

5S rRNA Gene Arrangements in Protists: A Case of Nonadaptive Evolution

Guy Drouin · Corey Tsang

Received: 13 April 2012 / Accepted: 25 June 2012 / Published online: 11 July 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Given their high copy number and high level of expression, one might expect that both the sequence and organization of eukaryotic ribosomal RNA genes would be conserved during evolution. Although the organization of 18S, 5.8S and 28S ribosomal RNA genes is indeed relatively well conserved, that of 5S rRNA genes is much more variable. Here, we review the different types of 5S rRNA gene arrangements which have been observed in protists. This includes linkages to the other ribosomal RNA genes as well as linkages to ubiquitin, splice-leader, snRNA and tRNA genes. Mapping these linkages to independently derived phylogenies shows that these diverse linkages have repeatedly been gained and lost during evolution. This argues against such linkages being the primitive condition not only in protists but also in other eukaryote species. Because the only characteristic the diverse genes with which 5S rRNA genes are found linked with is that they are tandemly repeated, these arrangements are unlikely to provide any selective advantage. Rather, the observed high variability in 5S rRNA genes arrangements is likely the result of the fact that 5S rRNA genes contain internal promoters, that these genes are often transposed by diverse recombination mechanisms and that these new gene arrangements are rapidly homogenized by unequal cross-overs and/or by gene conversions events in species with short generation times and frequent founder events.

Keywords Concerted evolution · 5S ribosomal RNA · Ribosomal RNA · Trans-spliced leader · Ubiquitin · snRNA · tRNA

Introduction

Ribosomal RNAs are the most abundant RNA molecules found in eukaryotic cells. In yeast, they account for about 80 % of all RNA molecules, with tRNAs and mRNAs making up the remaining 15 and 5 %, respectively (Warner 1999). This overwhelming abundance is not only the result of their efficient transcription but is also a consequence of the fact that the genes coding for these RNAs are almost invariably found in numerous copies in eukaryotic genomes (recently reviewed in Torres-Machorro et al. 2010). Although most eukaryotic species have more than 100 copies of both 5S rRNA genes and 18S, 5.8S and 28S rRNA genes, the number of 5S rRNA genes varies from only three in *Plasmodium falciparum* to as many as a million in *Euplotes eurystomus*. The number of rDNA units, coding for 18S, 5.8S and 28S rRNA genes, also varies from as few as 2 in *Theileria parva* to 5000 in *Acetabularia mediterranea*. Finally, there is no relationship between the number of 5S rRNA genes and the number of rDNA units. Although some species, such as *Saccharomyces cerevisiae* with its 100–200 linked copies of both 5S rRNA genes and rDNA units, have equal amounts of both types of genes, other species have either much more 5S rRNA genes or rDNA units. For example, whereas *Trypanosoma brucei* has almost 30 times more 5S rRNA genes than rDNA units (1,500 vs. 56, respectively), *Tetrahymena thermophila* has 30 times more rDNA units than 5S rRNA genes in its macronucleus (9,000 vs. 150, respectively; Torres-Machorro et al. 2010).

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9512-5) contains supplementary material, which is available to authorized users.

G. Drouin (✉) · C. Tsang
Département de biologie et Centre de recherche avancée en
génomique environnementale, Université d'Ottawa, 30 Marie
Curie, Ottawa, ON K1N 6N5, Canada
e-mail: gdrouin@science.uottawa.ca

Given their high copy number and high level of expression, one might expect that both their sequence and organization would be conserved during evolution (Drummond et al. 2005). Although the sequences of rRNA genes are indeed known to be very highly conserved, and are thus often used to infer the phylogenetic relationships of distantly related species, their organization has been shown to be quite variable (Woese et al. 1990; Moreira et al. 2007; Torres-Machorro et al. 2010). As recently reviewed by Torres-Machorro et al. (2010), ribosomal RNA genes can be dispersed, organized in tandem repeats with or without other genes or located in linear or circular extrachromosomal units. They can also contain introns, insertion sequences, transposable elements, protein-coding genes or additional spacers (Eickbush and Eickbush 2007; Torres-Machorro et al. 2010).

Here, we focus on the organization of 5S rRNA genes in protist genomes. Because eukaryotic 5S rRNA genes are transcribed by RNA polymerase III, whereas the other three eukaryotic rRNA genes (coding for 18S, 5.8S and 28S rRNA; henceforth referred to as rDNA genes or units) are transcribed by RNA polymerase I, one would not expect that these two types of genes would be associated with one another (Archambault and Friesen 1993; Vannini and Cramer 2012). This is what is observed in most eukaryotic species where 5S rRNA genes are either found dispersed as independent copies throughout genomes or found in one or more clusters of tandemly repeated copies (Gerbi 1985; Drouin and Moniz de Sá 1995). However, as described in the following, 5S rRNA and rDNA genes have repeatedly been found to be linked with one another. Furthermore, 5S rRNA genes are also often found linked to genes transcribed by RNA polymerase II, as well as the rDNA are genes transcribed by RNA polymerase III. Here, we argue that these diverse gene arrangements do not provide any selective advantage but simply represent the nonadaptive fixation of haphazard recombination events in species with short generation times and frequent founder events.

Data

We used three methods to obtain data concerning the organization of 5S rRNA genes in protist species. We first used PubMed to search published studies and recorded this information. As described in the Supplemental Tables, these studies used a variety of methods to obtain these data, including Southern blot analyses, PCR amplifications with or without subsequent sequencing and genome sequence analyses. We also searched entries in NCBI’s GenBank database (<http://www.ncbi.nlm.nih.gov/>) for linkages using ‘5S ribosomal RNA’ and ‘5S RNA’ as query terms for searching through GenBank’s protist nucleotide sequence

database. Finally, we used BLAST with a variety of 5S rRNA sequences to search protist sequences in NCBI’s GenBank database (Altschul et al. 1997).

