

# Cascades of convergent evolution: The corresponding evolutionary histories of euglenozoans and dinoflagellates

Julius Lukeš<sup>a,1</sup>, Brian S. Leander<sup>b</sup>, and Patrick J. Keeling<sup>b</sup>

<sup>a</sup>Biology Centre, Institute of Parasitology, Czech Academy of Sciences, and Faculty of Sciences, University of South Bohemia, 37005 České Budějovice, Czech Republic; and <sup>b</sup>Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Departments of Botany and Zoology, University of British Columbia, Vancouver, Canada V6T 1Z4

The majority of eukaryotic diversity is hidden in protists, yet our current knowledge of processes and structures in the eukaryotic cell is almost exclusively derived from multicellular organisms. The increasing sensitivity of molecular methods and growing interest in microeukaryotes has only recently demonstrated that many features so far considered to be universal for eukaryotes actually exist in strikingly different versions. In other words, during their long evolutionary histories, protists have solved general biological problems in many more ways than previously appreciated. Interestingly, some groups have broken more rules than others, and the Euglenozoa and the Alveolata stand out in this respect. A review of the numerous odd features in these 2 groups allows us to draw attention to the high level of convergent evolution in protists, which perhaps reflects the limits that certain features can be altered. Moreover, the appearance of one deviation in an ancestor can constrain the set of possible downstream deviations in its descendants, so features that might be independent functionally, can still be evolutionarily linked. What functional advantage may be conferred by the excessive complexity of euglenozoan and alveolate gene expression, organellar genome structure, and RNA editing and processing has been thoroughly debated, but we suggest these are more likely the products of constructive neutral evolution, and as such do not necessarily confer any selective advantage at all.

mitochondria | plastids | comparative genomics |  
molecular evolution | *Trypanosoma*

The vast majority of eukaryotes on the planet, in terms of both abundance and diversity, are microbial. Generalities about fundamental biological processes are based on knowledge of a few model organisms, yet many microeukaryotes have deviated well beyond these generalities over the course of evolutionary history, which is a reflection of the deep phylogenetic distances between eukaryotic lineages that are neither plants nor animals nor fungi. It is also clear that some groups of protists have broken more rules than others, and 2 diverse lineages that particularly stand out in this regard are the Euglenozoa and the Alveolata (Fig. 1). In members of both of these groups, fundamental structures and processes have substantially deviated from those of other eukaryotes; however, perhaps even more interestingly, both groups have frequently departed in the same general fashion, resulting in surprising levels of convergence that suggest limits to the ways these features can be altered.

The Euglenozoa is a monophyletic group within the Excavata consisting of single-celled flagellates composed of 2 major subgroups (kinetoplastids and euglenids), and 1 smaller subgroup (diploemids) (Fig. 1). Members of Euglenozoa have diverse modes of nutrition, including predation, parasitism, and photoautotrophy. Predatory euglenozoans are phylogenetically widespread within the group and tend to have diverse feeding apparatuses, feeding strategies and prey preferences (1). For instance, some predatory species are limited to small prey such as bacteria, whereas other species frequently consume larger prey, such as other eukaryotic cells. Photoautotrophy is re-

stricted to a specific subclade of euglenids and originated via secondary endosymbiosis between a predatory euglenid and a green algal prey (1). Parasitic and commensalic euglenozoans appear to have evolved independently several times within kinetoplastids (2), and some species (e.g., *Trypanosoma* and *Leishmania*) cause important human illnesses such as African sleeping sickness, Chagas's disease, and leishmaniasis.

The Alveolata, another monophyletic group of primarily single-celled eukaryotes that have adopted similarly diverse modes of life, is composed of 3 major subgroups: ciliates, apicomplexans, and dinoflagellates (Fig. 1). All 3 subgroups contain predatory and parasitic species, and only dinoflagellates and an unusual lineage called *Chromera* are known to contain fully integrated and photosynthetic plastids (3). Photosynthetic dinoflagellates play important roles as planktonic primary producers in oceanic ecosystems, and some of the lineages form symbiotic relationships with corals (e.g., *Symbiodinium*) and are critical for maintaining the health of reef systems around the world. Nonphotosynthetic plastids have independently evolved in some dinoflagellates and in apicomplexans, which are all obligate parasites of animals and a few are exceedingly important disease organisms of vertebrates (e.g., *Cryptosporidium*, *Toxoplasma*, and *Plasmodium*). Although plastids have not been definitively demonstrated in ciliates, several independent lineages in this group harbor photosynthetic symbionts that are intermittently replenished by feeding.

Both euglenozoans and alveolates have a reputation for "doing things their own way," which is to say that they have developed seemingly unique ways to build important cellular structures or carry out molecular tasks critical for their survival. Why such hotspots for the evolution of novel solutions to problems should exist in the tree of life is not entirely clear. However, the deeper we look into these groups, the more often it is found that they are also evolving strikingly similar mechanisms for achieving these essential biological functions. Significantly, however, there is a great weight of phylogenetic data that show these lineages are not closely related: of the 5 eukaryotic supergroups hypothesized to explain all eukaryotic diversity, alveolates and euglenozoans fall into 2 different supergroups, chromalveolates and excavates, respectively (Fig. 1). The support for these supergroups as a whole remains contentious (4–7), but

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "In the Light of Evolution III: Two Centuries of Darwin," held January 16–17, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS web site at [www.nasonline.org/Sackler.Darwin](http://www.nasonline.org/Sackler.Darwin).

