Review TRENDS in Ecology and Evolution Vol.20 No.12 December 2005

The tree of eukaryotes

Patrick J. Keeling[1], Gertraud Burger[2], Dion G. Durnford[3], B. Franz Lang[2], Robert W. Lee[4], Ronald E. Pearlman[5], Andrew J. Roger[6] and Michael W. Gray[6]

[1]Department of Botany, University of British Columbia, Vancouver, BC, Canada, V6S 1T4
[2]Centre Robert Cedergren for Bioinformatics and Genomics, Département de Biochimie, Université de Montréal, Montréal, QC, Canada, H3T 1J4
[3]Department of Biology, University of New Brunswick, Fredericton, NB, Canada, E3B 6E1
[4]Department of Biology, Dalhousie University, Halifax, NS, Canada, B3H 4J1
[5]Department of Biology, York University, Toronto, ON, Canada, M3J 1P3
[6]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS, Canada, B3H 1X5

Recent advances in resolving the tree of eukaryotes are converging on a model composed of a few large hypothetical 'supergroups', each comprising a diversity of primarily microbial eukaryotes (protists, or protozoa and algae). The process of resolving the tree involves the synthesis of many kinds of data, including single-gene trees, multigene analyses, and other kinds of molecular and structural characters. Here, we review the recent progress in assembling the tree of eukaryotes, describing the major evidence for each supergroup, and where gaps in our knowledge remain. We also consider other factors emerging from phylogenetic analyses and comparative genomics, in particular lateral gene transfer, and whether such factors confound our understanding of the eukaryotic tree.

Why search for the tree of life?

A well resolved phylogenetic tree, correctly describing the relationships among organisms, is an important tool that is used in many ways, often subconsciously. In the broadest sense, the tree is a means to harness biological information for interpretation or prediction. Processes of change and adaptation can only be understood in the context of a tree because the actions of these processes over time are lost to us without some way of reconstructing long-dead intermediate forms. Reconstructing such past events guides our understanding of modern biology in many ways: for example, recognition of the process of endosymbiosis in mitochondrial evolution has transformed our view of that organelle and its interactions with the rest of the cell. In terms of prediction, all comparative biology is based on the principle that the more closely two organisms are related to one another, the more they will resemble one another at the molecular, biochemical and morphological levels. This is not to say that change is not expected, but we make many assumptions about an organism based on the nature of its close relatives, and these assumptions are evolutionary ones.

Such predictive power is also important because it enables us to form at least simple expectations of the basic properties of an organism, expectations that provide a starting point to be tested. For these reasons, in addition to our desire to establish order, assembling the global tree of life has been a goal of biology ever since a tree-like structure of evolutionary history was first proposed. Here, we discuss advances in assembling the tree of eukaryotes, the kinds of evidence brought to bear on this level of phylogenetic diversity, and some of the factors that challenge this endeavour.

The tree of eukaryotes

For eukaryotes, relatively detailed schemes of evolutionary relationships have long been inferred using morphology and biochemistry. Even for microbial eukaryotes, this approach was successful in dividing diversity into a large number of distinct lineages that are still recognized in light of much molecular data. However, it was less successful in determining how these lineages are related to one another, because there are few shared derived characters to unite eukaryotic groups at this level (e.g. [1]).

Acquiring phylogenetic information has since been transformed by molecular biology, so that most phylogenetic trees are now inferred using molecular data. At first, this process seemed relatively straightforward: trees generated from a single gene, most commonly the small subunit ribosomal RNA gene (SSU rRNA) [2], appeared to provide a basic structure for the topology of eukaryotes, although many branches of the tree remained controversial. However, we are now more aware of the limitations of molecular phylogenies, and that gene sequences can sometimes be deceiving in the phylogenetic predictions that they make [3] (Box 1). The classic 'base' and 'crown' eukaryotic tree of the 1980s went through a period of deconstruction during the 1990s, when several protein-coding gene trees revealed important discrepancies (e.g. [4,5]). We are now in a period of rebuilding this tree using a wider variety of data, which are largely, but not entirely, molecular in nature, and are used in combination with several distinct strategies to examine large-scale phylogenetic questions (Box 2).

Currently, a hypothesis for the tree of eukaryotes looks something like Figure 1. In this scheme, five large

Corresponding author: Keeling, P.J. (pkeeling@interchange.ubc.ca).
Available online 10 October 2005

 0169-5347/$ - see front matter doi:10.1016/j.tree.2005.09.005

Box 1. Assembling the tree of eukaryotes

Every individual phylogeny is probably correct in some respects but misleading in others. Single-gene trees cannot be taken at face value for ancient relationships, so several methods have been used to 'assemble' the tree of eukaryotes from diverse data.