Types of Linkages

In protists, when 5S rRNA genes are found linked to rDNA genes, they are most commonly linked to rDNA units made up of 18S, 5.8S and 28S rRNA genes. Such linkages are observed in all protist groups so far studied, i.e., stramenopile, haptophyte, cryptophyte, metamonada, alveolate, euglenozoa and amoebzoa species (Figs. 1, 2, 3, 4

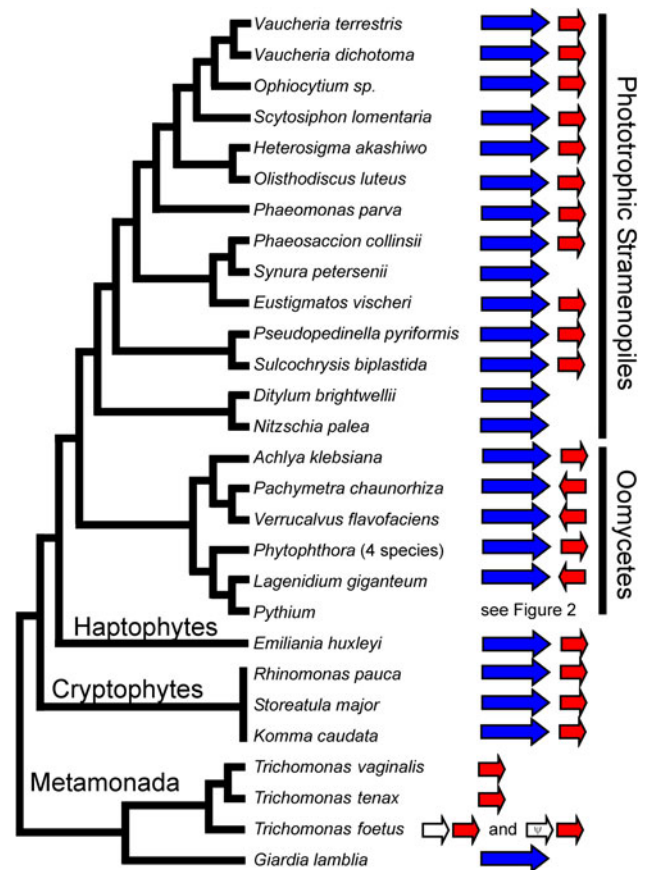


Fig. 1 Phylogenetic distribution of 5S genes arrangements in Stramenopile, Haptophyte, Cryptophyte and Metamonada species. *Blue arrows* represent rDNA units, *red arrows* represent 5S rRNA genes and *white arrows* indicate ubiquitin genes or pseudogenes (ψ). The presence of a *single arrow* represents genes that are tandemly repeated numerous times but not linked to any other genes. The presence of *two arrows* represents genes which are linked with one another to form a unit which is tandemly repeated numerous times. The *arrows* point in the direction of transcription. The presence of the word ‘and’ represents the fact that two types of units are found in this species. The references and accession numbers of the genes shown are listed in Supplemental Table 1. The phylogenetic relationships shown are based on the studies of Dawson and Pace (2002), Kleina et al. (2004), Baldauf (2008), Brown and Sorhannus (2010), Beakes et al. (2011) and Malik et al. (2011) (Color figure online)

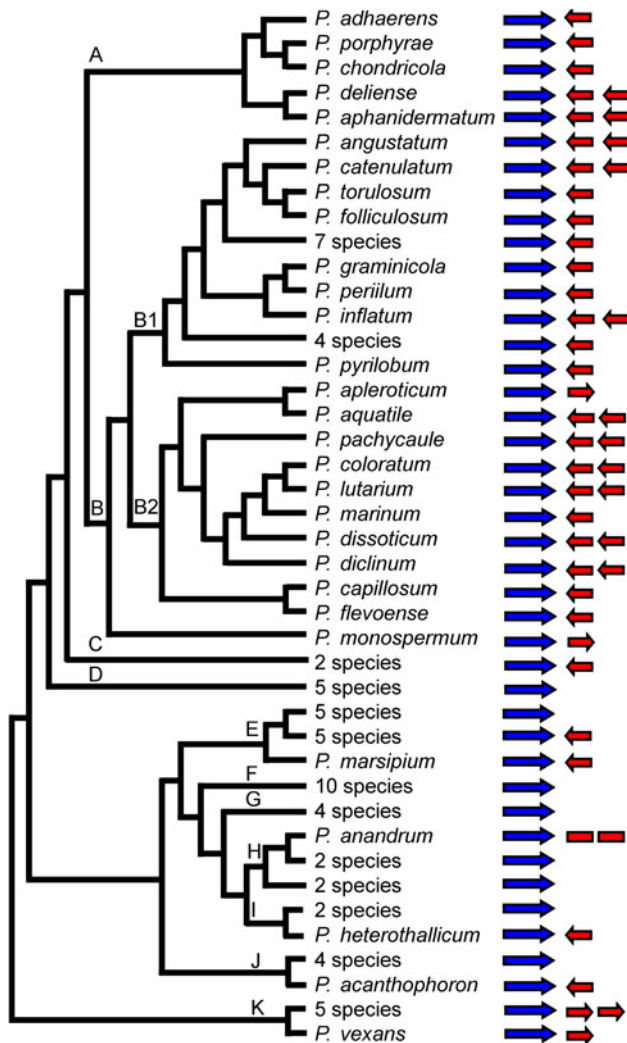


Fig. 2 Phylogenetic distribution of 5S genes arrangements in *Pythium* species. *Blue arrows* represent rDNA units and *red arrows* represent 5S rRNA genes. The presence of a *single arrow* represents genes that are tandemly repeated numerous times but not linked to any other genes. The presence of *two or three arrows* represents genes which are linked to form a unit which is tandemly repeated numerous times. The *arrows* point in the direction of transcription. *Red boxes* represent 5S rRNA genes for which the direction of transcription is not known. Note that most of these arrangements were deduced from PCR amplifications and that the two 5S rRNA genes found in some species might not both be functional (see Belkhiri and Klassen 1996). The gene arrangements are from Bedard et al. (2006) and the phylogenetic relationships shown are based on the study of Lévesque and de Cock (2004) (Color figure online)

and 5). In contrast, other 5S rRNA gene linkages are only observed in some protist taxa. For example, 5S rRNA gene linkages to spliced leader (SL) sequences have only been found in dinoflagellate and euglenozoa species and 5S rRNA gene linkages to tRNA genes have only been found in *Entamoeba* species (Figs. 3, 4 and 5). In extreme cases, some types of 5S rRNA gene linkages have only been found in single species. For example, 5S rRNA gene

