Rooting the tree of life: the phylogenetic jury is still out

Richard Gouy1,2, Denis Baurain1 and Hervé Philippe2,3

1Eukaryotic Phylogenomics, Department of Life Sciences and PhytoSYSTEMS, University of Liège, Liège 4000, Belgium
2Centre for Biodiversity Theory and Modelling, USR CNRS 2936, Station d’Ecologie Expérimentale du CNRS, Moulis 09200, France
3Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Quebec, Canada H3C 3J7

This article aims to shed light on difficulties in rooting the tree of life (ToL) and to explore the (sociological) reasons underlying the limited interest in accurately addressing this fundamental issue. First, we briefly review the difficulties plaguing phylogenetic inference and the ways to improve the modelling of the substitution process, which is highly heterogeneous, both across sites and over time. We further observe that enriched taxon samplings, better gene samplings and clever data removal strategies have led to numerous revisions of the ToL, and that these improved shallow phylogenies nearly always relocate simple organisms higher in the ToL provided that long-branch attraction artefacts are kept at bay. Then, we note that, despite the flood of genomic data available since 2000, there has been a surprisingly low interest in inferring the root of the ToL. Furthermore, the rare studies dealing with this question were almost always based on methods dating from the 1990s that have been shown to be inaccurate for much more shallow issues! This leads us to argue that the current consensus about a bacterial root for the ToL can be traced back to the prejudice of Aristotle’s Great Chain of Beings, in which simple organisms are ancestors of more complex life forms. Finally, we demonstrate that even the best models cannot yet handle the complexity of the evolutionary process encountered both at shallow depth, when the outgroup is too distant, and at the level of the inter-domain relationships. Altogether, we conclude that the commonly accepted bacterial root is still unproven and that the root of the ToL should be revisited using phylogenomic supermatrices to ensure that new evidence for eukaryogenesis, such as the recently described Lokiarcheota, is interpreted in a sound phylogenetic framework.

1. Introduction

Knowledge of the history of organisms is a prerequisite for the study of any evolutionary question. This explains why the evolutionary community has always been so committed to inferring phylogenies, resulting in a flood of species trees whenever new phylogenetic approaches were made available (e.g. cladistics in the 1960s; molecular data in the 1980s). More recently, the combined advances in sequencing technologies and computational methods have given a new impetus to the phylogenetic endeavour, as evidenced by the numerous studies trying to reconstruct (various parts of) the tree of life (ToL). At this point, it should be mentioned that phylogeny is only an approximation of the history of organisms. Several mechanisms are known to create full reticulations in species trees, including hybridization of related species, which is a recurrent phenomenon in numerous lineages such as flowering plants, and symbiogenesis (endosymbiosis of plastids and mitochondria), first suggested in 1905 by Mereschkowsky, albeit widely accepted only in the 1980s. Yet, exactly as Newton’s law of universal gravitation is a very powerful approximation, phylogeny remains extremely useful, especially to display evolutionary relatedness, though taking into account major reticulations, such as the α-proteobacterial origin of the mitochondrion,
invariably leads to cycles (or 'rings') in the ToL. Another mechanism, horizontal gene transfer (HGT), probably plays an important role in evolution (e.g. by allowing rapid adaptation) while creating partial reticulations. Even if the latter is more difficult to display on bifurcating trees, HGT events are several orders of magnitude less frequent than vertical gene transmission (VGT). In our opinion, this justifies sticking to phylogeny as the best synthetic representation of the history of organisms [1], with horizontal gene flows shown as superimposed thin lines when really massive, such as probably for hyperthermophilic bacteria [2–4].

A surprising contrast appears when comparing scientific inquiries on shallow and deep phylogenetic questions. Obviously, there are many more publications on genus-level phylogenies than on domain-level phylogenies, simply because the former are much more numerous than the latter. Hence, there are more ongoing debates about, for example, the sister group of land plants or the root of the animal tree than about intra-domain phylogenies (in particular, Bacteria) and the root of the ToL. Nevertheless, just like reconstructing the lifestyle of Magdalenians is more difficult than studying the habits of the Victorian era, inferring the deepest branches of the ToL is highly problematic. Consequently, this issue should still be a very hot topic, a topic that can be tackled only by the application of the most sophisticated and up-to-date methodology. Yet, a reality check shows that it is not the case. Instead, one of the most frequently cited references on the matter is a 25-year-old paper by Carl Woese and co-workers [6] (more than 3200 citations in Web of Science). For instance, the recent article describing the fascinatingly complex Lokiarchaeota [5] interprets them as an intermediate stage of the eukaryogenetic process, based on the bacterial rooting of the ToL that was inexplicably set in stone by that paper of Woese's tree (their fig. 1) also shows Microsporidia as the sister group of all remaining eukaryotes. Therefore, genomic data of the microsporidium *Mitosporidium daphniae*, especially its mitochondrial genome [7], could be, according to the same principle, interpreted as evidence that *Mitosporidium* represents an intermediate stage in the simplification of an ancestrally simple microsporidium into a complex eukaryote. However, thanks to their awareness of more recent references accounting for the heated debate that eventually led to recognize Microsporidia as Fungi [8,9], Haag and co-workers instead correctly interpret *Mitosporidium* as an intermediate stage in the simplification of a complex ancestral fungus into a simple microsporidium. Likewise, the complex Lokiarchaeota would be better interpreted as an intermediate stage in the simplification of a complex ancestral fungus into a simple microsporidium. Hence, a eukaryotic rooting is compatible with an ancestral (now extinct) prokaryotic life form.