Individual gene trees

Although they must be interpreted cautiously, single-gene trees continue to be valuable because, whereas one gene tree might fail to infer the relationship between, for example, animals and fungi, five other trees might resolve this affiliation robustly. By judiciously interpreting several individual trees (and incorporating other sorts of information), relationships emerge by congruence. The problem is that interpretations are open to differences of opinion among researchers, with no objective means to discriminate among these differing viewpoints.

Combining data

The most common approach in combining data for phylogenetic analysis is to concatenate sequences, resulting in a large amount of data from which a single tree is inferred. Early multi-gene studies treated such a concatenated data set as if it were one large 'super-gene', estimating phylogenies by standard methods [26]. However, this ignores the fact that the various genes being combined might have somewhat different properties, such as variant rates-across-sites distributions, different substitution models and, most importantly, distinct rates of evolution. Ignoring such gene-specific effects risks introducing systematic error in the phylogenetic estimation. Newer approaches incorporate estimation of gene-specific parameter sets during phylogenetic analysis [66,67].

A different approach is to combine estimated trees themselves into a larger 'supertrees'. Recently developed methods for supertree analysis are becoming increasingly popular, and are useful when the taxonomic representation of different gene sets is heterogeneous [68]. However, caution is warranted because sophisticated statistical techniques have yet to be developed for propagating the uncertainty in the original phylogenies into the final supertrees.

Assembling such data sets is a challenge and, even with whole-genome sequences, many taxa lack certain genes, so that concatenated gene sets generally feature 'missing data' [46]. Nevertheless, these approaches have been successful in several instances, most often at first using organelle (mitochondrial and chloroplast) gene data but more recently using nucleus-encoded sequences [27,28,66]. These successes notwithstanding, the nature of concatenated data is only partly understood, and some characteristics are potentially problematic because large data sets are thought to reduce stochastic error but to emphasize systematic errors (Box 2). In this respect, we would do well to remember that our over-confidence in early molecular phylogenies based on simplistic evolutionary models led to acceptance of several misleading conclusions, correction of which took many years.

Discrete molecular characters

Phylogenetic information that is not necessarily dependent on phylogenetic reconstruction (and associated problems) might be found within conserved insertions or deletions in gene sequences [23], intron positions, gene fusions or splits [7], or other complex molecular events. If these characters are highly conserved and shared by two or more organisms to the exclusion of others, they might well be informative, although each such character exhibits its own sources of error (e.g. convergence, recombination and/or parallel loss) and must be interpreted with as much caution as is a molecular phylogeny.

'supergroups' describe eukaryotic diversity, but the order of divergence among these groups is uncertain. Most of these organisms are microbial [protists (protozoa and algae)]. Remarkably, the core elements of three of these five groups have only been proposed during the past few years (and, accordingly, remain controversial), attesting to the scale of the changes that are reshaping our view of eukaryotic diversity. Although a certain degree of controversy remains for each supergroup, all five currently represent reasonable hypotheses based on available data. Nevertheless, additional supporting evidence is required for all five assemblages before they become universally accepted. Below, we review each supergroup and some of the evidence supporting them (the root of this tree is discussed elsewhere [6–8]; Box 3).

Box 2. How far back in time is molecular phylogeny able to reach?

The difficulties in inferring reliable molecular phylogenies over hundreds of millions to billions of years stem from two major sources: (i) random error; and (ii) systematic error.

Random error

Random error (or random noise) arises when the data contain too little information, resulting in a poorly resolved phylogeny. This problem is particularly acute over long timescales, because sites in molecules become saturated with multiple changes, erasing the deep phylogenetic signal. Some authors have recently shown that, if molecules evolve according to simple rules such as those of a Jukes-Cantor process (where all nucleotides or amino acids occur at equal frequencies and are interchanged with equal rates), historical information is lost rather rapidly and abruptly [69]. However, if molecules evolve in a more complex fashion (i.e. with variation in rates of evolution across sites and/or variation in rates at the same site over the tree), then phylogenetic information might persist over much longer timescales [69,70]. These more complex models do appear to describe real data significantly better than do simpler models [71]. Therefore, provided sufficient data are considered, it should be possible to infer ancient relationships. Of crucial importance, however, is the average rate of evolution of a given gene. For organismal divergences on a billion-year timescale, the more slowly a protein evolves (provided there is still some variation among the compared sequences) the more likely it is that it will retain ancient historical signal.

Systematic error

Systematic error describes the failure of a phylogenetic method to recover the correct tree, instead selecting a particular alternative topology as optimal, often with strong apparent statistical support. This typically occurs when the phylogenetic model used is overly simplistic, and it is particularly problematic for deep divergences, because most of the assumptions of even the most complex available phylogenetic models are violated by the true molecular evolutionary process over billions of years. Failure to model correctly rates-across-sites distributions, changing rates at sites over the tree, changing nucleotide and/or amino acid and/or codon usage among species and site-specific substitution properties all can lead to systematic error in tree estimation. The most common manifestation of this error is the erroneous grouping together of the most 'divergent' sequences, a phenomenon known as 'long branch attraction'. The development of more realistic models of molecular evolution is a rapidly expanding field of theoretical phylogenetics [72].