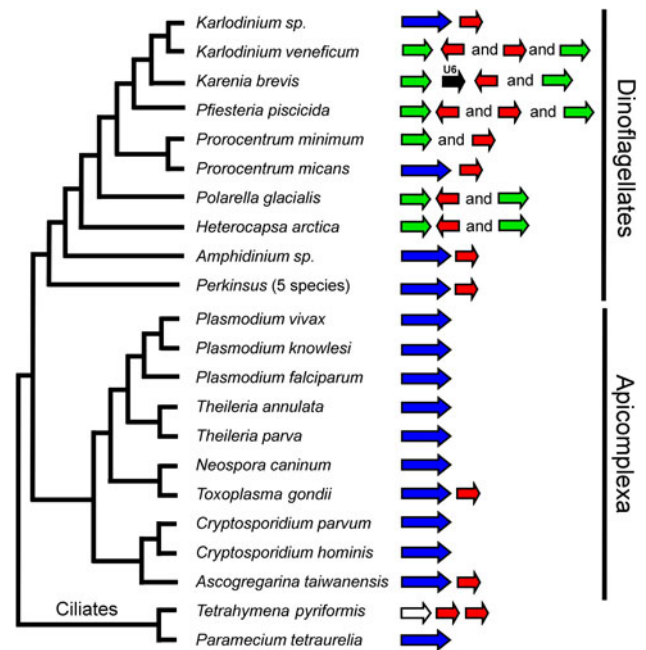


Fig. 3 Phylogenetic distribution of 5S genes arrangements in Alveolates (Ciliates, Apicomplexa and Dinoflagellates). *Blue arrows* represent rDNA units, *red arrows* represent 5S rRNA genes, *green arrows* represent spliced-leader genes, the *black arrow* represents U6 snRNA genes and the *white arrow* represents ubiquitin genes. The presence of a *single arrow* represents genes that are tandemly repeated numerous times but not linked to any other genes. The presence of *two or more arrows* represents genes which are linked to form a unit which is tandemly repeated numerous times. The *arrows* point in the direction of transcription. The presence of the word 'and' represents the fact that two types of units (or three types of units, when the word 'and' appears twice) are found in this species. The references and accession numbers of the genes shown are listed in Supplemental Table 2. The phylogenetic relationships shown are based on the work of Zhang et al. (2007), Baldauf (2008), Reeb et al. (2009), Janouškovec et al. (2010), Templeton et al. (2010), Bachvaroff et al. (2011) and Zhang et al. (2011) (Color figure online)

linkages to ubiquitin genes and pseudogenes have been found in only one of three *Trichomonas* species, the linkage of two 5S rRNA genes linked to ubiquitin genes has only been found in the ciliate species *Tetrahymena pyriformis*, 5S rRNA gene linkages to U6 snRNA genes and SL genes has only been found in the dinoflagellate species *Karenia brevis*, and 5S rRNA gene linkages to U5, U1, U2 snRNA genes and SL genes have only been found in the euglenid species *Entosiphon sulcatum* (Figs. 1, 3 and 4). In all cases, 5S rRNA genes are always found within spacer sequences and never within the other genes with which they are linked (Figs. 1, 2, 3, 4 and 5).

The first reports of 5S gene linkages were those of Rubin and Sulston (1973) and Maizels (1976) who showed that these genes were linked to rDNA units in *Saccharomyces cerevisiae* and *Dictyostelium discoideum*, respectively. Because these species were considered to be 'lower'

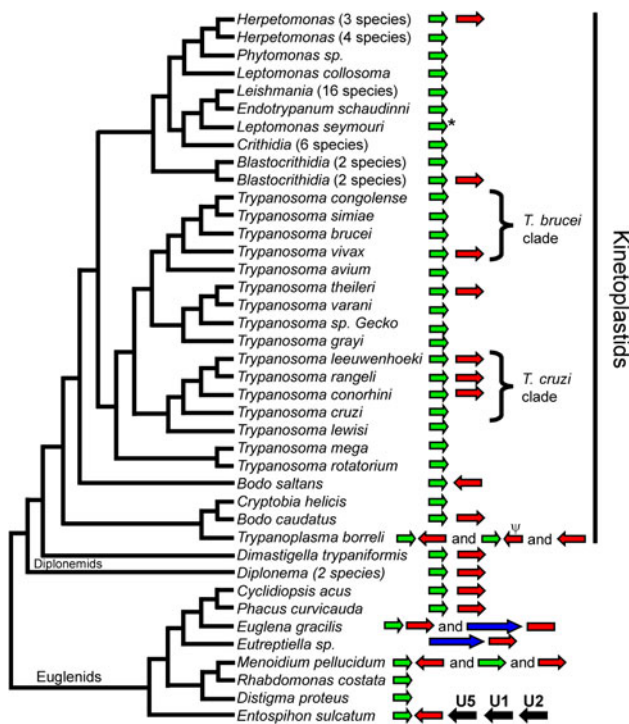


Fig. 4 Phylogenetic distribution of 5S genes arrangements in Euglenozoa (Kinetoplastids, Diplonemids and Euglenids) species. *Green arrows* represent spliced-leader sequences, *red arrows* represent 5S rRNA genes, *blue arrows* represent rDNA units and *black arrows* represent snRNA genes. The presence of a *single arrow* represents genes that are tandemly repeated numerous times but not linked to any other genes. The presence of *two or more arrows* represents genes which are linked to form a unit which is tandemly repeated numerous times. The *arrows* point in the direction of transcription. The presence of the word ‘and’ represents the fact that two types of units (or three types of units, when the word ‘and’ appears twice) are found in this species. *Red boxes* represent 5S rRNA genes for which the direction of transcription is not known and ψ indicates 5S pseudogenes. The start (*) next to the *L. seymouri* green arrow is to indicate that several undescribed isolates related to this species have been show to contain a 5S rRNA sequence within their splice-leader repeats (Westenberger et al. 2004). The references and accession numbers of the genes shown are listed in Supplemental Table 3. The phylogenetic relationships shown are based on the studies of Stevens et al. (1999), Frantz et al. (2000), Preisfeld et al. (2001), Hughes and Piontkivska (2003a, b), Hamilton et al. (2004), Moreira et al. (2004), Simpson and Roger (2004), von der Heyden et al. (2004), Baldauf (2008), D’Avila-Levy et al. (2009) and Deschamps et al. (2011) (Color figure online)

eukaryotes, and that their 5S rRNA gene arrangement was similar to that of eubacterial species (where 16S, 5S and 23S ribosomal RNAs are linked), this arrangement was at first interpreted as representing some sort of transitional stage between the linked eubacterial arrangement and the unlinked ‘higher’ eukaryote arrangement (Gerbi 1985; Drouin et al. 1987; Belkhirri et al. 1992; Howlett et al. 1992; Drouin and Moniz de Sá 1995). However, a large variety of 5S rRNA gene linkages have since been found and this interpretation is no longer tenable. As described earlier, in protists, this includes linkages to SL genes,

snRNA genes, ubiquitin genes and pseudogenes and tRNA genes. Furthermore, a large variety of 5S rRNA gene linkages are also found in multicellular eukaryotes. This includes linkages to ribosomal rRNA genes in land plant, nematode, spider and crustacean species, linkages to SL genes in nematode and hydrozoan species and linkages to histone and/or snRNA genes in fish, crustacean, insect and mollusk species (reviewed in Drouin and Moniz de Sá 1995; Pelliccia et al. 2001; Eirín-López et al. 2004; Cross and Rebordinos 2005; Manchado et al. 2006; Derelle et al. 2010; Cabral-de-Mello et al. 2011; Vierna et al. 2011; Wicke et al. 2011). The fact that 5S rRNA genes are sometimes found linked to a variety of other genes argues against the linkage of 5S rRNA genes and ribosomal RNA genes being the primitive condition not only in protists but also in other eukaryote species.

