Available online at www.sciencedirect.com
ScienceDirect
Parasitology International 57 (2008) 97–104
www.elsevier.com/locate/parint

Mini-review

Implications of the new eukaryotic systematics for parasitologists

Joel B. Dacks [a], Giselle Walker [b], Mark C. Field [a,*]

[a] The Molteno Building, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK
[b] University Museum of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Received 27 September 2007; received in revised form 15 November 2007; accepted 16 November 2007
Available online 4 December 2007

Abstract

An accurate understanding of evolutionary relationships is central in biology. For parasitologists, understanding the relationships among eukaryotic organisms allows the prediction of virulence mechanisms, reconstruction of metabolic pathways, identification of potential drug targets, elucidation of parasite-specific cellular processes and understanding of interactions with the host or vector. Here we consider the impact of major recent revisions of eukaryotic systematics and taxonomy on parasitology. The previous, ladder-like model placed some protists as early diverging, with the remaining eukaryotes "progressing" towards a "crown radiation" of animals, plants, Fungi and some additional protistan lineages. This model has been robustly disproven. The new model is based on vastly increased amounts of molecular sequence data, integration with morphological information and the rigorous application of phylogenetic methods to those data. It now divides eukaryotes into six major supergroups; the relationships between those groups and the order of branching remain unknown. This new eukaryotic phylogeny emphasizes that organisms including *Giardia*, *Trypanosoma* and *Trichomonas* are not primitive, but instead highly evolved and specialised for their specific environments. The wealth of newly available comparative genomic data has also allowed the reconstruction of ancient suites of characteristics and mapping of character evolution in diverse parasites. For example, the last common eukaryotic ancestor was apparently complex, suggesting that lineage-specific adaptations and secondary losses have been important in the evolution of protistan parasites. Referring to the best evidence-based models for eukaryotic evolution will allow parasitologists to make more accurate and reliable inferences about pathogens that cause significant morbidity and mortality.

Keywords: Protozoan parasites; Evolution; Informatics; Comparative genomics

1. Introduction

Systematics, the study and elucidation of the relationships between species and determination of their origins, is pivotal in biology. A reliable phylogeny can provide insights beyond simply demonstrating that one organism is closely or distantly related to another. A phylogenetic tree also allows inference of the primitive state in a given common ancestor. For protozoan parasites with seemingly simple biology, a phylogenetic tree can determine whether a specific lineage under consideration is truly primitive compared to its more complex relatives, or is minimized due to secondary losses of characteristics. Such mapping of transformations of character states over a tree clarifies evolutionary trends. For parasitologists studying protists, this can provide context for life cycles, pathogenic mechanisms and susceptibility to therapeutic or control interventions. Such information is gained via understanding the basic biology of the organism through ultrastructure, molecular cell biology and genomics and can provide clues to vaccine candidates and therapeutic targets.

Estimating an accurate phylogeny is far from trivial, and has occupied biologists literally for centuries. Methodology has developed from the morphological to molecular, beginning with analyses of single genes. The availability of fully sequenced genomes from diverse eukaryotes and the development of methods for obtaining and comparing homologues of a broad array of genes have extended molecular analysis to large

* Corresponding author. Tel.: + 44 1223 333734.
E-mail address: mcf34@cam.ac.uk (M.C. Field).

1383-5769/$ - see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.parint.2007.11.004

concatenated gene sets and reconstruction of ancestral states. This has led to a revolution in understanding of eukaryotic relationships and our view of parasite evolution.

2. The "Elitist" view of eukaryotic evolution: humans, plants and the "crown"

In the late 1980s and early 1990s, analyses of small subunit ribosomal RNA (SSU rRNA) genes [1,2] or selected protein coding genes [3], rooted by prokaryotic homologues of orthologous genes, produced a version of the eukaryotic tree still commonly, and unfortunately, adhered to by many today (Fig. 1). In this version, a few lineages of cytologically simple protists emerged from the base of the tree as deeply branching eukaryotes [4]. These included many prominent parasitic lineages such as the diplomonads (*Giardia*), parabasalids (*Trichomonas*), kinetoplastids (*Trypanosoma*) and the Microsporidia (*Nosema*). Above these lineages, representatives of various other groups of protists diverged away from the backbone of the tree in a ladder-like fashion, including the Apicomplexa (*Plasmodium*), the slime molds (*Dictyostelium*) and amoebae such as *Entamoeba*. The tree culminated in a crown radiation of taxa including animals, plants, Fungi and some further protists [4]. This model also appeared to agree with morphological data, since many of the organisms at the base of the tree also apparently lacked mitochondria. Because of this, these lineages were proposed to have evolved away from the main eukaryotic line prior to the acquisition of the mitochondria by endosymbiosis of an alpha proteobacterium [5,6]. The apparent lack of a Golgi complex, peroxisomes, introns and sexual cycles in these same taxa seemed to support a ladder of eukaryotic life model with its progression from the simple to the complex. Unfortunately this appealing model of eukaryotic relationships was wrong.

3. Relinquishing the crown

Evidence of several types have demonstrated that both the idea of ancient cytological simplicity and the ladder-like structure of the eukaryotic tree is incorrect, the product of both methodological artifact and "unparsimonious" evolution.

In every case that has been examined, the "mitochondria-lacking" taxa possess genes of mitochondrial origin encoded within their nuclear genomes [7–10]. In many cases these organisms also possess remnant mitochondrial organelles [11,12], which take either of two major forms: hydrogenosomes, a hydrogen-producing and energy generating organelle are found in parabasalids (*Trichomonas*) and some anaerobic ciliates and Fungi; and mitosomes, small organelles that do not appear to have a role in energy generation and are present in *Giardia*, *Entamoeba* and Microsporidia [13]. The presence of these organelles in formerly "amitochondriate" protists strongly suggests that the origin of mitochondria pre-dates the ancestor of extant eukaryotes and emphasizes that these organisms are not so simple after all [14].