In this paper, we first review the technical difficulties hindering phylogenetic inference as well as the recent methodological progresses on the matter, using the relatively recent (shallow) evolution of animals as our main case in point. Then, we explore the (sociological) reasons underlying the limited interest in accurately solving the rooting of the ToL, which is nonetheless fundamental to our understanding of prokaryogenesis and eukaryogenesis. Finally, we explore the potential avenues for a resolution of this issue.

### 2. The complexity of the evolutionary process makes phylogenetic inference difficult

A striking characteristic of phylogenetics, especially irritating for non-specialists, is that the ToL ‘evolves’ (i.e. the names and contents of clades change over time) and that several mutually incompatible solutions often coexist over long periods of time. The simplest explanation is that phylogenetics is an active field of science, which in itself is a positive fact. Importantly, contrary to the naive, yet commonly held view that open problems eventually get solved through the accumulation of more sequence data, incongruencies still persist in the genomic era (e.g. for streptophytes [11–16] or Bilateria [17–22]). Indeed, while phylogenomics helps in decreasing stochastic error (due to small sample sizes), it actually makes systematic error more apparent. Systematic error stems from methodological biases (i.e. model violations in a probabilistic framework) that cause the inference to converge towards an incorrect solution as more and more data are added. The most well-known case of this phenomenon is the infamous long-branch attraction (LBA) artefact, which was originally formalized to demonstrate the inconsistency of maximum parsimony when branch lengths are sufficiently unequal [23]. And even today, in spite of the widespread use of sophisticated methods and evolutionary models, numerous incongruencies in phylogenomics are still associated with long branches, corresponding to fast-evolving lineages and/or distant outgroups (e.g. Nematoda [24–29], Ctenophora [22,30,31] or Zygnematales [11,12,15,16]).

This difficulty is due to the formidable complexity of the underlying evolutionary process. Hence, all existing models, even the most sophisticated and computationally demanding ones, remain dramatically oversimplified. Phylogenetic inference can be schematically separated into three steps: (1) homology assessment, i.e. identifying (1a) homologous genes through database similarity searches and (1b) homologous positions through multiple alignment; (2) modelling of the substitution process, in order to detect the multiple substitutional events falling at the same positions (i.e. estimating the probabilities of mutation and fixation) and to infer the gene tree; and (3) inference of the species tree from the gene trees, i.e. taking into account incomplete lineage sorting (ILS), HGT and gene duplication/conversion. In theory, the three steps should be performed simultaneously, but this is computationally intractable (see the article of N. Lartillot in this issue [32]). In practice, they are thus performed separately, even if a few software packages are available for the joint inference of steps (1) and (2) [33,34] or steps (2) and (3) (see the article of B. Boussau and colleagues in this issue [35]). Nevertheless, computational limits imply that the joint evaluation of two or more steps is performed at the expense of using relatively simple methods within each step. For instance, the PHYLDAG software uses both a simplistic substitution model (homogeneous over time and across sites) and an incomplete gene history model (e.g. no gene conversion) [36]. To our knowledge, the relative performance of these joint approaches and of
the well-established supermatrix methods (which assume that steps (1) and (3) have been already solved) has not yet been carefully evaluated, in particular for ancient questions. Our bet is that the assessment of homologous characters (especially thanks to the removal of ambiguously aligned regions) and of orthologous genes is relatively accurate and does not constitute the most important issue in deep phylogenetics. In addition, supermatrix-based inference appears to be robust to the inclusion of paralogous [37] and xenologous (i.e. horizontally transferred) sequences (unpublished results), but sensitive to the substitution model (see below).

Therefore, from now on, we focus on the supermatrix approach (which we consider as the best one currently available, even if we acknowledge its limitations) and on the modelling of the substitution process.

