ELSEVIER

Genome diversity in microbial eukaryotes

Casey L. McGrath[1] and Laura A. Katz[1,2]

[1]Department of Biological Sciences, Smith College, Northampton, MA 01063, USA
[2]Program in Organismic and Evolutionary Biology, UMass-Amherst, Amherst, MA 01003, USA

The genomic peculiarities among microbial eukaryotes challenge the conventional wisdom of genome evolution. Currently, many studies and textbooks explore principles of genome evolution from a limited number of eukaryotic lineages, focusing often on only a few representative species of plants, animals and fungi. Increasing emphasis on studies of genomes in microbial eukaryotes has and will continue to uncover features that are either not present in the representative species (e.g. hypervariable karyotypes or highly fragmented mitochondrial genomes) or are exaggerated in microbial groups (e.g. chromosomal processing between germline and somatic nuclei). Data for microbial eukaryotes have emerged from recent genome sequencing projects, enabling comparisons of the genomes from diverse lineages across the eukaryotic phylogenetic tree. Some of these features, including amplified rDNAs, subtelomeric rDNAs and reduced genomes, appear to have evolved multiple times within eukaryotes, whereas other features, such as absolute strand polarity, are found only within single lineages.

Microbial eukaryotes are a diverse group of organisms characterized by many unusual genome features. These features challenge some of the concepts and assumptions about eukaryotic genome evolution that have emerged from studies of plants, animals and fungi (Box 1). Here, we demonstrate the dramatic diversity of genome structures of microbial eukaryotes using examples from both the nuclear and organellar genomes. Interpreting these examples in a phylogenetic context enables us to determine whether these features arose multiple times within eukaryotes or whether they had a single origin. Clearly, both the number of unusual genome structures and our ability to discern the evolutionary history of these genomic peculiarities will increase as more data from eukaryotic microbes become available.

The diversity of microbial eukaryotes

Interpreting the evolution of eukaryotic genomes requires knowledge of the evolutionary relationships among eukaryotes, particularly among the microbial lineages. Microbial eukaryotes, or protists, are defined loosely as eukaryotic organisms that are not plants, animals or fungi. Reconstructing eukaryotic phylogeny has proven difficult, in part because there are few morphological characters that can be used to resolve deep nodes [1]. Similarly, there is considerable discordance among single-gene genealogies and many eukaryotic groups are un- or undersampled [2–4]. Recent multigene analyses support the monophyly of several major clades, although deep nodes within eukaryotes remain unknown [5,6]. It is unclear to what extent the lack of resolution at deep nodes will be resolved when additional genes and taxa are sampled for molecular phylogenies.

Analyses of both molecular and morphological characteristics support the monophyly of the alveolates (ciliates, apicomplexans and dinoflagellates) and the Euglenozoa (euglenids and kinetoplastids; Figure 1, Box 2) [1,7,8]. Molecular data have also provided further support for groups with few morphological SYNAPOMORPHIES (see Glossary), including the stramenopiles (water molds, brown algae, diatoms and labyrinthulids) and opisthokonts (animals, fungi, microsporidians, choanoflagellates and ichthyosporeans) [1,7,9]. Similarly, analyses of multigene genealogies have led to a revised hypothesis of the acquisition of photosynthesis in eukaryotes, with a PRIMARY ENDOSYMBIOSIS occurring in the ancestor of the clade containing glaucocystophytes, red algae and green algae (including plants) [10–13]. All other photosynthetic eukaryotes (e.g. cryptomonads, chlorarachniophytes,

Glossary

Absolute strand polarity: genes found in clusters on only a single strand of the DNA of kinetoplastids.
Conjugation: a form of sex in which genetic material is transferred between two temporarily joined cells.
Episomal element: small, extrachromosomal piece of DNA.
Genome duality: the presence of two distinct types of genome (e.g. germline and somatic) within a given cell or organism.
Karyotype: the complement of chromosomes within an organism; refers to both number and length of chromosomes.
Nucleomorph: remnant nucleus from a secondary endosymbiosis.
Polycistronic transcription: the generation of a single RNA transcript containing multiple genes.
Primary endosymbiosis: the evolution of eukaryotic cell structures by a eukaryote engulfing a bacterium.
Secondary endosymbiosis: the evolution of eukaryotic cell structures by a eukaryote engulfing a eukaryote.
Spliceosomal introns: introns that are removed by the 'spliceosome', a ribonucleoprotein complex.
Subtelomeric: a location next to telomeres (repetitive sequences that mark the ends of eukaryotic chromosomes).
Synapomorphy: a shared, derived character state that unites members of a clade.
Unigenic: containing only a single gene.