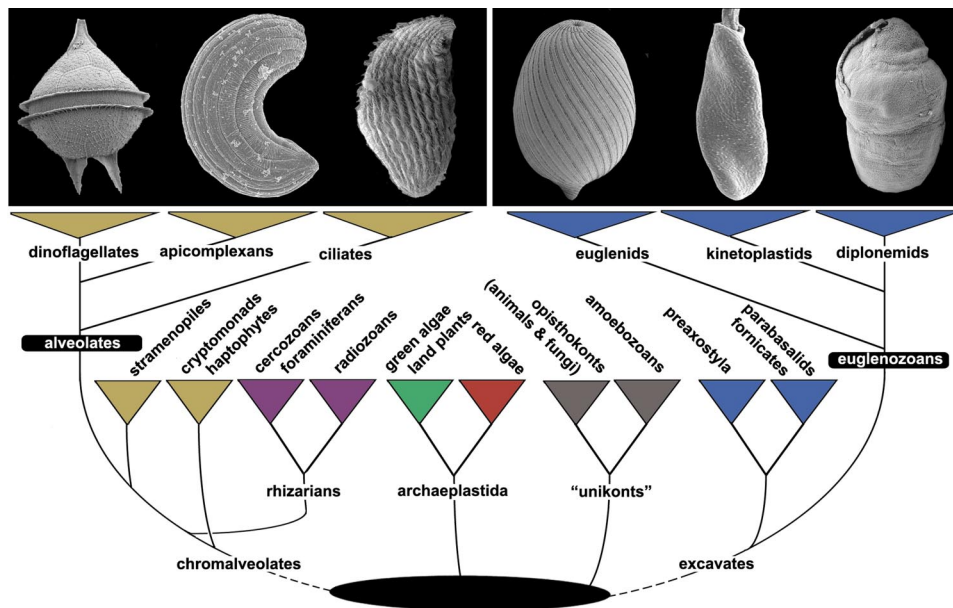
Author contributions: J.L., B.S.L., and P.J.K. designed research; J.L., B.S.L., and P.J.K. performed research; J.L., B.S.L., and P.J.K. analyzed data; and J.L., B.S.L., and P.J.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: [jula@paru.cas.cz](mailto:jula@paru.cas.cz).

This article contains supporting information online at [www.pnas.org/cgi/content/full/0901004106/DCSupplemental](http://www.pnas.org/cgi/content/full/0901004106/DCSupplemental).



**Fig. 1.** The tree of eukaryotes showing the relative positions of Alveolata and Euglenozoa. This tree is a reflection of many kinds of evidence that are summarized in refs. 4, 5, and 7. The Alveolata consists of 3 main groups, dinoflagellates, apicomplexans, and ciliates (illustrated in *Upper Left*; scanning electron micrographs of *Protoperidinium*, *Selenidium* and an unidentified ciliate), whereas the Euglenozoa consists of euglenids, kinetoplastids, and diplomonids (illustrated in *Upper Right*; scanning electron micrographs of *Lepocinclis*, *Leptomonas*, and *Diplonema*). Alveolates are members of the supergroup Chromalveolata, whereas euglenozoans are members of a different supergroup, the Excavata. Neither supergroup is without controversy (e.g., the rhizarian supergroup is shown within chromalveolates to reflect recent phylogenetic analyses), but the relationships of alveolates and euglenozoans to some other eukaryotic groups to the exclusion of one another, and the relationships within alveolates and euglenozoans, are both consistently well-supported by molecular analyses.

there is strong support from phylogenomics and many individual phylogenies and rare genomic characters for a specific relationship between alveolates and stramenopiles on one hand, and euglenozoans and heteroloboseans on the other hand (7). Moreover, no analysis of eukaryotic phylogeny has ever suggested they are closely related to one another. Still more significantly, the majority of the characteristics we discuss below are not universal to all members of either alveolates or euglenozoans, but rather appear to have evolved within a subgroup of each lineage. Altogether, the distribution of these characteristics can really only adequately be explained by convergent evolution. Below, we will examine some of these examples of convergence and what the cooccurrence of convergent traits may tell us about how they evolved.

### Convergent Evolution

Recognizing the independent origins of similar traits in distantly related lineages—convergent evolution—allows us to better understand how different environmental and intrinsic conditions have shaped the characteristics of organisms over time; each specific example of convergence reflects a fundamental biological problem and its possible solutions. The causes of convergent evolution are varied and can involve camouflage, mimicry, biomechanical optimization, molecular constraints, developmental canalization and character-state reversals. Examples of convergent evolution range from the biochemical level to the behavioral level and are best characterized within animals and land plants (8–11), which collectively represent only a small portion of the full tree of eukaryotes (4, 6). The occurrence and adaptive significance of convergent evolution in microbial eukaryotes, by contrast, is poorly understood, but it is clear from several examples that convergent traits can evolve over vast phylogenetic distances (6). Convergence in very distantly-related lineages is particularly compelling because the influence of homologous developmental programs (i.e., intrinsic conditions) in constraining subsequent evolution should be minimal if not absent altogether (6). Therefore, improved understanding of

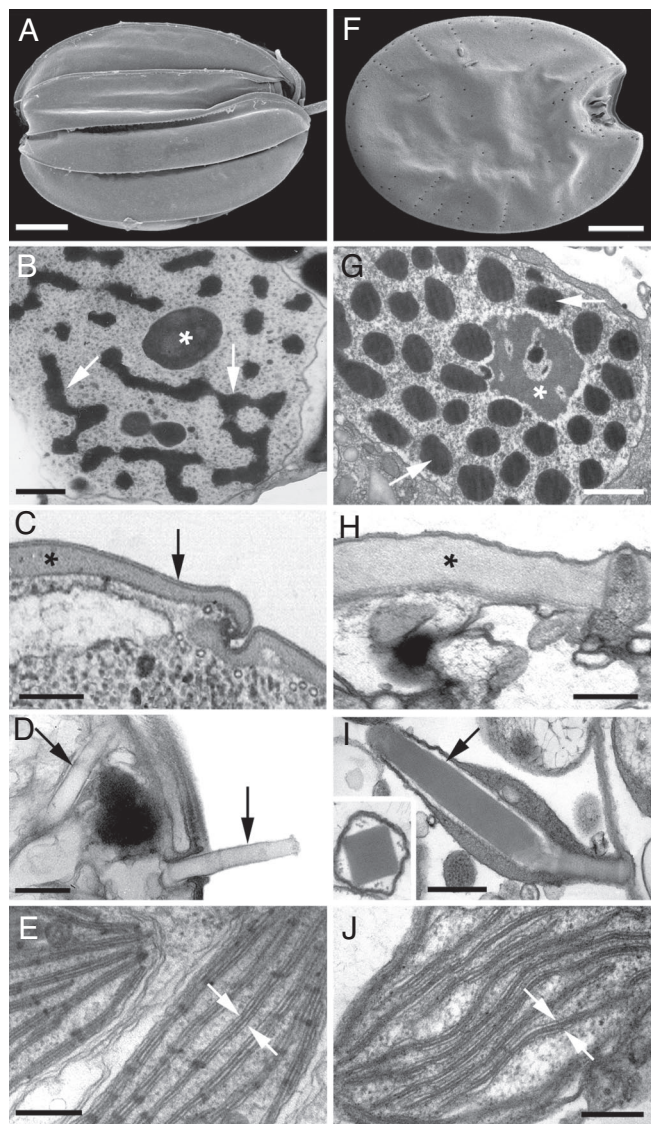
convergent evolution in distantly related microbes will provide a much broader framework for evaluating the forces of natural selection and the potential role of constructive neutrality during the evolution of ultrastructural systems and complex molecular processes.