While the presence of these mitochondrial homologues disrupts the interpretation of the eukaryotic tree, it does not contradict the stepwise model of evolution. Apart from the trivial, but often ignored, point that hypotheses of relatedness could only bear on the taxa sampled (many eukaryotic groups were not sampled until recently, e.g. by [15–18]), representation of eukaryotic groups was extremely unbalanced, with the vast majority of sequences in trees being from the taxa that seemed to belong to the "crown" group. When more of the "Archezoan" taxa were included in more complex analyses, formerly long branches began to appear higher in the tree, breaking up the clear 'basal branches + crown group' pattern, resolving clades more evenly and congruently with ultrastructure (e.g. [15,16,19]). It also became clear that methodological artifacts, particularly long-branch attraction (LBA), had severely affected the phylogenetic analyses: the realisation of this undermined that old view of eukaryotic systematics. LBA is the artifactual clustering of sequences that are divergent from the majority of the dataset, regardless of whether that divergence is due to rapid, but recent, evolution or slow accumulation of changes over time [20,21]. Genes from different organisms may evolve at different rates [22]. Not all sites within a gene evolve at the same rate either [23] or in the same way [24], with mutational saturation [25], compositional heterogeneity [26,27] and substitutional biases [28] also contributing to the selection of the incorrect tree [29]. Several studies, using a variety of methods showed that LBA was affecting analyses of the eukaryotic tree both for the protein and the SSU rRNA datasets [30–33, *inter alia*].

Compensating for LBA had two distinct effects. In the case of the Microsporidia, implementing maximum-likelihood methods, correcting for rate variation, and using new gene sequences less affected by rate variation, allowed a robust placement of these formerly deep-branching eukaryotes as a divergent fungal group [34–36]. This makes them no less interesting or pathogenic. It does, however, allow for their unique traits to be seen in the light of a very well-characterized group of organisms, hopefully enabling more effective study and eventual treatment. Similarly *Plasmodium*, kinetoplastids and *Entamoeba* all found new homes relatively quickly (see below). Other organisms such as *Giardia*, and *Trichomonas* have not been as fortunate [37]: resolution of their phylogenetic placement has taken years and is still on-going; the recent completion of the genomes of both of these organisms represents an important step towards resolution of this critical issue [38,39]. Generally, the use of more biochemically realistic models of sequence evolution reduces the support for the ladder-like topology in phylogenetic analyses [30–33]. Tellingly, as the most saturated positions of a dataset are removed, the support for resolution of the backbone recedes [25]. All of this is consistent with the observed resolution being due to artifact and not historical signal. By the time this period of phylogenetic deconstruction occurred in the late 1990s, the field was left with an unresolved assemblage of many eukaryotic groups and a fairly agnostic view of eukaryotic evolution [40].

Part of the problem was that the "elitist" view conflated two separate issues; the relationships between the various lineages and the rooting of the eukaryotic tree (which determines branching order from the base to the tips of the tree). By inclusion of