Variation of 5S rRNA Gene Linkages at Different Taxonomic Levels

The best way to study the evolution of 5S rRNA gene linkages is to map them to independently derived phylogenies (Drouin and Moniz de Sá 1995). These analyses show that 5S rRNA gene linkages are variable within large taxonomic groups, within genus and even within species (Figs. 1, 2, 3, 4 and 5).

In some large taxonomic groups, such as stramenopiles and apicomplexa (Figs. 1, 2 and 3), 5S rRNA genes are only found linked to rDNA units whereas in others, such as dinoflagellates and euglenids, they are found linked to a larger variety of other genes, such as SL and snRNA genes (Figs. 3 and 4). However, during evolution, the arrangement of 5S rRNA gene changes, even in taxonomic groups where 5S rRNA genes are only found linked to rDNA units. Whereas most autotrophic stramenopiles have 5S rRNA genes linked to rDNA units, three species, *S. petersenii*, *D. brightwellii* and *N. palea*, have lost this linkage (Fig. 1). Conversely, whereas most apicomplexa species do not have 5S rRNA genes linked to their rDNA units, the rDNA units of two species, *T. gondii* and *A. taiwanensis*, have independently acquired 5S rRNA genes (Fig. 3).

5S rRNA gene linkages are also variable within genus. Within the metamonada, 5S are not linked to other gene families in two Trichomonax species whereas they are linked to ubiquitin genes and pseudogenes in *T. foetus* (Fig. 1). In oomycetes, the well-studied *Pythium* genus shows that the 5S rRNA arrangement within rDNA units is highly variable even between closely related species (Fig. 2). This includes variation in the number of linked 5S rRNA genes, from zero to two, as well as variation in the orientation of these 5S rRNA genes. In dinoflagellates, one *Prorocentrum* species has its 5S rRNA genes linked to

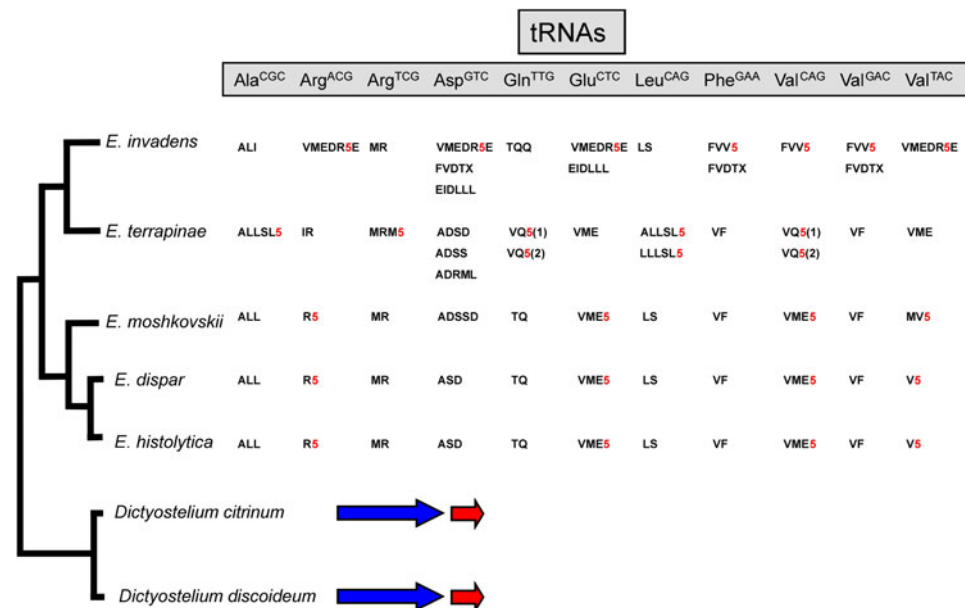


Fig. 5 Phylogenetic distribution of 5S genes arrangements in Amoebozoa species. For *Entamoeba* species, the arrangements of 5S genes relative to the tRNA tandemly repeated arrays are shown. In these arrays, each tRNA is represented using the single-letter amino acid code. 5S genes are represented by the number 5 shown in red. Numbers 1 and 2 in parentheses represent the fact that the VQ5 array is found in two different locations in the genome of *E. terrapinae*. X represents the presence of a gene encoding and unidentified small RNA common to all species. Each column represents the organization of a given tRNA genes in all five *Entamoeba* species. For example, 5S genes are not linked to Ala^{CGC} tRNA in any *Entamoeba* species except in *E. terrapinae*. Each row represents the linkages between 5S

and tRNA genes in each genome. For example, 5S genes are found in two different arrays in the *E. invadens* genome, i.e., those coding for VMEDR^{5E} and FW⁵. For simplicity, the orientation of the *Entamoeba* genes is not shown. For *Dictyostelium* species, the arrangements of 5S genes (red arrows) relative to the tandemly repeated ribosomal RNA genes (blue arrows) are shown. These two types of genes form a unit which is tandemly repeated numerous times in the genome of these two species. The arrows point in the direction of transcription. The references and accession numbers of the genes shown are listed in Supplemental Table 4. The phylogenetic relationships are from Clark et al. (2006) and Cole et al. (2010) (Color figure online)

rDNA units, whereas the other does not (Fig. 3). In kinetoplastids, closely related *Trypanosoma* species either have or do not have 5S rRNA genes linked to their SL genes (Fig. 4). In *Entamoeba* species, 5S rRNA genes are linked to different tRNA genes (Fig. 5).

The within species variation in 5S rRNA gene arrangement suggests that the linkage of 5S rRNA genes to other genes is a dynamic process. For example, the fact that some species, such as the dinoflagellate species *K. veneficum* and *P. piscicida* and the euglenid species *M. pellucidum*, have linked 5S rRNA and SL genes as well as independent clusters of tandemly repeated 5S rRNA genes and SL genes, suggests that neither the linked or unlinked arrangement has yet been fixed in these species but that both type of arrangements coexist. Such coexistence also supports the conclusion that presence or absence of such linkages has no functional purpose (see the following).