3. Progress in modelling the heterogeneities of the substitution process

It is necessary to model the substitution process because, at geological timescales, successive substitutions at the same position...
are the rule. These multiple substitutions first blur then erase and rewrite the original phylogenetic signal, and the resulting homoplasy prevents naive methods, such as similarity-based distances and maximum parsimony, from being consistent. Unfortunately, the substitution process is highly heterogeneous, both across sites and over time, thus making its efficient modelling particularly difficult. First, the mutational process varies across positions (e.g. the hypermutable methylated CpG) and over time (due to e.g. evolutionary changes in the efficiency of the DNA repair machinery). Second, and probably more importantly, the fixation probability of any given possible mutation also varies across sites, owing to functional constraints on the encoded products, and over time, mainly because of variable effective population size, changes in epistasis and variable environment.

The very first substitution model ever developed [38] made numerous assumptions of homogeneity and independence that simplified computation, only branch lengths being heterogeneous (i.e. the global substitution rate was allowed to vary). Since then, three major and three minor, yet significant, improvements have been proposed:

1. **Heterogeneity of substitutions among character states.** Some substitutions are obviously easier than others (e.g. transitions versus transversions or Asp → Glu versus Asp → Trp) and exchangeability matrices were rapidly introduced [39]. The General-Time-Reversible (GTR) model is now widely used for nucleotides, where it only requires eight parameters, but much less for amino acids because then it requires 208 parameters. Yet, when datasets are large, an amino acid GTR matrix has a better fit than empirical matrices (e.g. WAG and LG) [40].

2. **Heterogeneity of the substitution rate across sites.** Following the seminal observations of Uzzell & Corbin [41], various methods have been developed to handle the fact that some sites are more susceptible than others to accumulation of substitutions, and thus to generate artefacts. The gamma distribution appears as a good compromise between computational efficiency and biological realism. That is why it is now widely used. More refined models (such as mixture of gamma or Dirichlet processes) might nevertheless prove to be useful for solving difficult questions.

3. **Heterogeneity of the substitution process across sites.** The fact that only a few amino acids are possible at a given position (e.g. charged or hydrophobic amino acids) was established by biochemists a long time ago, but it has attracted the attention of phylogeneticists only recently [42,43]. This is surprising because the efficiency of the detection of multiple substitutions is much higher when the number of possible character states is reduced [25]. CAT-like models [43] use a Dirichlet process to affiliate individual sites to different CATegories defined by their character state frequencies. With hundreds to thousands of categories usually inferred in *a posteriori* analyses, the observed heterogeneity is very high, demonstrating both the biological relevance and the statistical significance of accounting for this aspect of the evolutionary process. As expected, the CAT–GTR model, and to a lesser extent the CAT model, has a much better fit to data, provided that a few thousand sites are considered. Accordingly, these models are also less sensitive to homoplasy and LBA artefacts [22,25].

4. **Separation of mutation and selection steps.** Codon models were proposed as early as 1994 [44,45]. Owing to their mechanistic modelling that contrasts with the phenomenological modelling of all other protein models, they are biologically more realistic. Yet, their computational slowness (due to the $61 \times 61$ matrix), combined with numerous simplifying assumptions, so far has limited their usefulness for phylogenetic inference. Nevertheless, recent improvements, in particular their coupling with the CAT model [46], make them promising.

5. **Heterogeneity of composition over time.** The existence of a compositional bias and its implication in reconstruction artefacts was also identified more than 20 years ago, based on ribosomal RNA alignments [47–49]. Various modelling approaches [47,50,51] have been proposed, but these are often computationally demanding. However, since the compositional bias is dominant at large evolutionary scales, it is better to address it when inferring deep phylogenies [8,52].

6. **Heterogeneity of rates within positions over time.** Because of epistasis, the probability of accepting a mutation at any given position is expected to vary along the branches of the tree, as demonstrated early on by Fitch & Markowitz [53]. In the 1990s, a renewed interest in the so-called ‘heterotachy’ led to the development of multiple models [54–56]. Surprisingly, however, the increase in statistical fit, albeit systematic, is not very important, and their impact on topology rather marginal [57].

Despite these significant improvements, incorrect phylogenies keep being published due to uncontrolled artefacts. This is because many problems remain to be solved. First, not all these improvements are jointly incorporated into a single model, the best models combining at most four out of six improvements at the expense of being tractable only for small datasets [50]. Since the first three are included in PHYLOBAYES [58], it is probably the most accurate and computationally tractable software available today. Second, numerous improvements are still needed to address the full spectrum of biological complexity. For instance, heterocepiel (the change of the substitution process at a position over time) is known to make the CAT model inconsistent [59]. Another example is the non-independence of sites, with the few models relaxing this assumption showing a better fit to data [60]. Importantly, future models should not try to account for all the subtleties of the evolutionary process but instead focus on the heterogeneities that are the most prone to generate phylogenetic artefacts.

