

Eukaryotic evolution, changes and challenges (Extracts)

T. Martin Embley & William Martin Vol 440|30 March 2006|doi:10.1038/nature04546

New findings have profoundly changed the ways in which we view early eukaryotic evolution, the composition of major groups, and the relationships among them. The changes have been driven by a flood of sequence data combined with improved, but by no means consummate, computational methods of phylogenetic inference.

Newer findings even call into question the very notion of a 'tree' as an adequate metaphor to describe the relationships among genomes. Placing eukaryotic evolution within a time frame and ancient ecological context is still problematic owing to the vagaries of the molecular clock and the paucity of Proterozoic fossil eukaryotes that can be clearly assigned to contemporary groups. Although the broader contours of the eukaryote phylogenetic tree are emerging from genomic studies, the details of its deepest branches, and its root, remain uncertain.

The universal tree based on small-subunit (SSU) ribosomal RNA¹ provided a first overarching view of the relationships between the different types of cellular life. The relationships among eukaryotes recovered from rRNA², backed up by trees of translation elongation factor (EF) proteins³, provided what seemed to be a consistent, and hence compelling, picture

The apparent agreement between molecules and morphology depicted the relative timing of the mitochondrial endosymbiosis as a crucial, but not ancestral, event in eukaryote phylogeny.

In a corollary to endosymbiotic theory called endosymbiotic gene transfer, during the course of mitochondrial genome reduction, genes were transferred from the endosymbiont's genome to the host's chromosomes, but the encoded proteins were reimported into the organelle where they originally functioned.

Hence, mitochondria, hydrogenosomes and mitosomes are viewed most simply as variations on a single theme, one that fits neatly within the framework provided by classical evolutionary theory. They are evolutionary homologues that share similarities because of common ancestry, but, like forelimbs in vertebrates, differ substantially in form and function across lineages owing to descent with modification.

Important similarities between *Trichomonas* hydrogenosomes and mitochondria include the use of common protein import pathways, conserved mechanisms of iron sulphur cluster assembly, conserved mechanisms of NAD⁺ regeneration, and conservation of a canonical ATP-producing enzyme of the mitochondrial Krebs cycle, succinate thiokinase.

Hence, a single common ancestry of mitochondria and hydrogenosomes sufficiently accounts for current observations.

Mitosomes were discovered in *Entamoeba* as mitochondrion-derived organelles that have undergone more evolutionary reduction than hydrogenosomes. They are also found in *Giardia* and microsporidia. Mitosomes seem to have no direct role in ATP synthesis. Mitosomes import proteins in a mitochondrial-like manner.

The discovery of mitochondrial homologues in *Giardia*, *Trichomonas* and microsporidians, which had been the best candidates for eukaryotes that primitively lacked mitochondria, has pinned the timing of the mitochondrial origin to the ancestor of all eukaryotes studied so far.

That issue hinges on efforts to construct reliable rooted phylogenetic trees depicting ancient eukaryotic relationships: a developing area of research that is fraught with difficulties. The tempo and mode of sequence evolution is far more complicated than is assumed by current mathematical models that are used to make phylogenetic trees. In computer simulations, where the true tree is known, model mis-specification can produce the wrong tree with strong support.

Different sites in molecular sequences evolve at different rates, and failure to accommodate this rate variation, something early methods failed to do, can lead to strongly supported but incorrect trees owing to a common problem called 'long-branch-attraction'. This occurs when branches that are long or 'fast evolving', relative to others in the tree, cluster together irrespective of evolutionary relationships. The molecular sequences of *Giardia*, *Trichomonas* and microsporidia often form long branches in trees and thus are particularly prone to this problem. The traditional models that placed microsporidia deep within trees assumed that all sequence sites evolved at the same rate, even though they clearly do not.

The position of *Giardia* and *Trichomonas* sequences at the base of eukaryotic molecular trees is also suspect, given that they also form long branches in the trees that place them in this way, and because other trees and models place them together as an internal branch of a rooted eukaryotic tree. Resolving which position is correct is particularly important, because *Giardia* and *Trichomonas* are still commonly referred to as 'early-branching' eukaryotes.

Given the evident uncertainties of such phylogenies, and the importance of the problem, the onus is on those who would persist in calling these species 'early branching' to show that trees placing them deep explain the data significantly better than trees that do not.

The usual way to root a phylogenetic tree is by reference to an outgroup; the rRNA and EF trees used prokaryotic sequences to root eukaryotes on either the *Giardia*, *Trichomonas* or microsporidia branch, but these rootings have not proved robust. The sequences of outgroups are often highly divergent compared to those of the ingroup, making it difficult to avoid model mis-specification and long-branch-attraction.

The coherence of animals, fungi and various unicellular eukaryotes (together called opisthokonts) is supported by phylogenetic trees and other characters.

Placing the eukaryote root between unikonts and bikonts would help to bring order to chaos, if it is correct. However, it assumes that the underlying tree over which the rooting character is mapped is known, when in fact the relationships especially for bikonts and many enigmatic protistan lineages remain uncertain.

On the basis of the unfolding tree, it would seem entirely possible, if not likely, that aerobic and anaerobic eukaryotes, harbouring mitochondrial homologues of various sorts, have co-existed throughout eukaryote history.

