The Concerted Evolution of 5S Ribosomal Genes Linked to the Repeat Units of Other Multigene Families

Guy Drouin and Mário Moniz de Sá
Biology Department, University of Ottawa

We review all instances in which the nuclear 5S rRNA genes of fungi, protist, nematode, and arthropod species have been reported to be linked to the tandemly repeated units of the rDNA, trans-spliced leader, and histone multigene families. The evolution of these gene arrangements is analyzed by mapping them to independently derived phylogenies. These analyses show that 5S rRNA genes have repeatedly become linked to diverse tandemly repeated gene families and that such linkages have also been subsequently inverted or lost in some species. These variable gene linkages are probably the result of stochastic gains and losses of variant repeat units, where functional 5S rRNA had transposed, by the mechanisms which are responsible for the concerted evolution of tandemly repeated multigene families. We discuss the possible mechanisms of 5S rRNA gene transposition and suggest that the characteristics of their promoter elements, transcription, and termination signals may allow functional copies of these genes to be fortuitously transposed through an RNA intermediate. We also review the evidence which shows that the linked 5S rRNA gene copies are transcribed. We conclude that the observed patterns of 5S rRNA gene linkages to the repeat units of other tandemly repeated multigene families have likely arisen due to fortuitous recombination events and are unlikely to represent the remnants of an eubacterial-like arrangement of rDNA operons or to have been established due to selective pressures.

Introduction

The organization and expression of the nuclear genes encoding ribosomal RNAs (rRNA) have been extensively studied. In eubacteria the rRNA genes coding for 16S, 23S, and 5S rRNA are usually organized in operons and are transcribed in this order by the same RNA polymerase. In contrast, eukaryotes have four different rRNA genes, and 5S rRNA genes are not transcribed by the same RNA polymerase as the three other rRNA genes. The 16S, 23S, and 5S rRNA genes of prokaryotes are homologous to the eukaryotic 18S, 28S, and 5S rRNA genes, respectively, whereas the 5.8S rRNA gene of eukaryotes is homologous to the eubacterial 18S-rRNA gene (reviewed in Gerbi 1985). The eukaryotic 18S, 5.8S, and 28S rRNA genes are grouped in that order to form RNA polymerase I transcription units (rDNA units), and multiple copies of these units are typically found clustered in long direct tandem arrays. The 5S rRNA genes are transcribed by RNA polymerase III and are also often found clustered in direct tandem arrays (Singer and Berg 1991, pp. 631-638).

Given that 5S rRNA genes and the other three rRNA genes are transcribed by different RNA polymerases, the observation that the 5S rRNA genes of many eukaryotic species are not linked to their rDNA units is not surprising. For example, 5S rRNA genes are not linked to the rDNA unit of several angiosperm plant species, several Drosophila species, the house cricket Acheta domesticus, Xenopus, and human (Hemleben and Grierson 1978; Beneš and Cave 1985; Tautz et al. 1987; Borisjuk and Hemleben 1993; Gonzalez et al. 1993). What is surprising is that 5S rRNA genes have in fact often been found linked to the rDNA units in the genome of several eukaryotic species (Gerbi 1985). The linkage of 5S rRNA genes to the rDNA units of Saccharomyces cerevisiae was first reported by Rubin and Sulston (1973), and these genes were later shown to be transcribed from the strand opposite to the other three rRNA species (Kramer et al. 1978). Evidence for the linkage of 5S genes to the rDNA units of Dictyostelium discoideum was first presented by Maizels (1976), and Hofmann et al. (1993) showed that these linked copies were transcribed from the same strand as the other three rRNA species. The finding of such linkages in "lower" eukaryote species has previously been taken to
represent some sort of transitional state between the typical eu-bacterial arrangement and the unlinked arrangement found in "higher" eukaryotic species (Gerbi 1985; Belkhiri et al. 1992).

