700242

end-joining),[23] is subject to the evolutionary forces of selection and drift. Most laterally acquired DNA is likely either neutral or deleterious, and the encoded genes degrade rapidly by point mutations and deletion, as has been documented.[13,24] Newly acquired genes may persist over longer evolutionary time-scales if they confer a new function that is adaptive, if they become essential by replacing an endogenous gene or pathway (described as "maintenance transfers" in ref. [22]), or if they are selfish genetic parasites (e.g., transposons).

All of this is entirely in agreement with well-accepted principles of molecular evolution; there is no sense to Martin's charge of Lamarckism. Nor have we claimed that eukaryotic LGT necessarily occurs at the same rates, or under the same circumstances, as in prokaryotes. Compared with LGT in prokaryotes, the study of LGT across eukaryotic diversity is in its infancy; not least because the taxon sampling for eukaryotic genomes remains a fraction of that of prokaryotes.[25] We still know very little about patterns of genome evolution for most eukaryotes (particularly microbes), let alone how the specific dynamics involved might affect the integration and retention of LGTs.

Rather than considering any of the possible mechanisms outlined above, Martin constructs an appeal to ridicule involving either non-random, targeted recombination of all genes involved in a specific pathway into the genome of the recipient eukaryote, and "mysterious" degradation of the rest of the donor genome; or outright hybridization of the two lineages into a "cybrid". These scenarios are then contrasted with the "simple" mechanisms of DNA exchange between prokaryotes, ignoring the well-documented eukaryotic mechanisms.

3. Are Estimated Frequencies of LGT Too High?

A key pillar of Martin's argument is that the rate of eukaryote LGTs into specific lineages cannot be as high as estimated, otherwise the cumulative effects of these LGTs should be significant and readily detectable. To illustrate this point, Martin discusses LGTs inferred in the genomes of 12 *Drosophila* species by Yoshida et al.[26] He argues their estimates of LGT are too high, stating: "I find claims of 0.5% LGT per fly genome very difficult to digest." However, Yoshida et al.[26] never made any such claim. They estimated LGT by flagging genes in a given *Drosophila* genome that displayed substantially greater similarity to non-metazoan versus metazoan homologs. Critically, arthropod and nematode sequences were deliberately removed from the within-metazoan databases prior to their analysis, so that genes in sister taxa would not be recovered as hits, allowing both older and younger potential transferred genes to be identified. Many of the flagged LGT candidates in each genome are likely orthologs in some, or all, of the other *Drosophila* genomes. Yoshida and colleagues never suggested that these LGTs were unique to each fly genome. Martin later seems to acknowledge that these may not be per-genome LGT estimates, but then presents three possible alternative interpretations: that all of the LGT are species-specific, that all of them were present in the *Drosophila* common ancestor, or that all of them were present in the last eukaryote common ancestor. The possibility that some of them were present in the common ancestor, while others were introduced into individual ancestral lineages at various time points—fulfilling Martin's criteria for lineage-specific LGT accumulation—is not discussed.

In any case, Martin uses the above example to derive a hypothetical conservative eukaryote LGT rate of one LGT per million years. He argues that this rate, although 20-fold lower than in bacteria, is still untenably high. If this rate were true, he claims, then approximately 700 LGTs should have accrued per each major animal lineage, and approximately 1600 for each eukaryotic supergroup, assuming that these two groups diversified 700 and 1600 million years ago, respectively. Martin contends that these numbers are far larger than what is actually observed in studies such as that of Ku et al.[27]

However, these extrapolations are invalid, because they are based on an obviously false model of genome evolution. For example, using Martin's reasoning and suggested LGT rate for bacteria, each major lineage of cyanobacteria—a group that is >2 billion years old[28]—should have accrued >40 000 distinct genes by LGT. Clearly this is false. Genomes do not grow in a linear fashion over time by LGT; genes are gained and lost. Yet, even assuming a limit of 5000 genes per bacterial genome still does not solve the problem, as every gene in each major cyanobacterial lineage must then have been replaced on average eight times over. If that were true, these genomes should show no more similarity to each other than to any other prokaryotic genome. Instead, major cyanobacterial lineages are clearly related to one another in both marker gene phylogenies and gene-content (e.g.,[29]). The key flaw in such extrapolations is that different genes in genomes are gained and lost at different rates.[30] Some genes are rarely transferred or lost, while others turn over rapidly. Any valid extrapolation of LGT rates to gene content differences requires knowledge of the distributions of the rates of gene gain and loss across different kinds of genes. Little is known about this in eukaryotes, so such extrapolations cannot currently be made.