Ultimately, the most promising approach to maximize the chances of recovering an accurate and robust tree of eukaryotes (Box 1) is to utilize information from as many slowly evolving proteins as possible in phylogenetic analyses using sufficiently accurate substitution models.

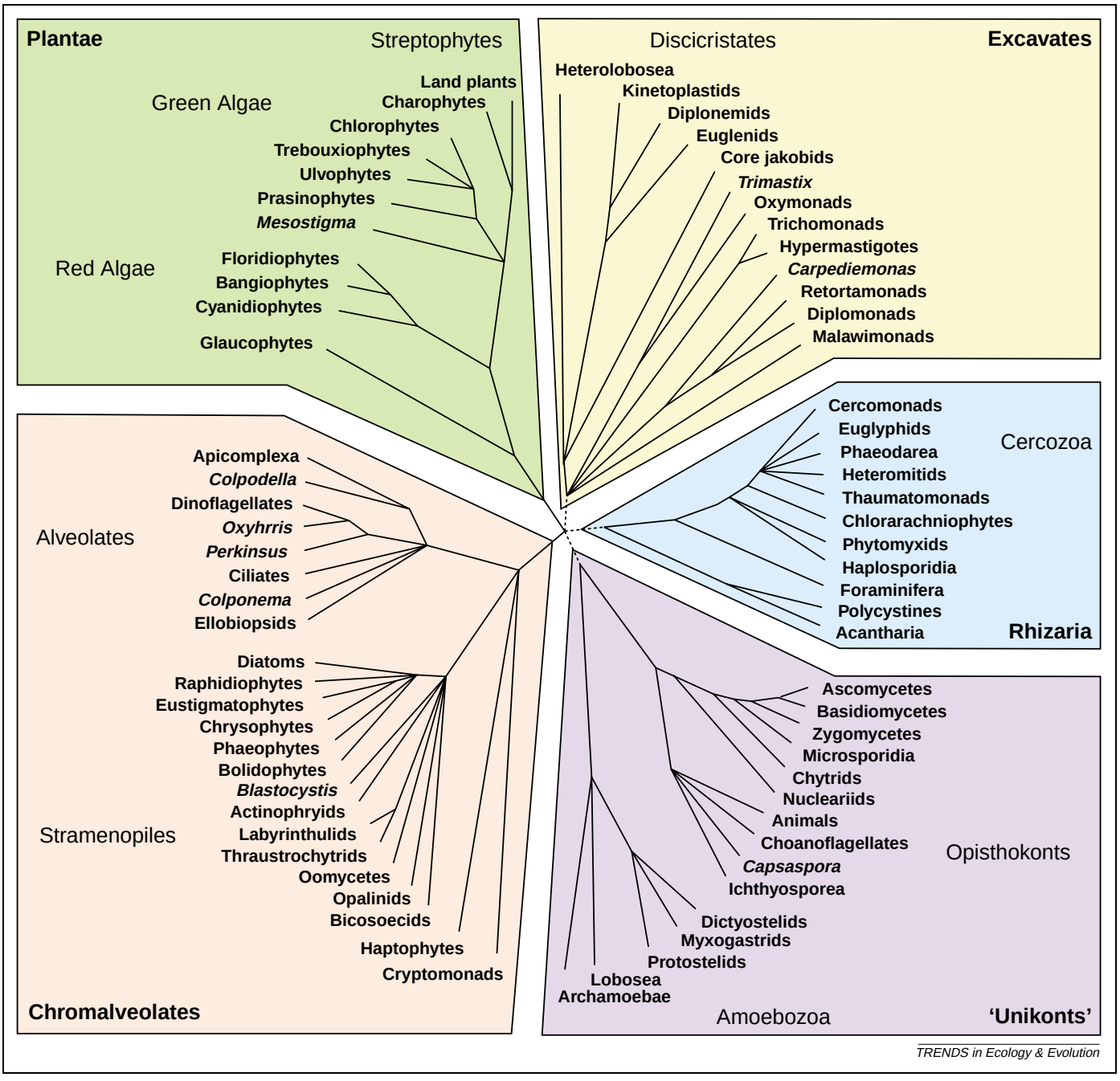

Figure 1. A tree of eukaryotes. The tree is a hypothesis composed from the various types of data discussed in the text, including molecular phylogenies and other molecular characters, as well as morphological and biochemical evidence. Five 'supergroups' are shown, each consisting of a diversity of eukaryotes, most of which are microbial (mostly protists and algae). Relationships are left unresolved (i.e. where several branches emerge simultaneously) when there is little or no evidence for the branching order. Other branches are shown dotted when there are only preliminary indications for this relationship. A handful of 'orphan' genera and two groups, Apusazoa and centrohelid Heliozoa, are not shown. There are few data from these organisms and they are not yet associated with any of these groups.

