Review in Advance first posted online on June 17, 2016. (Changes may still occur before final publication online and in print.)

Protists and the Wild, Wild West of Gene Expression: New Frontiers, Lawlessness, and Misfits

David Roy Smith[1] and Patrick J. Keeling[2]

[1]Department of Biology, University of Western Ontario, London, Canada N6A 5B7; email: dsmit242@uwo.ca

[2]Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, Canada V6T 1Z4; email: pkeeling@mail.ubc.ca

Annu. Rev. Microbiol. 2016. 70:161–78

The *Annual Review of Microbiology* is online at micro.annualreviews.org

This article's doi: 10.1146/annurev-micro-102215-095448

Keywords

constructive neutral evolution, mitochondrial transcription, plastid transcription, posttranscriptional processing, RNA editing, *trans*-splicing

Abstract

The DNA double helix has been called one of life's most elegant structures, largely because of its universality, simplicity, and symmetry. The expression of information encoded within DNA, however, can be far from simple or symmetric and is sometimes surprisingly variable, convoluted, and wantonly inefficient. Although exceptions to the rules exist in certain model systems, the true extent to which life has stretched the limits of gene expression is made clear by nonmodel systems, particularly protists (microbial eukaryotes). The nuclear and organelle genomes of protists are subject to the most tangled forms of gene expression yet identified. The complicated and extravagant picture of the underlying genetics of eukaryotic microbial life changes how we think about the flow of genetic information and the evolutionary processes shaping it. Here, we discuss the origins, diversity, and growing interest in noncanonical protist gene expression and its relationship to genomic architecture.

Contents

INTRODUCTION

The storage and expression of genetic information can exemplify the use of simple tools efficiently and effectively. One only needs to think of the beautiful and paradigmatic lactose operon or the lambda phage genetic switch, which are foremost models for teaching gene regulation. But as with many processes in biology, elegance is often the exception; closer inspection, broader sampling, and better understanding of the evolutionary processes that shape these systems can reveal a biological world that is more chaotic than elegant. Genomic architecture and gene expression typify this progression, and nowhere is the chaotic nature of evolutionary complexity more evident than in the genomes of microbial eukaryotes (protists), including their mitochondrial, plastid, and nuclear DNAs (mtDNAs, ptDNAs, and nucDNAs).

Protists are abundant and ubiquitous members of nearly all known ecosystems, and together they account for a large proportion of eukaryotic biodiversity (8, 91, 129). In fact, most major groups of eukaryotes are strictly composed of microbial species, and animals, fungi, and land plants evolved independently from protist ancestors (8). For many protists, genomic architecture is highly variable, especially in the mitochondrion and plastid (110), and protists can also harbor complicated transcriptional and translational jigsaw puzzles (5, 85). Beyond the need for the transcription to RNA and translation to protein, many genes require gratuitous RNA editing; *trans*-splicing of fragmented, scrambled exons; removal of introns within introns; and/or deciphering via nonstandard genetic codes. In some species, the levels of posttranscriptional processing are so extensive that given the DNA sequence alone it is not possible to distinguish coding from noncoding DNA or to deduce the resulting gene products (102). Taken together, protists can be veritable genetic circus acts, consistently breaking the rules of what was once thought to be axiomatic and generating questions and debate about the evolution and function of such extravagant expression systems (42, 72, 114). Although protists have long been models for investigating the expression of genes and proteins, their propensity toward unconventional transcriptional architectures and the fact that most microbial life is not maintained in culture collections have resulted

in barriers to their study (11, 22). Consequently, some of the most unusual modes of eukaryotic gene expression remain poorly understood and undercharacterized. But that is quickly changing.

Recently, there has been increased interest in protists, spurred on by a wider appreciation for their pivotal role in global biogeochemical cycles (11, 129) as well as by the introduction of high-throughput molecular sequencing technologies (83); highly sensitive proteomic methods (7); and sophisticated, user-friendly bioinformatics software (107). The last five years have seen large, international initiatives devoted to comprehending protist transcription, such as the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP), which assembled, annotated, and made publicly available the transcriptomes from hundreds of marine protists (55). Massive environmental RNA sequencing (metatranscriptomics) is also aiding research into microbial gene expression (76), as is single-cell transcriptomics (61), allowing for the acquisition of RNA data from species that are uncultivated or in complex culture. Proteomics, too, is disentangling protist gene expression, providing large-scale categorizations of proteins in a wide range of species and organelles, from the eyespot of *Chlamydomonas reinhardtii* (99) to the mitosome of *Giardia intestinalis* (51).

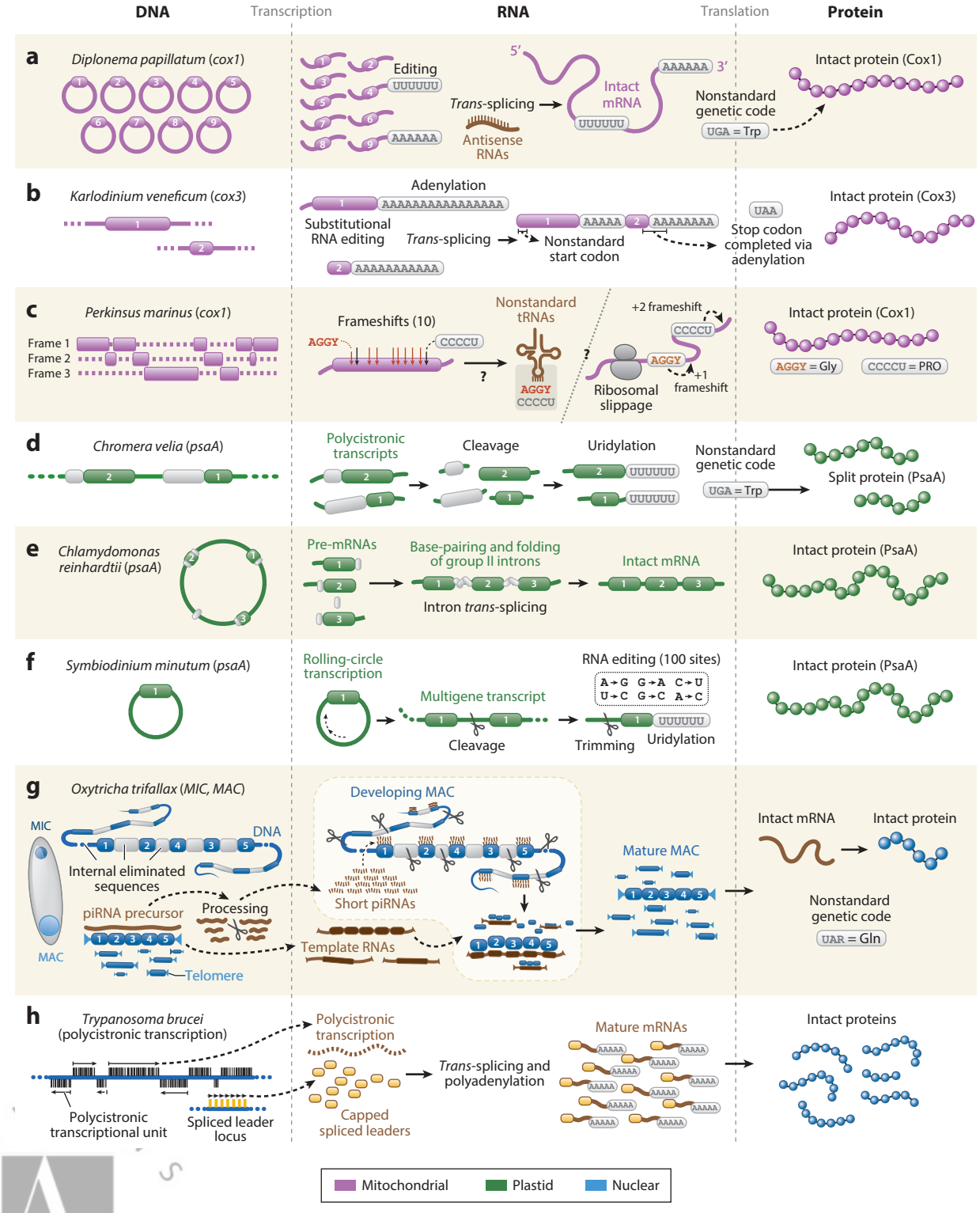
Ultimately, our current understanding of protist gene expression has come from the combination of many disciplines and approaches, particularly the blending of classical methods with next-generation technologies. The outcome is a complicated, labyrinthine picture of the underlying genetics of microbial life. Below, we highlight compelling examples of noncanonical transcription and translation in protist mitochondria, plastids, and nuclei. We discuss the roles of evolutionary ratchets in shaping convoluted expression systems and explore how high-throughput sequencing technologies have provided unprecedented amounts of data but have also resulted in a departure from more direct analyses of RNA and protein, which remain crucial for accurately characterizing gene expression.