Corresponding author: Laura A. Katz (Lkatz@Smith.edu).

http://tree.trends.com 0169-5347/$ - see front matter doi:10.1016/j.tree.2003.10.007

Table 1. Estimates of genome size and gene content[a,b]

Genome	Location/taxonomy	Mb	Estimated number of protein-coding genes
Homo sapiens mtDNA	Mitochondrion	0.016	13
Guillardia theta	**Euk: nucleomorph**	**0.51**	**511**
Rickettsia prowazekii	Bacterium	1.1	834
Encephalitozoon cuniculi	**Euk: microsporidia**	**2.5**	**1997**
Escherichia coli	Bacterium	4.6	4377
Giardia lamblia	**Euk: diplomonad**	**12**	**~ 5000**
Saccharomyces cerevisiae	**Euk: yeast**	**12**	**5885**
Plasmodium falciparum	**Euk: apicomplexan**	**30**	**6500**
Leishmania major	**Euk: kinetoplastid**	**33.6**	**8600**
Caenorhabditis elegans	Euk: animal	97	17085
Arabidopsis thaliana	Euk: plant	117	25 498
Paramecium tetraurelia	**Euk: ciliate**	**150**	**~ 30 000**
Drosophila melanogaster	Euk: animal	180	14 14 000
Homo sapiens	Euk: animal	3286	~ 30 000–40 000
Amoeba dubia	**Euk: amoeba**	**670 000**	**Unknown**

[a]Taxa in bold are microbial eukaryotes, Euk, eukaryotes.
[b]References for individual taxa described in text. Additional data from [65], and, for *Giardia*, from Hilary Morrison (pers. commun.).

dinoflagellates, diatoms, brown algae and euglenids) are the result of SECONDARY ENDOSYMBIOSIS, tertiary endosymbiosis and, perhaps, even quaternary endosymbiosis in which a non-photosynthetic eukaryotic ancestor engulfed a photosynthetic eukaryote [11–13].

Genome size

Even with the relatively limited data available from the genomes of microbial eukaryotes, it is evident that genome size varied tremendously during the evolution of the diverse lineages (Table 1). Extensive reviews of genome size can be found elsewhere (e.g. [14,15]) and only a few points key to microbial eukaryotes are described below.

Reduced eukaryotic genomes

The smallest complete nuclear genomes are found in microsporidians, intracellular parasites that infect many animal phyla [16–18]. Microsporidian genomes range in size from 2.3 Mb in *Encephalitozoon intestinalis* to 19.5 Mb in *Glugea atherinae* [19]. The completely sequenced genome of the microsporidian *Encephalitozoon cuniculi* is only 2.9 Mb, with an estimated 1997 protein-coding genes, less than half the number of genes found in *Escherichia coli* (4377). The reduction of the *E. cuniculi* genome involves not only fewer genes and smaller intergenic regions, but also smaller coding regions, including reduced rDNAs [16]. In a comparison of 350 protein-coding genes, >85% of *E. cuniculi* genes are shorter than their homologs in *Saccharomyces cerevisiae*, with a mean reduction of 14.6% [17,20]. These data indicate that the processes that generated reduced genomes in these parasites have impacted the lengths of both coding and noncoding sequences.

The smallest known nuclear genomes are found in the NUCLEOMORPHS of cryptomonads and chlorarachniophytes, and are approximately one-eighth of the size of the *E. coli* genome (Table 1). Nucleomorphs reside in cells that contain a distinct 'host' nucleus with a larger, more complete genome. Sequence analysis of the nucleomorph genomes of the cryptomonad (flagellate) *Guillardia theta* and the chlorarachniophyte (amoeboflagellate) *Bigelowiella natans* reveal that the nucleomorphs in each were obtained independently from a red and a green alga, respectively [21–23]. Both the cryptomonad and chlorarachniophyte nucleomorph genomes comprise three tightly packed chromosomes, ranging in size from 170 to 270 kb each [21,22,24]. Genes in both genomes are separated by only 65–75 nucleotides on average, and genes overlap in some cases [21]. SPLICEOSOMAL INTRONS are only 18–20 bp long in the chlorarachniophyte nucleomorphs, whereas cryptomonad nucleomorph introns are less frequent but generally larger (42–52 bp in *G. theta*) [21]. The similar pattern of genome reduction in the independently derived chlorarachniophyte and cryptomonad nucleomorph genomes indicates a striking convergence [21,24,25].

Expanded eukaryotic genomes

In contrast to the minimal genomes of nucleomorphs and microsporidians, ciliates have many protein-coding genes. For example, the genomes of *Paramecium tetraurelia* and *Oxytricha trifallax*, two relatively distantly related species, are estimated consistently to contain 25 000–40 000 protein-coding genes each [26–28], similar to current estimates for the number of genes in the human genome. The largest known eukaryotic genome is also found in a microbial eukaryote, the enigmatic *Amoebae dubia*, the genome of which is reportedly ~670 000 Mb in size [15,29]. (The genome size of *A. dubia* has yet to be estimated using current molecular techniques, so this estimate should be viewed with caution.) Given the bias in

Box 1. The big picture

- Understanding genome evolution requires synthesis of data from a broad range of organisms. Recent data about the genomes of microbial eukaryotes challenge the conventional views about genome evolution that have emerged from studies of plants, animals and fungi.
- Extreme examples of genome sizes, both large and small, can be found among microbial eukaryotes. Genome reduction has converged on similar structures in several lineages.
- The diversity of genome structures in eukaryotes includes some unusual features, such as subtelomeric rDNA, that have multiple origins, and others, such as highly fragmented mitochondrial genomes, that appear to have arisen only once. Further elaboration of both these features and eukaryotic relationships is required to understand the origin and maintenance of genomic structures.