Eukaryotic cells are built from a few core systems that have become tremendously diverse over the course of evolutionary history. Some systems are remarkably conserved, in particular fundamental molecular processes such as information flow or core metabolism, but even in these systems substantial modifications accumulated in some lineages. In other cases, conserved ancestral building blocks (such as the proteinaceous cytoskeleton involved in locomotion and feeding) are widely shared, but have been used in different ways with diverse outcomes. The origins of other components are less clear and likely more recent, but also show a great deal of morphological variation (examples include photoreception systems or surface armor). Taken together, the diversity of cellular and molecular systems in microbial eukaryotes is simply staggering, and some emerging patterns indicate that convergent evolution played a major role in shaping the overall organization of eukaryotic cells at all levels (6, 11).

Below, several features will be described, for which an excessive complexity is a common denominator. This is counterintuitive in single-celled organisms, especially when selective advantages for these complex structures and/or mechanisms remain elusive. We argue that the theory of constructive neutral evolution (12), which invokes nonselective factors such as excess capacities, can best account for their emergence.

### Cellular Organization of Euglenids and Dinoflagellates

The comparable combinations of ultrastructural features in euglenozoans and alveolates (Table S1) have been appreciated for decades (13, 14). For instance, the cells of benthic predatory species of euglenids and dinoflagellates are streamlined and dorsoventrally flattened and possess batteries of extrusive organelles, or extrusomes, that are similar in morphology and



**Fig. 2.** Electron micrographs showing convergent ultrastructural features in some euglenozoans (A–E) and dinoflagellates (F–J). (A) Scanning electron micrograph (SEM) of a benthic euglenid (*Ploetia*) showing a dorsoventrally compressed cell with broad longitudinally arranged pellicle strips. (B) Transmission electron micrograph (TEM) through the nucleus of a euglenid cell (*Phacus*) showing a large nucleolus (asterisk) and permanently condensed chromosomes (arrows). (C) TEM through the cell surface or pellicle of a euglenid cell showing the proteinaceous strips (asterisk) beneath the plasma membrane (arrow). (D) TEM through the cell surface of a euglenid cell (*Peranema*) showing mucocysts (arrows). (E) TEM through a euglenophyte cell (*Euglena*) showing chloroplasts with thylacoids in stacks of 3 (stack between the arrows). (F) SEM of a benthic dinoflagellate (*Prorocentrum*) showing a dorsoventrally compressed cell with broad thecal plates. (G) TEM through the nucleus of a dinoflagellate cell (*Apicoporus*) showing a large nucleolus (asterisk) and permanently condensed chromosomes (arrows). (H) TEM through the cell surface of a dinoflagellate cell (*Apicoporus*) showing the proteinaceous “dinoflagellate pellicle” (asterisk) that sits beneath the alveoli and plasma membrane (both missing in this preparation). (I) TEM through the cell surface of a dinoflagellate cell (*Polykrikos*) showing trichocysts (arrows). (Inset) TEM showing a cross section through a trichocyst of *Polykrikos*. (J) TEM through a dinoflagellate cell (*Apicoporus*) showing chloroplasts with thylacoids in stacks of 3 (stack between the arrows). (Scale bars: A and F, 10  $\mu\text{m}$ ; B and G, 1  $\mu\text{m}$ ; C, H, and J, 0.3  $\mu\text{m}$ ; D, E, and I, and 0.5  $\mu\text{m}$ .)

behavior (Fig. 2). The mucocysts of euglenids and the trichocysts of dinoflagellates are compact, linear bodies containing a highly organized latticed framework of carbohydrates. When these

bodies are released through discrete pores through the surface of the cell, the extrusomes become hydrated and rapidly extend in length as spear-like threads (Fig. 2D and I) (15). Although the origin and function of extrusomes in both groups is not clear, they probably play a role in escape responses, defense, and capturing prey cells.

Benthic euglenids and dinoflagellates, in particular, adhere to substrates and are capable of gliding motility using 2 heterodynamic flagella equipped with flagellar hairs (or mastigonemes). In both groups, the recurrent flagellum sits within a groove on the ventral surface of the cell and is oriented backwards. Euglenids and dinoflagellates also possess cytoskeletal elements (called “paraxial/paraflagellar rods,” which run in parallel to the 9 + 2 microtubular axonemes within each flagellum) that are not found in any other group of eukaryotes. A major difference between euglenids and dinoflagellates, however, is the structure, orientation and motility of the anterior flagellum. The anterior paraxial rod in euglenids is oriented on the ventral side of the axoneme, is stiff and held straight in front of the cell; the paraxial rod functions with the flagellar hairs to produce gliding forces (16). By contrast, the anterior flagellum of dinoflagellates forms a transverse loop or spiral around the circumference of the cell and usually sits within a transverse groove called the cingulum (Fig. 1). The coiled transverse flagellum bears hairs and a flagellar membrane that connects it to the base of the cingulum, and this entire apparatus is capable of producing forces on the surrounding medium that tend to spin the cell around its longitudinal axis.