4. **Improved phylogenies support organismal simplification at shallow depth**

These methodological improvements, along with enriched taxon samplings (sometimes the only way to avoid artefacts), better gene samplings and clever data removal strategies, have led to numerous revisions of the ToL, especially at an intermediate evolutionary scale (e.g. within Metazoa [17,18,61,62]). Strikingly, a major trend is visible in these revised phylogenies: morphologically simple organisms, once considered as akin to ancestral intermediates (‘living fossils’) in a gradual rise towards complex organisms, are often relocated within groups of complex organisms, thus implying that their simplicity is not primitive but secondarily
derived. In eukaryotes, ‘Archezoa’ (e.g. Microsporidia, Diplomonadida and Trichomonadida), which had been first recovered at the base of the rrNA tree [6], in apparent agreement with their lack of a mitochondrion, eventually turned out to be located (much) higher in the tree [8,9] and to possess degenerated mitochondria [63]. In animals, the very simple Myxozoa appear to be closely related to Medusozoa [64], while acelomate Platyhelminthes [24,25,27–29] and Acoelomorpha [65] have been shown to be closely related to Lophotrochozoa and Ambulacaria, respectively. Moreover, the mostly dull Urochordata are more closely related to Vertebrata than are the more complex Cephalochordata [66]. For all these phylogenetic errors, the methodological explanation is the same: morphological simplification is generally accompanied by an acceleration of the molecular evolutionary rate and by qualitative shifts in the substitution process. When simple models are used, this situation generates artefacts where the long branch of the (often distant) outgroup attracts the long branch of the simplified organisms, which erroneously results in a too basal location of the latter in the trees.

5. Deep phylogenetics and the prejudice of Aristotle’s Great Chain of Beings

This rapid overview of relatively recent phylogenies (i.e. within Eukaryota, which corresponds to a sub-clade of alpha-Proteobacteria, itself a sub-clade of Proteobacteria, itself a sub-clade of Bacteria) demonstrates that sophisticated approaches (and especially substitution models handling multiple heterogeneities) are mandatory for accurate phylogenetic inference and that morphologically simple organisms are the most difficult to correctly locate. These results have profound implications for deep phylogenies, which are by essence much more difficult to infer due to increased noise (more multiple substitutions, HGTs and heterogeneities) and to decreased signal (less homologous positions). Consequently, artefacts are much more likely to occur, especially when trying to position the simple prokaryotes (Archaea and Bacteria) with respect to Eukaryota.

Surprisingly, despite the flood of genomic data available since 2000, there has been almost no interest in inferring the root of the ToL (a dozen papers [67]) and only limited interest in the relationships within Bacteria and Archaea. More puzzlingly, with a few notable exceptions [52,68], these studies were almost always based on methods dating from the 1990s that have been shown to be inaccurate for much more recent questions! While a careful sociological study would be required to understand this baffling behaviour, our opinion is that it stems from the subliminal prevalence of Aristotle’s Great Chain of Beings, reinforced by the progressivism of the Age of Enlightenment, and from humans’ inclination for trends and ‘stories that go somewhere’, as pointed out by Gould [69]. An illustration of the strength of this prejudice is the recurrent use of scale-related wordings such as ‘higher plants’ or ‘lower animals’, a few per cent of manuscripts submitted to evolutionary journals comprising this inappropriate terminology (H. Philippe 2015, unpublished data). Another one is that assertions such as ‘eukaryotes arose from prokaryotes’ [70] are commonplace, whereas the evidence for this stance is both scarce and weak [10].

Aristotle’s prejudice is constantly revived by the fact that language shapes thought [71], an idea also known as the linguistic relativity principle (or Sapir–Whorf hypothesis) and that can be traced back to Wilhelm von Humboldt [72]. In particular, the words ‘prokaryotes’ (before nucleus) and ‘eukaryotes’ (true nucleus) make us more prone to accept that the former have preceded the latter, and thus to focus our attention on the origin of eukaryotes. Pace has made much of the idea that the word ‘prokaryote’ imposes a certain temporal directionality on the prokaryote/eukaryote dichotomy [73,74]. Two concepts were initially distinguished within the prokaryote–eukaryote dichotomy when R. Y. Stanier and C. B. van Niel introduced the concept of prokaryote in the early 1960s. The first one was organizational and referred to comparative cell structure, whereas the second one was phylogenetic and referred to a natural classification of the living world [75,76]. Thus, the definition of prokaryote is blurred. Do prokaryotes lump extant organisms without nuclear membranes (Archaea and Bacteria)? Or do they refer to some long-gone ancestors of eukaryotes? These are two different matters [77]. The last one is misleading for it gives a direction to evolution and allows us to think that extant eukaryotes emerged from ‘prokaryotes’ that still exist, so that eukaryotes are more ‘evolved’ than prokaryotes. As a case in point, searching for ‘eukaryogenesis’ in PubMed returns 53 articles (as of May 2015), while the related terms ‘prokaryogenesis’, ‘bacteriogenesis’ and ‘archaeogenesis’ do not yield any result. This is significant because, whatever the correct theory is, both eukaryogenesis and prokaryogenesis (including bacteriogenesis and archaeogenesis) have occurred during the evolution of life on Earth. Therefore, only a scenario that adequately addresses the two issues would be completely satisfactory. Indeed, the temptation to justify the lack of research about prokaryogenesis by equating the latter to the origin of the living cell not only takes the prefix ‘pro’ of prokaryotes in the literal meaning, but also lends credit to the mistaken view that contemporaneous Bacteria and Archaea are long-standing intermediate stages (i.e. surviving stem groups) on the path to Eukaryota.