Mechanisms: Transposition and Homogenization

The numerous types of 5S rRNA gene linkages observed at different taxonomic levels clearly show that, in protists,

there is no such thing as a universal primitive 5S rRNA gene arrangement. Rather, 5S rRNA genes move in and out of a variety of tandemly repeated gene families. This raises three questions: Is it the 5S rRNA genes, or the other genes, which move around, how do they move around and how are they homogenized throughout the gene families where they transpose to?

Given the wide variety of genes with which 5S rRNA genes are found associated with, such as ribosomal RNA, ubiquitin, histone, SL, snRNA and tRNA genes, the most parsimonious possibility is that it is the 5S rRNA genes that are transposed to other gene families and not the opposite. Furthermore, at least two possible mechanisms have been proposed to explain how 5S rRNA genes can be transposed to other loci: by the insertion of extrachromosomal covalently closed circular DNAs (cccDNAs) and by RNA-mediated transposition (reviewed in Drouin and Moniz de Sá 1995). cccDNAs are thought to be formed by homologous recombination between adjacent repeated sequences and cccDNAs containing 5S rRNA genes have been observed in many eukaryotic species including plant, invertebrate and vertebrate species (reviewed in Cohen and Segal 2009; Cohen et al. 2010). Given that 5S rRNA genes

are often found as tandemly repeated sequences in diverse eukaryotic genomes, it is therefore likely that such molecules exist in most eukaryotic genomes.

As discussed in more details in Drouin and Moniz de Sá (1995), RNA-mediated transposition of 5S rRNA genes is also a likely transposition mechanism because their promoter sequences are located within their coding regions. Therefore, RNA-mediated transposed 5S rRNA genes can still be expressed because they contain the necessary promoter sequences. In fact, such expressed reverse-transcribed 5S rRNA gene sequences have been found dispersed in the mouse and rat genomes (Drouin 2000). The fact that these genes have been shown to be expressed suggests that they are not deleterious and that their dispersed arrangement does not preclude them from being expressed. The 5S rRNA gene of *T. vivax*, which is linked to SL genes, also has a 16-pb long poly-A track at its 3'-end which suggest that it transposed to the SL locus through an RNA intermediate (Fig. 4; Roditi 1992). Therefore, 5S rRNA molecules retrotranscribed in cDNA molecules by an endogenous reverse transcriptase and inserted back into the genome is a likely mechanism of 5S rRNA gene transposition. In contrast, most of the other genes with which 5S rRNA genes are found linked with, such as ribosomal RNA genes and protein-coding genes, would most likely not be functional if they were retrotransposed because, given that their promoter sequences are located upstream from their coding regions, they would lack their necessary promoter sequences (Drouin and Moniz de Sá 1995; Paule and White 2000; Lenhard et al. 2012). The fact that we know of no examples of linkages between rDNA genes, ubiquitin, histone and/or splice leader genes support the suggestion that such genes do not frequently move around genomes.

As discussed further in the following, the reason why 5S rRNA genes are found linked with a wide variety of genes is likely not a functional reason because 5S rRNA genes have no obvious functional relationship with most of these genes (i.e., apart from their relationship with the other ribosomal RNA genes). However, the genes with which 5S rRNA genes are found linked with do have a common feature: They are all gene families that are encoded by tandemly repeated genes. This immediately suggests that the mechanisms by which 5S rRNA genes move in and out of a variety of tandemly repeated gene families are those involved in the concerted evolution of gene families: gene conversion and unequal crossingover (Dover 1982; Arnheim 1983; Dover 1993; Elder and Turner 1995; Graur and Li 2000). In the case of tandemly repeated gene families, it is often difficult to determine which of these two mechanisms is the most important for their concerted evolution because all repeats are almost identical (Eickbush and Eickbush 2007). Furthermore, both mechanisms can alter

gene copy numbers between generations (e.g., Szostak and Wu 1980; Pukkila and Skrzynia 1993). A change in copy numbers cannot therefore be assumed to necessarily represent the effect of unequal crossingover events. For example, gene conversions have been shown to be responsible for the gain and loss of rDNA units in the yeast genome (Gangloff et al. 1996). Smith (1974), using computer simulations, calculated that a tandemly repeated gene family having n genes would require n^2 unequal crossingover events to be homogenized. The homogenization of the rDNA gene family like that of *Saccharomyces cerevisiae*, which contains some 100–200 tandemly repeated rDNA units at a single locus on chromosome 12, would therefore only require some 10,000–40,000 generations (Bergeron and Drouin 2008). This predicted number of generations was later experimentally confirmed by the experimental work of Szostak and Wu (1980) who determined that some 56,000 generations were required to homogenize the rDNA repeat units of *S. cerevisiae*. For yeasts, this could represent as little as 15 years because, under laboratory conditions, they can have up to ten generations a day (Ganley and Kobayashi 2011). The recent study of Ganley and Kobayashi (2011) showed that a yeast rDNA unit can be gained or lost as frequently as once every cell division and that duplications and deletions occurred stochastically and at similar rates. They also uncovered that large rDNA deletions are frequent events which are likely to further increase the efficiency of homogenization. Rapid homogenization of plant ribosomal RNA units has also been observed (Kovarik et al. 2005). These fast rates of homogenization are likely responsible for the fact that, for example, so little intra-genomic variation is observed between the rDNA units of fungal species and the SL genes of trypanosome species (Ganley and Kobayashi 2007; Thomas et al. 2005). Such a high degree of homogenization is expected to be most pronounced for clustered tandemly repeated gene families because intrachromosomal exchanges are more frequent than interchromosomal exchanges (Schlötterer and Tautz 1994; Kuhn et al. 2011).

