REVIEW

doi:10.1038/nature12779

An archaeal origin of eukaryotes supports only two primary domains of life

Tom A. Williams[1], Peter G. Foster[2], Cymon J. Cox[3] & T. Martin Embley[1]

The discovery of the Archaea and the proposal of the three–domains 'universal' tree, based on ribosomal RNA and core genes mainly involved in protein translation, catalysed new ideas for cellular evolution and eukaryotic origins. However, accumulating evidence suggests that the three–domains tree may be incorrect: evolutionary trees made using newer methods place eukaryotic core genes within the Archaea, supporting hypotheses in which an archaeon participated in eukaryotic origins by founding the host lineage for the mitochondrial endosymbiont. These results provide support for only two primary domains of life—Archaea and Bacteria—because eukaryotes arose through partnership between them.

Since their discovery by Carl Woese and his co-workers in 1977, the Archaea have figured prominently in hypotheses for eukaryotic origins[1,2]. Although similar to Bacteria in terms of cell structure, molecular phylogenies for ribosomal RNA and a small core of genes, that mainly have essential roles in protein translation[3], suggested that the Archaea were more closely related to the eukaryotic nuclear lineage; that is, to the host cell that acquired the mitochondrion[4]. The idea that Archaea and eukaryotes are more closely related to each other than either is to Bacteria depends on analyses suggesting that the root of the tree should be placed on the bacterial stem, or within the Bacteria[5–12], implying that the prokaryotes—cells that lack a nucleus—are a paraphyletic group[13]. The main question now debated is whether core components of the eukaryotic nuclear lineage descend from a common ancestor shared with Archaea, as in the three-domains tree[14] (Fig. 1), which is also often called the 'universal tree' or 'tree of life'[15–17], or from within the Archaea, as proposed by archaeal-host hypotheses for eukaryotic origins[2]. The archaeal-host scenario with the greatest phylogenetic support is the eocyte hypothesis[18], which proposes a sister-group relationship between eukaryotes and the eocytes (or Crenarchaeota[14]), one of the major archaeal divisions (Fig. 1). However, the three-domains–eocyte debate remains controversial because different phylogenetic methods have delivered different results, often from the same data[19]. This disagreement is due, at least in part, to the difficulties associated with resolving ancient divergences in phylogenetic trees.

Challenges of reconstructing ancient relationships

A major issue in reconstructing ancient relationships is the strength and quality of historical signal remaining after the millions of years since the divergence of Archaea and eukaryotes. The earliest fossils identified as eukaryotic appeared by about 1.8 billion years ago[20]; over this enormous span of time, the accumulation of multiple substitutions in DNA and protein sequences might have erased any signal that would allow the relationship between archaeal and eukaryotic core genes to be established[21]. However, more recent simulations and empirical studies suggest that there are reasons to be cautiously optimistic that this is not the case: functional constraints vary across real DNA and protein sequences so that sites evolve at different rates[22–25]. Fast-evolving sites are indeed

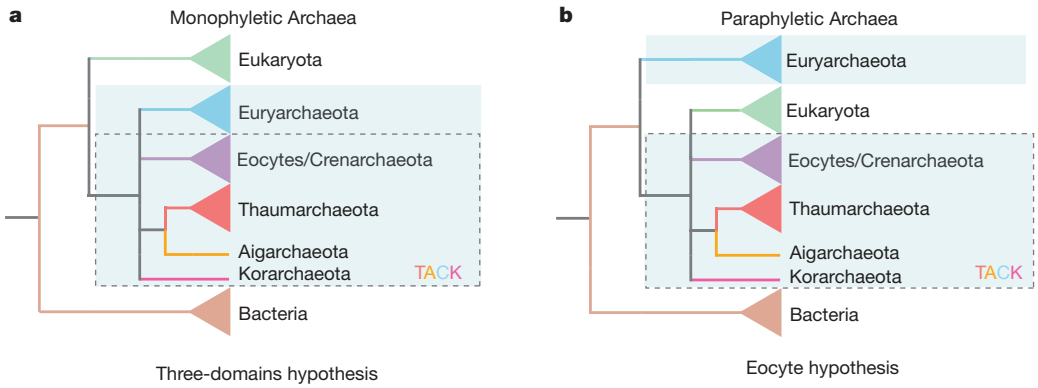

Figure 1 | Competing hypotheses for the origin of the eukaryotic host cell. **a**, The rooted three-domains tree[14] depicts cellular life divided into three major monophyletic groups or domains: the Bacteria, Archaea and Eukaryota—the latter representing the host lineage, sometimes also called the nuclear or nucleo-cytoplasmic lineage[5], that acquired the mitochondrial endosymbiont. In this tree the Archaea and Eukaryota are most closely related to each other because they share a common ancestor that is not shared with Bacteria. **b**, The rooted eocyte tree recovers the host-cell lineage nested within the Archaea as a sister group to the eocytes (which Woese *et al.*[14] called the Crenarchaeota); this implies that, on the basis of the small set of core genes, there are only two primary domains of life—the Bacteria and the Archaea. In its modern formulation shown here the eocyte hypothesis implies that the closest relative of the eukaryotic nuclear lineage is one, or all, of the TACK Archaea, which include newly discovered relatives of the eocytes/Crenarchaeota. Both trees have been traditionally rooted on the bacterial stem, consistent with some published analyses[5–8].