Many free-living euglenids and dinoflagellates engulf prey organisms using sophisticated feeding apparatuses positioned on the ventral side of the cell. Although the evolution of these apparatuses is a shared feature, the details of these ultrastructural systems are quite distinctive. For instance, there are a few chief components present in most of the predatory species of euglenids described so far, namely “rods” and “vanes.” Two feeding rods oriented longitudinally within the cell are composed of microtubules and amorphous proteinaceous material. These stiff elements provide structural support for gripping and internalizing prey cells and work in concert with 4–5 membranous vanes that are usually reinforced with additional microtubules (1). The vanes originate from the rods, form the inside core of the feeding apparatus, and create space within the apparatus by opening up in a pinwheel-like fashion; the same mechanism can cause the apparatus to protrude from the cell when feeding. By contrast, the diversity and complexity of feeding apparatuses in dinoflagellates probably reflect independent origins in different lineages within the group. The feeding apparatus in dinoflagellates can be simple pockets that unzip when prey is drawn into the cell, dynamic siphons that suck out the cytoplasm of prey cells in a straw-like fashion or expansive veils that completely envelop large filamentous prey and fold it methodically into manageable packets small enough to ingest. Different kinds of feeding apparatuses are often associated with different kinds of photoreceptive eyespots and ocelloids, suggesting that in some dinoflagellates, photoreceptors are adaptations for detecting and capturing photosynthetic prey. Some predatory euglenids with a rod-and-vane feeding apparatus also possess a photoreceptor system, as a putative stigma and photosensory swelling (17), and this combination of features may serve the same basic function as in dinoflagellates.

Another convergent similarity between benthic euglenids and dinoflagellates is the tendency to reinforce their cell surfaces with robust proteinaceous layers beneath the plasma membrane (Fig. 2C and H). Euglenids possess a distinctive (and synapomorphic) pellicle consisting of discontinuous strips that run longitudinally or helically over the entire cell surface (1). The strips articulate along their lateral margins, and in many euglenids these zones facilitate sliding between strips that produce rhythmic deformations in cell shape, called “euglenoid move-



5' end, of a SL already equipped with a methylated cap, followed by the polyadenylation at the 3' end (23).

Far less is known about the organization of dinoflagellate genomes. Due to the enormous size of their nuclear DNA, nearly all sequencing of dinoflagellate genes performed to date has focused on expressed sequence tags, which do not provide information on the context of the gene. Nevertheless, what little is known about dinoflagellate genomes suggests a fascinating parallel with kinetoplastids. It now appears that some genes are isolated in the genome, but others are organized as tandem repeats (20). These gene repeats are cotranscribed, resulting in polycistronic messages, and different from those of kinetoplastids because mRNAs have so far only been found to carry multiple copies of a single gene (20). These transcripts are apparently processed into monocistronic mRNAs, which are presumably the substrates for *trans*-splicing.

In kinetoplastids, the presence of polycistronic mRNAs, together with the absence of introns, is frequently argued to be an ancient holdover, frozen since their early divergence from other eukaryotes (25). However, this interpretation is flawed for several reasons, particularly because there is no evidence whatsoever for an ancient divergence of kinetoplastids (4). Nonetheless, the independent origin of the same features in dinoflagellates raises an intriguing alternative explanation, namely that the evolutionary origins of polycistronic mRNAs and *trans*-splicing are linked. This is all of the more compelling when one considers that both features are also found together in the nematode *Caenorhabditis elegans* (26). It is unlikely that this is either functionally advantageous or an evolutionary relict, but rather that the evolution of one feature preconditions the genome by removing deleterious effects of the second feature. For example, the establishment of widespread SL addition in a nuclear genome could precondition that genome for the subsequent establishment of polycistronic transcription. Polycistronic mRNAs that would otherwise be deleterious could flourish simply because the processing pathway eliminates their deleterious effect (the inability to translate all but the first cistron). SL addition appears to be universal in both dinoflagellates and kinetoplastids (in *C. elegans* 70% of mature mRNAs are produced through *trans*-splicing; 26). Polycistronic messages, however, are also near universal in kinetoplastids, whereas in dinoflagellates (and *C. elegans*) only a subset of genes are expressed on polycistronic mRNAs (20). Since so far only tandem duplications of closely related copies of the same gene are known in dinoflagellates, it would appear they may arise and dissolve continuously.

The functional impacts of SL addition and polycistronic transcription are also different in the 2 lineages. Posttranslational control may be somewhat restricted by the absence of sequence diversity at the 5' end of mRNAs, but more importantly a heavy use of polycistronic messages eliminates the possibility of transcription-level differentiation of expression of any genes within the same cluster. In kinetoplastids, there is only a handful of promoters and a marked paucity of transcription factors (25), unavoidably leading to the general lack of control over transcription initiation. Indeed, in the well-studied *T. brucei*, virtually all nuclear DNA seems to be permanently transcribed. Consequently, control levels in kinetoplastids are confined to RNA processing, export, and half-life, as well as translation and protein stability (27). This is a good illustration of how convergent processes differ in the details in different lineages. In this case, the kinetoplastids cotranscribe many different genes whereas dinoflagellates cotranscribe many copies of the same gene, and as a result transcription-level control is likely not so severely affected in the latter group.

### The Plastid: Three Membrane Plastids and Unique Targeting System

Plastids are known in both alveolates and euglenozoans to have been derived from secondary endosymbiosis: the uptake of a

eukaryotic alga by another eukaryote. In the Euglenozoa, plastids are derived from a green alga and are relatively restricted, being found in a subset of euglenids and nowhere else, namely the "euglenophytes" (28). In the Alveolata, plastids are derived from a red alga and are more widespread and ancient, being known in dinoflagellates and apicomplexans, and suspected of originating before the divergence of alveolates (5). As with nuclear genomes, plastids have evolved a number of unusual characteristics, some unique and some arising convergently. Euglenophyte plastid genomes are home to some unique self-splicing introns (29), whereas the dinoflagellate plastid genome has been massively reduced in coding content and broken into single gene minicircles with polyuridylylated transcripts (30). Curiously, both features are also found in kinetoplastid mitochondria (31).