NONCANONICAL GENE EXPRESSION IN MITOCHONDRIA AND PLASTIDS

At various points since their endosymbiotic origins over a billion years ago, mitochondria and plastids acquired transcriptional and translational quirks. In certain species these quirks are severe and multifaceted; in others they are minor or absent entirely. Sometimes they arose in parallel in diverse lineages and different genetic compartments; in other instances they have been restricted to a specific group or genome. But wherever these embellishments have popped up, they often appear to bestow no obvious selective benefit and instead seem to place a sizeable burden on the recipient, much like the unfettered expansion of government bureaucratic complexity. In the following paragraphs, and in **Figures 1** and **2**, we summarize some of the many bizarre forms of organelle gene expression exposed over the last three decades.

Multilayered Complexity of Organelle Transcription in Euglenozoans

Euglenozoan organelle genomes can contain layer upon layer of transcriptional encryption, the resolution of which requires extensive downstream processing, including RNA editing, unscrambling and rejoining of gene segments, and stepwise progressive splicing. Indeed, the journey from mtDNA to functional protein in the mitochondria of kinetoplastids—the archetype of bizarre gene expression—requires a complex interplay between many chromosomes, transcripts, proteins, and genetic compartments, and up to 90% of the codons within a mature mitochondrial transcript are established through RNA editing (70, 102). In *Trypanosoma brucei*, for example, the *cox3* primary transcript is gibberish until an RNA editing system, involving dozens of nucleus-encoded,



mitochondrion-targeted proteins and over 20 guide RNAs (gRNAs), mediates the insertion and deletion of about 550 and 40 uridine residues, respectively (10, 18, 30). The entire system requires hundreds of proteins and hundreds of gRNAs to express fewer than 20 genes and results in a genome that is fragmented into thousands of pieces and requires an elaborate filing system for accurate replication and segregation, all simply to produce proteins not unlike their ancestral form (70).

Trypanosome mitochondria are extreme but not unique. The mitochondrial genome of *Diplonema papillatum*, a sister euglenozoan to trypanosomes, comprises more than 75 miniature chromosomes, each with a small coding module that is joined with its partnered modules from neighboring chromosomes through *trans*-splicing (75, 125). This processing is likely facilitated through antisense RNAs and in some cases requires U-insertion RNA editing (124). The *cox1* coding modules from *D. papillatum* are dispersed over nine chromosomes and need to be transcribed independently before being linked together via eight splicing events into a single molecule (125). The *D. papillatum* mitochondrion also holds the current record for the largest number of uridines added at a single editing site: 26 (124).

Other euglenozoans break records for subtracting genetic elements. Expression of the *Euglena gracilis* plastid genome requires the removal of approximately 160 introns (one for every 900 nucleotides of ptDNA), including 15 twintrons (introns within introns), which need to be subtracted sequentially for accurate splicing (43), and some of these transcripts are also polyadenylated (130). As if that weren't enough, the plastid *rps18* gene has a multipartite twintron—an external intron interrupted by an internal intron containing two additional introns (25). The mitochondrial compartment of *E. gracilis* offers no return to normalcy with split ribosomal RNAs (rRNAs; described below) and what might be the evolutionary precursors for the gRNAs that direct post-transcriptional editing in kinetoplastids and diplonemids (32, 116).

Dinoflagellate Organelle Gene Expression: Not To be Outdone by Euglenozoans

Gene expression in euglenozoans is odd on many levels, so it is surprising that it is matched almost point for point by an unrelated lineage, the dinoflagellates. They, too, experience extensive organelle RNA editing, but of a nonhomologous substitutional type, affecting both the mitochondria (48, 126) and the plastids (5, 21), and where 11 of the 12 possible types of nucleotide substitution (A to C, A to G, etc.) have been identified (86, 96). Plastid transcripts also sport 3′ polyuridylated tails and are likely expressed by rolling-circle transcription of miniature circular chromosomes (4, 21, 133). Remarkably, the RNA editing and polyuridylation pathways in dinoflagellate plastids have infected replacement plastids, as displayed by *Karlodinium veneficum* and its close relative *Karenia mikimotoi*. These species substituted their ancestral peridinin-containing plastid for a

Figure 1

Diversity of mitochondrial (*purple*), plastid (*green*), and nuclear (*blue*) gene expression in various protists. The journey from DNA (*left*) to RNA (*middle*) to protein (*right*) is depicted for various mitochondrial (*a–c*), plastid (*d–f*), and hypothetical nuclear (*g*,*h*) genes, including ones encoding cytochrome *c* oxidase subunits I (*cox1*) and III (*cox3*) and photosystem I protein A1 (*psaA*). Many of the models shown are hypothetical, simplified, and not to scale; see primary references for details. Mitochondrial gene expression: euglenozoan *Diplonema papillatum* (*a*) (75, 125), dinoflagellate alga *Karlodinium veneficum* (*b*) (48), and perkinsid *Perkinsus marinus* (*c*) (79, 131). Plastid gene expression: chromerid alga *Chromera velia* (*d*) (49), green alga *Chlamydomonas reinhardtii* (*e*) (39), and dinoflagellate alga *Symbiodinium minutum* (*f*) (86). Nuclear gene expression: ciliate *Oxytricha trifallax* (*g*) (16, 28, 87, 88, 119) and kinetoplastid *Trypanosoma brucei* (*h*) (59, 66, 77, 118). Abbreviations: MAC, macronucleus; MIC, micronucleus; piRNA, Piwi-interacting small RNAs.