4. Are LGTs Really Just Differentially Retained Ancestral Genes?

One of Martin's favored explanations for non-contaminant prokaryotic LGT candidates is that they originated prior to the last eukaryotic common ancestor (LECA) from the mitochondrial endosymbiont, or were introduced into major photosynthetic eukaryotic lineages from the plastid endosymbiont (either by primary or secondary endosymbiosis). In either case, Martin argues that such genes were subsequently differentially lost many times in parallel in diverse eukaryotic lineages, thereby explaining their observed patchy distributions. However, this argument, when invoked without penalty to explain proposed LGTs, implies that genomes of ancestral eukaryotes were increasingly bloated with genes, and that gene content has been steadily decreasing during eukaryote evolution. This flawed inference—named the "genome of Eden" problem—was first pointed out in reference to skepticism about LGT in prokaryotes,[31] and the arguments apply with equal force to eukaryotes. Indeed, Szöllősi et al.[32] showed that probabilistic models that take into account only gene duplication and loss systematically overestimate gene contents of ancestral fungal genomes, in contrast to those that also take into account lateral gene transfer.

This is not the only problem with over-reliance on the "ancestral-but-differentially-lost" explanation for rare and patchily distributed genes in eukaryotes. Many genes argued to be LGTs in diverse eukaryotes confer traits that only make sense in terms of interactions with organisms that had not yet evolved at the time of LECA (such as defense against the metazoan immune system,[33] or plant-cell invasion mechanisms[9]).

5. Is There No Evidence for Cumulative Effects of LGT in Eukaryotes?

Martin's arguments against cumulative effects of LGT in eukaryotes rely heavily on the analyses of Ku et al.[27] that supposedly "found no evidence for lineage-specific acquisitions in eukaryotes."[1] However, the conclusions of this paper derive from questionable assumptions made in the analyses, and arbitrary exclusion of evidence.

Firstly, to be counted by Ku et al. as being of prokaryotic-origin, each gene had to be present in the genomes of at least two eukaryotes and five prokaryotes, restricting the number of recent lineage-specific LGTs in eukaryotes that could be inferred. Despite these parsimonious standards, 323 cases where eukaryotic homologues did not form a monophyletic group—some of which may be examples of LGT—were dismissed as being less important than the "monophyletic majority."[27]

Some genes with distribution patterns did, in fact, appear to be lineage-specific gains in eukaryotes (e.g., see labels a-e in Figure 2 of ref. [1]). These were argued, on the basis of several statistical tests, to either be ancestral eukaryotic genes or to have originated via primary or secondary plastid endosymbioses and subsequently been exclusively vertically inherited and differentially lost during eukaryote evolution. The first test relied on a comparison of phylogenies of genes of prokaryotic origin to phylogenies of eukaryote-specific genes; this test found no significant difference between these distributions of trees. Ku et al. interpret this to mean that all of these phylogenies of the prokaryote-derived genes reflect an origin by vertical inheritance. However, the possibility that a substantial fraction of the eukaryote-specific genes were also laterally transferred amongst eukaryotes seems not to have been considered. A second test relied on a comparison of eukaryotic gene phylogenies to null LGT distributions derived from random subtree pruning-regrafting operations on trees. Unlike true LGTs, these operations will frequently generate topologies that violate time-consistency,[34] making this an inappropriate null distribution. A third test (assessing sequence similarity to prokaryotic homologs of "lineage-specific" genes versus "more commonly distributed" genes) assumes that the rate of sequence evolution should be clock-like and comparable across functionally distinct sets of genes from different sources, also an unrealistic expectation. Even more concerningly, approximately 1000 genes apparently introduced into the Archaeplastida supergroup were assumed to have been acquired via the primary origin of chloroplasts from endosymbiotic cyanobacteria, despite the fact that the majority of their phylogenies do not show the expected cyanobacterial sister-group affinities of plastid-origin genes. In fact, no homologs were recovered for 356 of these genes in any of the 31 cyanobacteria sampled.[27] Yet, the alternative interpretation—that a substantial fraction of these genes correspond to LGTs from different bacterial donors—is dismissed by Ku et al. as absurd. This type of interpretation uncritically posits a plastid origin for all of these proteins at the base of the Archaeplastida lineage as a foregone conclusion.