Now 5S rRNA genes have been found linked to the tandemly repeated rDNA repeat units of several "higher" eukaryotic species, such as nematode and arthropod species, as well as to the tandemly repeated units of the trans-spliced leader (TSL) and histone multigene families. A schematic representation of the currently known linkage relationships of 5S rRNA genes to the repeat units of these multigene families in different taxonomic groups is shown in figure 1. This figure shows both the position of the linked 5S rRNA gene copies, as well as whether they are transcribed on the same or opposite strand as compared to the genes to which they are linked. Finding closely related species in whose genome the 5S rRNA genes are linked to different strands of their rDNA units would indicate either that they acquired their 5S rRNA genes independently or that a linked copy was subsequently inverted and homogenized throughout the rDNA multigene family of some species. Whereas the observed linkage of 5S rRNA genes is limited to rDNA repeat units in fungi, they are found linked to either rDNA or trans-spliced leader repeat units in different protist and nematode species and to either rDNA or histone gene repeat units in different crustacean species. They have also been found to be linked to the rDNA units of one spider species.

Clearly, it is difficult to explain such diverse 5S rRNA gene arrangements in diverse taxonomic groups by considering that they are the remnants of earlier evolutionary times. In the following section, we address this issue by reviewing all published reports of 5S rRNA gene linkages to the tandemly repeated units of these diverse multigene families. We have also included species in which there are no 5S rRNA gene linkages when it was relevant to our analysis of the evolution of these gene arrangements. However, we do not report the organization of 5S rRNA genes in these species (i.e., whether they are found as dispersed single copies or as clusters(s) with or without dispersed single-copy members) because it is not directly relevant to our analysis. Also, note that there are yet no published reports of 5S rRNA genes being linked to more than one other multigene family.

**Are the 5S rRNA Gene Linkages a Primitive or Derived Condition?**

As mentioned above, the linkage of 5S rRNA genes to the rDNA units of some "lower" eukaryote species was previously interpreted as representing a primitive condition. Similarly, the observation of the linkage of 5S rRNA genes to the rDNA unit of copepod (crustacean) species was also initially thought to represent a primitive condition in this group (Drouin et al. 1987). However, a later study showed that this arrangement, although present in several crustacean species, was not conserved in all related species (Drouin et al. 1992). If one is to understand the evolution of such gene arrangements, it is best to map them to independently derived phylogenies.

Figure 2 shows a phylogenetic tree based on parsimony analysis of 18S rRNA gene sequences to which the distribution of 5S rRNA gene linkages in the rDNA repeat units of fungi and oomycetes has been mapped. When known, the strand from which the 5S rRNA genes are transcribed is indicated. The oomycetes are now known not to be directly related to fungi but to have originated before the fungi (Förster et al. 1990). These two groups are represented on the same phylogenetic tree for convenience.

The evolutionary relationships of oomycetes species are poorly known, and they are represented as all originating from a single node (fig. 2). The fact that the linkage of 5S rRNA genes to the noncoding strand is observed in 14 Pythium species, as well as in 3 species from other genera, but that 5S rRNA genes are not linked to the rDNA units of 8 other Pythium species suggests either that the common ancestor to all Pythium species had linked 5S rRNA genes and that such linkage was lost in some species or that this ancestor did not have such linkage and it occurred later in the lineage, leading to the 14 other Pythium species as well as in the lineage(s) leading to the species of the three other genera.
The phylogenetic tree, based on parsimony analysis of 18S rDNA sequences, is an adaptation of the trees of Taylor et al. (1993), who themselves based them on previously published trees (Bruns et al. 1991, 1992; Berbee and Taylor 1992). Single boxes represent rDNA units where 5S genes are not linked, whereas double boxes represent rDNA units where 5S genes are linked. A plus sign (+) in the second box indicates that the coding region of the 5S gene copies is on the same strand as the coding strand of the other ribosomal genes, a minus (−) sign indicates that it is on the opposite strand, and a question mark (?) indicates that the orientation of the 5S genes relative to the other ribosomal genes has not been determined.