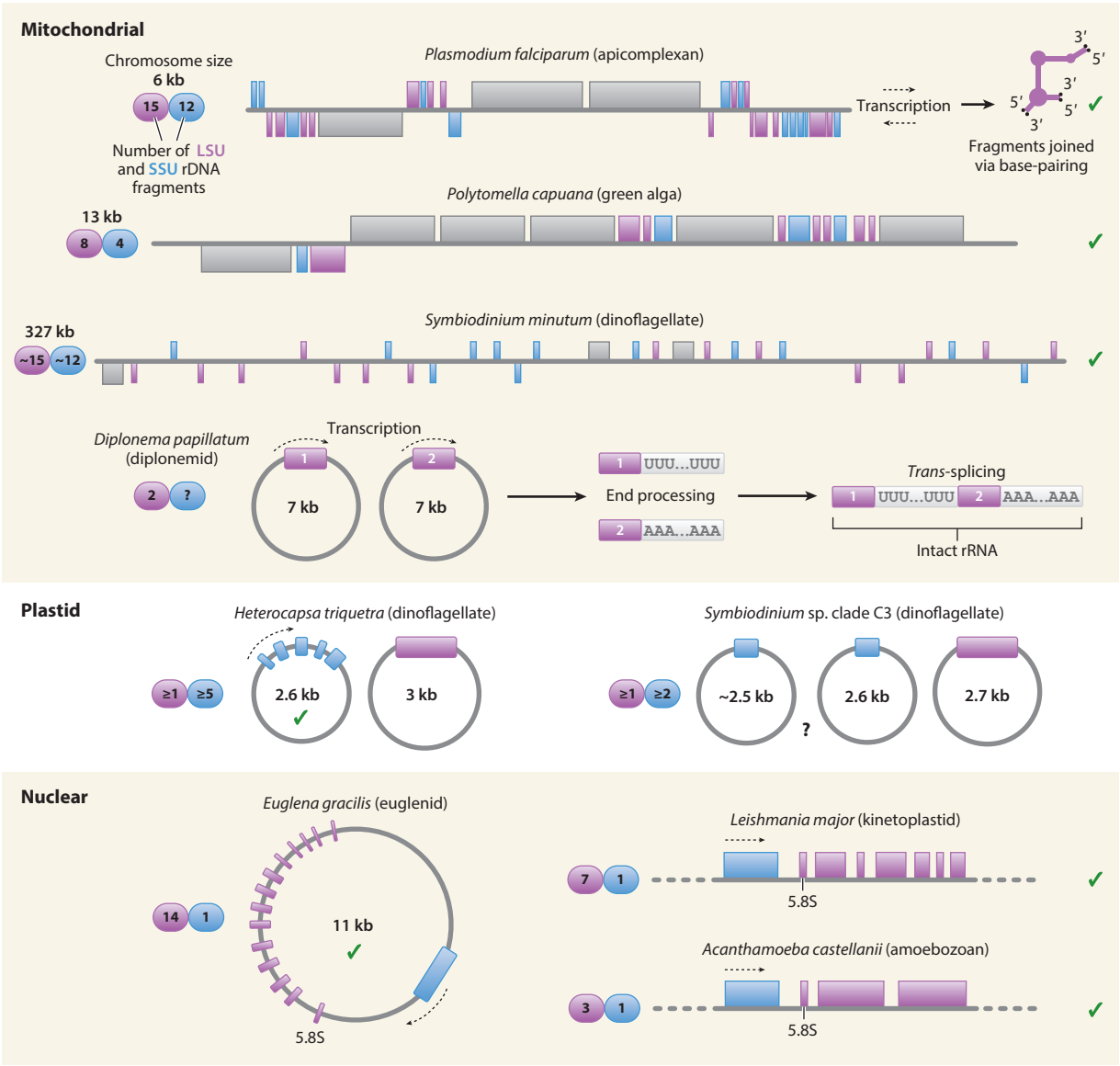


Figure 2

Fragmented large- and small-subunit (LSU and SSU) rRNAs from the mitochondrial, plastid, and nuclear genomes of various protists. Numbers of LSU and SSU rRNA-coding fragments are shown in purple and blue, respectively. Intervening genes (*gray boxes*), chromosome size (in kilobases; not to scale), and transcriptional polarity (*dashed arrows*) are shown. Transcribed rRNA fragments can come together to form a functional rRNA via secondary pairing interactions (*green checkmarks*) or through *trans*-splicing (*Diplonema papillatum*). For *Symbiodinium* sp. clade C3, the complete number of SSU rRNA fragments and how they are ultimately joined are unknown. For the nuclear genome, the LSU rRNA gene includes the 5.8S region.

plastid from haptophyte algae. This group displays neither RNA editing nor polyuridylation, but once housed in a dinoflagellate, the haptophyte-derived organelles acquired both characteristics (24, 96).

Expression in dinoflagellate mitochondria is, if anything, stranger still. In addition to undergoing widespread substitutional editing, mitochondrial transcripts are regularly 5′ polyuridylated and 3′ oligo-adenylated, and *trans*-spliced, and most lack canonical start and stop codons. The mitochondrial *cox3* transcript, for example, is fused with *cob* in some species (104), whereas in others it is fragmented and *trans*-spliced so that a species-specific number of bases from its polyadenylated tail are incorporated at the splice site (47, 48) and are essential for expression (**Figure 1**). The polyA tail is also used by several species to form a stop codon on *cox3*, whereas other genes appear to lack stop codons altogether and simply translate the tail as oligo-lysine (47, 104).

Translational Slippage in Mitochondria

The oyster parasite *Perkinsus marinus* is closely related to dinoflagellates, but its mode of mitochondrial expression has taken a different but equally perplexing route. Although RNA editing does not occur in *P. marinus* mitochondria, the *cox1* and *cob* genes contain multiple frameshifts that would ultimately lead to nonfunctional proteins if not amended (79, 131). These irregularities are retained in the corresponding mature transcripts and eventually corrected by programmed ribosomal slippage (79). Every time an in-frame AGG or CCC appears in the mRNA, the reading frame moves forward by one or two bases, respectively. The result is that AGGY codes for glycine and CCCCU for proline. How these frameshifts are introduced during translation is poorly understood, but it is hypothesized that stalled ribosomes skip the first bases of these codons or that specialized tRNAs recognize nontriplet codons (79). Whatever the mechanism, frameshifts have to be introduced 10 times within the *cox1* mRNA alone to yield a functional protein. In *Perkinsus chesapeaki*, the same system was found but with slight variation in the number of codon slippages that occur (131). *Perkinsus* spp. also appear to harbor a relic, nonphotosynthetic plastid, but it has lost its genome and gene expression system (79), which may be viewed as the ultimate in noncanonical expression diversity (see sidebar "Gene Expression–Less Organelles").

A parallel form of translational slippage in the mitochondria of a fungus has been described, but here the ribosome does not really skip but jumps a considerable distance. At over 80 sites in the *Magnusiomyces capitatus* mitochondrial genome, the ribosome encounters unassigned codons and secondary-structure-rich sequences similar to mobile elements in other fungi. From these cues, the ribosome recognizes takeoff and landing sites and uses these sites to bypass large segments of mRNA, resulting in intact and functional polypeptides (64).

GENE EXPRESSION–LESS ORGANELLES

Among the most radical changes to gene expression is to lose it altogether. In various lineages, mitochondria have been subverted into anaerobic organelles, called hydrogenosomes or mitosomes (45), and plastids have lost photosynthetic capabilities (50, 112). In some cases, these vestigial organelles retain a genome and gene expression (e.g., 82), but in other instances the genome has been completely lost (109). This has been known for some time for relic mitochondria (9, 45, 90), and more recently genome-lacking plastids have been identified (50, 112). These organelles are entirely dependent on genes in the nucleus, which is a radical change from their original state as free-living bacteria.