To become aware of how wording reinforces Aristotle’s prejudice, it is insightful to fantasize an alternative history of science, in which Édouard Chatton would not have coined the name ‘Prokaryota’. Instead, let us imagine that, impressed by the works of Mesenchkovsky on endosymbiosis and of Lwoff [78] on simplification in unicellular organisms, he would have proposed the evolutionary scheme shown in figure 1. Assuming that simple cells devoid of nucleus were derived from complex nucleated cells, he would have named them ‘Apokaryota’. Moreover, building on the idea that extant nucleated cells diversified after the mitochondrial endosymbiosis, he would have named the latter ‘Mitochondriophora’, reserving the names ‘Karyota’ for the common hybrid between prokaryote and eukaryote. Following this new terminology, we would not have prevented some researchers from arguing that Apokaryota are in fact ancestral to Mitochondriophora, exactly as some have proposed that Eukaryota actually preceded Prokaryota, the burden of the proof being just transferred on different shoulders.
6. On the persistent use of simple methods in deep phylogenetics

By looking at the phylogenetic studies published over the years, we are under the impression that the community shows a disproportionate interest in using ever more sequence data compared to using improved methods. Moreover, as aforementioned, this trend appears stronger for colleagues studying deep phylogenetic issues than for those interested in shallower questions. To flesh out this intuition, we searched Web of Science for phylogenetic studies published since 2005 and addressing either shallow or deep evolutionary issues. Our exact queries were ‘phylogenet* AND metazoa*’ and ‘phylogenom* AND (Bacteria OR Archaea)’, respectively. After a first screening of the numerous irrelevant articles, this allowed us to download two sets of PDF files: 93 about shallow phylogenies and 137 about deep phylogenies. We then examined each paper in turn to determine: (1) whether it was relevant for establishing our statistics about tree reconstruction practices; (2) whether the authors demonstrated an awareness of possible artefacts (through the use of keywords such as ‘long-branch attraction/LBA’, ‘artifact/artefact’, ‘non-phylogenetic signal’, ‘systematic error’, ‘homoplasy’, ‘saturation’); and (3) whether they had tried to reduce the systematic error by applying one of the three well-known approaches summarized in, for example, Philippe et al. [22]. As a reminder, these strategies are: (3a) varying the taxon sampling (e.g. inferring phylogenies with and without outgroups and/or fast-evolving lineages, replacing rogue organisms by slow-evolving relatives), (3b) removing fast-evolving (and/or biased) sites, based on preliminary rate or compositional analyses and (3c) using sophisticated substitution models (defined here as models heterogeneous across sites, such as CAT-like models, or over time, such as heterotachous/covarion models). The results of this quick bibliographic survey, limited to the relevant studies (69 ‘shallow studies’ and 57 ‘deep studies’), are shown in table 1 (see also the electronic supplementary material, tables S1 and S2 for individual paper analyses).

Strikingly, less than half the studies showed awareness of possible artefacts. In particular, only 36% (25/69) of the publications dealing with shallow phylogenies mentioned any of the key words of our list, while the situation was slightly worse for papers about deep phylogenetic issues (16/57 = 28%). Among the ‘shallow studies’ that effectively cared for artefacts, 76% (19/25) tried to do something to reduce the systematic error, a figure that was similar among ‘deep studies’ (13/16 = 81%). In both cases, the most common strategy was to use a heterogeneous substitution model (15/25 = 60% and 10/16 = 62%), an efficient approach that is also the easiest to implement. By contrast, site removal strategies were more often applied in ‘shallow studies’ (12/25 = 48%) than in ‘deep studies’ (6/16 = 37%), whereas varying the taxon sampling was three times more explored in ‘shallow studies’ (9/25 = 36%) than in ‘deep studies’ (2/16 = 12%), a low figure that might be due to the lack of alternative outgroups at the domain level. Interestingly, six publications (24%) dealing with shallow phylogenies did use the three approaches for controlling the artefacts, while only one publication (6%) trying to infer the ToL [80] was equally comprehensive according to our criteria. Altogether, our modest survey confirmed our initial intuition and indicated that there was room for improvement in deep phylogenetic inference without the need for any additional methodological