A Case of Nonadaptive Evolution

As humans, we tend to search for meaning every time we observe correlations or relationships. It is therefore not surprising that functional explanations have been proposed to explain the diverse 5S rRNA gene linkages observed. As mentioned earlier, given that the first 5S rRNA gene linkages observed in eukaryotes were with the ribosomal RNA genes of 'lower' eukaryotic species, this arrangement was at first interpreted as representing a transitional stage between the linked arrangement of ribosomal genes in eubacteria and the unlinked arrangement observed in most

'higher' eukaryotes. Given that 5S rRNA genes are now known to be linked to a wide variety of other genes, and that such linkages are observed in a wide variety of eukaryotic taxa, this explanation is no longer tenable. Other researchers also suggested that the observed 5S rRNA linkages were adaptive because they allowed the coordinated expression of the linked genes (Aksoy et al. 1992; Howlett et al. 1992; Andrews et al. 1987; Bedard et al. 2006; Wicke et al. 2011). However, the fact that 5S rRNA genes (transcribed by RNA polymerase III) have been found linked to ribosomal RNA genes (transcribed by RNA polymerase I), SL, ubiquitin and histone genes (transcribed by RNA polymerase II) as well as tRNA (transcribed by RNA polymerase III) and snRNA genes (transcribed by RNA polymerase II and III) does not support this hypothesis (Campbell et al. 2000; Thomas and Chiang 2006; Carter and Drouin 2009; Martínez-Calvillo et al. 2010). Furthermore, the fact that 5S rRNA genes are linked to both ribosomal RNA and SL genes in the genome of *Euglena gracilis* (Fig. 4) suggest that coordinated expression does not provide a selective advantage. Therefore, the absence of conservation in the arrangement of 5S rRNA genes, as well as the absence of support for the coordinated expression hypothesis, suggests that the arrangement of 5S rRNA genes is not subject to selective forces.

As mentioned earlier, the most obvious feature that 5S rRNA genes have with the genes with which they are found linked with is that all these other genes are found tandemly repeated in eukaryotic genomes. Because tandemly repeated gene evolve in a concerted fashion, a 5S rRNA gene sequence moving in or out of one of these repeats is quickly homogenized throughout all repeats of the gene family (see the preceding). For example, most autotrophic stramenopiles and most of their sister taxa, have their 5S rRNA genes linked to their ribosomal RNA genes (and in the same orientation; Fig. 1). However, in autotrophic stramenopiles, 5S rRNA genes were lost twice independently, once in the *S. petersenii* lineage and once in the lineage leading to *D. brightwellii* and *N. palea* (Fig. 1). Conversely, in the *Trichomonas* genus, 5S rRNA genes are not linked to any other genes in two species, but are linked to ubiquitin genes and pseudogenes in *T. foetus* (Fig. 1). Furthermore, changes do not only include 5S rRNA gene loss or gains but also their inversion. For example, in oomycetes (heterotrophic stramenopiles), the orientation of 5S rRNA genes relative to the ribosomal RNA genes was independently inverted repeatedly (Figs. 1 and 2).

Obviously, these deductions assume that the phylogenies on which we mapped the 5S rRNA gene arrangements are accurate. Although we cannot estimate how many of the phylogenetic relationships are accurate, the fact that different gene organizations are observed within well-supported (monophyletic) taxa demonstrates that such gene

organization changes do occur irrespective of the actual branching orders. For example, different gene organizations exist within the *Trichomonas*, *Pythium*, *Karodinnium*, *Porocentrum*, *Herpetomonas*, *Trypanosoma* (including the well-supported *T. brucei* and *T. cruzi* clades) and *Entamoeba* genus (Figs. 1, 2, 3, 4 and 5).

As discussed earlier, these changes can occur relatively quickly because only n^2 generations, where n is the number of repeats in a gene family, are required to homogenize a new variant. As these tandemly repeated gene families usually contain from about 100 to 300 members (Nelson et al. 1983; Muhich et al. 1987; Bellofatto et al. 1988; Keller et al. 1992; Keeling and Doolittle 1995), homogenization therefore requires some 10,000–100,000 generations, a relatively short evolutionary time for most protist species. Furthermore, the study of Zhang et al. (2009), one of the few studies that investigated intraspecific variation in the organization of 5S and SL genes, found that each of the six dinoflagellate species they investigated contained many different types of SL repeats containing complete or partial SL sequences, some of which also contained 5S rRNA and U6 snRNA genes (Fig. 3). Such extensive intraspecific variation shows that different gene organizations do co-exist in many species. Therefore, even though at any one time one of the variant might increase in frequency, often to the point where other organizations are not easily detectable (e.g., Bedard et al. 2006), other organizations may later become dominant. This is particularly true for many protozoan species where founder events, such as when a trypanosome species is transferred between vertebrate or invertebrate species living in the same environment, are likely to lead to the fixation of new gene organizations (Stevens et al. 2001; Stevens 2008). Such founder events not only likely explain the high diversity of these species but are also likely responsible for the numerous reversions in gene 5S rRNA gene organization observed during protist evolution (Westenberger et al. 2004; Fig. 1, 2, 3, 4 and 5). For example, the recurring presence and absence of 5S rRNA genes linked to the SL genes of trypanosome species (Fig. 4) likely do not represent independent loss and gains of 5S rRNA genes in these lineages, but simply a change in the abundance of pre-existing repeats with and without 5S rRNA genes.

In summary, the features of 5S rRNA genes, such as their high copy number and expression level, as well as their internal promoters that allow transposed copies to be expressed, allow them to be functional when a copy serendipitously transposes within an unrelated tandemly repeated gene family. Although they most likely transpose haphazardly throughout genomes, some of these unique transposition events become noticeable when they become homogenized within a tandemly repeated gene family. Furthermore, as recently emphasized by Lynch (2007a, b),

selection need not be invoked to explain numerous features of eukaryotic genomes. Rather, these features can readily be explained by the action of nonadaptive processes such as recombination and drift. The high variability in 5S rRNA genes arrangements observed between eukaryotic species therefore likely represents the nonadaptive fixation of haphazard recombination events in species with short generation times and frequent founder events.