[1]Institute for Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne NE2 4HH, UK. [2]Department of Life Sciences, Natural History Museum, London SW7 5BD, UK. [3]Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

quickly saturated but the slowest sites can still retain useful phylogenetic information, explaining why we are able to align some genes over the entire tree of life. Analyses of molecular sequences might therefore be able to distinguish between the alternative hypotheses for eukaryotic core gene origins, but the phylogenetic methods used and the types of data analysed are likely to be of critical importance in attempts to recover any historical signal[22–26].

The problems associated with phylogenetic reconstruction come into particularly sharp focus when comparing support for the three-domains and eocyte trees. The first studies to investigate this question generally recovered the three-domains tree, in which eukaryotes emerge as the sister group to Archaea[27,28], but the parsimony and distance methods used carried unrealistic assumptions, including constancy (homogeneity) of base compositions across lineages and of evolutionary rates across sites. These assumptions are clearly violated by key phylogenetic markers such as small subunit ribosomal RNA genes, which contain a mixture of fast- and slowly-evolving sites[29] and for which GC content varies widely among the three domains[12]. Compositional heterogeneity can cause phylogenetic error when not taken into account, because sequences of similar base or amino acid composition may group together in the tree even when they are not closely related[30–32]. Two pioneering studies used methods to mitigate possible convergence in the universal tree due to shared compositional biases in nucleotide sequences and, interestingly, both recovered an eocyte tree[33,34].

Long branch attraction (LBA) is another pervasive artefact in molecular phylogenies, in which sequences with long branches cluster together irrespective of their evolutionary history[25,35]. LBA is especially problematic for parsimony methods, but it can also affect probabilistic methods if the model ignores among-site rate variation, or is otherwise a poor fit to the data[36]. Trees for the rRNA and protein-coding genes used to infer relationships between domains often show evidence of long branches and are therefore susceptible to LBA. Some of the early attempts to mitigate the influence of LBA in inter-domain analyses also recovered eocyte trees, although with variable support. Evolutionary parsimony, a method designed to reduce the effect of long branches on the inferred tree, recovered an eocyte topology from rRNA sequences[37], although archaeal monophyly was favoured when a related method, compositional statistics, was used to analyse RNA polymerase sequences[38]. By contrast, analyses of rRNA and RNA polymerase using models that accounted for among-site rate variation supported the eocyte hypothesis over the three-domains tree[39]. To reconcile these results, Tourasse and Gouy[39] suggested that the three-domains tree might be a phylogenetic artefact caused by LBA between the long bacterial and eukaryotic branches, forcing an artefactual clustering of the shorter archaeal branches. In other words, the eocyte tree might be intrinsically more difficult to recover using simple methods, because it requires the clustering of the short branch leading to the eocytes/Crenarchaeota with the long eukaryotic branch.

Single-gene phylogenies often fail to strongly resolve the relationships between the domains[12,40], and so a number of studies have analysed concatenations of the core set of proteins conserved on all genomes. As already described, these genes largely function in translation and gene expression, and include many of the essential RNA and protein components of the ribosome. These cellular components have been called the 'genealogy-defining core'[3], the 'genetic core'[41] of cells or the 'functional core of genomes'[16], and their common history has been cited[3,16,41] as the strongest support for the three-domains tree. Testing the evolutionary origins of this small set of genes is therefore critical to the three-domains–eocyte debate. Interestingly, analyses using similar sets of concatenated core genes have yielded different conclusions, for example Katoh *et al.*[42] obtained an eocyte tree from a set of 39 universal proteins, whereas Ciccarelli *et al.*[43] analysed a similar set of proteins and obtained a three-domains tree. One reason for the conflicting results in this case may be the different methods used for making the sequence alignment: the order of alignment had previously been shown to dictate which tree (eocyte or three-domains) was recovered from elongation factor Tu sequences[44]. Ciccarelli *et al.*[43] aligned bacterial, archaeal and eukaryotic sequences separately before combining them into a single alignment. This stepwise procedure was criticized as potentially biasing the results towards a three-domains topology but also, when the individual alignments were combined, to have introduced alignment errors between domains[19]. Brown *et al.*[45] also inferred trees from a concatenation of a subset of 14 universally conserved proteins, but in this study the tree recovered depended on the phylogenetic method used; the three-domains topology was recovered using maximum parsimony, but model-based methods recovered an eocyte topology.

Over the past few years, phylogenetic models implemented in either a maximum likelihood or a Bayesian framework have continued to increase in sophistication by incorporating additional features of the evolutionary process. These include relaxing the assumptions of homogeneous amino acid or base composition across sites[46] or across branches of the tree[31]. These models seem to fit molecular sequence data much better than simpler models and this may make them less susceptible to LBA and other artefacts of model mis-specification[25]. Although relatively few analyses of the core gene set have used these models so far, all of them have recovered the eocyte tree, rather than the three-domains tree[12,22,47–49].