In contrast with the work on oomycetes, recent work has produced a well-supported phylogeny of fungi (Taylor et al. 1993). Mapping 5S rRNA gene arrangements to this phylogenetic tree clearly shows that the linkage relationships of these genes to rDNA repeat units is not conserved during evolution (fig. 2). Although most basidiomycetes have 5S rRNA genes linked to the coding strand of rDNA repeat units, one of the four Coprinus species has this gene linked to the noncoding strand rather than the coding strand as found in three other Coprinus species. Furthermore, Schizosaccharomyces pombe, an early ascomycete, does not have its 5S rRNA genes linked to its rDNA repeat units, whereas most ascomycete yeast species, Yarrowia lipolytica, does not have its 5S rRNA genes linked to its rDNA repeat units. Finally, most filamentous yeasts do not have their 5S rRNA genes linked to their rDNA repeat units, but here again there is a notable exception in that one loculoascomycete species has its 5S genes linked to the coding strand of its rDNA repeat units. Thus, the linkage of 5S rRNA genes to the rDNA repeat units of fungi species has not only occurred repeatedly during their evolution but might also have subsequently been lost or inverted in some of them.

In protozoan species, 5S rRNA genes have been found linked either to the rDNA repeat units or to the repeat units of the trans-spliced leader sequences (fig. 3). The phylogenetic relationships of the diverse species classified under the term protozoan has come a long way thanks to phylogenies based on nucleic acid sequences, but some of the proposed phylogenies are still controversial (see, e.g., Lake et al. 1988; Fernandes et al. 1993).
The presence of 5S rRNA genes in the rDNA repeat units of *Dictyostelium discoideum* and *Toxoplasma gondii* could represent independent insertions since these two species are only distantly related (Johnson and Bavister 1989). The highly variable pattern of 5S RNA gene linkages to the trans-spliced leader repeat units of the other protozoan species shown in figure 3 suggests that such linkages were established and lost repeatedly during their evolution.

The 5S rRNA gene linkages observed in "higher" eukaryotes, and the evolutionary relationships of these eukaryotes, are shown in figure 4. The most parsimonious way to explain the phylogenetic distribution of 5S rRNA gene linkages in nematodes is to assume that 5S rRNA genes became linked to rDNA repeat units in the *Meloidogyne arenaria* lineage, whereas they became linked to the trans-spliced leader repeat units in a common ancestor which gave rise to both the Rhabditida and Spirurida lineages. This linkage relationship was then lost in the common ancestor which gave rise to the Rhabditida and Strongylida lineages and either reestablished on the noncoding strand or inverted in the Rhabditida lineage (fig. 4A).

The phylogeny of arthropods has long been controversial, but current evidence favors the evolutionary relationships shown in figure 4B. This phylogeny suggests that several gain and loss events of 5S rRNA gene linkages must have occurred during the evolution of this group. Assuming that the chelicerate lineage predates both the hexapod and crustacean lineages, one has to infer that the 5S rRNA gene linkage to the rDNA repeat units of the spider *Araneus* species was lost in the lineage leading to the two other arthropod groups in order to explain that such an arrangement has not yet been found in any insect species and that 5S rRNA genes are linked to histone gene repeat units and not to the rDNA repeat units in species of the most ancient crustacean group, the branchiopods. The most parsimonious way to explain the arrangement of 5S rRNA genes in the other three crustacean groups is to assume that they are all linked to the noncoding strand of the rDNA repeat units and that this linkage was independently lost in the two copepod species which do not have 5S rRNA genes in their rDNA repeat units.