Acknowledgments We thank the associate editor and the two anonymous referees for their useful and constructive comments on a previous version of this manuscript. This study was supported by a discovery grant from the Natural Science and Engineering Research Council of Canada to G. D.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Aksoy S, Shay GL, Villanueva MS, Beard CB, Richards FF (1992) Spliced leader RNA sequences of *Trypanosoma rangeli* are organized within the 5S rRNA-encoding genes. *Gene* 113:239–243
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Andrews MT, Vaughn JC, Perry BA, Bagshaw JC (1987) Interspersion of histone and 5S RNA genes in *Artemia*. *Gene* 51:61–67
- Archambault J, Friesen JD (1993) Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol Rev* 57:703–724
- Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) *Evolution of genes and proteins*. Sinauer, Sunderland, MA, pp 38–61
- Bachvaroff TR, Handy SM, Place AR, Delwiche CF (2011) Alveolate phylogeny inferred using concatenated ribosomal proteins. *J Eukaryot Microbiol* 58:223–233
- Baldauf S (2008) An overview of the phylogeny and diversity of eukaryotes. *J Syst Evol* 46:263–273
- Beakes GW, Glockling SL, Sekimoto S (2011) The evolutionary phylogeny of the oomycete “fungi”. *Protoplasma* 249:3–19
- Bedard JE, Schurko AM, de Cock AW, Klassen GR (2006) Diversity and evolution of 5S rRNA gene family organization in *Pythium*. *Mycol Res* 110:86–95
- Belkhir A, Klassen GR (1996) Diverged 5s rRNA sequences adjacent to 5s rRNA genes in the rDNA of *Pythium pachycaule*. *Curr Genet* 29:287–292
- Belkhir A, Buchko J, Klassen GR (1992) The 5S ribosomal RNA gene in *Pythium* species: two different genomic locations. *Mol Biol Evol* 9:1089–1102
- Bellofatto V, Cooper R, Cross GA (1988) Discontinuous transcription in *Leptomonas seymouri*: presence of intact and interrupted mini-exon gene families. *Nucleic Acids Res* 16:7437–7456
- Bergeron J, Drouin G (2008) The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungi species. *Curr Genet* 54:123–131
- Brown JW, Sorhannus U (2010) A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One* 5:e12759
- Cabral-de-Mello DC, Cabrero J, López-León MD, Camacho JP (2011) Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. *Genetica* 139:921–931
- Campbell DA, Sturm NR, Yu MC (2000) Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol Today* 16:78–82
- Carter R, Drouin G (2009) Structural differentiation of the three eukaryotic RNA polymerases. *Genomics* 94:388–396
- Clark CG, Kaffashian F, Tawari B, Windsor JJ, Twigg-Flesner A, Davies-Morel MC, Blessmann J, Ebert F, Peschel B, Le Van A, Jackson CJ, Macfarlane L, Tannich E (2006) New insights into the phylogeny of *Entamoeba* species provided by analysis of four new small-subunit rRNA genes. *Int J Syst Evol Microbiol* 56:2235–2239
- Cohen S, Segal D (2009) Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res* 124:327–338
- Cohen S, Agmon N, Sobol O, Segal D (2010) Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mob DNA* 1:11
- Cole J, Anderson OR, Tekle YI, Grant J, Kat LA, Nerad T (2010) A description of a new “Amoebozoan” isolated from the American lobster, *Homarus americanus*. *J Eukaryot Microbiol* 57:40–47
- Cross I, Rebordinos L (2005) 5S rDNA and U2 snRNA are linked in the genome of *Crassostrea angulata* and *Crassostrea gigas* oysters: does the (CT)_n (GA)_n microsatellite stabilize this novel linkage of large tandem arrays? *Genome* 48:1116–1119
- D’Avila-Levy CM, Volotão AC, Araújo FM, de Jesus JB, Motta MC, Vermelho AB, Santos AL, Branquinha MH (2009) Bodo sp., a free-living flagellate, expresses divergent proteolytic activities from the closely related parasitic trypanosomatids. *J Eukaryot Microbiol* 56:454–458
- Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc Natl Acad Sci USA* 99:8324–8329
- Derelle R, Momose T, Manuel M, Da Silva C, Wincker P, Houliston E (2010) Convergent origins and rapid evolution of spliced leader trans-splicing in metazoa: insights from the ctenophora and hydrozoa. *RNA* 16:696–707
- Deschamps P, Lara E, Marande W, López-García P, Ekelund F, Moreira D (2011) Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids. *Mol Biol Evol* 28:53–58
- Dover GA (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299:111–117
- Dover GA (1993) Evolution of genetic redundancy for advanced players. *Curr Opin Genet Dev* 3:902–910
- Drouin G (2000) Expressed retrotransposed 5S rRNA genes in the mouse and rat genomes. *Genome* 43:213–215
- Drouin G, Moniz de Sá M (1995) The concerted evolution of 5S ribosomal genes linked to the repeat units of other multigene families. *Mol Biol Evol* 12:481–493
- Drouin G, Hofman JD, Doolittle WF (1987) Unusual ribosomal RNA gene organization in copepods of the genus *Calanus*. *J Mol Biol* 196:943–946
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA* 102:14338–14343
- Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* 175:477–485