Once again, however, 2 probably interconnected features have arisen in both groups. The vast majority of secondary plastids are bounded by 4 membranes. Most proteins in these plastids are encoded in the nucleus and are posttranslationally targeted to the organelle by way of a 2-part pathway beginning with the endomembrane system and followed by the original primary plastid targeting system. In dinoflagellates and euglenophytes, however, the plastid is novel in that it is bounded by 3 membranes rather than 4. It was argued that this may reflect a different mechanism of plastid uptake, specifically that in these lineages plastids arose through myzocytosis whereas other secondary plastids arose through endocytosis. Myzocytosis is a mode of predation where a cell pierces its prey and sucks the prey cytoplasm directly into a digestive vacuole, leaving the prey wall and membrane behind. Although not as common as endocytosis of whole prey cells, myzocytosis is known in both dinoflagellates and euglenozoans, leading to the suggestion that their plastids originated from a myzocytosed alga, and therefore lacked its plasma membrane (32). However, plastids in the closest relatives of dinoflagellates, apicomplexans and *Chromera*, are bounded by 4 membranes and have now been shown to be orthologous to the dinoflagellate plastid (3). Accordingly, in at least dinoflagellates, plastids must have originated in the same fashion as some 4-membrane counterparts and at one time been bounded by 4 membranes, which means that the origins of 3 membranes around the plastids in dinoflagellates and euglenophytes cannot be attributed to a shared, unusual mechanism such as myzocytosis.

Interestingly, the system used to target proteins to 3 membrane plastids is also different in subtle but important ways to that of canonical secondary plastids with 4 enveloping membranes, and the same variations have been adopted in dinoflagellates and euglenophytes. The N-terminal leaders that direct proteins to canonical secondary plastids include a signal peptide (to enter the endomembrane system) and a transit peptide (to cross the 2 plastid membranes), and are similar in secondarily derived red and green plastids. In dinoflagellates and euglenophytes, however, an additional hydrophobic domain is found following the transit peptide of some, but intriguingly not all proteins (33, 34). This domain is thought to anchor the proteins in the endomembrane, so as the protein moves through the Golgi apparatus the leader lays in the lumen but the mature protein remains in the cytosol (35, 36). The number of membranes and these unusual characteristics of targeting have both evolved convergently in dinoflagellates and euglenophytes, which suggests some link in how these 2 features evolved. Unfortunately, the mechanism by which proteins cross the membrane that is missing in both dinoflagellates and euglenophytes (the plasma membrane of the engulfed alga) is the most poorly understood step in the targeting pathway to canonical secondary plastids, so any specific model for preconditioning would be highly speculative.

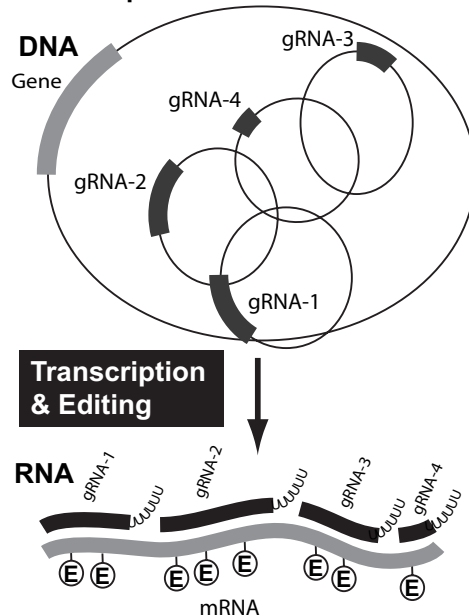
## The Mitochondrion: RNA Editing and Genome Breakdown

The mitochondrial genomes of dinoflagellates and kinetoplasts are both highly unorthodox, and once again have evolved some unique features and several common complex characteristics. The kinetoplastid mitochondrion contains uniquely structured, protein-rich mitochondrial ribosomes with a reduced RNA component, unusual fatty acid synthesis and respiratory complexes such as the prokaryotic-like complex I, alternative terminal oxidase, massive tRNA import, and incomplete Krebs cycle. The complex genome of the kinetoplastid mitochondrion is known as kinetoplast DNA or kDNA, its genes being subjected to unprecedented levels of RNA editing (Fig. 4) (31). Dinoflagellate mitochondria have received far less attention, but it is now emerging that their genomes have also evolved a number of highly unusual characteristics, including *trans*-splicing, tRNA import, fragmented rRNAs, the loss of start and stop codons, and an oligouridine tail (37, 38). Most strikingly, however, the structure of dinoflagellate mitochondrial genomes has also broken down into many fragments, the transcripts of which have high levels of RNA editing; however, as we discuss below, the details of both systems differ between kinetoplastids and dinoflagellates (Fig. 4).

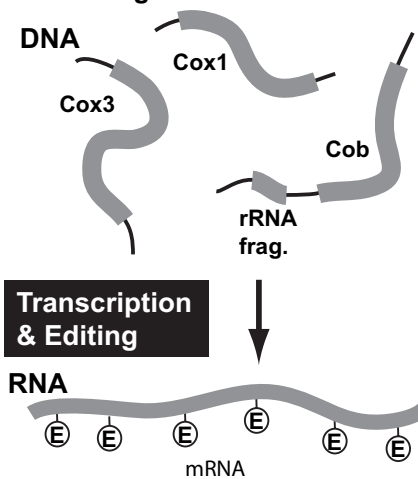
Within the Euglenozoa as a whole, mitochondrial genomes are generally odd. The euglenid mitochondrial genomes are experimentally refractive and remain poorly known (M. W. Gray, personal communication). The mitochondrion of related diplomonids was recently shown to harbor genome of unprecedented organization, with fragments of genes residing on minicircles, which are assembled in the correct order posttranscriptionally by means as yet unknown (39). Virtually nothing is known about the form or content of the giant kDNAs in bodonid flagellates, which are estimated to comprise millions of base pairs (40), whereas the kDNA networks of trypanosomatids are among the best studied and most complex mitochondrial genomes known. They are composed of circular DNA molecules that are relaxed and catenated into a single 3-dimensional network. These networks are composed of dozens of maxicircles, which are equivalents of classical mitochondrial genome, and thousands of minicircles (31) involved in editing, discussed below. The gene content of the maxicircle genome is not unusual, except for the complete absence of tRNA genes. tRNAs have been demonstrated to be imported from the cytosol into the tRNA-lacking organelle of *T. brucei*, so that the prokaryotic translation system of the mitochondrion must cope with imported eukaryotic tRNAs (41). The only exception is tRNA<sup>Met-i</sup>, the import of which is blocked because it cannot function in the prokaryotic system. Instead, tRNA<sup>Met</sup> formyl-transferase is present, which formylates the translation initiator tRNA<sup>Met-e</sup> upon import (42).