Noncanonical Genetic Codes

The universal genetic code is undeniably ancestral and highly conserved, but it is not quite universal, particularly within mitochondria (60). In the mtDNAs of kinetoplastids, diplonemids, and many red algae, for example, UGA (normally a stop codon) specifies tryptophan (121, 125), and dinoflagellate mitochondria regularly employ start codons alternative to the standard AUG (104, 126), as do some apicomplexan and ciliate mitochondrial genes (26). Novel code alterations are still being uncovered: The green alga *Pseudomuriella schumacherensis* was recently found to use UCG as a mitochondrial stop codon (35). In fact, most mitochondria, including those of animals, fungi, and plants, have experienced at least one mtDNA codon alteration, resulting in over a dozen mitochondrial codes and departure from the universal code. In contrast, the universal code still dominates in plastids, with sporadic examples of noncanonical codes in some apicomplexans and dinoflagellates (65, 80).

Connecting the Transcriptional Dots of Fragmented Genes

Introns are sometimes described as splitting genes into pieces, but most intron-containing genes are expressed as contiguous RNA molecules. However, many genes really are split in that they are physically separated in the genome (sometimes on different DNA molecules), resulting in disjointed, incomplete transcripts that need to be strung together after transcription (or after translation; 49). Mitochondria exhibit a proclivity for gene fragmentation, especially in their rRNAs, which have independently evolved fragmented architectures in diverse groups, including chlorophycean green algae, euglenozoans, and alveolates (6, 31, 116) (**Figure 2**).

The most extreme example of rRNA fragmentation yet described is in the malaria parasite *Plasmodium falciparum*, where the large and small subunit (LSU and SSU) mitochondrial rRNA genes have splintered into numerous coding modules randomly distributed across both strands of the genome and interspersed with protein-coding genes (31). Given their small sizes (23–190 nucleotides) and disorganization, it has taken years to identify the more than 25 rRNA modules, which are expressed via cleavage of long precursor polycistronic transcripts, undergo oligoadenylation, and come together through secondary pairing interactions to form functional rRNAs (31, 52). A similar situation exists in the mitochondria of ciliates (44), chlamydomonadalean algae (6, 27), and dinoflagellates (126); moreover, the latter group can also have split plastid rRNAs (21). As in *P. falciparum*, the fragmented rRNAs from these different groups and compartments are processed, sometimes from long polycistronic transcripts, and then joined via base pairing into complete rRNAs (6, 21, 27, 44).

In other systems, discontinuous rRNA genes are reassembled into a single covalently continuous molecule by *trans*-splicing. In *D. papillatum* this affects most genes, including the LSU rRNA, and appears to be mediated by antisense RNAs (124). Intron-mediated *trans*-splicing of organelle mRNAs is well documented for many different groups (39), particularly the plastid-encoded *psaA* gene, but it has surprisingly not yet been observed for bridging discontinuous organelle rRNAs (39). This is particularly surprising for dinoflagellate mitochondria, where a *trans*-splicing mechanism exists for ligating *cox3* mRNAs, but the fragmented mitochondrial rRNAs from this group are not *trans*-spliced (48).

SIMILAR PHENOMENA IN NUCLEAR GENE EXPRESSION: ALTERED CODES, FRAGMENTATION, *TRANS*-SPLICING, AND SLIPPAGE

Mitochondria and plastids have stretched the limits of genetic expression, but many of the same peculiarities are also found in nuclear genes, albeit at a lower frequency and to a lesser extent than

that typically seen in organelles (110). Other features of organelle expression are near absent from nuclei, such as RNA editing, which can occur in nuclear genomes (15) but not in the widespread or gratuitous form observed in organelles.

Noncanonical genetic codes have evolved many times in nuclear genomes—including those of yeasts, green algae, diplomonads, oxymonads, and especially ciliates where several code changes have occurred independently in different subgroups (17, 39, 46, 56, 57). The nonstandard code whereby the canonical stop codons UAG and UAA encode glutamine is the most common change in nuclei, but it has surprisingly not yet been found in organelles or bacteria, likely reflecting differences in translation (57).

Fragmented genes are rare but present in nuclear systems. Euglenozoans can have fragmented rRNAs, broken into as many as 14 pieces that function as a noncovalent network (**Figure 2**) (100, 127). In the diplomonad parasite *G. intestinalis*, some proteins are encoded by discontinuous pre-mRNAs, which are independently transcribed from disparate chromosomal regions and joined into a contiguous mRNA by spliceosome-mediated *trans*-splicing (54, 98). Nuclear tRNAs can be similarly disjointed. The 5′ halves of several tRNA genes in the red alga *Cyanidioschyzon merolae* are located downstream of their 3′ portions and are expressed via circular RNA intermediates (113); permuted tRNA genes have also been spotted in the nucDNAs of prasinophyte green algae and the nucleomorph genome of *Bigelowiella natans* (78)—nucleomorphs as a whole have bizarre forms of gene expression (see sidebar "Nucleomorph Gene Expression"). Lastly, spliceosomal introns can also form progressively spliced "stwintrons," like the group II twintrons of organelles (33).

Spliced-Leader *Trans*-Splicing and Polycistronic mRNA

Euglenozoans and dinoflagellates have idiosyncratic organelle gene expression systems, but they can undergo strange forms of nuclear expression as well. Trypanosome nuclear genomes, for instance, have large tracks of genes oriented on the same strand lacking canonical sequence-based promoters. Instead, RNA polymerase II initiates transcription at regions of modified histones upstream of gene tracts, often resulting in extremely long polycistronic transcripts, which can contain tens to hundreds of genes and cover large sections of the genome (77). These polycistronic transcripts are processed into gene-sized mRNAs by *trans*-splicing: The spliceosome initiates two *trans*-esterification reactions that cleave the genes from one another and add a short spliced-leader (SL) to the 5′ end of each gene (66, 118). Consequently, there is little or no control of expression at the level of transcription in trypanosomes, which is overcome by more strict control of transcript degradation and by genome organizational features (59). Other euglenozoans also appear to cap mRNAs with a SL (31, 34, 62), but apparently these systems have not reached the same extremes as those of trypanosomes.

NUCLEOMORPH GENE EXPRESSION

Nucleomorphs are vestigial nuclei arising from eukaryote-eukaryote endosymbioses (84). They are found in cryptophyte and chlorarachniophyte algae, originating from red and green algae, respectively. Despite their independent origins, nucleomorph genomes show remarkable convergence in architecture: They are reduced in size (~0.33–1.00 Mb) and composed of three linear chromosomes with a few hundred genes. Cryptophyte nucleomorphs have few or no introns (63), whereas chlorarachniophyte nucleomorphs contain hundreds of miniature spliceosomal introns (38, 103, 122). Transcriptomic data suggest that gene expression in nucleomorphs is a hotbed for genetic eccentricities and can be messy and inefficient (128).

Widespread use of SL *trans*-splicing also occurs in dinoflagellates (132). Like in trypanosomes, these systems have evolved polycistronic transcription for at least some genes, but the situation in dinoflagellates is poorly understood because their genomes are so large. The existence of SL *trans*-splicing likely preconditions expression systems for the emergence of polycistronic transcription because it allows for efficient and effective processing of multigene transcripts into functional mRNAs (71), although the presence of one by no means necessitates the presence of the other.