Box 2. Examples of eukaryotic lineages

There are an estimated 100–200 major eukaryotic lineages, of which plants, animals and fungi represent just three [1]. The remaining lineages are predominantly microbial, a few of which are described below (Figure 1 in main text).

Alveolates

Alveolates are united by the presence of 'alveolar sacs', which vary in function among lineages. Alveolates include three major lineages: ciliates, dinoflagellates and apicomplexans. Ciliates (e.g. *Tetrahymena* and *Paramecium*) are defined by the presence of dual genomes. Dinoflagellates are responsible for much of the photosynthesis in coral reefs and are the causative agents of many red tides. Apicomplexans are all parasitic, and include *Plasmodium falciparum*, the causative agent of malaria.

Euglenozoa

The Eulgenozoa contain two lineages: the predominantly photosynthetic euglenids and the kinetoplastids. Euglenids (e.g. *Euglena* and *Phacus*) are covered by a complex protein coat, or pellicle, that enables them to alter their shape. Kinetoplastids include *Leishmania* and *Trypanosoma*, genera that include the organisms that cause leishmaniasis or African sleeping sickness, respectively [69,70].

Stramenopiles

The stramenopiles or heterokonts – which include diatoms, water molds, brown algae and labyrinthulids – show few shared morphological characters. When present, flagella in stramenopiles are marked by hair-like projections. The silica shells of diatoms are known for their beauty and for their presence in diatomaceous earth used in gardening. The water mold *Phytophthora infestans* is famous for its devastating role in the Irish potato famines. Brown algae, including kelp, dominate the macroalgae of marine systems and can grow to >50m in length. Labyrinthulids, or slime nets, travel via a network of ectopic fibers and are predators on marine plants, including eelgrass.

Opisthokonts

Opisthokonts, organisms with a single posterior flagellum (when one is present), include the well-known animal and fungal lineages as well as several predominantly microbial groups. The choanoflagellates, or collared flagellates, are the probable sister lineage to the animals, and the chytrids are basal to the 'true' fungi. Also included in the opisthokonts are two groups of parasites: (i) the microsporidians, which include lineages that cause diarrhea in humans; and (ii) the ichthyosporeans (or DRIPs), which include several fish parasites.

Glaucocystophytes, red algae and green algae

The ancestor of this clade of eukaryotes is believed to have acquired photosynthesis through the primary acquisition of a plastid from an engulfed cyanobacterium. The clade contains two lineages with numerous macroscopic members (the red algae and green algae) plus the glaucocystophytes, which have retained peptidoglycan in their plastid membranes. Peptidoglycan is otherwise restricted to the cell walls of bacteria.

Cryptomonads and chlorarachniophytes

These two lineages of photosynthetic eukaryotes provide the two clearest examples of SECONDARY ENDOSYMBIOSIS of chloroplasts because they contain both a 'host' nucleus and a remnant nucleus (nucleomorph) from an engulfed eukaryote. Cryptomonads appear to be derived from a flagellate that engulfed a red algal cell, whereas chlorarachniophytes are descended from an amoeboflagellate that engulfed a green algal cell [10,24].

collecting data from organisms with relatively small genomes (e.g. many parasites), the general pattern of expansion and contraction of genome size in eukaryotes remains unknown.

Genome structures

In addition to the diversity in genome sizes, microbial eukaryotes also have a variety of novel genome structures. Several of these structures, including extrachromosomal or SUBTELOMERIC rDNA genes, hypervariable KARYOTYPES, and GENOME DUALITY, appear to have arisen multiple times, whereas others, such as ABSOLUTE STRAND POLARITY and highly-fragmented mitochondrial genomes, appear to be restricted to single lineages.

Unusual arrangements of rDNAs

Extrachromosomal rDNAs are amplified from chromosomal rDNA copies during the life cycles of several lineages of microbial eukaryotes, as well as in both animals and fungi [30–33]. In animals, which are one of the opisthokont lineages (Figure 1), extrachromosomal rDNAs are found in the oocytes of both insects and vertebrates [30,34], where they presumably meet the translation requirements of early development. Similarly, extrachromosomal rDNAs are also found in fungi [32] and both cellular [31] and acellular [35] slime molds. To determine whether the presence of extrachromosomal rDNAs is homologous among these groups requires further characterization of the phylogenetic distribution of this genome feature coupled with more detailed analyses of the mechanisms underlying the amplification of these genes.