The above results clearly show that 5S rRNA gene linkages were repeatedly established and lost during the evolution of eukaryotic genomes and that linked 5S rRNA gene copies might also have subsequently been inverted. The observation that 5S rRNA genes are variously found either linked or not linked to the repeat units of diverse tandemly repeated multigene families in all of the taxonomic groups reviewed here gives strong support to this conclusion. Although mapping the arrangement of these genes to independently derived phylogenies clearly shows that it has undergone frequent changes during evolution, the details of the particular changes suggested here will likely have to be reinterpreted when more data concerning both the arrangement of 5S rRNA genes from other species and more accurate phylogenies become available. For example, our interpretation of the evolution of 5S rRNA gene arrangement in arthropods is probably oversimplistic in that it assumes that the ancestral state of arthropods is represented by the arrangement observed in a single spider species.
and that this arrangement does not represent a derived condition. Ironically, given the available data and the trees to which the 5S rRNA gene arrangements were mapped, we have suggested (following the parsimony principle) the simplest hypothesis assuming that such arrangements are stable in time, whereas the evidence presented here shows that it is not so.

Possible Mechanisms of 5S rRNA Gene Transposition

How did 5S rRNA genes become linked to the diverse tandemly repeated multigene families reviewed above? Such recombination events could occur at either the DNA or RNA level. Illegitimate crossing-over at the DNA level or hitchhiking on transposable elements seem unlikely possibilities. It is difficult to see how either of these mechanisms would be so often associated with 5S rRNA genes compared to any other genes. Although repeated noncoding sequences are sometimes found on either side of the 5S rRNA gene coding regions (see, e.g., Drouin et al. 1987; Cruces et al. 1989), none of these repeated sequences have been reported to have similarity to transposable elements in the publications reviewed here.

An interesting, if somewhat unusual and yet unproven, possibility is that the recombination of 5S rRNA genes to other loci could be mediated by the insertion of extrachromosomal covalently closed circular DNAs (cccDNAs) containing 5S rRNA gene sequences. Such molecules have been found in many eukaryotic species, including mammals, chicken, Drosophila, and plants (reviewed in Renault et al. 1993). Several classes of cccDNAs have been found in D. melanogaster embryos, one of which contains a variable number of sequences homologous to 5S rRNA genes (Pont et al. 1987). They are thought to result from homologous recombination between adjacent repeats (Degrotte et al. 1990). Although the existence of such circular molecules in the germ-line of any species remains to be demonstrated, as does their ability to reinsert into chromosomes, such a mechanism of illegitimate recombination constitute an intriguing possibility which would have important evolutionary implications (Pont et al. 1987). Given that 5S rRNA genes are often found as tandemly repeated multigene families and that cccDNAs are the result of homologous recombination between adjacent chromosomal repeats, this mechanism could explain the relatively frequent transposition of 5S rRNA genes to other loci we have inferred from our phylogenetic analyses.

At the RNA level, an RNA mediated transposition of 5S rRNA genes analogous to the transposition of short interspersed repeated DNA elements (SINES) is another possibility. SINES can be classified into two groups (Deininger 1989). The first group consists of Alu-like elements which contain an internal RNA polymerase III promoter, an A-rich 3′-end on the strand corresponding to the transcript and are flanked by direct repeats. They are derived from 7SL RNA and are found only in primates. The second group consists of tRNA-like elements which contain an internal RNA polymerase III promoter but do not have 3′ A-rich regions and are not flanked by direct repeats. They are derived from tRNA and have been found in the genomes of protozoan, plant, insect, sea urchin, tortoise, newt, and fish species (Okada 1991; Deininger and Batzer 1993).

The characteristics of the promoter elements responsible for the expression of 5S rRNA genes make it likely that they could also transpose through an RNA intermediate (fig. 5). They are transcribed by RNA pol III, and most of the sequence information required for their expression and regulation resides within their coding region. Upstream regulatory elements can also play a role in the expression of 5S rRNA genes but, apart from the situation in human (see below), these elements have been shown to be quite short and simple, not strongly conserved, and in some cases not essential for their expression (fig. 5). It is therefore possible that a 5S rRNA cDNA copy could serendipitously insert downstream of sequences that could act as a 5′ promoter region and allow them to be transcribed. Such a possibility has also been proposed to explain why only a subset of the nearly one million Alu sequences in the human genome are transcribed and produce new Alu elements (Schmid and Maraiia 1992). In contrast, the human 5S rRNA genes have been shown to require a 12-base pair long 5′ promoter sequence for their expression (Nielsen et al. 1993). It could be that many of the 5S rRNA pseudogenes observed in humans were generated through RNA-mediated transposition events but, because of their need for more complex promoter elements in their 5′-flanking sequences, they could not be expressed (Wingender et al. 1988; Sørensen and Frederiksen 1991).