6. Are Phylogenetic Artefacts Responsible for Apparent Cases of LGT?

In both Ku et al. and Martin's essay, large numbers of "unexpected branching patterns in phylogenetic trees"[1] that might otherwise be viewed as evidence for LGT are explained away as phylogenetic artefacts. Of course, like any statistical inference procedure, phylogenetic estimation is subject to error. This can take the form of random error resulting from lack of strong historical signals in the data (leading to poorly supported erroneous branching patterns), or systematic error for cases in which evolutionary divergences are large and phylogenetic models are misspecified. Most credible cases of ancient eukaryote LGT have been proposed on the basis of phylogenetic trees showing strongly supported severe conflicts with the expected species phylogeny, with support assessed by bootstrap analysis, posterior probabilities and/or topology tests.[22] If such cases were artefacts, then, systematic error must be the cause. In these cases, the known conditions that induce phylogenetic artefacts—such as combinations of very short and long branches on the tree, coupled with clear violations of the phylogenetic model (e.g., amino acid compositional biases in some taxa,[35] site-pattern heterogeneity amongst sites,[36,37] or strong heterotachy effects[38,39])—should be evident. Yet, Ku et al.'s[27] assertion that "molecular phylogenetics sometimes simply fails" provides no details on the cause of the artefacts that are claimed to be plaguing hundreds of well-supported phylogenetic estimates. Surely some kind of evidence is required to support claims of widespread systematic error affecting phylogenetic analyses other than their disagreement with pre-conceived notions that all prokaryotic-origin genes should have either an archaeal, plastid, or mitochondrial ancestry. Furthermore, it is transparently circular to discount phylogenies as artefacts just because they conflict with these three possible origins, while simultaneously citing, as supporting evidence, the remaining subset of analyses that are consistent with them.

7. Are Anaerobic Energy Metabolism Enzymes Examples of LGT in Eukaryotes?

As an example of particularly problematic LGT inferences, Martin discusses the origins of anaerobic energy metabolism in eukaryotes. We, and others, have published numerous examples of putative LGTs in obligately, or facultatively, anaerobic protists. Many are examples of genes that are present in prokaryotic operons but in some eukaryotes are found fused into single open reading frames encoding multi-domain proteins.[21,40–42] In these cases, it is likely that the entire operon was transferred, and that some genes encoded on it were subsequently fused in a eukaryote. Clusters of genes encoding anaerobic ATP generation enzymes are also found in some eukaryotic chromosomes[41,43]

that might make multi-gene transfer possible between eukaryotes as well. But not all genes in a single pathway need be acquired at the same time; we have previously outlined a stepwise scenario in which genes involved in anaerobic ATP generation could have been acquired sequentially.[44]

However, Martin specifically takes issue with our discussion of the origins of the pyruvate:ferredoxin oxidoreductase (PFO) and pyruvate:NADP+ oxidoreductase (PNO) enzymes that function in anaerobic ATP generation. He suggests that the interpretation in our paper[45] is "that PFO entered the eukaryotic lineage via lateral acquisition long after mitochondria arose and was then distributed among diverse eukaryotic lineages via LGT." This interpretation, Martin contends, ignores an alternative plausible scenario involving ancient duplications prior to LECA followed by massive numbers of parallel secondary losses in most eukaryote lineages. This is incorrect. In fact, we explicitly highlighted "the overall lack of resolution in these trees, and the lack of a clear prokaryotic sister group to eukaryotes that could point to these enzymes being either of mitochondrial origin, or laterally transferred from a specific prokaryotic group."[45] The paragraph from which this quote is taken clearly referred to PFO, as well as to [FeFe]-hydrogenase. We further emphasized that it is unlikely that PFO, specifically, originated from an alphaproteobacterial (mitochondrial) source: it may therefore have been laterally acquired, or it may have been present in LECA, having originated from a different source. The data in this case are simply not clear enough to draw firm conclusions as to the origins of PFO, and we clearly stated this.

8. Conclusion

Throughout the essay, Martin portrays his position as one of healthy skepticism toward eukaryote LGT. However, his arguments repeatedly ignore plausible scenarios for eukaryote LGT origins and dynamics, in order to more easily dismiss them. Most troublingly, these arguments appear designed to be resistant to any evidence that LGT occurs in eukaryotes. If a suspected LGT appears in only one or a few sister taxa, then it is dismissed as either a contaminant, or—where this is clearly not the case—as a transitory event. If the gene is more broadly phylogenetically distributed, then it is argued to have been ancestrally present and to have originated from the mitochondrial or plastid endosymbiont genomes before being lost repeatedly, in parallel, in numerous lineages. The final argument is that molecular phylogenetics is flawed and artefact-prone. Taken together, these arguments do not simply dismiss existing eukaryote LGT claims, they create a standard by which no conceivable kind of data can be taken as evidence for significant LGT in eukaryotes.