Juggling Gene Expression with Two Distinct Nuclear Genomes

The oddities of nuclear gene expression are mostly poor reflections of their organelle counterparts. Protist nuclear genomes do, however, exhibit some eccentricities not found in mitochondria or plastids. Ciliates, for instance, have two distinct nuclear genomes: a dormant germline micronucleus (MIC) and a transcriptionally active somatic macronucleus (MAC) (92). The MIC genome architecture resembles a canonical nuclear genome in that it comprises a modest number of large chromosomes containing thousands of genes interspersed with long stretches of noncoding DNA (16). But MIC genes are mostly transcriptionally silent and are interrupted by nonintronic noncoding sequences called internal eliminated sequences (IES), which generally disrupt the reading frame (16). The MAC genome, on the other hand, has functional genes, but they are arranged on thousands of tiny, multicopy chromosomes (1, 119). One of the most extreme examples of this occurs in *Oxytricha trifallax*, where there is essentially one gene per MAC chromosome (119). The MAC is generated after sexual conjugation from a copy of the MIC, so functional genes must be generated by the programmed elimination of IES DNA, massive genomic reorganization, and the shattering and amplification of the chromosomes into small, multicopy pieces (87).

Functional MAC genes are reconstructed in ciliates epigenetically, but the mechanisms used to accomplish this can differ among species (20, 87, 101). In *Tetrahymena thermophila* and *Paramecium tetraurelia*, short noncoding RNAs are transcribed from across the entire MIC genome and transported to the parental MAC, where they can then bind to the pool of transcripts generated from the whole genome. The short RNAs that fail to bind to the parental MAC transcripts are sequestered to the developing MAC, where they single out genomic DNA (corresponding to IES) for excision and deletion, thus epigenetically passing along the genome content of the MAC (20, 101).

O. trifallax also employs RNA to epigenetically recreate the MAC genome, but the strategy differs from those of *T. thermophila* and *P. tetraurelia*. Both short and long noncoding RNAs are generated from the *O. trifallax* parental MAC genome, and they are transported to the developing MAC, where they target sequences not for deletion but for retention (28). These noncoding RNAs not only help identify and remove IES but also provide information for gene rearrangements, resulting in the unscrambling of MIC coding segments into coherent MAC gene sequences (88). This type of unscrambling mechanism would not be possible using only short RNAs, such as those found in *T. thermophila* and *P. tetraurelia*. Despite their differences, all of these systems use mechanisms derived from transposon defense to perform the actual elimination and rearrangement of DNA (14, 20, 120). Indeed, in *P. tetraurelia* some of the IES are clearly homologous to transposons (3).

Although the MIC genome has typically been considered transcriptionally inactive (except during sexual reproduction), this now appears to be an oversimplification: a number of germline-restricted protein-coding genes reside within IES and are not transmitted to the MAC but are clearly expressed and regulated (16). Further departures from conventional ciliate nuclear gene expression (12, 36) are now showing how truly dynamic the storage and decoding of genetic information can be.

MOLECULAR RUBE GOLDBERG MACHINES

Even when we begin to see how systems like the ciliate nuclear dualism have evolved, the question remains: Why did they evolve? Biologists have a propensity for seeing the world in an adaptive light, perhaps because at the levels of organization that we are used to dealing with, this is mostly reasonable. But there is increasing evidence that complexity and variation at deeper levels in the molecular world may also be a consequence of nonadaptive processes (42, 72, 73). Seeing the hand of natural selection in shaping the spear-like snout of a marlin or the streamlined feathers of a hawk is easy; explaining the origins of a noncanonical genetic code or jumbled rRNA is not so straightforward.

What advantage can there be to taking an intact, unadorned coding region and shattering it into dozens of unordered and independently transcribed pieces, only to undo all of this encryption and stitch the segments back together again at the RNA or amino acid level? What benefit can come from the addition of an intron into an intron that is already inside of an intron? Or from becoming dependent on an extensive and costly RNA-editing infrastructure that returns mRNAs to their ancestral state? Why alter a genetic code that has existed since the last universal common ancestor?

The evolution of unconventional and elaborate modes of gene expression has puzzled researchers for decades (37) and divided them along adaptive versus nonadaptive lines (72). For certain genetic eccentricities, there are step-by-step hypotheses explaining their evolution. There are multiple, well-regarded neutral hypotheses for reassigning codons in the genetic code (73, 89). But there are also adaptive alternatives, such as the genome-streamlining hypothesis, which argues that variation in the code is the result of selection for a reduced translational apparatus (2). Similarly, a variety of adaptationist explanations for RNA editing have been suggested, such as gene regulation, generating genetic variation, and/or mutational buffering (29, 40, 53). The existence of the MAC nuclear genome and its associated RNA infrastructure in *O. trifallax* has been suggested to provide an adaptive store of heritable variation (in addition to the MIC nucleus), which contributes to the evolutionary success of ciliates (87). The appearance and proliferation of introns has long been considered an adaptation for generating organismal complexity by exon shuffling and alternative splicing (37, 97).

These and other adaptationist arguments for genetic complexity might be valid, but most remain completely untested, several are inconsistent with vital aspects of the system, and in many cases cause is mistaken for effect. For example, suggesting that RNA editing evolved to buffer against deleterious mutations implies that the editing apparatus, which is tailored for specific sites, was put in place after the deleterious mutation(s) that it now corrects. However, this requires the deleterious mutation to persist in a population during the (probably lengthy) time required to evolve the elaborately complex and expensive compensatory mechanism. An alternative view is that the RNA editing apparatus already existed, potentially serving a different purpose altogether, and was ultimately fixed within the cell through fortuitous events (19, 42). Once the editing machinery for a single edit is fixed in the population (neutrally), the conditions are set for additional editing sites to arise more easily. Indeed, RNA editing in trypanosomes illustrates this most effectively because the gRNAs encode the ancestral sequences, so based on parsimony they should have existed before the mutations they now edit.

More and more attention is being given to nonadaptive models for the origins of cellular and molecular complexity (42, 68, 73). One general model is called constructive neutral evolution (CNE) (19, 42, 117): a ratchet-like process whereby neutral (or slightly deleterious) mutations that result in increased complexity are fixed by random processes, such as genetic drift. The ratchet in this model comes from the idea that the gains in complexity that are fixed neutrally are not easily

unfixed by random events. Although criticized by some (58; but see 114), CNE has been used to explain everything from the origins of RNA editing and intron splicing (19, 117) to the complexity of ribosomes and mitochondrial respiratory complexes (23, 69). The ratcheting of neutral mutations has also been invoked for major transitions in evolution (81), such as the shift from a unicellular to a multicellular existence (68). In essence, CNE and other nonadaptive evolutionary models, such as the mutational hazard hypothesis (73), argue that biological complexity is not necessarily driven by fine-tuning or sophistication, but instead can be the consequence of runaway bureaucracy—"biological parallels of nonsensically complex Rube Goldberg machines that are overengineered to perform a single task" (42).

One of the challenges of evaluating the different models for the evolution of molecular complexity in protists is a lack of population-genetics data. We presently know very little about the effective population sizes and mutation rates of protists. There is also a paucity of information on DNA repair and recombination. We do not even know if most microbial eukaryotes have sex, although most likely do (41, 115), and when sex is characterized, it can be complicated: Some species have many different mating types (13, 123).

Data on protist population dynamics, mutation, recombination, and DNA maintenance are fundamental to understanding the underlying processes shaping their genomes and gene expression systems. One way to gain insights into these processes is by studying genetic diversity within and among populations. Interesting relationships between genomic/transcriptional complexity and genetic diversity have already been observed in protists (73, 74, 111); and, overall, species with a propensity for genetic embellishments, such as RNA editing, have been shown to harbor low levels of silent-site nucleotide diversity, suggesting that they have small effective population sizes (74, 111). Other studies have found that inadequacies or idiosyncrasies in DNA maintenance processes can contribute to increased genomic complexity (110). These are based on a few relatively well-studied taxa, but population-genetics and mutational studies of protists will only get easier as massively parallel sequencing methods improve and our databases grow accordingly.