A potentially related genome structure occurs in the microbial eukaryotes *Entamoeba histolytica*, the causative agent of amoebiasis (or dysentery), and its close relative *Entamoeba dispar*. The rDNA genes of these two species are located on as many as 200 copies of a circular plasmid-like molecule, in contrast to the tandem arrays found in many eukaryotic genomes. However, unlike the rDNA amplification described above, no chromosomal copy of the rDNA genes has been found in *E. histolytica* [36]. This suggests that the 'plasmids' are not amplified products of a chromosomal locus, but instead are 'EPISOMAL ELEMENTS' that are maintained during the life cycle of *Entamoeba* [36–38]. Moreover, the extrachromosomal circles replicate independently and at least some of the molecules contain multiple replication origins [38]. This is unique, because no other 'plasmid' element in either prokaryotes or eukaryotes is known to initiate replication from widely dispersed locations.

There have been at least four independent origins of subtelomeric rDNAs in eukaryotes: in both cryptomonad and chlorarachniophyte nucleomorphs, in *Giardia lamblia*, and in some microsporidians. Each of the three chromosomes in both cryptomonad and chlorarachniophyte nucleomorphs end with inverted repeats comprising a single rDNA unit linked to a telomere [24,39]. The *G. lamblia* genome contains ~60 copies of a 5.6-Kb rDNA unit that is organized in tandem arrays near the telomeres of at least six chromosomes [40,41]. In the microsporidian *E. cuniculi*, each chromosome contains two subtelomeric rDNA units located ~15 Kb upstream of the chromosome

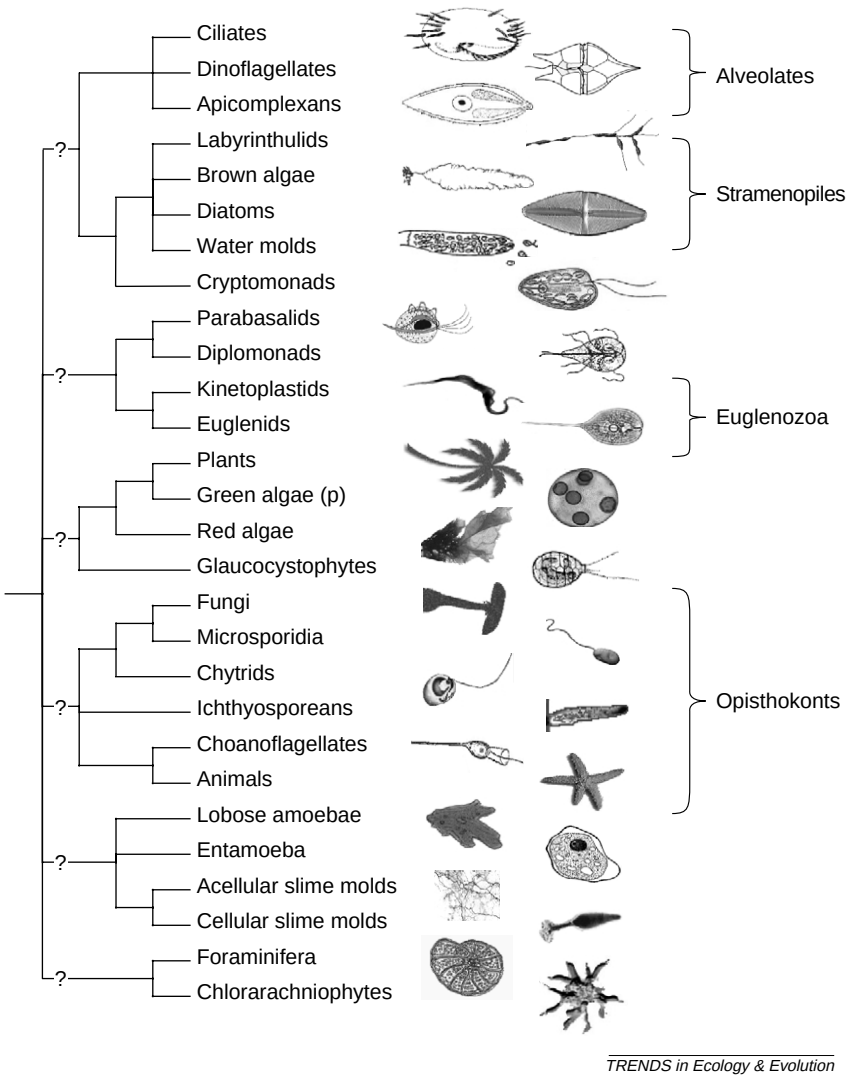

Figure 1. Hypothesis of eukaryotic relationships based on molecular data. Many branches, particularly at deep nodes, will probably change over the next decade as both more data and additional taxa are sampled. Green algae defined excluding plants are paraphyletic (p). Topology derived from [6,66,67].

ends [16,42,43]. Although the location and structure of rDNA units vary within microsporidians, rDNAs are only found within subtelomeric regions in both *E. cuniculi* and the diplomonad *G. lamblia*. It is unclear why subtelomeric rDNAs have evolved multiple times in all these diverse lineages.