Excavates

The excavates are a diverse group of protists, many of which are anaerobic and/or parasitic, the best known being *Trypanosoma*, *Giardia* and *Trichomonas*. Excavates are loosely united by a combination of molecular and morphological evidence [9]. To date, no single kind of evidence supports the entire group, but there are morphological similarities in cytoskeletal ultrastructure uniting a subset of excavates, and molecular phylogenetic data to support relationships for overlapping subsets. The two kinds of evidence considered together form a web that unites the entire group [9–11]. Several excavate groups (along with some taxa now known not to be related to excavates) were long considered to be ancient eukaryotic lineages that primitively lacked mitochondria [12]. However, with the exception of oxymonads and retortamonads, evidence for homologues of mitochondrial proteins and relict organelles has now been found in all of these taxa [13–16]. Whether these organisms should still be considered early eukaryotic lineages is less clear, but a recent hypothesis for the root of eukaryotes suggests they that are not [7]; currently, there is no evidence that directly supports an early divergence of these taxa from other eukaryotes [17].

Box 3. How old are eukaryotes?

The earliest fossil record for eukaryotes is notoriously controversial. Biomarker molecules and macro- and microfossils reminiscent of eukaryotes occur spottily in the fossil record from 1 billion–2.7 billion years ago (Bya), but whether these entities are truly the remains of ancient eukaryotes is debatable. The oldest fossil that is widely regarded to be of a 'crown' group eukaryote is a bangiophyte red alga from 1.2 Bya [73] and, although more candidates are emerging [74,75], most other plausible fossils of crown-group eukaryotes occur more recently than ~1 Bya [76].

This patchy fossil record has motivated several recent analyses aimed at dating the earliest divergences in the tree of eukaryotes using fossil-calibrated 'molecular dating' methods [77,78]. However, different analyses often give wildly differing divergence time estimates and remain controversial. For instance, using dozens to hundreds of protein genes and several dating methods, Hedges *et al.* [77] argued that the common ancestor of extant eukaryotes existed 2.3 Bya. By contrast, Douzery *et al.* [78] used 129 proteins to argue for a more recent divergence time of 0.95–1.26 Bya. The main differences between these studies include the assumed phylogeny of eukaryote diversity, as well as the specific 'dating' methods used and ways of assessing uncertainty. In addition, many aspects of molecular dating studies such as these have been roundly criticized in general [79], because often only a few fossil calibration points are used, error is compounded and is not properly accounted for, and, perhaps most damningly, all such ancient dates are extrapolated from fossil dates that are substantially much younger – in some cases two- to fivefold younger. Although new 'relaxed molecular clock' methods show promise for dealing with inherent rate changes in molecules across the tree of life [80], their statistical properties are poorly understood, and these methods are only as good as the models of molecular evolution used (Box 2). A more-complete fossil record and better molecular dating methods will be needed before the true age of eukaryotes can be determined with any certainty.

Rhizarians

Rhizaria is one of the most recently recognized eukaryotic supergroups [18,19], and is also distinguished by being united only by molecular data (i.e. no clearly homologous morphological character uniquely defines this group). Rhizarians are abundant in nature and are ecologically important, but few are known commonly. Cercozoa and Foraminifera, two large and diverse groups, comprise the core of this supergroup and are united by molecular trees based on actin [20], the largest subunit of RNA polymerase II [21], some SSU rRNA trees [19,22], as well as a unique insertion at the processing site of polyubiquitin [23]. The Rhizaria as a whole, including acantharid and polycistene radiolaria, is not as well supported as there are few data from the latter two groups; however, this supergroup is nevertheless defined by actin and some SSU rRNA trees [22,24]. Although this evidence is limited, it represents most of the available data from these organisms. Molecular gene trees have suggested alternative positions for some of these groups in the past (e.g. [25]), but with the subsequent availability of a broader sampling of sequences, molecular analyses have suggested no alternative. More evidence is required before this supergroup becomes universally accepted, but we predict that most new data will continue to support this union.