A FUTURE PAVED WITH RNA-SEQ, BUT LOOK OUT FOR POTHOLES

Besides simplifying and streamlining genomics and transcriptomics, next-generation sequencing (NGS) has provided a huge reserve of unexplored data. Vast quantities of raw sequencing data from diverse eukaryotic groups have been deposited in freely accessible repositories, such as GenBank's Sequence Read Archive (SRA) (**http://www.ncbi.nlm.nih.gov/sra**). As of early September 2015, the SRA contains $\sim 2 \times 10^{15}$ open-access NGS bases, including ~700,000 datasets from eukaryotes, a large fraction of which are RNA-seq. Only about 5% of these datasets come from eukaryotes that are not animals, land plants, or fungi, but that still leaves thousands of NGS projects from protists, including ones for closely related species or distinct isolates of the same species, many of which are transcriptomic surveys (55).

These transcriptomic data are obviously a valuable resource for studying nuclear gene expression, but are data from organelle genomes as readily available? The high copy number per cell and elevated expression levels of organelle DNAs mean their transcripts are well-represented in eukaryotic RNA-seq experiments (94, 106), and because mitochondrial and plastid intergenic regions are often transcribed (4, 93), near-complete organelle genome assemblies are available (67, 95). Consequently, the SRA contains billions of unanalyzed reads from hundreds of microbial eukaryotes and includes both nuclear and organelle sequences, making it an excellent, untapped resource for investigating unusual and poorly understood forms of gene expression.

But relying too heavily on bulk sequence data for the identification and characterization of novel gene expression systems is rarely sufficient for understanding how complex genomes function.

Accurate and detailed characterizations of gene expression tend to involve a combination of massive sequencing with more demanding methods. But it is these technically challenging techniques that are sometimes lacking from contemporary transcriptome research, particularly that focusing on organelle genetics (108). High-throughput sequencing and bioinformatics have removed barriers for identifying where potentially complex expression systems might exist and understanding how they might operate, but additional information is required to really understand the machinery involved in posttranscriptional processing, modes of transcription, and the underlying DNA/RNA maintenance processes.

Moving forward, studies of protist expression systems need to combine NGS with molecular-biology-focused methods. This in combination with data on the population genetics and mutation rates, as well as a more unified understanding of cytonuclear interactions and coevolution (105) will indeed lead to an exciting synthesis.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by Discovery Grants to D.R.S. and P.J.K. from the Natural Sciences and Engineering Research Council (NSERC) of Canada. P.J.K. is a fellow of the Canadian Institute for Advanced Research.

LITERATURE CITED

1. Aeschlimann SH, Jönsson F, Postberg J, Stover NA, Petera RL, et al. 2014. The draft assembly of the radically organized *Stylonychia lemnae* macronuclear genome. *Genome Biol. Evol.* 6:1707–23
2. Andersson SGE, Kurland CG. 1995. Genomic evolution drives the evolution of the translation system. *Biochem. Cell Bio.* 73:775–87
3. Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury JM, et al. 2012. The *Paramecium* germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. *PLOS Genet.* 8:e1002984
4. Barbrook AC, Dorrell RG, Burrows J, Plenderleith LJ, Nisbet RER, et al. 2012. Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*. *Plant Mol. Biol.* 79:347–57
5. Barbrook AC, Howe CJ, Kurniawan DP, Tarr SJ. 2010. Organization and expression of organellar genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365:785–97
6. Boer P, Gray MW. 1988. Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55:399–411
7. Boersema PJ, Kahraman A, Picotti P. 2015. Proteomics beyond large-scale protein expression analysis. *Curr. Opin. Biotechnol.* 34:162–70
8. Burki F. 2014. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* 6:a016147
9. Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, et al. 2013. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in rhizaria. *Curr. Biol.* 23:1541–47
10. Carnes J, Trotter JR, Peltan A, Fleck M, Stuart K. 2008. RNA editing in *Trypanosoma brucei* requires three different editosomes. *Mol. Cell. Biol.* 28:122–30
11. Caron DA, Worden AZ, Countway PD, Demir E, Heidelberg KB. 2009. Protists are microbes too: a perspective. *ISME J.* 3:4–12

12. Carradec Q, Götz U, Arnaiz O, Pouch J, Simon M, et al. 2015. Primary and secondary siRNA synthesis triggered by RNAs from food bacteria in the ciliate *Paramecium tetraurelia*. *Nucleic Acids Res.* 43:1818–33
13. Cervantes MD, Hamilton EP, Xiong J, Lawson MJ, Yuan D, et al. 2013. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. *PLOS Biol.* 11:e1001518
14. Chalker DL, Yao MC. 2011. DNA elimination in ciliates: Transposon domestication and genome surveillance. *Annu. Rev. Genet.* 45:227–46
15. Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, et al. 1987. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* 238:363–66
16. Chen X, Bracht JR, Goldman AD, Dolzhenko E, Clay DM, et al. 2014. The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell* 158:1187–98
17. Cocquyt E, Verbruggen H, Leliaert F, De Clerck O. 2010. Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Mol. Biol. Evol.* 27:2052–61
18. Corell RA, Feagin JE, Riley GR, Strickland T, Guderian JA, et al. 1993. *Trypanosoma brucei* minicircles encode multiple guide RNAs which can direct editing of extensively overlapping sequences. *Nucleic Acids Res.* 21:4313–20
19. Covello P, Gray M. 1993. On the evolution of RNA editing. *Trends Genet.* 9:265–68
20. Coyne RS, Lhuillier-Akakpo M, Duharcourt S. 2012. RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence? *Biol. Cell* 104:309–25
21. Dang Y, Green BR. 2009. Substitutional editing of *Heterocapsa triquetra* chloroplast transcripts and a folding model for its divergent chloroplast 16S rRNA. *Gene* 442:73–80
22. del Campo J, Sieracki ME, Molestina R, Keeling P, Massana R, et al. 2014. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* 29:252–59
23. Doolittle WF. 2012. Evolutionary biology: a ratchet for protein complexity. *Nature* 481:270–71
24. Dorrell RG, Howe CJ. 2012. Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. *PNAS* 109:18879–84
25. Drager RG, Hallick RB. 1993. A complex twintron is excised as four individual introns. *Nucleic Acids Res.* 21:2389–94
26. Edqvist J, Burger G, Gray MW. 2000. Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J. Mol. Biol.* 297:381–93
27. Fan J, Schnare MN, Lee RW. 2003. Characterization of fragmented mitochondrial ribosomal RNAs of the colorless green alga *Polytomella parva*. *Nucleic Acids Res.* 31:769–78
28. Fang W, Wang X, Bracht JR, Nowacki M, Landweber LF. 2012. Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell* 151:1243–55
29. Farajollahi S, Maas S. 2010. Molecular diversity through RNA editing: A balancing act. *Trends Genet.* 26:221–30
30. Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53:413–422
31. Feagin JE, Harrell MI, Lee JC, Coe KJ, Sands BH, et al. 2012. The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. *PLOS ONE* 7:e38320
32. Flegontov P, Gray MW, Burger G, Lukeš J. 2011. Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? *Curr. Genet.* 57:225–32
33. Flipphi M, Fekete E, Ág N, Scazzocchio C, Karaffa L. 2013. Spliceosome twin introns in fungal nuclear transcripts. *Fungal Genet. Biol.* 57:48–57
34. Frantz C, Ebel C, Paulus F, Imbault P. 2000. Characterization of *trans*-splicing in Euglenoids. *Curr. Genet.* 37:349–55
35. Fučíková K, Lewis PO, González-Halphen D, Lewis LA. 2014. Gene arrangement convergence, diverse intron content, and genetic code modifications in mitochondrial genomes of Sphaeropleales (Chlorophyta). *Genome Biol. Evol.* 6:2170–80
36. Gao F, Roy SW, Katz LA. 2015. Analyses of alternatively processed genes in ciliates provide insights into the origins of scrambled genomes and may provide a mechanism for speciation. *mBio* 6:e01998-14
37. Gilbert W. 1978. Why genes in pieces? *Nature* 271:501
38. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *PNAS* 103:9566–71