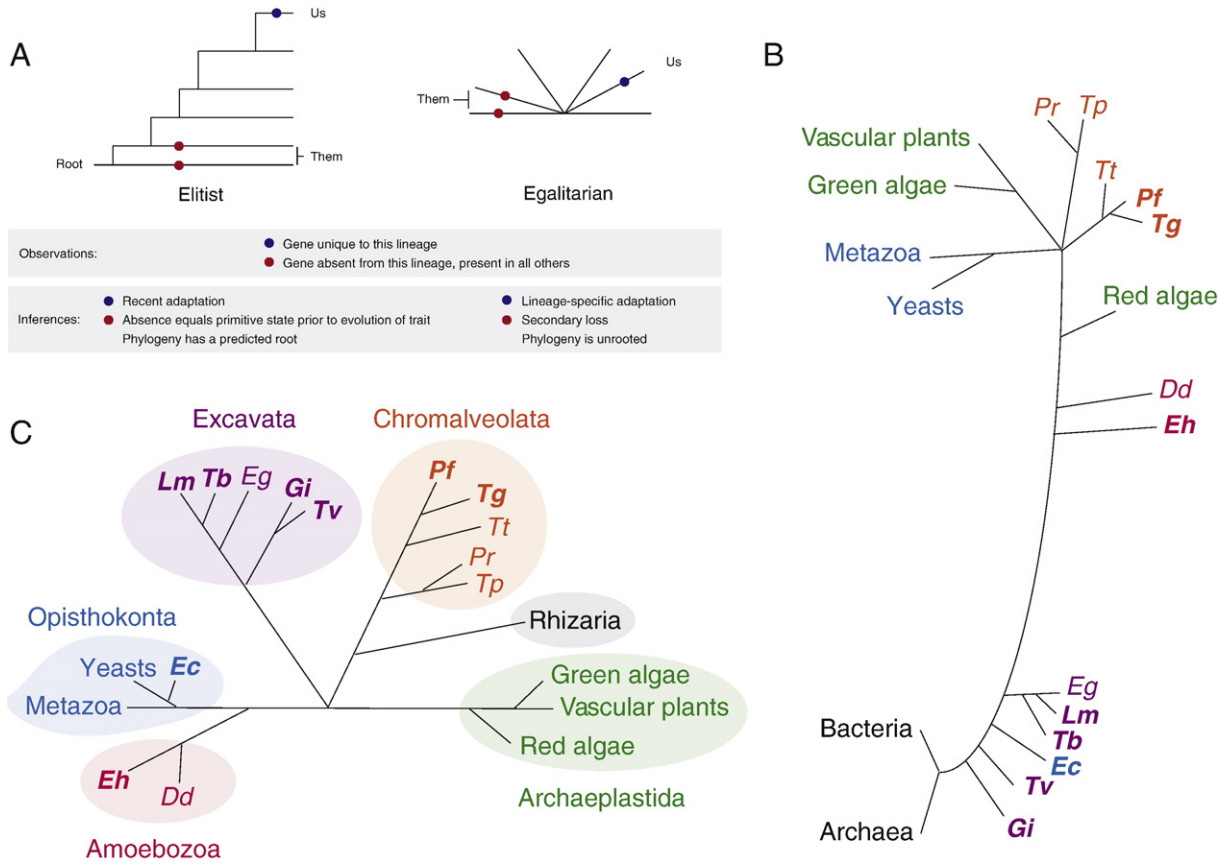

Fig. 1. : Comparisons of evolutionary models and interpretive implications. Panel A: The Egalitarian and Elitist models. The classical tree of life model, termed here "Elitist" as mammals and *Homo* in particular are placed at the apex. This representation is based on the assumption that the root is placed amongst taxa containing protists, and is supported by rRNA data, together with other evidence, and in a sense mirrors the great chain of being of the 19th century. Critically the topology implies that taxa close to the root have existed for a longer time (and hence are ancient systems) than those at the apex, which are considered as modern. The major problem with interpretation of this model is that many workers assume that once a lineage has speciated from the upward course of evolution, it ceases to evolve, and in some manner reflects life on earth at the point of speciation. This contrasts with the more recent "Egalitarian" view, which critically has an unresolved root and a "big bang" event essentially producing all eukaryotic supergroups rapidly in a time-unresolved event. Here all taxa are more easily recognised as equally ancient (or modern). This topology is supported by more extensive sequence and morphological evidence than the Elitist tree, *albeit* with several issues concerning precise placement of the root remaining unresolved at this time [13]. Blue and red dots indicate the presence or absence, respectively, of genes (or traits) of interest, while "them" refers to parasitic pathogens and "us" to *Homo.* The very critical point here is that blue events under the Egalitarian model are interpreted as lineage-specific evolution, rather than as necessarily recent events, while red events, i.e. absences, are best interpreted as the result of secondary losses rather than evidence for an ancestral or primitive state. This has a profound impact on our views of the sophistication of parasitic systems, their biology and strategies by which to control them. Panel B: Example Elitist tree topology, based on data from rRNA data and loosely adapted from Sogin [4]. Panel C; Egalitarian tree based on Adl et al. [23] and others. Here supergroups are named and indicated by shaded lozenges. Critically, present data do not allow a root to be determined. Note also that parasitic taxa are clearly interspersed by non-parasitic species. In both panels B and C specific taxa are indicated by two letter abbreviations of the Linnean names, while general taxa, e.g. vascular plants, are not specified. Taxa represented are Dd; *Dictyostelium discoidium*, Ec; *Encephalitozoon cuniculi*, Eg; *Euglena gracilis,* Eh; *Entamoeba histolytica*, Gi; *Giardia intestinalis*, Lm; *Leishmania major*, Pf; *Plasmodium falciparum*, Pr; *Phytophthora ramorum*, Tb; *Trypanosoma brucei*, Tg; *Toxoplasma gondii*, Tp; *Thalassiosira pseudonana*, Tt; *Tetrahymena thermophila* and Tv; *Trichomonas vaginalis.* Taxon supergroups are colour-coded and parasitic taxa are in bold.

sequences from prokaryotes, which truly had evolved away from eukaryotes a long time ago, the rapidly evolving protistan lineages were artifactually attracted to these outgroups. More recently analyses have separated these questions, some using diverse types of data to resolve the relative relationships between eukaryotes and others to place the root of the eukaryotic tree.

4. The new phylogeny of eukaryotes: a more egalitarian view

A combination of morphological and molecular studies have begun to resolve the majority of eukaryotes into six 'supergroups' described in a synthesis of recent literature by Simpson and Roger and formalized by Adl et al. [41,42]. These six groups are illustrated in Fig. 1C. Beginning from the left, the opisthokonts contain the metazoa, the Fungi and various single-celled relatives including the nucleariid amoebae and fish-pathogens, the Ichthyosporea. This supergroup is well supported by single gene [43], multigene [44,45] and morphological characters, as well as the presence of a diagnostic insertion in the EF1 alpha gene [46]. The supergroup Amoebozoa unites many, but not all, lobose amoeboid taxa including the slime molds (Mycetozoa), the lobose amoebae *sensu stricto* such as *Amoeba proteus* and *Chaos*, the "archamoebae" i.e. pelobionts (*Mastigamoeba* and *Pelomyxa*) and entamoebae (including *Entamoeba histolytica*), as well as many taxa of indeterminate

affinities such as the parasite *Acanthamoeba castellanii*. The supergroup Archaeplastida consists of the red algae, green algae and vascular plants, and the glaucophytes [41]. Members of these groups all possess primary plastids, the result of endosymbiosis of cyanobacteria. Single and multigene phylogenies support the unity of these lineages [47,48], and although it appears likely that the acquisition of plastids in his supergroup happened only once, some controversy remains [49]. The supergroup Rhizaria includes taxa with fine, filose, sometimes granular pseudopodia, including Cercozoa, Foraminifera, *Gromia*, phytomyxids (including the significant pathogen *Plasmodiophora brassicae*), and Haplosporidia (including parasites of the genera *Haplosporidium* and *Urosporidium*) [50]. Recent results from multigene analysis supports the Rhizaria as a group [45,51,52]. These are the four least controversial supergroups, some with clear morphological synapomorphies, some supported by rare genetic characters such as diagnostic insertions in genes, and each resolved by studies from single gene and multigene concatenated datasets.