- Eirín-López JM, Fernanda Ruiz M, González-Tizón AM, Martínez A, Sánchez L, Méndez J (2004) Molecular evolutionary characterization of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with “orphan” features. *J Mol Evol* 58:131–144
- Elder JF Jr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol* 70:297–320
- Frantz C, Ebel C, Paulus F, Imbault P (2000) Characterization of trans-splicing in Euglenoids. *Curr Genet* 37:349–355
- Gangloff S, Zou H, Rothstein R (1996) Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO J* 15:1715–1725
- Ganley AR, Kobayashi T (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res* 17:184–191
- Ganley AR, Kobayashi T (2011) Monitoring the rate and dynamics of concerted evolution in the ribosomal DNA repeats of *Saccharomyces cerevisiae* using experimental evolution. *Mol Biol Evol* 28:2883–2891
- Gerbi SA (1985) Evolution of ribosomal DNA. In: MacIntyre RJ (ed) *Molecular evolutionary genetics*. Plenum, New York, pp 419–517
- Graur D, Li W-H (2000) *Fundamentals of molecular evolution*, 2nd edn. Sinauer, Sunderland, MA
- Hamilton PB, Stevens JR, Gaunt MW, Gidley J, Gibson WC (2004) Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int J Parasitol* 34:1393–1404
- Howlett BJ, Brownlee AG, Guest DI, Adcock GJ, McFadden GI (1992) The 5S ribosomal RNA gene is linked to large and small subunit ribosomal RNA genes in the oomycetes, *Phytophthora vignae*, *P. cinnamomi*, *P. megasperma f.sp. glycinea* and *Saprolegnia ferax*. *Curr Genet* 22:455–461
- Hughes AL, Piontkivska H (2003a) Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. *Mol Biol Evol* 20:644–652
- Hughes AL, Piontkivska H (2003b) Molecular phylogenetics of Trypanosomatidae: contrasting results from 18S rRNA and protein phylogenies. *Kinetoplastid Biol Dist* 28:15
- Janoušková J, Horák A, Oborník M, Lukeš J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA* 107:10949–10954
- Keeling PJ, Doolittle WF (1995) Concerted evolution in protists: recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis*. *J Mol Evol* 41:556–562
- Keller M, Tessier LH, Chan RL, Weil JH, Imbault P (1992) In Euglena, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated. *Nucleic Acids Res* 20:1711–1715
- Kleina P, Bettim-Bandinelli J, Bonatto SL, Benchimol M, Bogo MR (2004) Molecular phylogeny of Trichomonadidae family inferred from ITS-1, 5.8S rRNA and ITS-2 sequences. *Int J Parasitol* 34:963–970
- Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS (2005) Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics* 169: 931–944
- Kuhn GC, Küttler H, Moreira-Filho O, Heslop-Harrison JS (2011) The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Mol Biol Evol* 29:7–11
- Lenhard B, Sandelin A, Carninci P (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* 13:233–245
- Lévesque CA, de Cock AW (2004) Molecular phylogeny and taxonomy of the genus *Pythium*. *Mycol Res* 108:1363–1383
- Lynch M (2007a) The origins of genome architecture. Sinauer, Sunderland, MA
- Lynch M (2007b) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci USA* 104:8597–8604
- Maizels N (1976) Dictyostelium 17S, 25S, and 5S rDNAs lie within a 38,000 base pair repeated unit. *Cell* 9:431–438
- Malik SB, Brochu CD, Bilic I, Yuan J, Hess M, Logsdon JM Jr, Carlton JM (2011) Phylogeny of parasitic Parabasalia and free-living relatives inferred from conventional markers vs. Rpb1, a single-copy gene. *PLoS One* 6(6):e20774
- Manchado M, Zuasti E, Cross I, Merlo A, Infante C, Rebordinos L (2006) Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea senegalensis*: a new linkage to the U1, U2, and U5 small nuclear RNA genes. *Genome* 49:79–86
- 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
- Moreira D, López-García P, Vickerman K (2004) An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol* 54:1861–1875
- Moreira D, von der Heyden S, Bass D, López-García P, Chao E, Cavalier-Smith T (2007) Global eukaryote phylogeny: combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol* 44:255–266
- Muhich ML, Hughes DE, Simpson AM, Simpson L (1987) The monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, contains a transcriptionally active, multicopy mini-exon sequence. *Nucleic Acids Res* 15:3141–3153
- Nelson RG, Parsons M, Barr PJ, Stuart K, Selkirk M, Agabian N (1983) Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphans in *Trypanosoma brucei*. *Cell* 34:901–909
- Paule MR, White RJ (2000) Transcription by RNA polymerases I and III. *Nucleic Acids Res* 28:1283–1298
- Pelliccia F, Barzotti R, Bucciarelli E, Rocchi A (2001) 5S ribosomal and U1 small nuclear RNA genes: a new linkage type in the genome of a crustacean that has three different tandemly repeated units containing 5S ribosomal DNA sequences. *Genome* 44:331–335
- Preisfeld A, Busse I, Klingberg M, Talke S, Ruppel HG (2001) Phylogenetic position and inter-relationships of the osmotrophic euglenids based on SSU rDNA data, with emphasis on the Rhabdomonadales (Euglenozoa). *Int J Syst Evol Microbiol* 51:751–758
- Pukkila PJ, Skrzynia C (1993) Frequent changes in the number of reiterated ribosomal RNA genes throughout the life cycle of the basidiomycete *Coprinus cinereus*. *Genetics* 133:203–211
- Reeb VC, Peglar MT, Yoon HS, Bai JR, Wu M, Shiu P, Grafenberg JL, Reyes-Prieto A, Rümmele SE, Gross J, Bhattacharya D (2009) Interrelationships of chromalveolates within a broadly sampled tree of photosynthetic protists. *Mol Phylogenet Evol* 53:202–211
- Roditi I (1992) *Trypanosoma vivax*: linkage of mini-exon (spliced leader) and 5s ribosomal RNA genes. *Nucleic Acids Res* 20:1995
- Rubin GM, Sulston JE (1973) Physical linkage of the 5S cistrons to the 18S and 28S ribosomal RNA cistrons in *Saccharomyces cerevisiae*. *J Mol Biol* 79:521–530
- Schlötterer C, Tautz D (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol* 4:777–783

- Simpson AG, Roger AJ (2004) Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol Phylogenet Evol* 30:201–212
- Smith GP (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symp Quant Biol* 38:507–513
- Stevens JR (2008) Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite* 15:226–232
- Stevens JR, Teixeira MM, Bingle LE, Gibson WC (1999) The taxonomic position and evolutionary relationships of *Trypanosoma rangeli*. *Int J Parasitol* 29:749–757
- Stevens JR, Noyes HA, Schofield CJ, Gibson W (2001) The molecular evolution of Trypanosomatidae. *Adv Parasitol* 48:1–56
- Szostak JW, Wu R (1980) Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* 284:426–430
- Templeton TJ, Enomoto S, Chen WJ, Huang CG, Lancto CA, Abrahamsen MS, Zhu G (2010) A genome-sequence survey for *Ascogregarina taiwanensis* supports evolutionary affiliation but metabolic diversity between a Gregarine and Cryptosporidium. *Mol Biol Evol* 27:235–248
- Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* 41:105–178
- Thomas S, Westenberger SJ, Campbell DA, Sturm NR (2005) Intragenomic spliced leader RNA array analysis of kinetoplastids reveals unexpected transcribed region diversity in *Trypanosoma cruzi*. *Gene* 352:100–108
- Torres-Machorro AL, Hernández R, Cevallos AM, López-Villaseñor I (2010) Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? *FEMS Microbiol Rev* 34:59–86
- Vannini A, Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 45:439–446
- Vierna J, Jensen KT, Martínez-Lage A, González-Tizón AM (2011) The linked units of 5S rDNA and U1 snDNA of razor shells (Mollusca: bivalvia: pharidae). *Heredity* 107:127–142
- von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T (2004) Ribosomal RNA phylogeny of bodonid and diplomid flagellates and the evolution of euglenozoa. *J Eukaryot Microbiol* 51:402–416
- Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24:437–440
- Westenberger SJ, Sturm NR, Yanega D, Podlipaev SA, Zeledón R, Campbell DA, Maslov DA (2004) Trypanosomatid biodiversity in Costa Rica: genotyping of parasites from Heteroptera using the spliced leader RNA gene. *Parasitology* 129:537–547
- Wicke S, Costa A, Muñoz J, Quandt D (2011) Restless 5S: the rearrangement(s) and evolution of the nuclear ribosomal DNA in land plants. *Mol Phylogenet Evol* 61:321–332
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87:4576–4579
- Zhang H, Bhattacharya D, Lin S (2007) A three-gene dinoflagellate phylogeny suggests monophyly of prorocentrales and a basal position for amphidinium and heterocapsa. *J Mol Evol* 65:463–474
- Zhang H, Campbell DA, Sturm NR, Lin S (2009) Dinoflagellate spliced leader RNA genes display a variety of sequences and genomic arrangements. *Mol Biol Evol* 26:1757–1771
- 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