Unikonts

'Unikont' is a controversial name for the union of two individually well supported groups: amoebozoans and opisthokonts [18]. Overall, unikonts includes animals and fungi, as well as some amoebae (e.g. *Entamoeba*), slime molds (e.g. *Dictyostelium*), and a few parasitic protists. Evidence supporting the group as a whole comes from phylogenies based on concatenates of four genes [26], 123 genes [27] and 129 genes [28], several individual gene phylogenies (e.g. [4]) and the presence of an internal duplication of one domain of phosphofructokinase [29]. The presence of stand-alone genes encoding dihydrofolate reductase (DHFR) and thymidylate synthase (TS), which are found as a fusion protein in other eukaryotes [7], is also consistent with this grouping. However, because this might be the ancestral state of eukaryotes, the significance of this character is dependent on the placement of the root of the eukaryotic tree. Evidence for the relationships among animals and their unicellular relatives (together termed 'Holozoa'), and the affinity of Holozoa and Fungi (together termed 'opisthokonts') is much stronger. The opisthokont lineage is supported by insertions in elongation factor-1α and enolase [30], as well as many individual [30,31] and concatenated gene phylogenies [26–28]. Amoebozoa, in turn, is also supported as a group by several individual and concatenated gene phylogenies [27], and partially by the presence of fused genes encoding cytochrome oxidase 1 and 2 in the mitochondrial DNA of slime molds and lobose amoebae [32].

Chromalveolates

Chromalveolates account for much of the diversity of algae (e.g. kelps, diatoms, coccolithophirds and dinoflagellates), as well as incorporating several major protist groups (e.g. apicomplexans and ciliates). The group unites the well supported alveolates with the more contentious chromists, and was originally proposed based on the presence of secondary plastids of red algal origin in many chromalveolate groups [33]. Alveolates are one of the most firmly established protist assemblages, being supported by a variety of nuclear gene trees and concatenated analyses (e.g. [34]). Many gene trees, including ones based on SSU rRNA and analyses of individual and concatenated nuclear protein-coding genes, also support various subgroups of chromalveolates, in particular alveolates and heterokonts [26,28,35,36]. An analysis of five concatenated plastid-encoded genes representing broad taxon sampling has also supported the grouping of heterokonts with haptophytes and cryptomonads (chromists) [37,38], whereas an analysis of 14 photosystem genes unites cryptomonads and heterokonts [39] (although alveolates cannot be examined with these latter plastid data because the Apicomplexa do not have these genes and the dinoflagellate homologues are either highly divergent or not yet characterized). Two nucleus-encoded plastid-targeted proteins have also been shown to have unusual evolutionary histories that support a common origin for chromalveolate plastids. These proteins are glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and fructose-

1,6-bisphosphate aldolase (FBA), the plastid copies of which have both been lost and replaced through duplication of the cytosolic form and/or by lateral gene transfer [40,41]. These are rare events and, because the chromalveolate plastid targeted are all closely related to one another to the exclusion of all other homologues known (plastid or otherwise), these replacements are interpreted as having occurred only once in the common ancestor of the affected organisms.

Plantae

Members of Plantae are distinguished by the presence of plastids derived by primary endosymbiosis: it was through the ancestor of this supergroup that eukaryotes first acquired photosynthesis. Although some controversy remains about the relative branching order among glaucophytes, red algae and green algae (the sister relationship of plants and green algae is beyond doubt), the relationships within several of these groups are emerging and there is an increasing body of evidence underpinning the supergroup as a whole. First, phylogenies of many plastid genes support this relationship (e.g. [42]). More recently, several nuclear gene trees have confirmed this affiliation, including well supported trees based on concatenates of six genes [43] and 143 genes [44]. Mitochondrial gene phylogenies also support the red algae and green algae as a group [45], as do analyses of 13 concatenated nuclear genes [43]. Overall, Plantae is the best supported of the five groups and we also know more about the relationships between its subgroups than we do of any other supergroup.

Dissenting evidence

We have presented a hypothesis for a global phylogeny of eukaryotes where eukaryotic diversity is sequestered into five large groups. Several forms of consistent evidence support each of these groups but, not surprisingly, there are dissenting observations. Even some of the best-supported groups, for example opisthokonts, have been questioned in recent analyses [46]. Indeed, every major lineage of eukaryotes is contradicted by at least one molecular phylogeny, but this does not mean that they are all incorrect.

First, it is important when evaluating phylogenetic evidence to distinguish between data that fail to resolve a relationship (i.e. negative evidence) and data that support an alternative relationship (i.e. positive evidence). In other cases, even strongly supported incongruence can change with a better understanding of the data or more thorough analyses. To take an example from the tree of eukaryotes, an analysis of 42 plastid genes was used to argue against the monophyly of chromalveolates because a diatom and a cryptomonad each branch with a different red alga [47]. At face value, this result is strongly supported, but a recent re-analysis of 41 of these genes showed that there is more going on [39]. When all genes were analyzed together, the initial result was recovered; however, when the data were partitioned in various ways, removal of the fastest evolving genes (primarily ribosomal protein sequences) resulted in the two chromalveolates forming a group. Moreover, topology tests show that the separation of chromalveolates in the larger data set is not significant, whereas their union in the data set based on slow-evolving genes is. Presumably, as databases increase in coverage, analytical methods improve and our theoretical and empirical understanding of the data increases, some conflicting cases will be found either to be insignificant or to converge on a single answer. This is not to say that different trees for different genes do not exist: they surely do, and do so for a variety of reasons (one of which we discuss below) that can be important in assembling and interpreting the tree of eukaryotes.