More contentious are the last two supergroups, the excavates and the chromalveolates, both of which contain important parasitic lineages. The chromalveolates include three major protistan groups which all contain members with chloroplasts derived from a red algal endosymbiosis [53]. The first major group, the alveolates, has long been recognised on morphological grounds [54]; it harbours parasites such as *Plasmodium*, *Toxoplasma* and *Cryptosporidium*. These apicomplexans are reliably placed as related to the dinoflagellates (another group with parasitic representatives) and the ciliates [55,56]. In turn, the alveolates are thought to be related to the stramenopiles, another group initially recognised on morphological grounds [57] that includes diatoms, brown algae and the oomycetes. The third major group in the chromalveolates includes the haptophytes, kathablepharids and cryptophytes [51,58]. The support joining alveolates, stramenopiles and the cryptophyte–haptophyte lineage is variable [59]. Recent analyses have placed the Rhizaria as robustly sister to the group of chromists and alveolates (Fig. 1C) which even raises the question of whether there should be six supergroups or five [45,51,60].

The Excavata, or excavates, contains many parasitic lineages, including *Giardia*, *Trichomonas*, and *Trypanosoma* [61]. As yet no molecular analysis has resolved all ten proposed excavate lineages as a monophyletic group. However, five of the excavate taxa possess a ventral feeding groove, which is structurally supported by a complex set of cytoskeletal elements and argues for the monophyly of the organisms possessing it [61]. Single-gene molecular data also unites groove-bearing organisms with other excavates that lack the groove structure as a whole but that do possess some of the homologous cytoskeletal elements [61]. Recent multigene concatenations have provided some higher-level resolution within the excavates [62,63]. Since the excavate group is thought to contain many taxa strongly affected by LBA, it is not surprising that establishing the validity of this group and the phylogenetic placement of its composite membership continues to be challenging.

The final unresolved question concerns the placement of the root of the eukaryotic tree. Because of problems with systematic and stochastic phylogenetic artefacts, an alternate strategy involving rare genetic characters such as gene duplications or fusions, indels and endosymbiotic events has been used to exclude the root from certain positions within the tree. This approach has led to several competing hypotheses, supported by apparently mutually inconsistent evidence (for more discussion and relevant data see [64] and references therein; [65–69]. Whatever the root, it will clearly require a sophisticated explanation, based on multiple characters and a resolved set of interrelationships amongst eukaryotes at the highest level.

5. What the new eukaryotic systematics means to parasitologists

The new view of eukaryotic systematics does more than displace *Homo sapiens* further from a pre-eminent position within the tree of life. This re-organization of relationships and especially the removal of the root of eukaryotes from organisms such as *Trichomonas*, *Giardia*, *Trypanosoma, Entamoeba* and the Microsporidia has several important implications for parasitologists when inferring traits or interpreting their data.

Firstly, that these parasites belong to different supergroups, separated by many intervening non-parasitic taxa, emphasises that parasitism is relatively rare in the unicellular eukaryotes, and has arisen independently many times. The spaces between well known parasitic taxa are becoming ever more filled with knowledge of their relatives [41,70]. All amoebae are not necessarily related: for example, *Entamoeba* is in the Amoebozoa and the veterinary parasite *Monocercomonas* in the excavates along with the amoeboflagellate *Naegleria* [71]. Consequently, inferences drawn from one parasite cannot necessarily be mapped to others, even those with similar morphology. In contrast, where mechanisms between two divergent taxa do appear similar, as in the control of the *var* gene expression in *Plasmodium* and VSG genes in *T. brucei*, the implications of convergent evolution are profound. Placement of variable surface antigens at telomeres and the restriction of active genes within limited sub-nuclear domains specifically to constrain transcriptional activity to one or limited numbers of variable surface antigen genes is common to both *Plasmodium* and *T. brucei* antigenic variation strategies [72,73]. Apparent convergent evolution of a specific mechanism for antigenic variation suggests multiple independent origins, and implies that this strategy may be the only effective mechanism possible for achieving antigenic variation in the context of blood-borne infection. Additionally, the phylogenetic separation of the parasites contained within the Amoebozoa, Excavata, and Chromalveolata highlights the importance of the choice of model experimental systems and indicates that each group must be characterized to considerable depth in its own right, as generalizations may frequently not be relevant.

Secondly, the phylogeny has serious implications for the idea of primitive and ancient parasites. To begin with, a common misconception in the parasitological literature is that organisms nearest to the base of the tree are more ‘ancient’ than humans, yeast or plants [74–76]. This is patently false, regardless of which tree is considered, but is more obvious with the revised phylogeny. All eukaryotes living today are the same evolutionary

age; some lineages may have diverged or speciated away from the remaining eukaryotes earlier than others but that does not make the organisms in them "older". Nor does early speciation imply that these organisms have remained evolutionarily frozen and thus represent a primitive or ancestral state. This is especially true of parasites that have co-evolved with their more recently speciated animal hosts. Even early-branching eukaryotes, when a properly rooted tree is determined, will have had the same time to have evolved their own adaptations as all other eukaryotes.

Both in terms of organelles and genes, the evidence coming from the mapping of characters on the new phylogeny, as well as from comparative cytology and genomics, support the conclusion that nearly all cases of 'primitive absence' are better explained as secondary loss or failure to identify the relevant homologue. This is not only true of mitochondria as discussed early, but also of other supposedly absent features that suggested various eukaryotes as primitively simple. For example, the absence of a stacked Golgi complex in multiple lineages is clearly the result of secondary loss on several occasions, and despite the morphological evidence, molecular data indicate that much of the function of the Golgi complex is likely retained in the "Golgi-less" lineages [76–78]. Secondary loss also appears to explain the absence or paucity of introns [79], meiotic machinery [80], peroxisomes [81] and nucleoli [82], amongst others. Crucially, reconstruction of the gene complement of the eukaryotic ancestor suggests a complicated cell [83–85] and implies that where a gene is absent in a particular parasite it has most likely been lost because homologues are present in the non-parasitic close relatives [86]. *Giardia*, *Trichomonas*, trypanosomes and apicomplexans are not ancient, primitive or living fossils that can provide insights into the cell biology of ancient and extinct eukaryotes *per se*. They can, however, provide excellent insight into the impact of streamlining and adaptation on the evolution of parasitic systems if one uses comparative genomics in a rigorous way, taking care about statements of absence.