New archaeal lineages and eukaryotic origins

In addition to improvements in phylogenetic methods, the diversity of molecular sequences from organisms related to the eocytes has also increased considerably, driven by the ease with which sequences from uncultured prokaryotes can now be sampled from the environment using molecular methods[50,51]. Improved sampling can have positive effects on phylogenetic reconstruction, particularly when it helps to break up long branches[52]. Recently discovered relatives of the eocytes include the Korarchaeota[53], the Thaumarchaeota[54] and the Aigarchaeota[55]; the 'TACK' superphylum was subsequently proposed as an informal group to encompass these four taxa[47]. To date, the studies including TACK sequences have supported a version of the eocyte hypothesis extended to recognize this improved sampling, rather than the three-domains tree[47,48,56]. In this extended sense, the eocyte hypothesis implies that the closest relative of the eukaryotic nuclear lineage is one, or all, of the TACK Archaea. If this tree is correct, then an important place to look for prokaryotic homologues of eukaryotic cellular componentry should be among the TACK phyla. Consistent with this prediction, members of this group encode homologues of a number of key eukaryotic genes (Fig. 2 and Supplementary Table 1), including actin[57] and tubulin[58] (the essential components of the eukaryotic cytoskeleton), a ubiquitin protein-modification system[55] and a number of genes involved in transcription and translation[47,59]. However, no single characterized TACK genome possesses all of these features[47,57,58], implying that gene loss, and potentially horizontal gene transfer (HGT), have contributed to the patterns of gene sharing on contemporary archaeal and eukaryotic genomes[60,61].

Which history do universal trees represent?

In their seminal papers, Woese and Fox[1,4] recognized that the rRNA tree represented only one component, the host for the mitochondrial endosymbiont, in the composite origins of the eukaryotic cell. That composite nature has been confirmed by comparative genomics, which has demonstrated that eukaryotic genomes contain a mixture of genes with different origins[13,62–66]. Some genes are ancestrally present in all three groups or unique to eukaryotes, but many others appear to have origins through gene transfers from different Bacteria, including the endosymbiotic progenitors of mitochondria and plastids, and relatively few—including the core set of conserved proteins we have been discussing—have affinities with the Archaea. From these data it is clear that no one tree is sufficient to describe the history of all of the genes on modern eukaryotic genomes[67,68]. However, even though this fact is now widely documented, the three-domains tree is often still called the 'tree of life' or 'universal tree' in textbooks[15] and reviews[16,17].

The sequencing of genomes from across the tree of eukaryotes is beginning to provide a clearer picture of the impact on eukaryotic genomes of HGT from prokaryotes[65]. These data suggest that the acquisition of bacterial genes, at least by microbial eukaryotes, has been an ongoing