Transcription termination of RNA polymerase III transcripts has been shown to be relatively simple and to consist of one or two oligo (dT)₃ clusters 3′ of the coding region and that these termination signals are usually part of the primary transcript (Campbell and Setzer 1992). Thus, the reverse transcription and integration of 5S rRNA gene transcripts could in theory produce transposed copies of these genes which would contain not only their internal promoter elements but also their transcription termination signals.

Of the two possible mechanisms suggested here, RNA-mediated transposition could be more likely. It would account for the observation that a single 5S rRNA gene copy is invariably found linked to the repeat units of other tandemly repeated multigene families, whereas illegitimate recombination of a cccDNA molecules con-
FIG. 5.—DNA sequence elements involved in the expression of 5s rRNA genes in *Drosophila melanogaster* (Sharp and Garcia 1988), *Neurospora crassa* (Tyler 1987), *Saccharomyces cerevisiae* (Challice and Segall 1989), and *Homo sapiens* (Nielson et al. 1993). Thick boxes represent transcribed regions, whereas thin boxes represent 5'-flanking regions. All 5s rRNA genes contain internal control regions (ICRs) and, with the exception of the yeast ICR (Taylor and Segall 1985), share in common two functionally distinct domains referred to as the A-box and the C-box. The ICR III and ICR IV of *D. melanogaster* are functionally equivalent to the A-box and the C-box, respectively. The Box D of *N. crassa* is not functionally equivalent to the D-box in *H. sapiens* (Tyler 1987). The SSE (start site element) found in *S. cerevisiae* is a region where deletions affect the expression level, but only position +1 is essential for transcription (Challice and Segall 1989). Positions of the A-box and C-box in human genes were determined by sequence similarity to that of *Xenopus borealis* (Pieler et al. 1987). The ICR of all 5s rRNA genes contain the common binding site (C-box) for the specific and distinguishing transcription factor of 5s rRNA genes, TFIIIA. Although the transcription factor responsible for binding to RNA polymerase III (TFIIB) binds to upstream noncoding DNA sequence, this binding is not specific but depends on protein-protein interactions with other transcription factors which bind to the internal promoter sequences (Wolffe 1991; Geiduschek and Kassavetis 1992). Upstream regulatory elements can also play a role in the expression of 5s rRNA genes (Selker et al. 1986; Tyler 1987; Reynolds and Azer 1988; Sharp and Garcia 1988; Challice and Segall 1989; Felgenhauer et al. 1990; Oei and Pieler 1990; Nielson et al. 1993). However, in all cases, except mammals, these elements have been shown to be quite short and simple. In addition, these 5'-flanking sequences can be diverse and not strongly conserved, even within a single species (Selker et al. 1986; Geiduscheck and Fassavetis 1992). In some cases the 5'-flanking sequences are not essential for expression of 5s rRNA genes (Worthington et al. 1981; Wingender et al. 1988; Wolfe 1991; Geiduschek and Kassavetis 1992). Transcription termination of 5s rRNA genes by RNA polymerase III is relatively simple, involving one or two oligo (dT)₃ clusters 3' of the coding region (Bogenhagen and Brown 1981; Campbell and Setzer 1992). These poly-T termination signals (Tₙ) are often part of the primary transcript (Campbell and Setzer 1992).
their essential internal promoters, as well as their transcription termination signals, and that the limited upstream regulatory regions they may require could often be present serendipitously at their insertion site.