39. Glanz S, Kück U. 2009. *Trans*-splicing of organelle introns—a detour to continuous RNAs. *Bioessays* 31:921–34
40. Gommans WM, Mullen SP, Maas S. 2009. RNA editing: a driving force for adaptive evolution? *Bioessays* 31:1137–45
41. Goodenough U, Heitman J. 2014. Origins of eukaryotic sexual reproduction. *Cold Spring Harb. Perspect. Biol.* 6:a016154
42. Gray MW, Lukeš J, Archibald JM, Keeling PJ, Doolittle WF. 2010. Irremediable complexity? *Science* 330:920–21
43. Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, et al. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21:3537–44
44. Heinonen TY, Schnare MN, Young PG, Gray MW. 1987. Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. Biol. Chem.* 262:2879–87
45. Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philos. Trans. R. Soc. B* 365:713–27
46. Horowitz S, Gorovsky MA. 1985. An unusual genetic code in nuclear genes of *Tetrahymena*. *PNAS* 82:2452–55
47. Jackson CJ, Norman JE, Schnare MN, Gray MW, Keeling PJ, et al. 2007. Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol.* 5:41
48. Jackson CJ, Waller RF. 2013. A widespread and unusual RNA *trans*-splicing type in dinoflagellate mitochondria. *PLOS ONE* 8:e56777
49. Janouškovec J, Sobotka R, Lai DH, Flegontov P, Koník P, et al. 2013. Split photosystem protein, linear-mapping topology and growth of structural complexity in the plastid genome of *Chromera velia*. *Mol. Biol. Evol.* 30:2447–62
50. Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, et al. 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *PNAS* 112:10200–7
51. Jedelský PL, Doležal P, Rada P, Pyrih J, Smíd O, et al. 2011. The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PLOS ONE* 6:e17285
52. Ji YE, Mericle BL, Rehkopf DH, Anderson JD, Feagin JE. 1996. The *Plasmodium falciparum* 6 kb element is polycistronically transcribed. *Mol. Biochem. Parasitol.* 81:211–23
53. Jobson RW, Qiu YL. 2008. Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biol. Dir.* 3:43
54. Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T. 2011. Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated *trans*- splicing. *Curr. Biol.* 21:311–315
55. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, et al. 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol.* 12:e1001889
56. Keeling PJ, Doolittle WF. 1996. A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.* 15:2285–90
57. Keeling PJ, Leander BS. 2003. Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J. Mol. Biol.* 326:1337–49
58. Keeling PJ, Leander BS, Lukeš J. 2010. Reply to Speijer: Does complexity necessarily arise from selective advantage? *PNAS* 107:E26
59. Kelly S, Kramer S, Schwede A, Maini PK, Gull K, et al. 2012. Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open Biol.* 2:120033
60. Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: Evolvability of the genetic code. *Nat. Rev. Genet.* 2:49–58
61. Kolisko M, Boscaro V, Burki F, Lynn DH, Keeling PJ. 2014. Single-cell transcriptomics for microbial eukaryotes. *Curr. Biol.* 24:R1081–82
62. Kuo RC, Zhang H, Zhuang Y, Hannick L, Lin S. 2013. Transcriptomic study reveals widespread spliced leader *trans*-splicing, short 5′-UTRs and potential complex carbon fixation mechanisms in the euglenoid alga *Eutreptiella* sp. *PLOS ONE* 8:e60826

63. Lane CE, van den Heuvel K, Kozera C, Curtis BA, Parsons BJ, et al. 2007. Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *PNAS* 104:19908–13
64. Lang BF, Jakubkova M, Hegedusova E, Daoud R, Forget L, et al. 2014. Massive programmed translational jumping in mitochondria. *PNAS* 111:5926–31
65. Lang-Unnasch N, Aiello DP. 1999. Sequence evidence for an altered genetic code in the *Neospora caninum* plastid. *Int. J. Parasitol.* 29:1557–62
66. Lasda EL, Blumenthal T. 2011. *Trans*-splicing. *Wiley Interdiscip. Rev. RNA* 2:417–34
67. Le P, Fisher PR, Barth C. 2009. Transcription of the *Dictyostelium discoideum* mitochondrial genome occurs from a single initiation site. *RNA* 15:2321–30
68. Libby E, Ratcliff WC. 2014. Ratcheting the evolution of multicellularity. *Science* 346:426–27
69. Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. 2011. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* 63:528–37
70. Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, et al. 2002. Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell* 4:495–502
71. Lukeš J, Leander BS, Keeling PJ. 2009. Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. *PNAS* 106:9963–70
72. Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. *PNAS* 104:8597–604
73. Lynch M. 2007. *The Origins of Genome Architecture*. Sunderland, MA: Sinauer
74. Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–4
75. Marande W, Burger G. 2007. Mitochondrial DNA as a genomic jigsaw puzzle. *Science* 318:415
76. Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, et al. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *PNAS* 109:E317–25
77. Martínez-Calvillo S, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela RG, Figueroa-Angulo EE. 2010. Gene expression in trypanosomatid parasites. *J. Biomed. Biotechnol.* 2010:525241
78. Maruyama S, Sugahara J, Kanai A, Nozaki H. 2010. Permuted tRNA genes in the nuclear and nucleomorph genomes of photosynthetic eukaryotes. *Mol. Biol. Evol.* 27:1070–76
79. Masuda I, Matsuzaki M, Kita K. 2010. Extensive frameshift at all AGG and CCC codons in the mitochondrial cytochrome c oxidase subunit 1 gene of *Perkinsus marinus* (Alveolata; Dinoflagellata). *Nucleic Acids Res.* 38:6186–94
80. Matsumoto T, Ishikawa SA, Hashimoto T, Inagaki Y. 2011. A deviant genetic code in the green alga-derived plastid in the dinoflagellate *Lepidodinium chlorophorum*. *Mol. Phylogenet. Evol.* 60:68–72
81. Maynard Smith JM, Szathmary E. 1995. *The Major Transitions in Evolution*. Oxford, UK: Oxford Univ. Press
82. McFadden GI, Reith ME, Munholland J, Lang-Unnasch N. 1996. Plastid in human parasites. *Nature* 381:482
83. Metzker ML. 2010. Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11:31–46
84. Moore CE, Archibald JM. 2009. Nucleomorph genomes. *Ann. Rev. Genet.* 43:251–64
85. Moreira S, Breton S, Burger G. 2012. Unscrambling genetic information at the RNA level. *Wiley Interdiscip. Rev. RNA* 3:213–28
86. Mungpakdee S, Shinzato C, Takeuchi T, Kawashima T, Koyanagi R, et al. 2014. Massive gene transfer and extensive RNA editing of a symbiotic dinoflagellate plastid genome. *Genome Biol. Evol.* 6:1408–22
87. Nowacki M, Shetty K, Landweber LF. 2011. RNA-mediated epigenetic programming of genome rearrangements. *Annu. Rev. Genomics Hum. Genet.* 12:367–89
88. Nowacki M, Vijayan V, Zhou Y, Schotanus K, Doak TG, et al. 2008. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* 451:153–58
89. Osawa S, Jukes TH. 1989. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.* 28:271–78
90. Palmer JD. 1997. Organelle genomes: going, going, gone! *Science* 275:790–91
91. Pawlowski J. 2013. The new micro-kingdoms of eukaryotes. *BMC Biol.* 11:40
92. Prescott DM. 1994. The DNA of ciliated protozoa. *Microbiol. Rev.* 58:233–67