6. Comparative genomics; determining commonalities and unique features in parasites

Parasitologists are often as interested in the genes that are absent from a genome as much as they are in those that are present. Absence of a gene product may provide important insight into minimized systems, the exploitation of a specific pathway or a potential drug target. For example, *T. gondii* is unable to synthesize purines and has a total reliance on host purine metabolism [87], while the presence of the serum resistance-associated antigen gene in some subspecies of *T.*

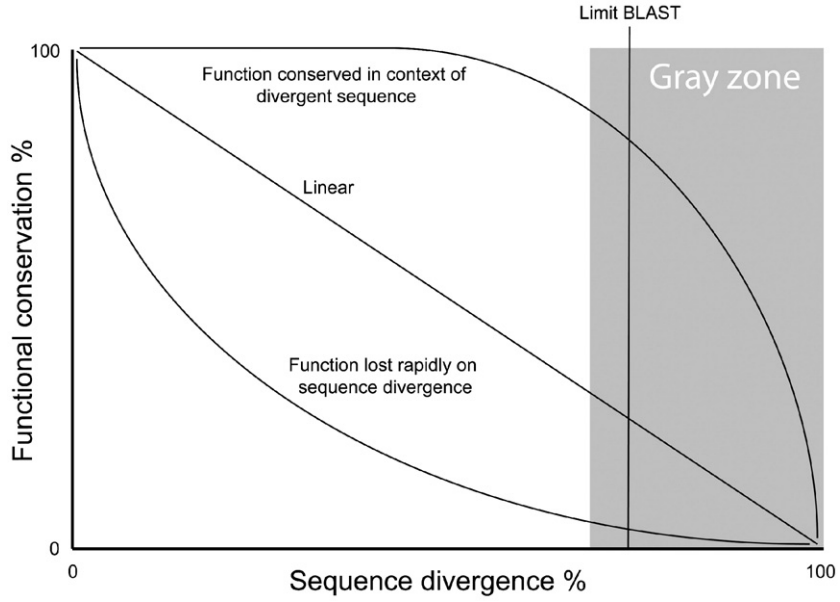

Fig. 2. : The problem with homology searching. Nearly all search algorithms are essentially text string-based, and while there may be additional aspects to how these searches are conducted (e.g. global versus local alignments, significance matrices, masks, hidden Markov models, annotation extraction and rarely, consideration of 3D-structure), all of these algorithms will eventually fail when the divergence between a query sequence and the desired homologue falls below a statistically determined level of similarity. This issue is particularly of importance to parasitologists, and is frequently under-appreciated by many in the mainstream cell biology community. That the divergence between the protozoan organisms of interest and many eukaryotic model systems can be huge, due both to deep-evolutionary splits and rapid evolutionary rates in many parasites compounds the difficulty in reliable identification of genes. Considering protein sequence data, the point at which detection fails is dependent on the nature of the protein; specifically the manner in which change of function maps to sequence divergence. For example, while it is possible that such a relationship is linear (middle line), it is far more likely that functional equivalence may be retained between two sequences even in the face of considerable sequence divergence (upper line), or functional conservation may be more rapidly lost (lower line). As the point of most searches is to identify genes encoding the same or related function, the manner of mapping sequence to function is critical. BLAST (or any other search algorithm) fails at some point to detect a statistically significant relationship between two sequences, but it is clear that the point of failure is both dependent on the nature of the sequences being investigated and frequently occurs prior to true loss of functional relatedness, i.e. a false negative. This constitutes a gray zone, a region of uncertainty in which it can be very challenging to "make a call" on the presence or absence of a specific factor. Recent experience by us using the Smith-Waterman algorithm has indicated that the false-negative rate for BLAST is quite significant, and may approach 25% for some systems (MCF and JBD, unpublished data). Of course, the precise definition of function, or functional equivalence, is itself a somewhat elusive concept in many instances.

brucei has provided major insights into host range and resistance mechanisms [88]. Recent efforts have sought to determine whether there are specific genes associated with the distinct clinical manifestations of infection with various *Leishmania* species [89].

The theoretical ability to predict absence of a genomic component is one of the most powerful advances that genomics has provided. Comparative genomics attempts to predict function, based on identification within genomes of interest of orthologous or paralogous sequence relationships. From this, one should be able to build a comprehensive picture of a parasitic system including what is not present because the genomic sequence, theoretically, is the list of all genes in a genome. Combined with a resolved set of eukaryotic relationships, this allows mapping of the evolution and rise of particular character states. Such analysis is very powerful but there are pitfalls that await the unwary.

In practice, true absence is difficult to determine, and failure to identify a gene of interest may be due to explanations other than the gene not being present in the genome in question. Data limitations (an incomplete genome), BLAST failure or inappropriate choice of query sequence, are possible factors. In fact, few eukaryotic genomes are truly complete, many languishing at the 95–99% coverage level, and for financial and technical reasons are likely never to progress beyond this point. BLAST failure is particularly of concern in parasitology. Often opisthokont genes are used as search queries, since these are frequently the homologues that have been functionally characterized. However, these may be highly dissimilar at the sequence level from parasites within the excavate or chromalveolate supergroups. As BLAST searches for regions of local similarity, this issue is compounded both by divergence and differential rates of sequence evolution (Fig. 2). Regardless of the sophistication of a search method (BLAST, Smith-Waterman, structural, HMM, PSI-BLAST, etcetera), at some level homology or orthology becomes impossible to statistically determine with any confidence. A potent example of the dangers of over-interpretation due to absence of data was the suggestion of massive lateral gene transfer (LGT) from prokaryotes into the human genome. The vast majority of LGT candidates were, in fact, present in a wide variety of eukaryotic genomes, but this artifact only became apparent when additional genomic sequences became available [90].