It is certainly necessary to treat new sequence data with caution, and to carefully consider the possibilities of contamination or widespread secondary loss. This need for caution has been highlighted by recent controversies over the proposed number of laterally transferred genes in tardigrades and humans (e.g.,[46–49]). However, it should remain possible for us to become convinced by sufficient evidence for LGTs into eukaryotes. Categorical claims that there is "no evidence" for lineage-specific LGT in eukaryotes should, therefore, be tempered by these concerns, and weighed against the dozens of well-verified reports of exactly that.[22]

Abbreviations

LGT, lateral gene transfer; PFO, pyruvate:ferredoxin oxidoreductase; PNO, pyruvate:NADP+ oxidoreductase.

Acknowledgements

We would like to thank W. Ford Doolittle and Iñaki Ruiz-Trillo for helpful comments on the manuscript. M.M.L. is supported by a European Research Council Consolidator grant (ERC-2012-Co-616960) to Prof. Iñaki Ruiz-Trillo. L.E. is funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 704263. C.W.S. is supported by a European Molecular Biology Organization long-term fellowship (ALTF-997-2015) and the Natural Sciences and Engineering Research Council of Canada (NSERC) postdoctoral research fellowship (PDF-487174-2016). A.J.R. is supported by the Canada Research Chairs Program and grants from the Canadian Institutes for Health Research (MOP-142349) and NSERC (2016-016792).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

eukaryote genomes, horizontal gene transfer, lateral gene transfer

Received: December 13, 2017
Revised: February 6, 2018
Published online:

[1] W. F. Martin, *BioEssays* **2017**, *39*, 1700115.
[2] J. Filée, *Virology* **2014**, *53*, 466.
[3] S. Schaack, C. Gilbert, C. Feschotte, *Trends Ecol. Evol*. **2010**, *25*, 537.
[4] A. Monier, A. Chambouvet, D. S. Milner, V. Attah, R. Terrado, C. Lovejoy, H. Moreau, A. E. Santoro, E. Derelle, T. A. Richards, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7489.
[5] L. Gallot-Lavallée, G. Blanc, *Viruses* **2017**, *9*, E17.
[6] N. Yutin, Y. I. Wolf, E. V. Koonin, *Virology* **2014**, *38*, 466.
[7] B. Lacroix, V. Citovsky, *MBio* **2016**, *7*, e00863.
[8] J. C. Dunning Hotopp, M. E. Clark, D. C. Oliveira, J. M. Foster, P. Fischer, M. C. Muñoz Torres, J. D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R. V. Nene, J. Shepard, J. Tomkins, S. Richards, D. J. Spiro, E. Ghedin, B. E. Slatko, H. Tettelin, J. H. Werren, *Science* **2007**, *317*, 1753.
[9] T. Nakayama, K. Ishida, *Curr. Biol*. **2009**, *19*, R284.
[10] J. Ropars, R. C. Rodriguez de la Vega, M. Lopez-Villavicencio, J. Gouzy, E. Sallet, E. Dumas, S. Lacoste, R. Debuchy, J. Dupont, A. Branca, T. Giraud, *Curr. Biol*. **2015**, *25*, 2562.
[11] J. C. Slot, A. Rokas, *Curr. Biol*. **2011**, *21*, 134.
[12] T. L. Friesen, E. H. Stukenbrock, Z. Liu, S. Meinhardt, H. Ling, J. D. Faris, J. B. Rasmussen, P. S. Solomon, B. A. McDonald, R. P. Oliver, *Nat. Genet*. **2006**, *38*, 953.
[13] A. Strese, A. Backlund, C. Alsmark, *BMC Evol. Biol*. **2014**, *14*, 119.
[14] D. Soanes, T. A. Richards, *Annu. Rev. Phytopathol*. **2014**, *52*, 583.