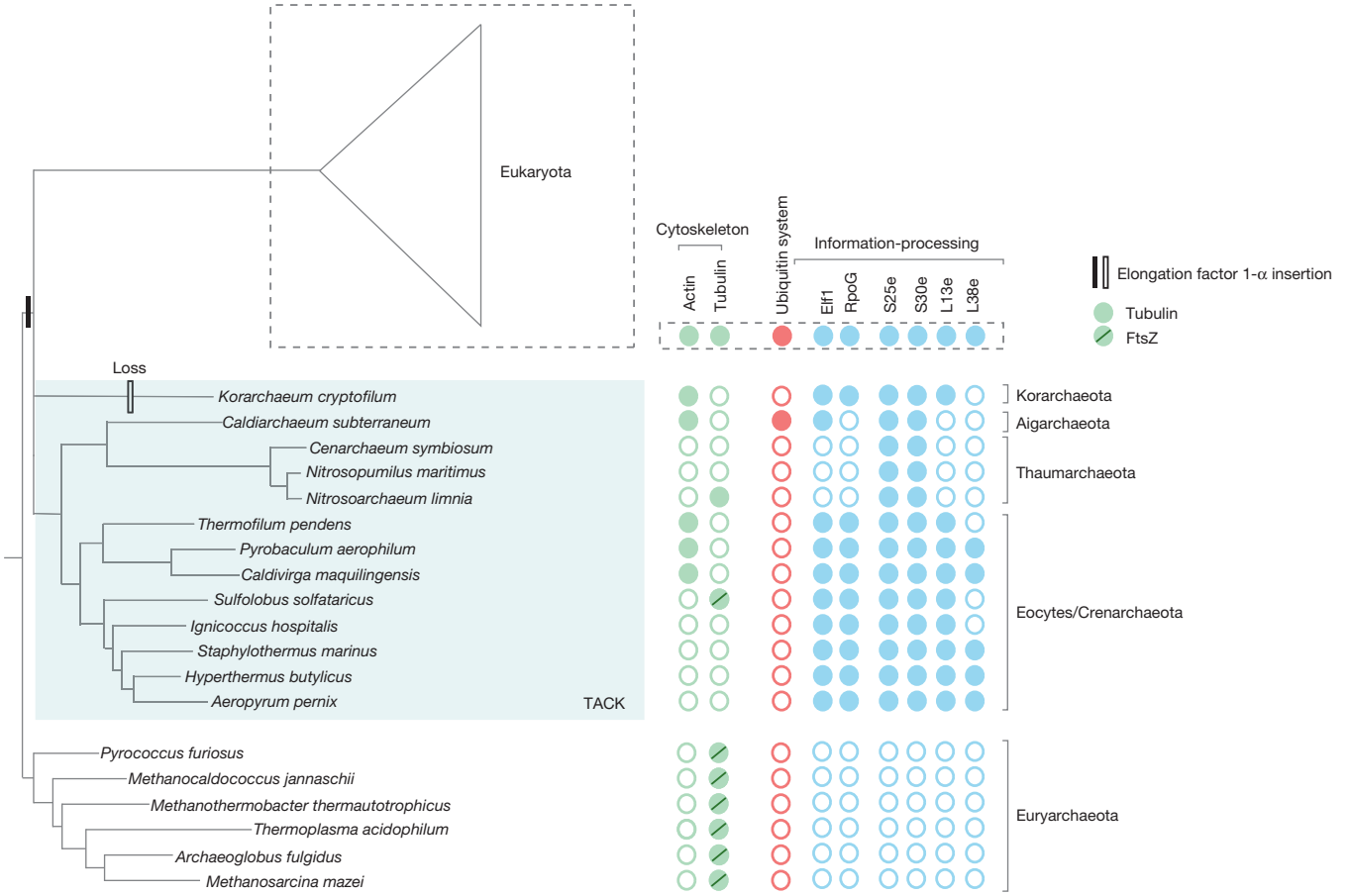

Figure 2 | Archaeal links in the origin of eukaryotes. A schematic tree depicting the relationships between Archaea and the eukaryotic nuclear lineage, consistent with recent analyses of core genes using new methods[47–49] and rooted using the Bacteria as the outgroup[5–11]. The phylogenetic position of *Korarchaeum* was not consistently resolved in these different analyses and hence is depicted as part of a polytomy. Genome analyses have detected homologous genes in Archaea and eukaryotes that are consistent with them sharing a common ancestor to the exclusion of Bacteria. Many of these patterns of gene sharing do not distinguish between the rooted three-domains or eocyte trees, as they are expected to occur under both hypotheses. Recently published analyses of the genomes of TACK Archaea, however, have increased the number of homologues shared with eukaryotes and some of these are relevant to ideas about eukaryotic origins and the evolution of their unique features. These include putative orthologues of actin[57] and tubulin[58], which in eukaryotes form the core of the cytoskeleton, as well as components of a ubiquitin protein-modification system in *Caldiarchaeum subterraneum*[55]. Distant homologues of some of these genes have also been detected in Euryarchaeota[104,105], but they cluster outside the eukaryote/TACK clade in phylogenetic trees[57,58,106]. We have followed existing usage[58] in distinguishing between the FtsZ-like tubulin family members found in some Archaea and the eukaryote-like tubulin homologue found in *Nitrosoarchaeum*. Several eukaryotic genes involved in transcription and translation have prokaryotic homologues or conserved sequence features that have been found so far only among the TACK Archaea. These include four ribosomal proteins[47], the RNA polymerase subunit RpoG[59], the elongation factor Elf1 (ref. 107), and a short amino acid insertion[108] in the broadly conserved elongation factor 1-α that has only been found in TACK Archaea and eukaryotes as indicated by the vertical bar. Accession numbers and additional details are provided in Supplementary Tables 1 and 2.

process that extends beyond the initial injection of genes provided by the mitochondrial and plastid endosymbionts. From the perspective of ongoing HGT, the existence of any coherent vertical signal for ancient relationships may seem surprising. However, the impact of HGT on the core genes used to reconstruct the tree of life appears rather limited. Although cases of HGT have been reported[69,70], these occur mainly within rather than between domains, and at present there is little evidence that they have generally perturbed inferences of inter-domain relationships[3,12,41,69]. In addition to genuine cases of HGT, poorly fitting phylogenetic models may also lead to disagreements between gene trees[25,26]: recent work has shown that improving the fit of phylogenetic models[48] or integrating the signal from different genes through joint inference of gene and species trees[71,72] can reduce the level of incongruence and the number of inferred HGT events.

The reasons why core genes involved in transcription, translation, and related processes might be transferred (that is, fixed) less frequently than genes for metabolic pathways are currently understood in terms of their degree of functional integration into cells. Their gene products are often found in large subcellular complexes and therefore tend to have more interaction partners than genes for metabolic pathways; as a result, horizontal replacement of these genes is more likely to disrupt important cellular interactions and thus to be opposed by negative selection[66,73,74]. In essence, the universal core might be the largest coherent set of vertically inherited genes that can be tracked across the history of cellular life[3], and as such represents a key resource for tracing the emergence of the eukaryotic cellular lineage. Under the rooted three-domains hypothesis[14], that ancestral lineage is as old as the Archaea. By contrast, the eocyte hypothesis predicts that eukaryotes are a relatively young group because their core genes originated from within the Archaea[18].

The origin of eukaryotes in light of other data

In principle, it might be possible to determine the order of events relevant to eukaryotic origins, or at least to exclude some scenarios, using the fossil and biogeochemical record. However, this record is very incomplete and subject to deep and sometimes heated controversy. The first fossil that is indisputably eukaryotic is of a bangiophyte red alga dated to between 1.2 billion and 720 million years ago[75], but earlier microfossils with a possible eukaryotic origin are found in rocks dated to approximately 1.8 billion

years ago[20]. These data are consistent with molecular dating analyses that place the last common ancestor of eukaryotes at between 1.9 and 1.7 billion years ago[76]. An earlier origin for eukaryotes had been suggested on the basis of the presence of sterane biomarkers in 2.7-billion-year-old rocks[77], but these were subsequently shown to be contaminants from younger rocks[78,79]. An early origin for Archaea has been inferred on the basis of the presence of biological methane, today produced only by methanogenic Euryarchaeota, in rocks that are 3.5 billion years old[80]. Analyses of microfossils and stromatolites—modern versions of which harbour complex bacterial communities[81]—in 3.4-billion-year-old rocks suggest the presence of photosynthetic bacteria[82–84]. Thus, on the data available, Bacteria and Archaea may pre-date eukaryotes in the fossil record by almost 2 billion years.