Transcription studies have shown that the linked 5S rRNA genes are indeed functional in several species, such as *Saccharomyces cerevisiae*, in which 5S rRNA genes are linked to the noncoding strand of rDNA repeat units, and the kinetoplastid protozoa *Trypanosoma rangeli*, in which 5S rRNA genes are only found linked to the coding strand of the trans-spliced leader repeat units (Challice and Segall 1989; Aksoy et al. 1992). In *T. borreli*, 5S rRNA genes are found dispersed in the genome as well as on the noncoding strand of two different trans-spliced leader clusters, named the T1 and T2 clusters. The 5S rRNA gene copies linked to the T1 repeat units have been shown to be transcribed, whereas the copies linked to the T2 repeat units are not (Maslov et al. 1993). The 5S rRNA genes of the nematode species *Brugia malayi*, *Caenorhabditis elegans*, and *C. briggsae*, which are found linked to either the coding or noncoding strand of trans-spliced leader repeat units, have also been shown to be transcribed (Nelson and Honda 1985; Honda et al. 1986; Takacs et al. 1988; Nelson and Honda 1989). In the nematode *Melodogynie arenaria*, only the 5S rRNA gene copies which are linked to the noncoding strand of the 9-kb rDNA repeat units seem to be transcribed, whereas the 5S rRNA gene copies which are linked to the noncoding strand of the 5-kb rDNA repeat units are likely pseudogenes (Vahidi et al. 1991).

Indirect evidence for the likely transcription of linked 5S rRNA genes can also be inferred from the fact that genomic Southern blot hybridization experiments suggest that these genes are not found outside the repeat units of the multigene families to which they are linked. In fact, copies of 5S rRNA genes have not been found outside the trans-spliced repeat units of the protozoan species *Herpetomonas* spp., *Toxoplasma gondii*, and *T. rangeli*, whereas functional 5S rRNA genes have not been found outside the rDNA repeat units of *Dictyostelium discoideum* (Aksoy et al. 1992; Guay et al. 1992; Hofmann et al. 1993). 5S rRNA genes are found exclusively within the trans-spliced repeat units of the nematode species *B. malayi* and *C. briggsae*, whereas they are found exclusively in the histone repeat units of the crustacean species *Artemia salina* (Andrews et al. 1987; Takacs et al. 1988; Nelson and Honda 1989). Thus, transcription studies and the observations that 5S rRNA genes are only found as linked copies in the genome of some species demonstrate that linked 5S rRNA genes often constitute functional copies of these genes.

**Concerted Evolution**

It has often been observed that the repeat units of tandemly repeated multigene families (of which the rDNA, trans-spliced leader, and histone multigene families reviewed here are examples) are generally more similar to each other within a given species than would be expected if they had been evolving independently, a phenomenon usually referred to as the concerted evolution of these multigene families. Unequal crossing-over is generally thought to be the most important mechanism responsible for the concerted evolution of tandemly repeated multigene families, although gene conversion might also be involved. Such concerted evolution requires not only that new variant repeats be homogenized within a multigene family by mechanisms of DNA transfer but also that they become fixed within a species by random genetic drift (Dover 1982; Arnheim 1983; Ohta 1983; Dover and Tautz 1986). Computer simulations have indeed shown that repeated cycles of unequal crossing-over can homogenize any given variant repeat throughout these multigene families as a consequence of random genetic drift (Smith 1976).

Apart from the 5S rRNA gene linkage in the nematode species *Panagrellus redivivus* (see below), 5S rRNA gene linkages to the multigene families reviewed here seem to have been homogenized to all the repeat units of these multigene families. In all these cases, hybridization studies to either genomic DNA Southern blots or to PCR amplified fragments have shown that the 5S rRNA probes hybridized to the same fragments as probes of the multigene family to which they were found to be linked. Given that the genomic DNAs used in these experiments were often obtained from several individuals, the fact that both type of probes hybridized to the same discrete DNA fragments implies that the linkage of 5S rRNA genes has not only been homogenized within the multigene families but has also been fixed in the species. On the other hand, the conclusions of hybridization studies should be interpreted cautiously. Since the number of the most prevalent arrangement type could often be much larger than the corresponding variants, hybridization studies could not be sensitive enough to reveal rare variants which would have acquired or lost a 5S rRNA gene copy.