Hypervariable karyotypes

Several microbial eukaryotes exhibit heterogeneity in karyotype within a single species. Intriguingly, many of these eukaryotes are parasitic, and it is possible that the karyotypic variability is related to evasion of host immune systems.

Diplomonads display extensive karyotype variability within species. In fact, it has been impossible so far to identify a basic karyotype for the diplomonad *G. lamblia* owing to the substantial heterogeneity among isolates [44–47]. Homologous chromosomes in different isolates vary in size by hundreds of Kb pairs because of whole and partial chromosome duplications, subtelomeric loss followed by duplication, and extensive internal duplications among chromosomes of different strains [44,45]. In spite of this heterogeneity, there is a core region in each *Giardia* chromosome that remains stable, whereas the subtelomeric regions, which contain the rDNA units, are hypervariable [46]. Genomic plasticity in diplomonads might enable species to evolve new drug-resistant phenotypes through relatively fast mutations of genes [47].

Some microsporidians also have variable karyotypes. Intraspecific variation has been described in four microsporidians: *G. atherinae*, *Vavraia oncoperae*, *E. cuniculi* and *Encephalitozoon hellem* [48]. The genome of *E. cuniculi*, for example, is highly plastic, as evidenced by chromosomal size polymorphisms among strains [42,49]. The variation in karyotypes is due predominantly to DNA rearrangements near the chromosome ends. Similarly, karyotype variation reported in yeast genomes is a result of subtelomeric rearrangements [50].

Dual genomes

Ciliates and some foraminiferans are unique among microbial eukaryotes in that they exhibit genome duality,

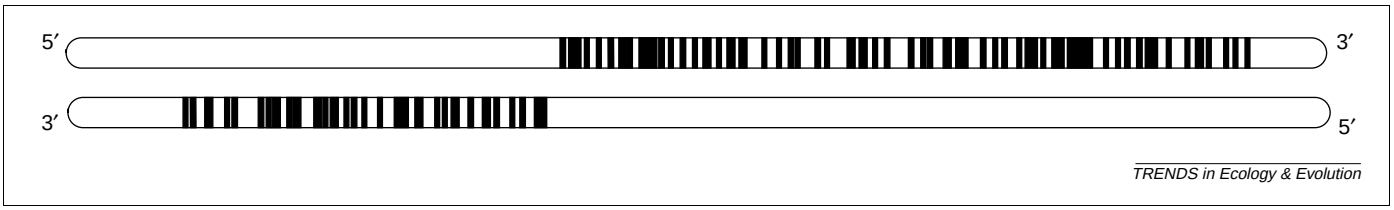

Figure 2. Absolute strand polarity in kinetoplastids: genes are arrayed in clusters on one strand or the other. The black bars represent the 79 genes characterized from chromosome 1 of *Leishmania major*, 29 of which are on one strand and 50 of which are on the other [55,68].

with two distinct types of nuclei within each cell [51–54]. In ciliates, where this system has been studied relatively well, the micronuclear germline genome is involved in CONJUGATION, whereas the somatic macronuclear genome is the site of the majority of transcription [52–54]. Macronuclear genomes are highly processed, such that zygotic chromosomes are fragmented, some segments are eliminated and the remaining chromosomes are amplified. In species of *Tetrahymena*, fragmentation of as few as five micronuclear chromosomes produces up to 200 unique molecules in the macronucleus, each of which is amplified ~60 times [26,52].

Members of the ciliate classes Spirotrichea and Phyllopharyngea, as well as the sister orders Armophorida and Clevelandellida, process their zygotic nuclei extensively to generate gene-sized macronuclear 'chromosomes' [54]. In some of the relatively well-studied spirotrichs, 95% or more of the micronuclear sequence is eliminated during the development of the macronucleus and the ~120 micronuclear chromosomes fragment into as many as 24 000 different gene-sized chromosomes in the macronucleus [53,54]. Furthermore, each of these highly processed macronuclear chromosomes is then amplified 950–15 000 times [53,54]. The ciliate groups with extensive chromosomal fragmentation (generating gene-sized macronuclear chromosomes) are polyphyletic, suggesting that the mechanisms underlying chromosomal fragmentation in ciliates are highly plastic or have evolved multiple times [54].