What can be done? To find a reliable homologue, certainly more sensitive search methods are powerful. Smith-Waterman [91], a more global sequence similarity search algorithm than BLAST, is in our hands able to detect up to 25% additional paralogues in many protistan genomes in searches where BLAST fails (unpublished data). PSI-BLAST can also increase sensitivity but also the possibility for false positives and so must be used with care [92]. Also, once a reliable homologue has been found in one organism using a functionally characterised query sequence, then that new homologue can be used as the query to search in genomes from the same supergroup. Finally, motif searches and inspection of annotated genes based on the presence of a conserved domain are helpful in establishing valid homology. However, it is critical that a reverse BLAST is carried out back to a genome which has a functionally characterised homologue to validate the relationship and discriminate between orthology, paralogy and a chance hit that is in fact erroneous. Even better is to use phylogenetic reconstruction to robustly confirm an orthologous relationship in the cases of a paralogous gene family. Ultimately, however, one has to accept that a sequence has simply "not been found". Using this statement, instead of a strong statement of absence, leaves open the possibility that improved algorithms, availability of more closely related genomes, or even additional sequencing, may ultimately uncover a sequence that could not be detected previously.

Absence can still be predicted, with a few caveats. For example, if all components of a complex are missing from a given genome, that is suggestive of legitimate absence. Similarly, if the same component is missing from several members of the same evolutionary grouping, e.g. all components of the TRAPPII tethering complex are missing from the excavates sampled [93], then it can be more safely predicted that the gene is legitimately not present. Here again accurately estimating the phylogeny of eukaryotes becomes important.

7. Concluding remarks — knowing the enemy

Knowledge of the systematics of parasites is immensely important for parasitologists, and the old rRNA model has served us well. The revision of this tree, as we move into the post-genomic era for many protozoan pathogens, is set to inform both research strategy and precise experimental design. Ongoing efforts in phylogenetics are confidently expected to bring forward further surprises. Comparative genomics is clearly one of the more important emerging tools that parasitologists can use to study and combat disease, but the correct interpretation of these data depends triply on an accurate knowledge of species relationships, on understanding the limitations of the algorithms used to compare gene sequences, and in appreciating the careful and cautious considerations that interpretation of data obtained from these studies requires. Protozoan parasites, as highly divergent organisms, remain a massive challenge to human health, but with a new view of their relationships we are coming to know our enemy perhaps rather better.

Acknowledgements

This work was supported by a Wellcome Trust grant to MCF and fellowship support to JBD from the Canadian Institutes of Health Research and the Wellcome Trust, as well as from the Parke Davis Foundation. GW was supported by the John Stanley Gardiner Fund of the University of Cambridge's Zoology Department.

References

[1] Vossbrinck CR, Maddox JV, Friedman S, Debrunner-Vossbrinck BA, Woese CR. Ribosomal RNA sequence suggests Microsporidia are extremely ancient eukaryotes. Nature 1987;326:411–4.

[2] Leipe DD, Gunderson JH, Nerad TA, Sogin ML. Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol Biochem Parasitol 1993;59:41–8.