93. Rackham O, Shearwood AMJ, Mercer TR, Davies SM, Mattick JS, et al. 2011. Long noncoding RNAs are generated from the mitochondrial genome and regulated by nuclear-encoded proteins. *RNA* 17:2085–93
94. Raz T, Kapranov P, Lipson D, Letovsky S, Milos PM, et al. 2011. Protocol dependence of sequencing-based gene expression measurements. *PLOS ONE* 6:e19287
95. Rehkopf DH, Gillespie DE, Harrell MI, Feagin JE. 2000. Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. *Mol. Biochem. Parasitol.* 105:91–103
96. Richardson E, Dorrell RG, Howe CJ. 2014. Genome-wide transcript profiling reveals the coevolution of plastid gene sequences and transcript processing pathways in the fucoxanthin dinoflagellate *Karlodinium veneficum*. *Mol. Biol. Evol.* 31:2376–86
97. Rogozin IB, Carmel L, Csuros M, Koonin EV. 2012. Origin and evolution of spliceosomal introns. *Biol. Dir.* 7:6150–57
98. Roy SW, Hudson AJ, Joseph J, Yee J, Russell AG. 2012. Numerous fragmented spliceosomal introns, AT-AC splicing, and an unusual dynein gene expression pathway in *Giardia lamblia*. *Mol. Biol. Evol.* 29:43–49
99. Schmidt M, Geßner G, Luff M, Heiland I, Wagner V, et al. 2006. Proteomic analysis of the eyespot of *Chlamydomonas reinhardtii* provides novel insights into its components and tactic movements. *Plant Cell* 18:1908–30
100. Schnare MN, Gray MW. 1990. Sixteen discrete RNA components in the cytoplasmic ribosome of *Euglena gracilis*. *J. Mol. Biol.* 215:73–83
101. Schoeberl UE, Mochizuki K. 2011. Keeping the soma free of transposons: programmed DNA elimination in ciliates. *J. Biol. Chem.* 286:37045–52
102. Simpson L, Shaw J. 1989. RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* 57:355–66
103. Slamovits CH, Keeling PJ. 2009. Evolution of ultra-small spliceosomal introns in highly reduced nuclear genomes. *Mol. Biol. Evol.* 26:1699–705
104. 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.* 372:356–68
105. Sloan DB. 2015. Using plants to elucidate the mechanisms of cytonuclear co-evolution. *N. Phytol.* 205:1040–46
106. Smith DR. 2013. RNA-Seq data: a goldmine for organelle research. *Brief. Funct. Genomics* 12:454–56
107. Smith DR. 2015. Buying in to bioinformatics: an introduction to commercial sequence analysis software. *Brief. Bioinform.* 16:700–9
108. Smith DR. 2016. The past, present and future of mitochondrial genomics: Have we sequenced enough mtDNAs? *Brief. Funct. Genomics* 15:47–54
109. Smith DR, Asmail SR. 2014. Next-generation sequencing data suggest that certain nonphotosynthetic green plants have lost their plastid genomes. *N. Phytol.* 204:7–11
110. Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes but significant differences at the extremes. *PNAS* 11:10177–84
111. Smith DR, Lee RW. 2010. Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Mol. Biol. Evol.* 27:2244–56
112. Smith DR, Lee RW. 2014. A plastid without a genome: Evidence from the nonphotosynthetic green alga *Polytomella*. *Plant Phys.* 164:1812–19
113. Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, et al. 2007. Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. *Science* 318:450–53
114. Speijer D. 2010. Constructive neutral evolution cannot explain current kinetoplastid panediting patterns. *PNAS* 107:E25
115. Speijer D, Lukeš J, Eliáš M. 2015. Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *PNAS* 112:8827–34
116. Spencer DF, Gray MW. 2011. Ribosomal RNA genes in *Euglena gracilis* mitochondrial DNA: fragmented genes in a seemingly fragmented genome. *Mol. Genet. Genomics* 285:19–31
117. Stoltzfus A. 1999. On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49:169–81

118. Sutton RE, Boothroyd JC. 1986. Evidence for *trans* splicing in trypanosomes. *Cell* 47:527–35
119. Swart EC, Bracht JR, Magrini V, Minx P, Chen X, et al. 2013. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLOS Biol*. 11:e1001473
120. Swart EC, Nowacki M. 2015. The eukaryotic way to defend and edit genomes by sRNA-targeted DNA deletion. *Ann. N. Y. Acad. Sci.* 1341:106–14
121. Tablizo FA, Lluisma AO. 2014. The mitochondrial genome of the red alga *Kappaphycus striatus* ("Green Sacol" variety): complete nucleotide sequence, genome structure and organization, and comparative analysis. *Mar. Genomics* 18:155–61
122. Tanifuji G, Onodera NT, Brown MW, Curtis BA, Roger AJ, et al. 2014. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. *BMC Genomics* 15:374
123. Umen JG. 2013. Genetics: swinging ciliates' seven sexes. *Curr. Biol.* 23:R475–77
124. Valach M, Moreira S, Kiethega GN, Burger G. 2014. *Trans*-splicing and RNA editing of LSU rRNA in *Diplonema* mitochondria. *Nucleic Acids Res.* 42:2660–72
125. Vlcek C, Marande W, Teijeiro S, Lukeš J, Burger G. 2011. Systematically fragmented genes in a multipartite mitochondrial genome. *Nucleic Acids Res.* 39:979–88
126. Waller RF, Jackson CJ. 2009. Dinoflagellate mitochondrial genomes: Stretching the rules of molecular biology. *Bioessays* 31:237–45
127. White TC, Rudenko G, Borst P. 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. *Nucleic Acids Res.* 14:9471–89
128. Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *PNAS* 102:10936–41
129. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, et al. 2015. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347:1257594
130. Záhonová K, Hadariová L, Vacula R, Yurchenko V, Eliáš M, et al. 2014. A small portion of plastid transcripts is polyadenylated in the flagellate *Euglena gracilis*. *FEBS Lett.* 588:783–88
131. Zhang H, Campbell DA, Sturm NR, Dungan CF, Lin S. 2011. Spliced leader RNAs, mitochondrial gene frameshifts and multi-protein phylogeny expand support for the genus *Perkinsus* as a unique group of alveolates. *PLOS ONE* 6:e19933
132. Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, et al. 2007. Spliced leader RNA *trans*-splicing in dinoflagellates. *PNAS* 104:4618–23
133. Zhang Z, Green BR, Cavalier-Smith T. 1999. Single gene circles in dinoflagellate chloroplast genomes. *Nature* 400:155–59