Strand polarity and polycistronic gene clusters

Kinetoplastid chromosomes exhibit a unique feature termed absolute strand polarity. The genes of kinetoplastids are arranged into large clusters arrayed on only one strand of the chromosomes, with no intervening genes on the other strand [55–58] (Figure 2). The genes on chromosome 1 of *Leishmania major*, for example, are organized into two large clusters, with the first 29 genes on one DNA strand and the other 50 genes on the second strand (Figure 2) [55]. Moreover, trypanosomes and *Leishmania* have POLYCISTRONIC TRANSCRIPTION, a feature that is generally found only in prokaryotes [56,57,59]. However, in contrast to prokaryotic clusters, genes in kinetoplastids do not cluster into prokaryote-like operons of genes with similar function [55–57,59]. Not surprisingly, given the linkage of the transcription of so many genes, regulation of expression in kinetoplastids is primarily post-transcriptional.

Unusual structures of organellar genomes

Mini- and maxicircles in the mitochondrial genomes of kinetoplastids

Kinetoplastids are defined by another unique genome feature that is found in their unusually structured mitochondria: kinetoplastid mitochondrial genomes exist as concatenated mini- and maxicircles [60,61]. Some of the maxicircles contain incomplete genes that require RNA editing to produce open reading frames, and at least part of the RNA editing is templated by sequences on minicircles [62]. The molecular mechanisms underlying the replication of these complex organellar genomes are still under investigation [60,61]. The phenomenon of RNA editing occurs among numerous eukaryotic and prokaryotic lineages, although the mechanisms vary among lineages and only kinetoplastids use minicircles as guide RNAs. The discovery of RNA editing in kinetoplastids exemplifies the importance of studies of the genomes of microbial eukaryotes.

Fragmented linear chromosomes in mitochondria of Amoebidium

Another unusual organellar genome arrangement in microbial eukaryotes is the highly fragmented mitochondrial genome of the ichthyosporean *Amoebidium*, an opisthokont [63]. In contrast to the single (or small number of) circular or linear molecule(s) typical of most mitochondrial genomes, the *Amoebidium* mitochondrial genomes contains several hundred distinct 0.3–8.3-Kb linear chromosomes [63]. These chromosomes fall into three categories: (i) small molecules with no identified coding regions, (ii) medium-sized molecules that encode a single gene, and (iii) larger molecules containing multiple genes. All chromosomes studied so far also contain terminal repeat structures [63]. The phylogenetic distribution of this organellar genome feature is not known, largely because of the lack of data from other ichthyosporeans.

Unigenic minicircles in dinoflagellate chloroplasts

UNIGENIC minicircles are a unique genome structure that has been reported in the chloroplasts of peridinean dinoflagellates. The chloroplast genes of these dinoflagellates occur on 2–3-Kb minicircles, which contain generally only one gene plus an origin of replication and a promoter [64]. This is in striking contrast to the 120–200-Kb genomes found in most chloroplasts. The origin of minicircles occurred probably only once in the ancestor of all extant photosynthetic dinoflagellates [64]. Zhang *et al*. [64] provide two possible models for the origin of chloroplast minicircles: (i) duplicative transposition of

replicon origin sequences throughout the chloroplast genome followed by deletions, and (ii) differential deletion within a multicopy population of chloroplast chromosomes.

Summary

Our understanding of the tremendous diversity in genome size and structure in microbial eukaryotes stems from studies of relatively few eukaryotic lineages. Interpreting these studies in light of the current reconstruction of eukaryotic phylogeny indicates that some features have multiple origins, whereas others probably arose only once. Intriguingly, features that arose multiple times in eukaryotes, such as subtelomeric rDNAs and genome duality, are probably subject to some form of positive darwinian selection and require further research from a comparative viewpoint. Furthermore, the tremendous diversity among microbial eukaryotes demonstrates that general principles of eukaryotic genome evolution based on studies of plants, animals and fungi should be interpreted with extreme caution. As data emerge from additional microbial lineages, the number of examples of unusual genome structures will no doubt increase. Future studies must combine examination of genome structures from diverse lineages with further elaboration of the eukaryotic tree of life. Combining such approaches is essential for the elucidation of the tempo and mode of genome evolution in eukaryotes.

Acknowledgements

This work is supported by two NSF grants (DEB-0092908 and DEB-0079325) to L.A.K.