[15] F. Savory, G. Leonard, T. A. Richards, *PLoS Pathog*. **2015**, *11*, e1004805.
[16] C. G. McCarthy, D. A. Fitzpatrick, *mSphere* **2016**, *1*, e00195.
[17] E. Vikeved, A. Backlund, C. Alsmark, *PLoS Negl. Trop. Dis*. **2016**, *10*, e0004326.
[18] C. Alsmark, P. G. Foster, T. Sicheritz-Ponten, S. Nakjang, T. M. Embley, R. P. Hirt, *Genome Biol*. **2013**, *14*, R19.
[19] A. Faddeeva-Vakhrusheva, K. Kraaijeveld, M. F. L. Derks, S. Y. Anvar, V. Agamennone, W. Suring, A. A. Kampfraath, J. Ellers, G. Le Ngoc, C. A. M. van Gestel, J. Marien, S. Smit, N. M. van Straalen, D. Roelofs, *BMC Genomics* **2017**, *18*, 493.
[20] F. Husnik, J. P. McCutcheon, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E5416.
[21] A. D. Tsaousis, S. Ollagnier de Choudens, E. Gentekaki, S. Long, D. Gaston, A. Stechmann, D. Vinella, B. Py, M. Fontecave, F. Barras, J. Lukeš, A. J. Roger, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10426.
[22] F. Husnik, J. P. McCutcheon, *Nat. Rev. Microbiol*. **2017**, https://doi.org/10.1038/nrmicro.2017.137.
[23] D. Wang, J. N. Timmis, *Genome Biol. Evol*. **2013**, *5*, 1060.
[24] M. A. Campbell, M. Staats, J. A. van Kan, A. Rokas, J. C. Slot, *Mycologia* **2013**, *105*, 1126.
[25] S. J. Sibbald, J. M. Archibald, *Nat. Ecol. Evol*. **2017**, *1*, 145.
[26] 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.
[27] 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.
[28] P. Sanchez-Baracaldo, J. A. Raven, D. Pisani, A. H. Knoll, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7737.
[29] V. Kunin, L. Goldovsky, N. Darzentas, C. A. Ouzounis, *Genome Res*. **2005**, *15*, 954.
[30] Y. I. Wolf, K. S. Makarova, A. E. Lobkovsky, E. V. Koonin, *Nat. Microbiol*. **2016**, *2*, 16208.
[31] W. F. Doolittle, Y. Boucher, C. L. Nesbo, C. J. Douady, J. O. Andersson, A. J. Roger, *Philos. Trans. R Soc. Lond. B Biol. Sci*. **2003**, *358*, 39.
[32] G. J. Szöllősi, A. A. Davin, E. Tannier, V. Daubin, B. Boussau, *Philos. Trans. R Soc. Lond. B Biol. Sci*. **2015**, *370*, 20140335.
[33] L. Eme, E. Gentekaki, B. Curtis, J. M. Archibald, A. J. Roger, *Curr. Biol*. **2017**, *27*, 807.
[34] J. P. Doyon, V. Ranwez, V. Daubin, V. Berry, *Brief Bioinform*. **2011**, *12*, 392.
[35] P. G. Foster, C. J. Cox, T. M. Embley, *Philos. Trans. R Soc. Lond. B Biol. Sci*. **2009**, *364*, 2197.
[36] H. C. Wang, K. Li, E. Susko, A. J. Roger, *BMC Evol. Biol*. **2008**, *8*, 331.
[37] N. Lartillot, H. Brinkmann, H. Philippe, *BMC Evol. Biol*. **2007**, *7*, S4.
[38] Y. Zhou, N. Rodrigue, N. Lartillot, H. Philippe, *BMC Evol. Biol*. **2007**, *7*, 206.
[39] Y. Inagaki, E. Susko, N. M. Fast, A. J. Roger, *Mol. Biol. Evol*. **2004**, *21*, 1340.
[40] M. C. Posewitz, P. W. King, S. L. Smolinski, L. Zhang, M. Seibert, M. L. Ghirardi, *J. Biol. Chem*. **2004**, *279*, 25711.
[41] C. W. Stairs, A. J. Roger, V. Hampl, *Mol. Biol. Evol*. **2011**, *28*, 2087.
[42] J. O. Andersson, A. J. Roger, *Eukaryot. Cell* **2002**, *1*, 304.
[43] M. M. Leger, R. M. Gawryluk, M. W. Gray, A. J. Roger, *PLoS ONE* **2013**, *8*, e69532.
[44] C. W. Stairs, M. M. Leger, A. J. Roger, *Philos. Trans. R Soc. Lond. B Biol. Sci*. **2015**, *370*, 20140326.
[45] M. M. Leger, L. Eme, L. A. Hug, A. J. Roger, *Mol. Biol. Evol*. **2016**, *33*, 2318.
[46] G. Koutsovoulos, S. Kumar, D. R. Laetsch, L. Stevens, J. Daub, C. Conlon, H. Maroon, F. Thomas, A. A. Aboobaker, M. Blaxter, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5053.
[47] S. L. Salzberg, *Genome Biol*. **2017**, *18*, 85.
[48] I. Eyres, C. Boschetti, A. Crisp, T. P. Smith, D. Fontaneto, A. Tunnacliffe, T. G. Barraclough, *BMC Biol*. **2015**, *13*, 90.
[49] 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* **2015**, *112*, 15976.