Interestingly, Pukkila and Skrzynia (1993) have recently shown that the number of rDNA repeat units in *Coprinus cinereus* undergoes frequent expansions and contractions, which presumably represent the frequent occurrence of unequal crossing-overs. They could be involved in the concerted evolution of the different 5S rRNA gene arrangements observed in the different species of this genus where three *Coprinus* species have 5S rRNA genes linked on the coding strand of their rDNA repeat units, whereas they are linked on the noncoding strand of rDNA repeat units in a fourth *Coprinus* species (Cassidy and Pukkila 1987; fig. 2).
In the genome of the nematode species *P. redivivus*, some trans-spliced leader repeat units contain linked 5S rRNA gene copies whereas others do not (Bektesh et al. 1988; fig. 4A). In this species, the screening of a genomic library produced 129 clones which hybridized only to trans-spliced leader sequences, 29 which hybridized only to 5S rRNA gene sequences, and only 7 which hybridized to both the trans-spliced leader and 5S probes. In contrast, performing the same experiment on 10,000 *Caenorhabditis elegans* genomic phages revealed 47 clones which hybridized to both the trans-spliced leader and 5S probes and none which hybridized to only the trans-spliced leader or the 5S probe. The situation in *P. redivivus* could thus represent a transitional stage of the homogenization of this multigene family in which either the linked or unlinked arrangement has not yet been homogenized throughout this multigene family and/or not fixed in this species (Dover 1993).

The linked 5S rRNA genes likely originated from the dispersed copies of these genes which are still present in the genome of many species. In fact, hybridization of 5S rRNA probes to genomic Southern blots have often revealed that dispersed copies were also present in the genome of species in which they are found linked to the repeat units of other multigene families. For example, dispersed 5S rRNA gene copies have been reported in the genomes of the yeast *Saccharomyces cerevisiae*, the protozoan *Trypanoplasma borreli*, the nematode *C. elegans*, and several copepod species (Piper et al. 1984; Nelson and Honda 1985; Drouin et al. 1987, 1992; Maslov et al. 1993). Here again, the conclusions of hybridization studies should be interpreted cautiously. They can easily lead to the conclusion that dispersed copies are absent from the genome of a given species when they are not. For example, a “normal” exposure of genomic Southern blots of *C. elegans* hybridized with a 5S rRNA probe revealed a single strong signal corresponding to the 5S rRNA genes present within the trans-spliced leader cluster of this species, whereas an overexposure of the same blot was necessary to reveal their presence outside this cluster (Nelson and Honda 1985). It is not clear whether such overexposures were always performed in the studies reviewed here.

Previous studies were able to show that concerted evolution had taken place by showing that some restriction enzyme sites had been fixed in all the family members of one species and were absent in the family members of a related species. For example, the nontranscribed spacer (now called the *intergenic spacer*) of the human rDNA units all have an *HpaI* site which is not present in the rDNA units of chimpanzee (Arneheim et al. 1980). Other studies have shown that different variant subrepeats have been fixed in different species. For example, the rDNA intergenic spacer of *Drosophila melanogaster* contains several copies of a 240–base pair repeat, whereas the rDNA intergenic spacer of *D. virilis* contains 220–base pair repeats showing little sequence similarity to the 240–base pair repeats of *D. melanogaster* (Tautz et al. 1987). An interesting example of the insertion and homogenization of an RNA-mediated transposition event has recently been reported by Gonzalez et al. (1993). While sequencing a copy of the ~40-kb human rDNA repeat units, they found that its intergenic spacer contained a 2-kb processed pseudogene derived from the cell cycle protein cdc27hs. Southern blot analysis showed that this linked processed pseudogene has been homogenized to all the 200 human rDNA units which arc found as five tandemly repeated unit clusters located on different chromosomes.