In light of the uncertainties for dating eukaryotic origins in the geological record, much attention has focused on the historical record revealed by the ultrastructure of the eukaryotic cell and in particular on the timing of the mitochondrial endosymbiosis[85]. When the three primary kingdoms and three-domains tree were originally proposed[1,14] some contemporary eukaryotes called 'archezoans'[85,86] were hypothesized to descend from eukaryotic lineages that never had mitochondria[85,86], providing modern-day evidence for the emergence of nucleated cells before the mitochondrial endosymbiosis. The archezoans included the obligate intracellular parasites Microsporidia and a number of parasitic microaerophilic protists including *Entamoeba*, *Giardia* and *Trichomonas*[85,86]. However, representatives of all of these groups have now been shown to possess a mitochondrial homologue, either a hydrogenosome or mitosome, sharing common ancestry with classical mitochondria[2,87]. These results imply that the mitochondrion was acquired before the radiation of known eukaryotes; therefore, the observation that the mitochondrion descends from an endosymbiotic member of the alphaproteobacteria[64,88] provides strong evidence that the origin of eukaryotes postdates the origin of that bacterial group[2,89]. A relatively late origin of eukaryotes compared to Bacteria is consistent with the best evidence from the geological record and with either the three-domains or eocyte tree rooted on the bacterial stem or within the Bacteria[5–11]. Moreover, if all eukaryotes have both mitochondria and a nucleus, then we can no longer be sure which structure arose first during evolution: in other words, the host cell that acquired the mitochondrion need not have already possessed a nucleus. Indeed, there are now well-argued hypotheses suggesting that the acquisition of the mitochondrion was the key event that sparked the prokaryote-to-eukaryote transition[90,91]. In any case, the failure of the Archezoa hypothesis removes a key obstacle to theories that propose a prokaryotic host for the mitochondrial endosymbiont, including hypotheses that are consistent with the eocyte tree[2].

The origin of eukaryotic cell membranes

The plasma membranes of Bacteria and eukaryotes predominantly contain phospholipids in which fatty acids are covalently bound to *sn*-glycerol-3-phosphate via an ester linkage. By contrast, Archaea—including the few TACK Archaea studied so far—predominantly contain phospholipids with isoprenoid chains linked to *sn*-glycerol-1-phosphate via an ether bond[92,93]. This pattern is most parsimoniously explained on the rooted three-domains tree by inferring a switch to using mainly glycerol isoprenoid ethers along the archaeal stem, with eukaryotes retaining the ancestral type. This transition may have been driven by a need to maintain membrane function at the high temperatures and acidic conditions of the habitats occupied by early Archaea[92,94]. A commonly voiced challenge to the eocyte hypothesis—and all archaeal host models for eukaryotic origins—is how to explain the reversion of the archaeal-host membrane to a bacterial-type plasma membrane.

In fact, most of the genes needed for the synthesis of both types of lipid are common to all three groups, suggesting that neither the transition from ester to ether lipids in the common ancestor of Archaea, nor the subsequent reversion along the eukaryotic stem, would require radical genomic change[95,96]. Archaeal-type ether lipids have been detected in some Bacteria and phospholipids based on *sn*-glycerol-1-phosphate are found in certain endomembrane components of eukaryotes, suggesting that the distinctions among contemporary membranes may not be as sharp as once thought; there is still much to be discovered about the natural diversity of lipid membranes[93,95–98]. Moreover, recent experiments have indicated that artificial membranes containing mixtures of bacterial and archaeal lipids are stable[99], demonstrating the potential for natural mixed-membrane intermediate stages. Given these considerations, the reversion to bacterial-type membranes in eukaryotes might be explained as part of the same process whereby ancestral archaeal pathways were replaced by bacterial equivalents to yield the metabolic similarities observed between Bacteria and contemporary eukaryotes[62–65,95,96,100]. This transition need not have greatly affected membrane function: in the Haloarchaea, which have obtained a large number of bacterial genes by HGT, transporters derived from Bacteria appear to function normally in the archaeal plasma membrane[101].

Conclusions

Ancient phylogenies provide a fascinating window into the distant past, but are difficult to build and interpret – as evidenced by the first thirty years of debate over the tree of life in the era of molecular phylogenetics. Evolutionary biologists now have access to more data and better phylogenetic methods than ever before, although there is still much room for improvement and many uncertainties remain. These caveats apply equally to all attempts to infer ancient relationships, affecting not only the debate over whether the three-domains or eocyte tree best depicts the history of core eukaryotic genes, but also the placement of the universal root[5,9–12,21] and the relationships among major eukaryotic phyla[26,102,103]. The pioneering analyses of molecular sequence data led by Carl Woese and his co-workers culminated in the three-domains tree recognizing the Archaea, Bacteria and Eukaryota as the three primary domains of cellular life. Although evidence of widespread HGT means that no single tree can depict the history of all genes on prokaryotic and eukaryotic genomes, the three-domains tree holds a special place in biology. It appears in most textbooks and reviews, where it is often called the 'universal tree' and the 'tree of life'. But support for the iconic three-domains tree has waned with improvements in phylogenetic methods and taxon sampling. Within the limits of methods and data, a version of the eocyte tree is now the best-supported hypothesis for the origin of the subset of genes that mainly function in translation and appear to be most resistant to HGT. The placement of these genes, and by extension the eukaryotic nuclear lineage, within the Archaea is consistent with only two primary lineages and with hypotheses for a symbiogenic origin for eukaryotes involving an archaeon and one or more bacterial partners. The eocyte tree, if correct, suggests that the TACK Archaea, currently a relatively unexplored group, might contain additional clues as to the origin of complex eukaryotic structures. It also rejects the hypothesis that eukaryotes are a primordial cellular lineage, leaving only two candidate primary domains, Archaea and Bacteria, and it identifies a key piece of the puzzle—the host lineage—in the chimaeric origins of the eukaryotic cell.