[3] Hashimoto T, Nakamura Y, Kamaishi T, Nakamura F, Adachi J, Okamoto K, et al. Phylogenetic place of mitochondrion-lacking protozoan, *Giardia lamblia*, inferred from amino acid sequences of elongation factor 2. Mol Biol Evol 1995;12:782–93.
[4] Sogin ML. Early evolution and the origin of eukaryotes. Cur Op Gen Dev 1991;1:457–63.
[5] Margulis L. Origin of eukaryotic cells; evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. New Haven: Yale University Press; 1970.
[6] Cavalier-Smith T. A 6-kingdom classification and a unified phylogeny. In: Schenk HEA, Schwemmler W, editors. Endocytobiology II. Berlin: Walter de Gruyter; 1983. p. 1027–34.
[7] Clark CG, Roger AJ. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc Natl Acad Sci U S A 1995;92:6518–21.
[8] Bui ET, Bradley PJ, Johnson PJ. A common evolutionary origin for mitochondria and hydrogenosomes. Proc Natl Acad Sci U S A 1996;93: 9651–6.
[9] Germot A, Philippe H, Le Guyader H. Evidence for the loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. Mol Biochem Parasitol 1997;87:159–68.
[10] Roger AJ, Svard SG, Tovar J, Clark CG, Smith MW, Gillin FD, et al. 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 1998;95:229–34.
[11] Tovar J, Fischer A, Clark CG. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. Mol Microbiol 1999;32:1013–21.
[12] Tovar J, Leon-Avila G, Sanchez LB, Sutak R, Tachezy J, Van Der Giezen M, et al. Mitochondrial remnant organelles of *Giardia* function in iron–sulphur protein maturation. Nature 2003;426:172–6.
[13] van der Giezen M, Tovar J, Clark CG. Mitochondrion-derived organelles in protists and Fungi. Int Rev Cytol 2005;244:175–225.
[14] Embley TM, Martin W. Eukaryotic evolution, changes and challenges. Nature 2006;440:623–30.
[15] Silberman JD, Simpson AG, Kulda J, Cepicka I, Hampl V, Johnson PJ, et al. Retortamonad flagellates are closely related to diplomonads — implications for the history of mitochondrial function in eukaryote evolution. Mol Biol Evol 2002;19:777–86.
[16] Simpson AG, Roger AJ, Silberman JD, Leipe DD, Edgcomb VP, Jermiin LS, et al. Evolutionary history of "early-diverging" eukaryotes: the excavate taxon Carpediemonas is a close relative of *Giardia*. Mol Biol Evol 2002;19: 1782–91.
[17] Cavalier-Smith T, Chao EE. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. J Mol Evol 2003;56: 540–63.
[18] Cavalier-Smith T, Chao EE, Oates B. Molecular phylogeny of Amoebozoa and the evolutionary significance of the unikont *Phalansterium*. European Journal of Protistology 2004;40:21–48.
[19] Dacks JB, Silberman JD, Simpson AG, Moriya S, Kudo T, Ohkuma M, et al. Oxymonads are closely related to the excavate taxon *Trimastix*. Mol Biol Evol 2001;18:1034–44.
[20] Hendy M, Penny D. A framework for the quantitative study of evolutionary trees. Syst Zool 1989;38:297–309.
[21] Holland BR, Penny D, Hendy MD. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock — a simulation study. Syst Biol 2003;52:229–38.
[22] Lockhart PJ, Steel MA, Barbrook AC, Huson DH, Charleston MA, Howe CJ. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol Biol Evol 1998;15:1183–8.
[23] Yang Z. Among-site variation and its impact on phylogenetic analyses. TREE 1996;11:367–72.
[24] Lockhart PJ, Larkum AW, Steel M, Waddell PJ, Penny D. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc Natl Acad Sci U S A 1996;93:1930–4.
[25] Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, et al. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc R Soc Lond B Biol Sci 2000;267:1213–21.
[26] Philippe H, Germot A. Phylogeny of eukaryotes based on ribosomal RNA: long-branch attraction and models of sequence evolution. Mol Biol Evol 2000;17:830–4.
[27] Foster PG, Hickey DA. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J Mol Evol 1999;48: 284–90.
[28] Lartillot N, Brinkmann H, Philippe H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol 2007;7(Suppl 1):S4.
[29] Gaut BS, Lewis PO. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol Biol Evol 1995;12:152–62.
[30] Dacks JB, Marinets A, Doolittle WF, Cavalier-Smith T, Logsdon Jr JM. Analyses of RNA Polymerase II genes from free-living protists: phylogeny, long branch attraction, and the eukaryotic big bang. Mol Biol Evol 2002;19:830–40.
[31] Philippe H, Adoutte A. The molecular phylogeny of Eukaryota: solid facts and uncertainties. In: Coombs G, et al, editor. Evolutionary relationships among Protozoa. London: Chapman & Hall; 1998. p. 25–56.
[32] Roger AJ, Sandblom O, Doolittle WF, Philippe H. An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. Mol Biol Evol 1999;16:218–33.
[33] Stiller JW, Hall BJ. Long-branch attraction and the rDNA model of early eukaryotic evolution. Mol Biol Evol 1999;16:1270–9.
[34] Edlind TD, Li J, Visvesvara GS, Vodkin MH, McLaughlin GL, Katiyar SK. Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. Mol Phylogenet Evol 1996;5:359–67.
[35] Keeling PJ, Doolittle WF. Alpha-tubulin from early-diverging eukaryotic lineages and the evolution of the tubulin family. Mol Biol Evol 1996;13: 1297–305.
[36] Hirt RP, Logsdon Jr JM, Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci U S A 1999;96: 580–5.
[37] Keeling PJ. Genomics. Deep questions in the tree of life. Science 2007;317: 1875–6.
[38] Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science 2007;315:207–12.
[39] Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, et al. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 2007;317:1921–6.
[40] Dacks JB, Doolittle WF. Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. Cell 2001;107:419–25.
[41] Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 2005;52:399–451.
[42] Simpson AG, Roger AJ. The real 'kingdoms' of eukaryotes. Curr Biol 2004;14:R693–6.
[43] Wainright PO, Patterson DJ, Sogin ML. Monophyletic origin of animals: a shared ancestry with the Fungi. Soc Gen Physiol Ser 1994;49:39–53.
[44] Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 2000;290:972–7.
[45] Rodriguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW, Philippe H, et al. Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. Curr Biol 2007;17:1420–5.
[46] Baldauf SL, Palmer JD. Animals and Fungi are each other's closest relatives: congruent evidence from multiple proteins. Proc Natl Acad Sci U S A 1993;90:11558–62.
[47] Moreira D, Le Guyader H, Phillippe H. The origin of red algae and the evolution of chloroplasts. Nature 2000;405:69–72.
[48] Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, et al. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 2005;15:1325–30.
[49] Larkum AW, Lockhart PJ, Howe CJ. Shopping for plastids. Trends Plant Sci 2007;12:189–95.
[50] Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, et al. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci U S A 2004;101:8066–71.
[51] Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. Phylogenomic analysis supports the monophyly of cryptophytes and