Lateral gene transfer and the tree of eukaryotes

Lateral, or horizontal, gene transfer (LGT) is the movement of genetic information between two distantly related genomes (i.e. not sexual recombination within one species), resulting in a genome that contains an expressed, functional gene from a foreign source. Determining the scale of LGT is important for interpreting molecular phylogenies and the distribution of molecular characteristics in any group. The debate over the impact of LGT on prokaryote evolution has a long history, and this sometimes-fiery exchange has been stoked considerably by bacterial genomics. At one extreme, it is claimed that bacterial evolution is not tree-like, owing to the pervasiveness of LGT [48,49], or that LGT is the driving force of bacterial innovation [50]. The opposing pole suggests that the importance of LGT has been overstated [51–54] and that various factors limit the likelihood of transfer of certain kinds of gene, so that some genes might represent organismal evolution whereas others might not [55,56]. An intuitive example is the complexity hypothesis, which suggests that genes for proteins involved in large complexes are less likely to be transferred [55].

Debate over the role of LGT in eukaryotes has lagged behind the prokaryotic debate, probably as a result of the lag in eukaryotic comparative genomics. Several individual cases of LGT have been described in eukaryotes, many with interesting functional or ecological implications (e.g. [49,57–60]). There have also been a few recent reports suggesting more widespread LGT in certain eukaryotes, some events within a lineage and others between distantly related organisms [61–64]. Altogether, the emerging picture is that LGT has affected some eukaryotes more than others, and some genes more than others. However, there are currently few data from the lineages where this process is expected to be most active. In animals, for instance, the separation of germ and soma should reduce the impact of LGT, but this is not the case in microbial eukaryotes.

Currently, many of the best examples of LGT in eukaryotes involve genes derived from bacteria, because these transfers are easiest to spot. The process of eukaryote–eukaryote transfer might be more common if such gene products integrate more easily than do prokaryotic ones, but eukaryote–eukaryote transfer is also more difficult to detect, especially in uncertain or poorly sampled parts of the tree. This is because the best way to detect LGT is by identifying genes whose phylogenies depart from our expectations based on the accepted tree of the organisms in which they are found.

Simplistically, this might seem like an impossible paradox: LGT 'erases' the organismal tree from our sight, so how can we use the tree to find LGT? This paradox is only true if LGT is sufficiently frequent in the case of all genes. We currently have no indication that the frequency of LGT in eukaryotes is sufficient to eradicate the underlying process of bifurcation that is necessary to recover a tree of organismal relationships. That said, there are also few data from the eukaryotes that are most likely to have high levels of at least recent LGT, a state of affairs rapidly being corrected by comparative genomics.

The future of the tree of eukaryotes

Comparative genomics will have a dramatic and positive effect on our understanding of the tree itself, for two reasons. First, large-scale concatenated data sets covering a broad variety of eukaryotes require more data than are realistically generated by a targeted gene-by-gene approach, but such data will emerge from whole-genome and EST sequencing projects. Second, non-tree-based characters (such as gene fusions, insertions and gene replacements) typically are not intentionally sought, but are discovered by sampling done for other reasons. The next generation of molecular analyses has the potential to: (i) buttress support for the five supergroups, or perhaps topple one or two of them in favor of some alternative; (ii) resolve the order between and among supergroups; (iii) reveal the position of the last few taxa not tentatively assigned to a supergroup in Figure 1 (e.g. Apusozoa and some members of Heliozoa); and (iv) provide further clues as to the root of the eukaryotic tree. In addition, several finer-scale issues within each supergroup are particularly interesting. For instance, resolving the branching order of the animals and their closest protist allies will reveal much about the nature of the unicellular ancestor of animals [65]. Other parts of the tree will be as informative about processes such as the origins of parasitism, the effects of mitochondrial degeneration, plastid origins, and patterns of evolution of pathways as diverse as intermediary metabolism, bioenergetics, signaling and splicing, other RNA processing pathways (e.g. editing, modification and interference) and the genetic code.

To take full advantage of these new data, it will be wise to re-examine the lessons learned from the early days of molecular phylogeny: for those of us determined to uncover the past, Santayana's adage that 'those who cannot remember the past are condemned to repeat it' might be especially apt.

Acknowledgements

The authors are members of the Protist EST Program (PEP), supported by funding from Genome Canada, Génome Québec, Genome Atlantic and the Atlantic Canada Opportunities Agency (Atlantic Innovation Fund).