Received 17 June; accepted 14 October 2013.

1. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74,** 5088–5090 (1977).
A landmark paper that, together with ref. 4, reported the discovery of the Archaea and discussed its far-reaching implications for early evolution.
2. Embley, T. M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440,** 623–630 (2006).
3. Woese, C. R. On the evolution of cells. *Proc. Natl Acad. Sci. USA* **99,** 8742–8747 (2002).
4. Woese, C. R. & Fox, G. E. The concept of cellular evolution. *J. Mol. Evol.* **10,** 1–6 (1977).
5. Doolittle, W. F. & Brown, J. R. Tempo, mode, the progenote, and the universal root. *Proc. Natl Acad. Sci. USA* **91,** 6721–6728 (1994).
6. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl Acad. Sci. USA* **86,** 9355–9359 (1989).
Together with ref. 7, this paper presented the first evidence for rooting the tree of life on the bacterial stem, but see ref. 5 for a still-relevant discussion of these analyses and other contemporary ideas about early evolution.

7. Gogarten, J. P. *et al.* Evolution of the vacuolar H+-ATPase: implications for the origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **86,** 6661–6665 (1989).
8. Dagan, T., Roettger, M., Bryant, D. & Martin, W. Genome networks root the tree of life between prokaryotic domains. *Genome Biol. Evol.* **2,** 379–392 (2010).
9. Lake, J. A., Skophammer, R. G., Herbold, C. W. & Servin, J. A. Genome beginnings: rooting the tree of life. *Phil. Trans. R. Soc. B* **364,** 2177–2185 (2009).
10. Skophammer, R. G., Servin, J. A., Herbold, C. W. & Lake, J. A. Evidence for a gram-positive, eubacterial root of the tree of life. *Mol. Biol. Evol.* **24,** 1761–1768 (2007).
11. Cavalier-Smith, T. Rooting the tree of life by transition analyses. *Biol. Direct* **1,** 19 (2006).
12. Cox, C. J., Foster, P. G., Hirt, R. P., Harris, S. R. & Embley, T. M. The archaebacterial origin of eukaryotes. *Proc. Natl Acad. Sci. USA* **105,** 20356–20361 (2008).
The first of a series of recent papers demonstrating that analyses of core genes using new phylogenetic models favour the eocyte tree rather than the three-domains tree.
13. Doolittle, W. F. & Zhaxybayeva, O. in *The Prokaryotes: Prokaryotic Biology and Symbiotic Associations* (ed. Rosenberg, E.) (Springer, 2013).
A very clear discussion about the issues facing the integration of phylogenetics and classification given the evidence for extensive lateral gene transfer.
14. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87,** 4576–4579 (1990).
Woese and colleagues present their arguments for the rooted three-domains tree of life.
15. Madigan, M. T., Martingo, J. M., Stahl, D. A. & Clark, D. P. *Brock Biology of Microorganisms* 13th edn (Benjamin Cummings, 2010).
16. Pace, N. R. Time for a change. *Nature* **441,** 289 (2006).
17. Pace, N. R. Mapping the tree of life: progress and prospects. *Microbiol. Mol. Biol. Rev.* **73,** 565–576 (2009).
18. Lake, J. A., Henderson, E., Oakes, M. & Clark, M. W. Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl Acad. Sci. USA* **81,** 3786–3790 (1984).
This paper presents comparisons of ribosomal structure in Bacteria, Archaea and eukaryotes, providing the initial motivation for the eocyte hypothesis.
19. Gribaldo, S., Poole, A. M., Daubin, V., Forterre, P. & Brochier-Armanet, C. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nature Rev. Microbiol.* **8,** 743–752 (2010).
20. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. R. Soc. B* **361,** 1023–1038 (2006).
21. Philippe, H. & Forterre, P. The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49,** 509–523 (1999).
22. Foster, P. G., Cox, C. J. & Embley, T. M. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Phil. Trans. R. Soc. B* **364,** 2197–2207 (2009).
23. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* **53,** 711–723 (2001).
24. Ho, S. Y. & Jermiin, L. Tracing the decay of the historical signal in biological sequence data. *Syst. Biol.* **53,** 623–637 (2004).
25. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7** (suppl. 1), S4 (2007).
26. Philippe, H. *et al.* Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* **9,** e1000602 (2011).
27. Gouy, M. & Li, W. H. Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. *Nature* **339,** 145–147 (1989).
28. Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51,** 221–271 (1987).
29. Olsen, G. J. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harb. Symp. Quant. Biol.* **52,** 825–837 (1987).
30. Foster, P. G. & Hickey, D. A. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* **48,** 284–290 (1999).
31. Foster, P. G. Modeling compositional heterogeneity. *Syst. Biol.* **53,** 485–495 (2004).
32. Hirt, R. P. *et al.* Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96,** 580–585 (1999).
33. Lake, J. A. Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc. Natl Acad. Sci. USA* **91,** 1455–1459 (1994).
34. Yang, Z. & Roberts, D. On the use of nucleic acid sequences to infer early branchings in the tree of life. *Mol. Biol. Evol.* **12,** 451–458 (1995).
An important early contribution demonstrating that modelling changing nucleotide composition in RNA sequences from different species supported the eocyte tree.
35. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27,** 401–410 (1978).
36. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nature Rev. Genet.* **13,** 303–314 (2012).
37. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331,** 184–186 (1988).
38. Sidow, A. & Wilson, A. C. Compositional statistics: an improvement of evolutionary parsimony and its application to deep branches in the tree of life. *J. Mol. Evol.* **31,** 51–68 (1990).
39. Tourasse, N. J. & Gouy, M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phylogenet. Evol.* **13,** 159–168 (1999).
40. Yutin, N., Makarova, K. S., Mekhedov, S. L., Wolf, Y. I. & Koonin, E. V. The deep archaeal roots of eukaryotes. *Mol. Biol. Evol.* **25,** 1619–1630 (2008).
41. Harris, J. K., Kelley, S. T., Spiegelman, G. B. & Pace, N. R. The genetic core of the universal ancestor. *Genome Res.* **13,** 407–412 (2003).
42. Katoh, K., Kuma, K. & Miyata, T. Genetic algorithm-based maximum-likelihood analysis for molecular phylogeny. *J. Mol. Evol.* **53,** 477–484 (2001).
43. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311,** 1283–1287 (2006).
44. Lake, J. A. The order of sequence alignment can bias the selection of tree topology. *Mol. Biol. Evol.* **8,** 378–385 (1991).
45. Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28,** 281–285 (2001).
46. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21,** 1095–1109 (2004).
One of the most notable improvements in phylogenetic modelling in the last decade, providing a Bayesian framework for accommodating across-site compositional heterogeneity—a key feature of molecular sequence data.
47. Guy, L. & Ettema, T. J. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol.* **19,** 580–587 (2011).
48. Williams, T. A., Foster, P. G., Nye, T. M., Cox, C. J. & Embley, T. M. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. R. Soc. Lond. B* **279,** 4870–4879 (2012).
49. Lasek-Nesselquist, E. & Gogarten, J. P. The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol. Phylogenet. Evol.* **69,** 17–38 (2013).
50. Pester, M., Schleper, C. & Wagner, M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* **14,** 300–306 (2011).
51. Lloyd, K. G. *et al.* Predominant archaea in marine sediments degrade detrital proteins. *Nature* **496,** 215–218 (2013).
52. Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47,** 9–17 (1998).
53. Elkins, J. G. *et al.* A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl Acad. Sci. USA* **105,** 8102–8107 (2008).
54. Brochier-Armanet, C., Boussau, B., Gribaldo, S. & Forterre, P. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Rev. Microbiol.* **6,** 245–252 (2008).
55. Nunoura, T. *et al.* Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39,** 3204–3223 (2011).
56. Kelly, S., Wickstead, B. & Gull, K. Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. R. Soc. Lond. B* **278,** 1009–1018 (2011).
57. Ettema, T. J., Lindas, A. C. & Bernander, R. An actin-based cytoskeleton in archaea. *Mol. Microbiol.* **80,** 1052–1061 (2011).
58. Yutin, N. & Koonin, E. V. Archaeal origin of tubulin. *Biol. Direct* **7,** 10 (2012).
59. Koonin, E. V., Makarova, K. S. & Elkins, J. G. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol. Direct* **2,** 38 (2007).
60. Csurös, M. & Miklos, I. Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic birth-and-death model. *Mol. Biol. Evol.* **26,** 2087–2095 (2009).
61. Wolf, Y. I., Makarova, K. S., Yutin, N. & Koonin, E. V. Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* **7,** 46 (2012).
62. Ribeiro, S. & Golding, G. B. The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15,** 779–788 (1998).
Together with ref. 63, this paper presented some of the first tree-based evidence that eukaryotes are genomic chimaeras containing some genes that are most similar to those of Bacteria and others to Archaea.
63. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95,** 6239–6244 (1998).
64. Esser, C. *et al.* A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21,** 1643–1660 (2004).
65. Alsmark, C. *et al.* Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. *Genome Biol.* **14,** R19 (2013).
66. Cotton, J. A. & McInerney, J. O. Eukaryotic genes of archaebacterial origin are more important than the more numerous eubacterial genes, irrespective of function. *Proc. Natl Acad. Sci. USA* **107,** 17252–17255 (2010).
67. Dagan, T. & Martin, W. The tree of one percent. *Genome Biol.* **7,** 118 (2006).
68. Doolittle, W. F. & Bapteste, E. Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl Acad. Sci. USA* **104,** 2043–2049 (2007).
69. Williams, D. *et al.* A rooted net of life. *Biol. Direct* **6,** 45 (2011).
70. Creevey, C. J., Doerks, T., Fitzpatrick, D. A., Raes, J. & Bork, P. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. *PLoS ONE* **6,** e22099 (2011).
71. Boussau, B. *et al.* Genome-scale coestimation of species and gene trees. *Genome Res.* **23,** 323–330 (2013).
72. Szollősi, G. J., Boussau, B., Abby, S. S., Tannier, E. & Daubin, V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc. Natl Acad. Sci. USA* **109,** 17513–17518 (2012).
73. Cohen, O., Gophna, U. & Pupko, T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol. Biol. Evol.* **28,** 1481–1489 (2011).