Within alveolates, mitochondrial genome evolution has also taken more than its share of strange turns. Although the circular mitochondrial genome of ciliates is undistinguished in both form and content, the genomes in apicomplexans and dinoflagellates are both highly reduced and often scrambled (37, 38, 43). These lineages have the smallest mitochondrial genomes known, with most species examined with just 3 protein-coding genes: *cox1*, *cox3*, and *cob* (strictly speaking, the dinoflagellate *Oxyrrhis* has only 2 genes since *cob* and *cox3* are expressed as a fusion) (37). The only other coding regions are small fragments of rRNAs. These do not amount to an entire copy of either large or small subunit rRNAs, so fragments are although to be important and the functional RNAs assembled by base pairing interactions. As with kinetoplastids, no tRNAs are encoded in these genomes, and they have been shown to be imported into apicomplexan mitochondria. Moreover, apicomplexans also block the import of tRNA<sup>Met-i</sup>, and use tRNA<sup>Met</sup> formyl-transferase to formylate the translation initiator tRNA<sup>Met-e</sup>. Indeed, kinetoplastids and api-

### A Kinetoplastid Mitochondria



### B Dinoflagellate Mitochondria



**Fig. 4.** Convergent organization and processing in kinetoplastid and dinoflagellate mitochondria. (A) In kinetoplastids the genome is composed of maxi- and minicircles. mRNAs expressed from maxicircles (gray) require extensive RNA editing, specifically the insertion and deletion of uridine residues. The process is mediated by several multiprotein complexes, but the information comes from small gRNAs (black), that are encoded on minicircles. In dinoflagellates, the genome is also fragmented but into multiple potentially linear pieces encoding 1 or more genes or gene fragments. Only 3 genes are encoded in the genome (*cox1*, *cox3*, and *cob*), and mRNAs are subject to extensive RNA editing. In this case, however, bases are substituted rather than inserted or deleted.

complexans have independently evolved very similar tRNA import mechanisms to cope with this unique lack of tRNAs (44). In apicomplexans, the 3 protein-coding genes map to a linear, tandem repeat with rRNA fragments interspersed (43). In dinoflagellates, the same coding regions are present, but the organization is much more complex. Here, multiple copies of each gene are found in various orientations on linear chromosomes of varying size. In some species, all possible permutations of 3 genes are adjacent, whereas in others chromosomes seem to contain copies of only 1 gene. Chromosomes also contain rRNA

fragments, and substantial noncoding regions, and some have been shown to have structurally complex ends characterized by families of repeats (37, 38, 45).

In kinetoplastids, the evolution of the complex genome organization is tightly linked to how genes are expressed, and specifically to RNA editing. The genes, such as they are, are encoded on the maxicircles and expressed as polycistronic mRNAs, but after processing into monocistrons these messages are then massively altered by the insertion and deletion of uridine residues (up to 553 insertions and 89 deletions in a single mRNA). Editing is mediated by hundreds of small guide (g) RNAs in an elaborate process involving numerous multisubunit protein complexes (31, 46). The gRNAs that contain the information that directs editing are encoded on the minicircles, so the breakup of the genome into 2 chromosome types is likely linked to the evolution of editing.

In dinoflagellates, RNA editing has also been found to be widespread, but the process is mechanistically different and in no way related to the breakdown of the genome structure. Here, transcripts are edited at  $\approx 2\%$  of their positions via substitutional editing, as opposed to insertion/deletion editing (38, 47, 48). Although A to G is the most common substitution, several others have been observed (U  $\rightarrow$  C, G  $\rightarrow$  C, G  $\rightarrow$  A, A  $\rightarrow$  C, and C  $\rightarrow$  U), suggesting a highly flexible and sophisticated editing mechanism (38, 48). Fragments of edited gene sequences have been found in dinoflagellate mitochondrial genomes, prompting the suggestion that they employ gRNAs similar to that of kinetoplastids (49). However, the genomes are prone to recombination, so the significance of these fragments remains unclear; overall, there is no direct evidence for any particular editing mechanism at present. It is worth noting that mitochondrial transcripts in dinoflagellates have substantial polyadenylated tails, a feature linked to the editing process in kinetoplastids (50), and generally very rare in organelles.

The limited data further indicate that uridine insertion type of RNA editing might even coexist with *trans*-splicing in diplomonads (51). We predict that the extreme diversity of editing types documented in the dinoflagellate mitochondrion (48) also requires poorly understood albeit complex protein machinery that is the result of constructive neutrality similar to that described above for the kinetoplastid mitochondrion.

## Conclusions

The deeper we look at protist biology, the greater the variety we discover in how cells can accomplish fundamental processes. Not only do protists represent the majority of the phylogenetic tree of eukaryotes, and therefore the greatest evolutionary diversity, but they also have pushed the limits of many biological systems and bending the “rules” of biology (such as the central dogma) far beyond what we see in the better studied multicellular eukaryotes. The alveolates and the euglenozoans may be “hotspots” for the generation of diverse solutions to fundamental processes, but it is also possible that they only appear this way because they are among the best studied protist groups. Other odd protists abound, but we know next to nothing about many of them, particularly at the molecular and genomics levels. All this is presently changing, and to interpret genomic diversity in eukaryotes we will have to set aside many of our preconceptions.