The relationships between major groups of eukaryotes are uncertain because of the lack of agreement between different proteins and different analyses;

It has been suggested that the lack of resolution in deeper parts of the eukaryotic tree stems from an evolutionary 'big bang' or rapid radiation for eukaryotes, perhaps driven by the mitochondrial endosymbiosis. However, both theory and computer simulations suggest that a lack of resolution at deeper levels is to be expected given sparse data, our assumptions about sequence evolution, and the limitations of current phylogenetic methods. Thus, loss of historical signal provides a simple null hypothesis for the observed lack of resolution in deeper parts of the eukaryotic tree.

Eukaryotic cell organization is more complex than prokaryotic, boasting, inter alia, a nucleus with its contiguous endoplasmic reticulum, Golgi, flagella with a '9b2' pattern of microtubule arrangement, and organelles surrounded by double membranes. There are no obvious precursor structures known among prokaryotes from which such attributes could be derived, and no intermediate cell types known that would guide a gradual evolutionary inference between the prokaryotic and eukaryotic state. Accordingly, thoughts on the topic are diverse, and new suggestions appear faster than old ones can be tested.

Biologists have traditionally derived the complex eukaryotic state from the simpler prokaryotic one. In recent years, even that has been called into question, as some phylogenies have suggested that prokaryotes might be derived from eukaryotes. However, the ubiquity of mitochondrial homologues represents a strong argument that clearly polarizes the prokaryote-to-eukaryote transition: because the common ancestor of contemporary eukaryotes contained a mitochondrial endosymbiont that originated from within the proteobacterial lineage, we can confidently infer that prokaryotes arose and diversified before contemporary eukaryotes, the only ones whose origin requires explanation, did. This view is consistent with microfossil and biogeochemical evidence.

The rooted rRNA tree is vastly more visible to non-specialists.

For every analysis of a eukaryotic informational gene that recovers the neomuran topology, a different analysis of the same molecule(s) has recovered the eocyte tree, with the latter being favoured by more sophisticated phylogenetic analyses and by a shared amino-acid insertion in eocyte and eukaryotic elongation factor 1-a.

The problems associated with the phylogenetic relationships discussed above are exacerbated at such deep levels, and there is currently neither consensus on this issue nor unambiguous evidence that would clarify it.

Individual analyses of informational genes arrive at fundamentally different interpretations.

An issue that still requires clarification concerns the origin of thousands of eukaryotic operational genes that are clearly eubacterial, but not specifically *alpha*-proteobacterial, in origin.

Lateral Gene Transfer (LGT) among prokaryotes figures into the origin of eukaryotic operational genes in a fundamental manner that is often overlooked. Most claims of outright LGT to ancestral eukaryotes (that is, from donors distinct from the mitochondrion) implicitly assume a static chromosome model in which prokaryotes do not exchange genes among themselves; finding a eukaryotic gene that branches with a group other than *alpha*-proteobacteria is taken as evidence for an origin from that group.

The ancestor of mitochondria possessed an as yet unknown collection of genes.

Diversified unicellular microfossils of uncertain phylogenetic affinity (acritarchs), but widely accepted as eukaryotes, appear in strata of 1.45 billion years (Gyr) of age, providing a minimum age for the group.

Bangiomorpha, a fossilized multicellular organism virtually indistinguishable in morphology from modern bangiophyte red algae, has been found in strata of 1.2 Gyr of age, placing a lower bound on the age of the plant kingdom.

A wide range of molecular clock estimates of eukaryote age have been reported, but these are still uncertain, being contingent both on the use of younger calibration points and on the phylogenetic

model and assumed tree. At present, a minimum age of eukaryotes at ,1.45Gyr and a minimum age of the plant kingdom at ,1.2Gyr seem to be criteria that the molecular clock must meet.

The classical view of early eukaryote evolution posits two main ecological stages: (1) the early emergence and diversification of anaerobic, amitochondriate lineages, followed by (2) the acquisition of an oxygen-respiring mitochondrial ancestor in one lineage thereof and the subsequent diversification of aerobic eukaryotic lineages. Concordant with that view, mitochondrial origins have traditionally been causally linked to the global rise in atmospheric oxygen levels and an assumed 'environmental disaster' for cells lacking the mitochondrial endosymbiont, providing a selective force (oxygen detoxification) for the acquisition of the mitochondrion.

Two observations challenge this model. First, it is now clear that the contemporary anaerobic eukaryotes did not branch off before the origin of mitochondria. Second, new isotope studies indicate that anaerobic environments persisted locally and globally over the past 2Gyr. That oxygen first appeared in the atmosphere at 2Gyr ago is still generally accepted, but it is now thought that, up until about 600Myr ago, the oceans existed in an intermediate oxidation state, with oxygenated surface water (where photosynthesis was occurring), and sulphide-rich (sulphidic) and oxygen-lacking (anoxic) subsurface water. Hence, the 'oxygen event' in the atmosphere should be logically decoupled from anoxic marine environments, where anaerobic eukaryotes living on the margins of an oxic world could have flourished, as they still do today.

In the past, phylogenetic trees have produced **a particular view of early eukaryote history that was appealing, but turned out to be wrong** in salient aspects. Simply testing whether a model used to make a tree actually fits the data would do much to restore confidence in the merits of deep phylogenetic analyses.

The available phylogenetic findings from genomes are not fully consistent with any current hypothesis for eukaryote origins, the underlying reasons for which, being biological, methodological or both, are as yet unclear. Genomes must surely bear some testimony to eukaryotic origins, but new approaches and **more rigorous attention to the details of phylogenetic inference will be required to decipher the message.**