<table>
<thead>
<tr>
<th>no. studies</th>
<th>artefact awareness</th>
<th>taxon sampling</th>
<th>site removal</th>
<th>heterogeneous model</th>
</tr>
</thead>
<tbody>
<tr>
<td>shallow phylogenies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>4</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>1</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>6</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>69</td>
<td>25</td>
<td>9</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>deep phylogenies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>2</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>1</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>1</td>
<td>y</td>
<td>y</td>
<td>y</td>
<td>y</td>
</tr>
<tr>
<td>57</td>
<td>16</td>
<td>2</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

6. On the persistent use of simple methods in deep phylogenetics

By looking at the phylogenetic studies published over the years, we are under the impression that the community shows a disproportionate interest in using ever more sequence data compared to using improved methods. Moreover, as aforementioned, this trend appears stronger for colleagues studying deep phylogenetic issues than for those interested in shallower questions. To flesh out this intuition, we searched Web of Science for phylogenetic studies published since 2005 and addressing either shallow or deep evolutionary issues. Our exact queries were ‘phylogenet* AND metazoa*’ and ‘phylogenom* AND (Bacteria OR Archaea)’, respectively. After a first screening of the numerous irrelevant articles, this allowed us to download two sets of PDF files: 93 about shallow phylogenies and 137 about deep phylogenies. We then examined each paper in turn to determine: (1) whether it was relevant for establishing our statistics about tree reconstruction practices; (2) whether the authors demonstrated an awareness of possible artefacts (through the use of keywords such as ‘long-branch attraction/LBA’, ‘artifact/artefact’, ‘non-phylogenetic signal’, ‘systematic error’, ‘homoplasy’, ‘saturation’); and (3) whether they had tried to reduce the systematic error by applying one of the three well-known approaches summarized in, for example, Philippe et al. [22]. As a reminder, these strategies are: (3a) varying the taxon sampling (e.g. inferring phylogenies with and without outgroups and/or fast-evolving lineages, replacing rogue organisms by slow-evolving relatives), (3b) removing fast-evolving (and/or biased) sites, based on preliminary rate or compositional analyses and (3c) using sophisticated substitution models (defined here as models heterogeneous across sites, such as CAT-like models, or over time, such as heterotachous/covarion models). The results of this quick bibliographic survey, limited to the relevant studies (69 ‘shallow studies’ and 57 ‘deep studies’), are shown in table 1 (see also the electronic supplementary material, tables S1 and S2 for individual paper analyses).

Strikingly, less than half the studies showed awareness of possible artefacts. In particular, only 36% (25/69) of the publications dealing with shallow phylogenies mentioned any of the key words of our list, while the situation was slightly worse for papers about deep phylogenetic issues (16/57 = 28%). Among the ‘shallow studies’ that effectively cared for artefacts, 76% (19/25) tried to do something to reduce the systematic error, a figure that was similar among ‘deep studies’ (13/16 = 81%). In both cases, the most common strategy was to use a heterogeneous substitution model (15/25 = 60% and 10/16 = 62%), an efficient approach that is also the easiest to implement. By contrast, site removal strategies were more often applied in ‘shallow studies’ (12/25 = 48%) than in ‘deep studies’ (6/16 = 37%), whereas varying the taxon sampling was three times more explored in ‘shallow studies’ (9/25 = 36%) than in ‘deep studies’ (2/16 = 12%), a low figure that might be due to the lack of alternative outgroups at the domain level. Interestingly, six publications (24%) dealing with shallow phylogenies did use the three approaches for controlling the artefacts, while only one publication (6%) trying to infer the ToL [80] was equally comprehensive according to our criteria. Altogether, our modest survey confirmed our initial intuition and indicated that there was room for improvement in deep phylogenetic inference without the need for any additional methodological
development. This is especially true for studies dealing with issues buried deeply in the ToL, where model violations, and thus artefacts, are expected to be much more frequent.

Rooting the ToL cannot be achieved using an outgroup. A clever way to get around this problem is to resort to universal duplicated paralogous genes, namely genes duplicated before the last universal common ancestor (LUCA), which are present in at least two copies in the three domains of life [81–83]. Half a dozen of such gene pairs were identified and put to use in the 1990s, most often with methods that we now consider as inaccurate. As a consequence, conflicting results were obtained (see table 1 in [67]). In 1999, one of us (H.P.) published several papers on the rooting of the ToL, one of them introducing a new method (the S/F method) that hinted at a possible eukaryotic root [84]. When looking at the subsequent publications citing this work (see the electronic supplementary material, table S3 for individual paper analyses), an interesting pattern appears: the majority of the citations are due to the new method and not to the unorthodox result. Hence, for the 121 citations of Brinkmann & Philippe [84] that we analysed in table 1 in [67]), seven different outgroups were considered, thus resulting in seven different supermatrices, each one containing 36 species and 2016 unambiguously aligned amino acid positions. We then analysed all supermatrices using RAXML [85] and PHYLOBAYES [59] under four different substitution models: LG + Γ, GTR + Γ, CAT + Γ and CATGTR + Γ. Bootstrap proportions (LG and GTR) and posterior probabilities (PPs; CAT and CATGTR) for the monophyly of bivalves (upper tree) and for an alternative (LBA) topology, in which fast-evolving bivalves are attracted by the outgroup (lower tree), were computed from 100 bootstrap pseudo-replicates or from two replicate chains per outgroup/model combination, each one run for 10 000 cycles. The burnin was set to 1000 cycles. In the associated table, outgroups are sorted by descending phylogenetic relatedness to molluscs, not evolutionary distance, to illustrate the fact that the latter parameter is the one that really drives the results. To see this, compare PPs for Hymenoptera versus Maxillopoda, two arthropod clades.