References

1 Brugerolle, G. and Taylor, F.J.R. (1977) Taxonomy, cytology and evolution of the Mastigophora. In *Proceedings of the Fifth International Congress of Protozoology* (Hutner, S.H., ed.), pp. 14–28, Pace University
2 Sogin, M.L. (1991) Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* 1, 457–463
3 Philippe, H. and Adoutte, A. (1998) The molecular phylogeny of Eukaryota: solid facts and uncertainties. In *Evolutionary Relationships Among Protozoa* (Coombs, G.H. *et al*., eds), pp. 25–56, Chapman & Hall
4 Baldauf, S.L. and Doolittle, W.F. (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl. Acad. Sci. U. S. A.* 94, 12007–12012
5 Keeling, P.J. and Doolittle, W.F. (1996) Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. *Mol. Biol. Evol.* 13, 1297–1305
6 Simpson, A.G. and Roger, A.J. (2002) Eukaryotic evolution: getting to the root of the problem. *Curr. Biol.* 12, R691–R693
7 Stechmann, A. and Cavalier-Smith, T. (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297, 89–91
8 Arisue, N. *et al*. (2005) Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.* 22, 409–420
9 Simpson, A.G. (2003) Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). *Int. J. Syst. Evol. Microbiol.* 53, 1759–1777
10 Simpson, A.G. *et al*. (2002) Evolutionary history of 'early-diverging' eukaryotes: the excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol. Biol. Evol.* 19, 1782–1791
11 Simpson, A.G. and Patterson, D.J. (2001) On core jakobids and excavate taxa: the ultrastructure of *Jakoba incarcerata*. *J. Eukaryot. Microbiol.* 48, 480–492
12 Cavalier-Smith, T. (1983) A 6-kingdom classification and a unified phylogeny. In *Endocytobiology II: Intracellular Space as Oligogenetic* (Schenk, H.E.A. *et al*., eds), pp. 1027–1034, Walter de Gruyter & Co
13 Bui, E.T. *et al*. (1996) A common evolutionary origin for mitochondria and hydrogenosomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9651–9656
14 Roger, A.J. *et al*. (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 95, 229–234
15 Tovar, J. *et al*. (2003) Mitochondrial remnant organelles of *Giardia* function in iron–sulphur protein maturation. *Nature* 426, 172–176
16 Dyall, S.D. *et al*. (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304, 253–257
17 Roger, A.J. (1999) Reconstructing early events in eukaryotic evolution. *Am. Nat.* 154, S146–S163
18 Cavalier-Smith, T. (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* 52, 297–354
19 Cavalier-Smith, T. (2003) Protist phylogeny and the high-level classification of Protozoa. *Eur. J. Protistol.* 39, 338–348
20 Keeling, P.J. (2001) Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol. Biol. Evol.* 18, 1551–1557
21 Longet, D. *et al*. (2003) Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. *Int. J. Syst. Evol. Microbiol.* 53, 1735–1739
22 Nikolaev, S.I. *et al*. (2004) The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8066–8071
23 Archibald, J.M. *et al*. (2002) A novel polyubiquitin structure in Cercozoa and Foraminifera: evidence for a new eukaryotic supergroup. *Mol. Biol. Evol.* 20, 62–66
24 Polet, S. *et al*. (2004) Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. *Protist* 155, 53–63
25 Pawlowski, J. *et al*. (1996) Early origin of foraminifera suggested by SSU rRNA gene sequences. *Mol. Biol. Evol.* 13, 445–450
26 Baldauf, S.L. *et al*. (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977
27 Bapteste, E. *et al*. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostellium, Entamoeba, and Mastigamoeba*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1414–1419
28 Philippe, H. *et al*. (2004) Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 21, 1740–1752
29 Stechmann, A. and Cavalier-Smith, T. (2003) The root of the eukaryote tree pinpointed. *Curr. Biol.* 13, R665–R666