References

1 Patterson, D.J. (1999) The diversity of eukaryotes. *Am. Nat.* 154, S96–S124
2 Katz, L.A. (2002) Lateral gene transfers and the evolution of eukaryotes: theories and data. *Int. J. Syst. Evol. Microbiol.* 52, 1893–1900
3 Katz, L.A. (1999) The tangled web: gene genealogies and the origin of eukaryotes. *Am. Nat.* 154, S137–S145
4 Roger, A.J. (1999) Reconstructing early events in eukaryotic evolution. *Am. Nat.* 154, S146–S163
5 Bapteste, E. *et al.* (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1414–1419
6 Baldauf, S.L. *et al.* (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290, 972–977
7 Taylor, F.J.R. (1999) Ultrastructure as a control for protistan molecular phylogeny. *Am. Nat.* 154, S125–S136
8 Sogin, M.L. *et al.* (1996) Ancestral relationships of the major eukaryotic lineages. *Microbiologia* 12, 17–28
9 Cavalier-Smith, T. (1993) Kingdom Protozoa and its 18 phyla. *Microbiol. Rev.* 57, 953–994
10 Palmer, J.D. (2003) The symbiotic birth and spread of plastids: how many times and whodunit? *J. Phycol.* 39, 4–11
11 Archibald, J.M. and Keeling, P.J. (2002) Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends Genet.* 18, 577–584
12 McFadden, G.I. (2001) Chloroplast origin and integration. *Plant Physiol.* 125, 50–53
13 McFadden, G.I. (2001) Primary and secondary endosymbiosis and the origin of plastids. *J. Phycol.* 37, 951–959
14 Petrov, D.A. (2001) Evolution of genome size: new approaches to an old problem. *Trends Genet.* 17, 23–28
15 Cavalier-Smith, T. (1985) Introduction: the evolutionary significance of genome size. In *The Evolution of Genome Size* (Cavalier-Smith, T., ed.), pp. 1–36, John Wiley & Sons
16 Vivares, C.P. and Metenier, G. (2001) The microsporidian *Encephalitozoon*. *BioEssays* 23, 194–202
17 Vivares, C.P. *et al.* (2002) Functional and evolutionary analysis of a eukaryotic parasitic genome. *Curr. Opin. Microbiol.* 5, 499–505
18 Vivares, C.P. and Metenier, G. (2000) Towards the minimal eukaryotic parasitic genome. *Curr. Opin. Microbiol.* 3, 463–467
19 Biderre, C. *et al.* (1995) Evidence for the smallest nuclear genome (2.9 Mb) in the microsporidium *Encephalitozoon cuniculi*. *Mol. Biochem. Parasitol.* 74, 229–231
20 Katinka, M.D. *et al.* (2001) Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414, 450–453
21 Gilson, P.R. and McFadden, G.I. (2002) Jam-packed genomes – a preliminary, comparative analysis of nucleomorphs. *Genetica* 115, 13–28
22 Douglas, S. *et al.* (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* 410, 1091–1096
23 Cavalier-Smith, T. (2002) Nucleomorphs: enslaved algal nuclei. *Curr. Opin. Microbiol.* 5, 612–619
24 Maier, U.G. *et al.* (2000) The nucleomorph genomes of cryptophytes and chlorarachniophytes. *Protist* 151, 103–109
25 Gilson, P.R. and McFadden, G.I. (1997) Good things come in small packages: the tiny genomes of chlorarachniophyte endosymbionts. *BioEssays* 19, 167–173
26 Prescott, D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.* 58, 233–267
27 Sperling, L. *et al.* (2002) Random sequencing of paramecium somatic DNA. *Eukaryot. Cell* 1, 341–352
28 Prescott, D.M. *et al.* (2002) Coding properties of macronuclear DNA molecules in *Sterkiella nova* (*Oxytricha nova*). *Protist* 153, 71–77
29 Friz, C.T. (1968) The biochemical composition of the free-living Amoebae *Chaos chaos*, *Amoeba dubia* and *Amoeba proteus*. *Comp. Biochem. Physiol.* 26, 81–90
30 Kubrakiewicz, J. (2002) Extrachromosomal rDNA amplification in the oocytes of *Polystoechotes punctatus* (Fabricius) (Insecta-Neuroptera-Polystoechotidae). *Arthropod. Struct. Dev.* 31, 23–31
31 Glockner, G. *et al.* (2002) Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* 418, 79–85
32 Huber, D.H. and Rustchenko, E. (2001) Large circular and linear rDNA plasmids in *Candida albicans*. *Yeast* 18, 261–272
33 Pauli, U. *et al.* (1988) Unusual nucleoprotein structures in the vicinity of the replication origins of the extrachromosomal rDNA of the slime mold *Physarum polycephalum*. *Biochem. Cell Biol.* 66, 935–941
34 Mais, C. *et al.* (2002) On the formation of amplified nucleoli during early *Xenopus* oogenesis. *J. Struct. Biol.* 140, 214–226
35 Ferris, P.J. *et al.* (1983) Inheritance of extrachromosomal rDNA in *Physarum polycephalum*. *Mol. Cell. Biol.* 3, 635–642
36 Bagchi, A. *et al.* (1999) Lack of a chromosomal copy of the circular rDNA plasmid of *Entamoeba histolytica*. *Int. J. Parasitol.* 29, 1775–1783
37 Paul, J. *et al.* (2002) Close sequence identity between ribosomal DNA episomes of the nonpathogenic *Entamoeba dispar* and pathogenic *Entamoeba histolytica*. *J. Biosci.* 27, 619–627
38 Bhattacharya, S. *et al.* (1998) The ribosomal DNA plasmids of *Entamoeba*. *Parasitol. Today* 14, 181–185
39 Gilson, P.R. (2001) Nucleomorph genomes: much ado about practically nothing. *Genome Biol.* 2, 10221–10225
40 Leblancq, S.M. *et al.* (1991) Frequent rearrangements of ribosomal rRNA-encoding chromosomes in *Giardia lamblia*. *Nucleic Acids Res.* 19, 4405–4412
41 Hou, G.Y. *et al.* (1995) Structure of a frequently rearranged ribosomal RNA encoding chromosome in *Giardia lamblia*. *Nucleic Acids Res.* 23, 3310–3317
42 Brugere, J.F. *et al.* (2000) *Encephalitozoon cuniculi* (microspora) genome: physical map and evidence for telomere-associated rDNA units on all chromosomes. *Nucleic Acids Res.* 28, 2026–2033
43 Brugere, J.F. *et al.* (2000) Occurrence of subtelomeric rearrangements in the genome of the microsporidian parasite *Encephalitozoon cuniculi*, as revealed by a new fingerprinting procedure based on two-dimensional pulsed field gel electrophoresis. *Electrophoresis* 21, 2576–2581