74. Jain, R., Rivera, M. C. & Lake, J. A. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 3801–3806 (1999).
75. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26,** 386–404 (2000).
76. Parfrey, L. W., Lahr, D. J., Knoll, A. H. & Katz, L. A. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* **108,** 13624–13629 (2011).
77. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. *Science* **285,** 1033–1036 (1999).
78. Rasmussen, B., Fletcher, I. R., Brocks, J. J. & Kilburn, M. R. Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* **455,** 1101–1104 (2008).
79. Fischer, W. W. Biogeochemistry: life before the rise of oxygen. *Nature* **455,** 1051–1052 (2008).
80. Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era. *Nature* **440,** 516–519 (2006).
81. Papineau, D., Walker, J. J., Mojzsis, S. J. & Pace, N. R. Composition and structure of microbial communities from stromatolites of Hamelin Pool in Shark Bay, Western Australia. *Appl. Environ. Microbiol.* **71,** 4822–4832 (2005).
82. Allwood, A. C. *et al.* Controls on development and diversity of Early Archean stromatolites. *Proc. Natl Acad. Sci. USA* **106,** 9548–9555 (2009).
83. Tice, M. M. & Lowe, D. R. Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* **431,** 549–552 (2004).
84. Schopf, J. W. Fossil evidence of Archaean life. *Phil. Trans. R. Soc. B* **361,** 869–885 (2006).
85. Cavalier-Smith, T. Eukaryotes with no mitochondria. *Nature* **326,** 332–333 (1987).
86. Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H.E.A.) 1027–1034 (De Gruyter, 1983).
87. van der Giezen, M., Tovar, J. & Clark, C. G. Mitochondria-derived organelles in protists and fungi. *Int. Rev. Cytol.* **244,** 175–225 (2005).
88. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396,** 133–140 (1998).
89. Horner, D. S., Hirt, R. P., Kilvington, S., Lloyd, D. & Embley, T. M. Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc. R. Soc. Lond. B* **263,** 1053–1059 (1996).
90. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467,** 929–934 (2010).
91. Martin, W. & Koonin, E. V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440,** 41–45 (2006).
92. Lombard, J., Lopez-Garcia, P. & Moreira, D. The early evolution of lipid membranes and the three domains of life. *Nature Rev. Microbiol.* **10,** 507–515 (2012).
93. Pitcher, A. *et al.* Core and intact polar glycerol dibiphytanyl glycerol tetraether lipids of ammonia-oxidizing archaea enriched from marine and estuarine sediments. *Appl. Environ. Microbiol.* **77,** 3468–3477 (2011).
94. van de Vossenberg, J. L., Driessen, A. J. & Konings, W. N. The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2,** 163–170 (1998).
95. Boucher, Y., Kamekura, M. & Doolittle, W. F. Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Mol. Microbiol.* **52,** 515–527 (2004).
96. Lombard, J., Lopez-Garcia, P. & Moreira, D. An ACP-independent fatty acid synthesis pathway in archaea: implications for the origin of phospholipids. *Mol. Biol. Evol.* **29,** 3261–3265 (2012).
97. Guldan, H., Matysik, F. M., Bocola, M., Sterner, R. & Babinger, P. Functional assignment of an enzyme that catalyzes the synthesis of an archaea-type ether lipid in bacteria. *Angew. Chem. Int. Edn Engl.* **50,** 8188–8191 (2011).
98. Tan, H. H., Makino, A., Sudesh, K., Greimel, P. & Kobayashi, T. Spectroscopic evidence for the unusual stereochemical configuration of an endosome-specific lipid. *Angew. Chem. Int. Edn Engl.* **51,** 533–535 (2012).
99. Shimada, H. & Yamagishi, A. Stability of heterochiral hybrid membrane made of bacterial *sn*-G3P lipids and archaeal *sn*-G1P lipids. *Biochemistry* **50,** 4114–4120 (2011).
Reports the production of stable heterochiral membranes containing a mixture of bacterial- and archaeal-type lipids, demonstrating the feasibility of natural mixed membranes.
100. Martin, W. & Muller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392,** 37–41 (1998).
101. Nelson-Sathi, S. *et al.* Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc. Natl Acad. Sci. USA* **109,** 20537–20542 (2012).
102. Hampl, V. *et al.* Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc. Natl Acad. Sci. USA* **106,** 3859–3864 (2009).
103. Song, S., Liu, L., Edwards, S. V. & Wu, S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl Acad. Sci. USA* **109,** 14942–14947 (2012).
104. Lindås, A. C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J. & Bernander, R. A unique cell division machinery in the Archaea. *Proc. Natl Acad. Sci. USA* **105,** 18942–18946 (2008).
105. Makarova, K. S., Yutin, N., Bell, S. D. & Koonin, E. V. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nature Rev. Microbiol.* **8,** 731–741 (2010).
106. Blombach, F. *et al.* Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea. *Biol. Direct* **4,** 39 (2009).
107. Daniels, J. P., Kelly, S., Wickstead, B. & Gull, K. Identification of a crenarchaeal orthologue of Elf1: implications for chromatin and transcription in Archaea. *Biol. Direct* **4,** 24 (2009).
108. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257,** 74–76 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by a Marie Curie postdoctoral fellowship to T.A.W. T.M.E. acknowledges support from the European Research Council Advanced Investigator Programme and the Wellcome Trust. We thank J. Archibald for comments on the manuscript.

Author Contributions T.A.W., P.G.F., C.J.C. and T.M.E. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.E. (Martin.Embley@ncl.ac.uk).