30 Baldauf, S.L. and Palmer, J.D. (1993) Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. U. S. A.* 90, 11558–11562
31 Wainright, P.O. *et al.* (1993) Monophyletic origins of the metazoa: an evolutionary link with fungi. *Science* 260, 340–322
32 Gray, M.W. *et al.* (2004) Mitochondria of protists. *Annu. Rev. Genet.* 38, 477–524
33 Cavalier-Smith, T. (1998) A revised six-kingdom system of life. *Biol. Rev. Camb. Philos. Soc.* 73, 203–266
34 Fast, N.M. *et al.* (2002) Re-examining alveolate evolution using multiple protein molecular phylogenies. *J. Eukaryot. Microbiol.* 49, 30–37
35 Harper, J.T. *et al.* (2005) On the monophyly of the chromalveolates using a six-protein phylogeny of eukaryotes. *Int. J. Syst. Evol. Microbiol.* 55, 487–496
36 Van de Peer, Y. and De Wachter, R. (1997) Evolutionary relationships among eukaryotic crown taxa taking into account site-to-site variation in 18S rRNA. *J. Mol. Evol.* 45, 619–630
37 Yoon, H.S. *et al.* (2002) A single, ancient origin of the plastid in the Chromista. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15507–15512
38 Yoon, H.S. *et al.* (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809–818
39 Hagopian, J.C. *et al.* (2004) Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of rhodoplasts and their relationship to other plastids. *J. Mol. Evol.* 59, 464–477
40 Fast, N.M. *et al.* (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* 18, 418–426
41 Patron, N.J. *et al.* (2004) Gene replacement of fructose-1,6-bisphosphate aldolase (FBA) supports a single photosynthetic ancestor of chromalveolates. *Eukaryot. Cell* 3, 1169–1175
42 Turner, S. *et al.* (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* 46, 327–338
43 Moreira, D. *et al.* (2000) The origin of red algae and the evolution of chloroplasts. *Nature* 405, 69–72
44 Rodríguez-Ezpeleta, N. *et al.* (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae and glaucophytes. *Curr. Biol.* 15, 1325–1330
45 Burger, G. *et al.* (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* 11, 1675–1694
46 Philip, G.K. *et al.* (2005) The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. *Mol. Biol. Evol.* 22, 1175–1184
47 Martin, W. *et al.* (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12246–12251
48 Doolittle, W.F. (1999) Lateral genomics. *Trends Cell Biol.* 9, M5–M8
49 Doolittle, W.F. *et al.* (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 39–57
50 Ochman, H. *et al.* (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304
51 Eisen, J.A. (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* 10, 606–611
52 Kurland, C.G. (2000) Something for everyone. Horizontal gene transfer in evolution. *EMBO Rep.* 1, 92–95
53 Salzberg, S.L. *et al.* (2001) Microbial genes in the human genome: lateral transfer or gene loss? *Science* 292, 1903–1906
54 Kurland, C.G. *et al.* (2003) Horizontal gene transfer: a critical view. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9658–9662
55 Jain, R. *et al.* (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3801–3806
56 Kurland, C.G. (2005) What tangled web: barriers to rampant horizontal gene transfer. *Bioessays* 27, 741–747
57 Fast, N.M. *et al.* (2003) Bacterial catalase in the microsporidian *Nosema locustae*: implications for microsporidian metabolism and genome evolution. *Eukaryot. Cell* 2, 1069–1075
58 Keeling, P.J. and Inagaki, Y. (2004) A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1alpha. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15380–15385
59 Richards, T.A. *et al.* (2003) Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* 154, 17–32
60 Andersson, J.O. (2005) Lateral gene transfer in eukaryotes. *Cell Mol. Life Sci.* 62, 1182–1197
61 Andersson, J.O. *et al.* (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr. Biol.* 13, 94–104
62 Archibald, J.M. *et al.* (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7678–7683
63 Bergthorsson, U. *et al.* (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424, 197–201
64 Loftus, B. *et al.* (2005) The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433, 865–868
65 Lang, B.F. *et al.* (2002) The closest unicellular relatives of animals. *Curr. Biol.* 12, 1773–1778
66 Delsuc, F. *et al.* (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375
67 Nylander, J.A. *et al.* (2004) Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67
68 Bininda-Emonds, O.R.P. (2004) The evolution of supertrees. *Trends Ecol. Evol.* 19, 315–322
69 Penny, D. *et al.* (2001) Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.* 53, 711–723
70 Mossel, E. and Steel, M. (2005) How much can evolved characters tell us about the tree that generated them?. In *Mathematics of Evolution & Phylogeny* (Gascuel, O., ed.), pp. 384–412, Oxford University Press
71 Galtier, N. (2001) Maximum-likelihood phylogenetic analysis under a covarion-like model. *Mol. Biol. Evol.* 18, 866–873
72 Steel, M. (2005) Should phylogenetic models be trying to 'fit an elephant'? *Trends Genet.* 21, 307–309
73 Butterfield, N. (2000) *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26, 386–404
74 Butterfield, N.J. (2005) Probable proterozoic fungi. *Paleobiology* 31, 165–182
75 Javaux, E. *et al.* (2004) TEM evidence for eukaryotic diversity in the mid-Proterozoic. *Geobiology* 2, 121–132
76 Porter, S.M. (2004) The fossil record of early eukaryotic diversification. *Paleontol. Soc. Papers* 10, 35–50
77 Hedges, S.B. *et al.* (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4, 2
78 Douzery, E.J. *et al.* (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. U. S. A.* 101, 15386–15391
79 Graur, D. and Martin, W. (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20, 80–86
80 Welch, J.J. and Bromham, L. (2005) Molecular dating when rates vary. *Trends Ecol. Evol.* 20, 320–327