Comparing the alveolates and the euglenozoans is also appealing because they have broken many of the same rules in the same general way. Because they are so distant on the phylogenetic tree of eukaryotes (4, 7), convergence between the 2 groups would ultimately be influenced only by intrinsic factors of a very basic nature (i.e., that are likely common to most or all eukaryotes) (6). In contrast, where multiple aspects of a system have all converged similarly, it is likely that the convergent appearance of one new characteristic can be a strong factor in the

convergent evolution of others. Even if these characteristics are not obligatorily functionally linked, their evolution may be tightly linked. For example, polycistronic mRNAs can exist without a SL, but they are evolutionarily linked because adding the SL allows the polycistronic mRNA to function. Conversely, one can imagine other ways to get a polycistronic mRNA to function without SL processing (e.g., changes to translation initiation), but because no such system is known, these are evidently less likely than the advent of SL processing. In other words, within the limited universe of acceptable changes, one change closes some possibilities, but opens new ones as well.

So why have protists in general and alveolates and euglenozoans in particular engaged in so much evolutionary experimentation? Many characteristics discussed here have been considered individually and concluded to be ancient relicts, going back even so far as the RNA world, or to have been favored by selection over the canonical way of accomplishing the same task (52, 53). We find neither of these arguments to be particularly compelling given the narrow distribution of these characters in nature, and their often extreme complexity. For example, dozens of nuclear-encoded proteins are required for *T. brucei* to edit just 12 mRNAs (31, 46, 50). Despite considerable controversy, no obvious evolutionary advantage has ever been demonstrated for this type of editing, and such possible advantages that have been proposed (e.g., the generation of 2 proteins from 1 gene) (53) are more than outweighed by the demonstrated cost (i.e., “save” 1 gene at the cost of dozens of genes). We argue that constructive neutral evolution offers a more compelling explanation (12, 54). This is a very simple and intuitive way of explaining complexity in biological systems, but one that has not received much attention. Briefly, it is possible for a biological system to increase in complexity (that is, to increase the number of components or interactions needed to sustain the system) by making a series of neutral changes that collectively do not affect fitness. Pan-editing is often thought of as an error correcting system, but as Stoltzfus (12) pointed out the duplicated information (e.g., gRNAs) must have been created before the mutations they are correcting, or they too would carry the mutations—so the error-then-solution model is backwards. Instead, if a gratuitous duplication of information took place first (i.e., the origin of a gRNA), then a subsequent mutation could be neutralized by the presence of the duplicated information needed to change it. The fixation of such a mutation would render the gRNA essential, and would also allow for further mutations as long as the gRNAs could mediate their reversal. This last point is important because it would bias the system against the loss of the gRNA since mutations at many sites will further establish the gRNA as essential, whereas only complete reversion to the original sequence could render it unnecessary. Overall, the editing activity and the sites that are edited will coevolve, and the complexity of the system will inevitably grow while conferring no real selective advantage (for many other case studies and much greater detail) (see refs. 12 and 54).

Within this framework, together with the recognition that the evolution of an unusual character can be an intrinsic factor in the subsequent evolution of additional, specific characters, a complex cellular system may be explained simply by identifying the event(s) that preconditioned the cell for such a system. Convergence may offer a glimpse into these conditions by revealing how characters are linked when the same events are played out multiple times.

**ACKNOWLEDGMENTS.** We thank Mona Hoppenrath and Susan Breglia for providing the images in Figs. 1 and 2F, respectively. This work was supported by the Grant Agency of the Czech Republic Grant 204/09/1667 (to J.L.); Ministry of Education of the Czech Republic Grants LC07032, 2B06129, and 600766580 (to J.L.); a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution (to B.S.L. and P.J.K.); and the Canadian Institute for Advanced Research (J.L., B.S.L., and P.J.K.).