haptophytes and the association of Rhizaria with chromalveolates. Mol Biol Evol 2007;24:1702–13.
[52] Burki F, Pawlowski J. Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. Mol Biol Evol 2006;23:1922–30.
[53] Cavalier-Smith T. Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Euk Micro 1999;46:347–66.
[54] Taylor FJ. Flagellate phylogeny: a study in conflicts. J Protozool 1976;23:28–40.
[55] Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML. Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol 1991;45:147–54.
[56] Fast NM, Xue L, Bingham S, Keeling PJ. Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol 2002;49:30–7.
[57] Patterson DJ. Stramenopiles: chromophytes from a protistan perspective. In: Green JC, Leadbeater BSC, Diver WL, editors. The Chromophyte Algae: Problems and Perspectives. Oxford: Clarendon Press; 1989. p. 357–79.
[58] Patron NJ, Inagaki Y, Keeling PJ. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol 2007;17:887–91.
[59] Bodyl A. Do plastid-related characters support the chromalveolate hypothesis? J Phycol 2005;41:712–9.
[60] Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, et al. Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE 2007;2:e790.
[61] Simpson AG. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol 2003;53:1759–77.
[62] Simpson AG, Inagaki Y, Roger AJ. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol 2006;23:615–25.
[63] Hampl V, Horner DS, Dyal P, Kulda J, Flegr J, Foster PG, et al. Inference of the phylogenetic position of oxymonads based on nine genes: support for Metamonada and Excavata. Mol Biol Evol 2005;22:2508–18.
[64] Moestrup O. The flagellate cytoskeleton: introduction of a general terminology for microtubular flagellar roots in protists. In: Leadbeater BSC, Green JC, editors. The Flagellates: Unity, Diversity and Evolution. London: Taylor and Francis; 2000. p. 69–94.
[65] Moestrup O, Sengco M. Ultrastructural studies on *Bigellowiella natans*, gen. et sp. nov, a chlorarachniophyte flagellate. J Phycol 2001;37:624–46.
[66] Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Higashiyama T. Phylogenetic implications of the CAD protein complex from the primitive red alga *Cyanidioschyzon merolae* (Cyanidiales, Rhodophyta). J Phycol 2005;41:652–7.
[67] Richards TA, Cavalier-Smith T. Myosin domain evolution and the primary divergence of eukaryotes. Nature 2005;436:1113–8.
[68] Stechmann A, Cavalier-Smith T. Rooting the eukaryote tree by using a derived gene fusion. Science 2002;297:89–91.
[69] Stechmann A, Cavalier-Smith T. The root of the eukaryote tree pinpointed. Curr Biol 2003;13:R665–6.
[70] Patterson DJ. The diversity of eukaryotes. Am Nat 1999;154:S96–S124.
[71] Hampl V, Cepicka I, Flegr J, Tachezy J, Kulda J. Critical analysis of the topology and rooting of the parabasalian 16S rRNA tree. Mol Phylogenet Evol 2004;32:711–23.
[72] Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, et al. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. Cell 2005;121:13–24.
[73] Horn D, Barry JD. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. Chromosom Res 2005;13:525–33.
[74] Tazeh NN, Bangs JD. Multiple motifs regulate trafficking of the LAMP-like protein p67 in the ancient eukaryote *Trypanosoma brucei*. Traffic 2007;8:1007–17.
[75] Davis CA, Brown MP, Singh U. Functional characterization of spliceosomal introns and identification of U2, U4, and U5 snRNAs in the deep-branching eukaryote *Entamoeba histolytica*. Eukaryot Cell 2007;6:940–8.
[76] Marti M, Regos A, Li Y, Schraner EM, Wild P, Muller N, et al. An ancestral secretory apparatus in the protozoan parasite *Giardia intestinalis*. J Biol Chem 2003;278:24837–48.
[77] Dacks JB, Davis LA, Sjogren AM, Andersson JO, Roger AJ, Doolittle WF. Evidence for Golgi bodies in proposed 'Golgi-lacking' lineages. Proc Biol Sci 2003;270(Suppl 2):S168–71.
[78] Ghosh SK, Field J, Frisardi M, Rosenthal B, Mai Z, Rogers R, et al. Chitinase secretion by encysting *Entamoeba invadens* and transfected *Entamoeba histolytica* trophozoites: localization of secretory vesicles, endoplasmic reticulum, and Golgi apparatus. Infect Immun 1999;67:3073–81.
[79] Vanacova S, Yan W, Carlton JM, Johnson PJ. Spliceosomal introns in the deep-branching eukaryote *Trichomonas vaginalis*. Proc Natl Acad Sci U S A 2005;102:4430–5.
[80] Ramesh MA, Malik SB, Logsdon Jr JM. A phylogenomic inventory of meiotic genes; evidence for sex in Giardia and an early eukaryotic origin of meiosis. Curr Biol 2005;15:185–91.
[81] Schluter A, Fourcade S, Ripp R, Mandel JL, Poch O, Pujol A. The evolutionary origin of peroxisomes: an ER-peroxisome connection. Mol Biol Evol 2006;23:838–45.
[82] Xin DD, Wen JF, He D, Lu SQ. Identification of a *Giardia* krr1 homolog gene and the secondarily anucleolate condition of *Giaridia lamblia*. Mol Biol Evol 2005;22:391–4.
[83] Hartman H, Fedorov A. The origin of the eukaryotic cell: a genomic investigation. Proc Natl Acad Sci U S A 2002;99:1420–5.
[84] Dacks JB, Field MC. Eukaryotic cell evolution from a genomic perspective: the endomembrane system. In: Hirt RP, Horner DS, editors. Organelles, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics. London: CRC Press; 2004. p. 309–34.
[85] Field MC, Gabernet-Castello C, Dacks JB. Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. Adv Exp Med Biol 2007;607:84–96.
[86] Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, et al. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A 2005;102:1548–53.
[87] Chaudhary K, Ting LM, Kim K, Roos DS. *Toxoplasma gondii* purine nucleoside phosphorylase biochemical characterization, inhibitor profiles, and comparison with the *Plasmodium falciparum* ortholog. J Biol Chem 2006;281:25652–8.
[88] Pays E, Vanhollebeke B, Vanhamme L, Paturiaux-Hanocq F, Nolan DP, Perez-Morga D. The trypanolytic factor of human serum. Nat Rev Microbiol 2006;4:477–86.
[89] Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat Genet 2007;39:839–47.
[90] Stanhope MJ, Lupas A, Italia MJ, Koretke KK, Volker C, Brown JR. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. Nature 2001;411:940–4.
[91] Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol 1981;147:195–7.
[92] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–402.
[93] Koumandou VL, Dacks JB, Coulson RM, Field MC. Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. BMC Evol Biol 2007;7:29.