7. Inability of current methods to prevent long-branch attraction artefacts

To show how sticking to simple methods in deep phylogenetics is doomed to failure, we illustrate that artefacts easily keep occurring with the sophisticated inference methods available today, even for shallow questions. Let us examine the tree of Bivalvia in the presence of Gastropoda, two molluscan groups whose monophylogenies are well established. To trigger the artefacts, we chose to study concatenated mitochondrial proteomes, because these of some Bivalvia (Pteriomorphia) have evolved much faster than those of others (Unionoida), and to include outgroups of decreasing relatedness (from Annelida to Fungi). As shown in figure 2, all models perform equally well as long as the outgroup is close (Annelida), but become sensitive to LBA when the outgroup distance gets larger, either due to old divergence (Fungi) or to fast evolutionary rate (Hymenoptera). As expected, site-heterogeneous models (CAT + Γ and CATGTR + Γ) perform slightly better than site-homogeneous models (LG + Γ and GTR + Γ). However, the key difference here is not the substitution model used, but the taxon sampling (outgroup distance), which is precisely the parameter that is almost fully constrained when rooting the ToL (owing to the existence of only three domains and a few anciently duplicated genes). Several important model violations are known to affect mitochondrial genes: (i) heterogeneous amino acid composition across taxa [50], (ii) heterotachy [86] and (iii) heteroplicity [59]. These model violations are due to variations in the substitution process over time and initially stem from a change in functional

<table>
<thead>
<tr>
<th>model outgroup topology</th>
<th>LG</th>
<th>GTR</th>
<th>CAT</th>
<th>CATGTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelida Bivalvia</td>
<td>95</td>
<td>96</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Hymenoptera Bivalvia</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Maxillopoda Bivalvia</td>
<td>99</td>
<td>100</td>
<td>95</td>
<td>99</td>
</tr>
<tr>
<td>Myriapoda Bivalvia</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Echinodermata Bivalvia</td>
<td>81</td>
<td>79</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Portiera Bivalvia</td>
<td>38</td>
<td>46</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fungi Bivalvia</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2. Our best substitution models cannot yet address difficult phylogenetic issues, even at shallow depth. We assembled supermatrices by concatenating the translated mitochondrial genomes (12 genes) of nine slow-evolving bivalves (Unionoida), nine fast-evolving bivalves (Pteriomorphia), nine gastropods and nine outgroups. Seven different outgroups were considered, thus resulting in seven different supermatrices, each one containing 36 species and 2016 unambiguously aligned amino acid positions. We then analysed all supermatrices using RAXML [85] and PHYLOBAYES [59] under four different substitution models: LG + Γ, GTR + Γ, CAT + Γ and CATGTR + Γ. Bootstrap proportions (LG and GTR) and posterior probabilities (PPs; CAT and CATGTR) for the monophyly of bivalves (upper tree) and for an alternative (LBA) topology, in which fast-evolving bivalves are attracted by the outgroup (lower tree), were computed from 100 bootstrap pseudo-replicates or from two replicate chains per outgroup/model combination, each one run for 10 000 cycles. The burnin was set to 1000 cycles. In the associated table, outgroups are sorted by descending phylogenetic relatedness to molluscs, not evolutionary distance, to illustrate the fact that the latter parameter is the one that really drives the results. To see this, compare PPs for Hymenoptera versus Maxillopoda, two arthropod clades.
constraints (e.g. relaxed selection). This means that long branches not only retain less phylogenetic signal but also bear a misleading signal, hence the observed LBA artefacts. This illustrates how easily our best phylogenetic methods (here Bayesian inference under the CATGTR + G model) can still be misled when model violations are large. In fact, this is precisely what happens when one tries to root the ToL [87]: the outgroup is incredibly distant (i.e. a paralogous gene with a very different function, which favours heterotachy and heteropercy) while substitution rates for any marker are far from constant over billions of years. To this respect, we do expect major accelerations for informational genes on the branch connecting Mitochondriophora (Eukaryota) to Apokaryota (Archaeae+Bacteria), at the very least because of the absence/presence of transcription/translation coupling. Other events, such as the adaptation to hyperthermophily or the (possible) loss of the nucleus, should also have led to major shifts of the functional constraints and thus to drastic changes in the evolutionary properties of each site over time. Considering that Bacteria always display an extremely long branch in unrooted gene trees [87] and that current methods are unable to resolve similar but much more recent issues (such as the monophyly of Bivalvia, figure 2), it is rather perplexing that the traditional bacterial rooting is taken for granted by so many colleagues in the field.