44 Upcroft, P. and Upcroft, J.A. (1999) Organization and structure of the *Giardia* genome. *Protist* 150, 17–23
45 Le Blancq, S.M. *et al.* (1992) Spontaneous chromosome rearrangements in the protozoan *Giardia lamblia* – estimation of mutation rates. *Nucleic Acids Res.* 20, 4539–4545
46 Le Blancq, S.M. and Adam, R.D. (1998) Structural basis of karyotype heterogeneity in *Giardia lamblia*. *Mol. Biochem. Parasitol.* 97, 199–208
47 Le Blancq, S.M. (1994) Chromosome rearrangements in *Giardia lamblia*. *Parasitol. Today* 10, 177–179
48 Delarbre, S. *et al.* (2001) Genetic diversity in the microsporidian *Encephalitozoon hellem* demonstrated by pulsed-field gel electrophoresis. *J. Eukaryot. Microbiol.* 48, 471–474
49 Biderre, C. *et al.* (1999) Molecular karyotype diversity in the microsporidian *Encephalitozoon cuniculi*. *Parasitology* 118, 439–445
50 Kellis, M. *et al.* (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423, 241–254
51 Lee, J.J. and Anderson, R.A. (1991) *Biology of Foraminifera*, Academic Press
52 Yao, M.C. *et al.* (2002) Genome-wide rearrangements of DNA in ciliates. In *Mobile DNA II* (Craig, N.L. *et al.*, eds), pp. 730–758, ASM Press
53 Jahn, C.L. and Klobutcher, L.A. (2002) Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.* 56, 489–520
54 Katz, L.A. (2001) Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. *Int. J. Syst. Evol. Microbiol.* 51, 1587–1592
55 Myler, P.J. *et al.* (2001) The *Leishmania* genome project: new insights into gene organization and function. *Med. Microbiol. Immunol. (Berl.)* 190, 9–12
56 Myler, P.J. *et al.* (2000) Genomic organization and gene function in *Leishmania*. *Biochem. Soc. Trans.* 28, 527–531
57 Myler, P.J. and Stuart, K.D. (2000) Recent developments from the *Leishmania* genome project. *Curr. Opin. Microbiol.* 3, 412–416
58 Worthey, E.A. *et al.* (2003) *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res.* 31, 4201–4210
59 Vanhamme, L. and Pays, E. (1995) Control of gene expression in trypanosomes. *Microbiol. Rev.* 59, 223–240
60 Klingbeil, M.M. *et al.* (2001) Unlocking the secrets of trypanosome kinetoplast DNA network replication. *Protist* 152, 255–262
61 Lukes, J. *et al.* (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell* 1, 495–502
62 Simpson, L. *et al.* (2000) Evolution of RNA editing in trypanosome mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6986–6993
63 Burger, G. *et al.* (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl. Acad. Sci. U. S. A.* 100, 892–897
64 Zhang, Z.D. *et al.* (2002) Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins. *Mol. Biol. Evol.* 19, 489–500
65 Copley, R.R. *et al.* (2002) Genome and protein evolution in eukaryotes. *Curr. Opin. Chem. Biol.* 6, 39–45
66 Stechmann, A. and Cavalier-Smith, T. (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297, 89–91
67 Katz, L.A. (2002) Protists. In *Encyclopedia of Evolution* (Pagel, M., ed.), pp. 956–961, Oxford University Press
68 Myler, P.J. *et al.* (1999) *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2902–2906
69 Cavalier-Smith, T. (1981) Eukaryote kingdoms – seven or nine. *BioSystems* 14, 461–481
70 Simpson, A.G.B. (1997) The identity and composition of the Euglenozoa. *Arch. Protistenkd.* 148, 318–328