The data reviewed here show that extensive concerted evolution need not be limited to specific nucleotides, subrepeats, or pseudogenes but that transposed 5S rRNA gene copies can remain functional following their insertion, homogenization, and fixation within diverse tandemly repeated multigene families in the genome of different species.

**Conclusions**

The data reviewed here show that 5S rRNA genes repeatedly became linked to different tandemly repeated multigene families during evolution. Although the mechanism responsible for their transposition still has to be established, the mechanisms by which they would subsequently spread to all members of a tandemly repeats multigene family after having transposed in a single repeat are well known. Such new gene arrangements have not only often become fixed in several species but have also subsequently been lost in related species, presumably through the homogenization and fixation of variant repeats which did not acquire or had lost their 5S rRNA gene copy.

The observation that 5S rRNA genes are found linked either to the coding or noncoding strand of diverse tandemly repeated multigene families clearly shows that the linkage of these genes to the rDNA repeat units of several eukaryotic species is unlikely to represent the remnant of a eubacterial-like arrangement of rDNA operons. The diversity of these linkages, as well as their lack of conservation during evolution, make it unlikely that they provide any selective advantage by allowing transcriptional coregulation of 5S rRNA genes and of the genes to which they are linked, as suggested by some authors (Andrews et al. 1987; Aksoy et al. 1992). Furthermore, even when 5S rRNA genes are encoded by the same strand as the other genes, they are unlikely to be cotranscribed with them since 5S rRNA genes are transcribed by RNA-polymerase III, whereas the other...
rRNA genes are transcribed by RNA-polymerase I, and histone genes are transcribed by RNA-polymerase II. Trans-spliced leader genes are usually believed to be transcribed by RNA-polymerase II, but it has been suggested that they might be transcribed by RNA-polymerase III in Trypanosoma brucei (Grondal et al. 1989; Nelson and Honda 1989; Nilsen et al. 1989; Aksoy et al. 1992). Finally, 5S rRNA gene transcription is now thought to be regulated mainly by the availability of its TFIIIA transcription factor, which binds to their C-box internal promoter (Neigeborn and Warner 1990).

Although we reviewed the linkage of 5S rRNA genes to tandemly repeated gene families, we suspect that they probably transpose equally frequently to other loci. Their apparent overrepresentation in tandemly repeated multigene families may be due to the bias introduced by the interest of molecular evolutionists, for whom rRNA genes are by far the most common genes used to infer species phylogenies, and of molecular biologists interested in the function and evolution of trans-spliced leader and histone gene sequences. As is the case with SINES, where recently transposed copies have been identified by comparing the sequences of orthologous loci from related species, we suspect that recently transposed single copies of 5S rRNA genes will also be found in the same fashion (see, e.g., Murata et al. 1993).

The origin of eu- karyotic cells has been marked by the acquisition of a separate promoter for the nuclear 5S rRNA genes which are now transcribed by RNA polymerase III. This freed them from their original dependence on the eubacterial-like rDNA operons for transcription and allowed them to move to different chromosomal location(s). Today, they are still often found located at loci independent of the rDNA loci in the genome of several eukaryotic species, an observation that is still often thought as being universal. We hope this review has brought some qualification to this common belief and has provided yet another convincing example of the dramatic effects that fortuitous recombination events can have on the structure of eukaryotic genomes when coupled with random genetic drift.

Acknowledgments

We would like to thank Linda Bonen, Donal A. Hickey, and Douglas A. Johnson as well as Thomas H. Eickbush and two anonymous reviewers for their comments on a previous version of the manuscript. G.D.'s work is supported by a research grant from the Natural Sciences and Engineering Research Council of Canada.

LITERATURE CITED


Hofmann, J., T. WInckler, A. HANENKAMP, M. BUENBERGER, G. SCHUMANN, R. MARSCHALEK, and T. DINGWELL.
Concerted Evolution of Linked 5S rRNA Genes 491


THOMAS H. EICKBUSH, reviewing editor

Received July 21, 1994

Accepted November 18, 1994