8. Difficulty to root the tree of life using anciently duplicated genes

We re-examined the case of one anciently duplicated gene pair, the elongation factor: EF-Tu delivers aminoacyl-tRNAs to the A site of the ribosome, while EF-G catalyses the translocation of the peptidyl-tRNA. Even if these two functions are quite different, as shown by the fact that only the GTPase domain can be aligned, this disadvantage is compensated by the preservation of mitochondrial/plastid copies and, more importantly, by the absence of other inter-domain gene transfers. We used the CATGTR + G model, which appears to be the less sensitive to LBA [65], albeit the limited number of positions available in the EF alignment (198) prevents it from working at its best, not because of its large number of parameters that might cause over-fitting (see N. Lartillot’s paper in this issue [32]), but because of the
9. Conclusion

Our results (figures 2 and 3) demonstrate that the root of the ToL is currently unknown, chiefly because published phylogenies are plagued by tremendous model violations and associated LBA artefacts. Nevertheless, properly addressing this issue is key to make progress in our understanding of archaeogenesis, bacteriogenesis and eukaryogenesis. Indeed, we argue that the current consensus about a bacterial root for the ToL is the product of the prejudice of Aristotle’s Great Chain of Being, in which simple organisms are ancestors of more complex life forms. By contrast, our Apokaryota/Mitochondriophora stance builds on the many examples where advances in phylogenetic inference have relocated morphologically simple organisms higher in the ToL. However, we acknowledge that a non-bacterial rooting of the ToL would not necessarily entail that our unorthodox scenario is correct. Indeed, an archaean rooting or, probably more likely, an intra-domain (within Archaea or within Bacteria) rooting cannot yet be ruled out.

Since stochastic and systematic errors have more impact on rooting the ToL than on resolving any of its parts, rooting strategies should be first validated on shallower issues of similar difficulty, such as the monophyly of Bivalvia studied in figure 2. In our opinion, it is unwise to directly apply new approaches, as clever as they might be, to locate the root of the ToL [89–92] without an extensive prior validation on difficult questions with known answers, in particular using very distant outgroups (or without outgroup in the case of non-reversible/non-stationary models). The needed test datasets are straightforward to assemble by subsampling already published complete datasets. Following this reasonable prerequisite, we argue that the supermatrix approach remains the method of choice for rooting the ToL, as it is the most widely used and validated strategy.

To take advantage of the best-fitting models, a relatively large number of characters is necessary, which cannot be obtained using single genes only (e.g. figure 3). However, the concatenation of the few anciently duplicated genes (elongation factors, ATPases, SRP, tRNA-synthetases, etc.) should be possible, as long as the xenologous copies are removed, a task that is within reach thanks to the plethora of complete genomes available today.

While this phylogenetic approach is absolutely required, it will not provide us with a definitive answer, rather the opposite. In the best case, it will locate the root, probably with limited statistical support, which we will need to take into account when developing new evolutionary scenarios. However, beyond being compatible with a correctly rooted ToL, these scenarios will have to fulfil a number of additional constraints, such as:

(i) to provide an explanation for the length heterogeneity observed between major branches (e.g. the long branches at the base of Bacteria and Eukaryota);
(ii) to accommodate palaeontological, genomic, biochemical and cellular knowledge;
(iii) to explain equally well the emergence of the three major cellular types (bacterial, eukaryotic and archaeal, the latter group likely being paraphyletic), instead of only addressing eukaryogenesis;
(iv) to provide transitional steps that are evolutionarily simple and plausible, rather than just proposing that simple organisms are more complex ones.

In this respect, the study of the recently discovered, yet uncultured, Lokiarchaeota [5], an archaeal group featuring several eukaryotic-specific genes (many of them potentially involved in complex membrane remodelling), opens new avenues for completely rethinking the fascinating question of the origin of the three domains of life. Nevertheless, we hope that these will be pursued once freed from the prejudice of Aristotle’s Great Chain of Beings.

**Competing interests** We declare we have no competing interests.

**Funding** We gratefully acknowledge the financial support provided by the Labex TULIP and by SFRD-12/04 (Fonds Spéciaux du Conseil de la Recherche de l’Université de Liège) and the Réseau Québécois de Calcul de Haute Performance for computational resources.

**References**