1. Leander BS, Esson HJ, Breglia SA (2007) Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays* 29:987–1000.
2. Simpson AGB, Stevens JR, Lukeš J (2006) The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* 22:168–174.
3. Obornik M, Janouškovec J, Chrudimský T, Lukeš J (2009) Evolution of the apicoplast and its hosts: From heterotrophy to autotrophy and back again. *Int J Parasitol* 39:1–12.
4. Keeling PJ, et al. (2005) The tree of eukaryotes. *Trends Ecol Evol* 20:670–676.
5. Keeling PJ (2009) Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J Euk Microbiol* 56:1–8.
6. Leander BS (2008) A hierarchical view of convergent evolution in microbial eukaryotes. *J Euk Microbiol* 55:59–68.
7. Hampl V, et al. (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “supergroups.” *Proc Natl Acad Sci USA* 106:3859–3864.
8. Conway-Morris S, Gould SJ (1998) Showdown on the Burgess Shale. *Nat Hist* 107:48.
9. Zakon HH (2002) Convergent evolution on the molecular level. *Behav Evol* 59:250–261.
10. Emery NJ, Clayton NS (2004) The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* 306:1903–1907.
11. Arndt J, Reznick D (2008) Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends Ecol Evol* 23:26–32.
12. Stoltzfus A (1999) On the possibility of constructive neutral evolution. *J Mol Evol* 49:169–181.
13. Taylor FJR (1987) *The Biology of Dinoflagellates* (Blackwell Scientific, Oxford).
14. Bouck GB, Ngo H (1996) Cortical structure and function in euglenoids with reference to trypanosomes, ciliates, and dinoflagellates. *Int Rev Cytol* 169:267–318.
15. Hausmann K (1978) Extrusive organelles in protists. *Int Rev Cytol* 52:197–268.
16. Saito A, et al. (2003) Gliding movement in *Peranema trichophorum* is powered by flagellar surface motility. *Cell Motil Cytoskel* 55:244–253.
17. Leander BS, Triemer RE, Farmer MA (2001) Character evolution in heterotrophic euglenids. *Eur J Protistol* 37:337–356.
18. McEwan M, Humayun R, Slamovits CH, Keeling PJ (2008) Nuclear genome sequence survey of the dinoflagellate *Heterocapsa triquetra*. *J Euk Microbiol* 55:530–535.
19. Berriman M, et al. (2005) The genome of the African trypanosome, *Trypanosoma brucei*. *Science* 309:416–422.
20. Bachvaroff TR, Place AR (2008) From stop to start: Tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS One* 3:e2929.
21. Zhang H, et al. (2007) Spliced leader RNA trans-splicing in dinoflagellates. *Proc Natl Acad Sci USA* 104:4618–4623.
22. Slamovits CH, Keeling PJ (2008) Widespread recycling of processed cDNAs in dinoflagellates. *Curr Biol* 18:R550–R552.
23. Campbell DA, Thomas S, Sturm NR (2003) Transcription in kinetoplastid protozoa: Why be normal? *Microbes Infect* 5:1231–1240.
24. Ghedin E, et al. (2004) Gene synteny and evolution of genome architecture in trypanosomatids. *Mol Biochem Parasitol* 134:183–191.
25. Gunzl A, Vanhamme L, Myler PJ (2007) Transcription in Trypanosomes: A different means to the end. In *Trypanosomes After the Genome*, eds Barry D, McCulloch R, Mottram J, Acosta-Serrano A (Horizon Bioscience, Norfolk, UK).
26. Graber JH, Salisbury J, Hutchins LN, Blumenthal T (2007) *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. *RNA* 13:1409–1426.
27. Clayton CE (2002) Life without transcriptional control? From fly to man and back again. *EMBO J* 21:1881–1888.
28. Leander BS (2004) Did trypanosomatid parasites have photosynthetic ancestors? *Trends Microbiol* 12:251–258.
29. Copertino DW, Hallick RB (1993) Group-II and group-III introns of twintrons—potential relationships with nuclear pre-messenger RNA introns. *Nucl Acids Res* 18:467–471.
30. Wang Y, Morse D (2006) Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*. *Nucl Acids Res* 34:613–619.
31. Lukeš J, Hashimi H, Ziková A (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates. *Curr Genet* 48:277–299.
32. Schnepf E, Deichgraber G (1984) “Myzocytosis,” a kind of endocytosis with implications to compartmentalization in endosymbiosis. *Naturwissenschaften* 71:218–219.
33. Patron NJ, Waller RF, Archibald JM, Keeling PJ (2005) Complex protein targeting to dinoflagellate plastids. *J Mol Biol* 384:1015–1024.
34. Durnford DG, Gray MW (2006) Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryot Cell* 5:2079–2091.
35. Sulli C, Fang ZW, Muchhal U, Schwartzbach SD (1999) Topology of *Euglena chloroplast* protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J Biol Chem* 274:457–463.
36. Nassoury N, Cappadocia M, Morse D (2003) Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* 116:2867–2874.
37. Slamovits CH, Saldarriaga JF, Larocque A, Keeling PJ (2007) The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *J Mol Biol* 372:356–368.
38. Nash EA, Nisbet RER, Barbrook AC, Howe CJ (2008) Dinoflagellates: A mitochondrial genome all at sea. *Trends Genet* 24:328–335.
39. Marande W, Lukeš J, Burger G (2005) Unique mitochondrial genome structure in diplomemids, the sister group of kinetoplastids. *Eukaryot Cell* 4:1137–1146.
40. Lukeš J, Jirků M, Avliyakov N, Benada O (1998) Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. *EMBO J* 17:838–846.
41. Crousaz-Esseiva A, Naguleswaran A, Hemphill A, Schneider A (2004) Mitochondrial tRNA import in *Toxoplasma gondii*. *J Biol Chem* 279:42363–42368.
42. Tan THP, Bochud-Allemann N, Horn EK, Schneider A (2002) Eukaryotic-type elongator tRNA<sup>Met</sup> of *Trypanosoma brucei* becomes formylated after import into mitochondria. *Proc Natl Acad Sci USA* 99:1152–1157.
43. Feagin JE (2000) Mitochondrial genome diversity in parasites. *Int J Parasitol* 33:371–390.
44. Bouzaidi-Tiali N, Aeby E, Charriere F, Pusnik M, Schneider A (2007) Elongation factor 1a mediates the specificity of mitochondrial tRNA import in *T. brucei*. *EMBO J* 20:4302–4312.
45. Jackson CJ, et al. (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol* 5:41.
46. Hashimi H, Ziková A, Panigrahi AK, Stuart KD, Lukeš J (2008) TbRGG1, a component of a novel multiprotein complex involved in kinetoplastid RNA editing. *RNA* 14:970–980.
47. Lin SJ, Zhang HA, Spencer DF, Norman JE, Gray MW (2002) Widespread and extensive editing of mitochondrial mRNAs in dinoflagellates. *J Mol Biol* 320:727–739.
48. Zhang H, Lin S (2008) mRNA editing and spliced-leader RNA trans-splicing groups *Oxyrrhis*, *Noctiluca*, *Heterocapsa*, and *Amphidinium* as basal lineages of dinoflagellates. *J Phycol* 44:703–711.
49. Nash EA, et al. (2007) Organisation of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol Biol Evol* 24:1528–1536.
50. Etheridge RD, Aphasizheva I, Gershon PD, Aphasizhev R (2008) 3' adenylation determined mRNA abundance and monitors completion of RNA editing in *T. brucei* mitochondria. *EMBO J* 27:1596–1608.
51. Marande W, Burger G (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science* 318:415.
52. Speijer D (2007) Evolutionary aspects of RNA editing. In *RNA Editing*, ed Goringher HU (Springer, Berlin).
53. Ochsenreiter T, Cipriano M, Hajduk SL (2008) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS One* 3:e1566.
54. Covello PS, Gray MW (1993) On the evolution of RNA editing. *Trends Genet* 9:265–268.