Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm

Alejandro P. Rooney* and Todd J. Ward

Microbial Genomics and Bioprocessing Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604

Edited by Masatoshi Nei, Pennsylvania State University, University Park, PA, and approved February 17, 2005 (received for review December 23, 2004)

In eukaryotes, the primary components of the ribosome are encoded by multicopy nuclear ribosomal RNA (rRNA) genes: 28S, 18S, 5.8S, and 5S. Copies of these genes are typically localized within tandem arrays and homogenized within a genome. As a result, nuclear rRNA gene families have become a paradigm of concerted evolution. In filamentous fungi of the subphylum Pezizomycotina, 5S rRNA genes exist as a large and dispersed multigene family, with between 50 and 100 copies per genome. To determine whether these genes defy the concerted evolution paradigm, we examined the patterns of evolution of these genes by using sequences from the complete genomes of four species. Analyses of these sequences revealed (i) multiple 5S gene types within a genome, (ii) interspecies clustering of gene types, (iii) multiple identical gene types shared among species, (iv) multiple pseudogenes within a genome, and (v) presence/absence variation of individual 5S copies in comparisons of closely related species. These results demonstrate that the 5S family in these species is characterized by birth-and-death evolution under strong purifying selection. Furthermore, our results suggest that birth-and-death evolution occurs at different rates in the genera examined, and that the multiplication and movement of 5S genes across the genome are highly dynamic. As such, we hypothesize that a mechanism resembling retroposition controls 5S rRNA gene amplification, dispersal, and integration in the genomes of filamentous fungi.

5S rRNA | retroposition

Eukaryotes possess four nuclear ribosomal RNA (rRNA) genes: 28S/26S, 18S, 5.8S, and 5S. These genes are repeated several hundred to a few thousand times in the genome and are nearly always arranged either in a single large tandem repeat array or in multiple tandem arrays found on one or a few chromosomes. Brown et al. (1) first proposed the idea that rRNA gene copies evolve “horizontally,” meaning that a mutation that arises in one copy spreads to all other copies, eventually homogenizing all of them. Later, this process was called “concerted evolution” (2). Because concerted evolution theory could explain the observed lack of genetic variability among rRNA gene copies in many different species, it eventually became dogma that it was the fundamental mode of rRNA multigene family evolution (3–5).

Although the overwhelming majority of studies have shown that most rRNA multigene families evolve in a strictly concerted manner (meaning that all gene copies in the family are homogenized), some exceptional cases have been identified in actinobacteria, oak trees, flatworms, and apicomplexan protists (6–9). In the case of actinobacteria, errors during DNA replication are assumed to be the cause when horizontal transfer can be ruled out (10). In the flatworm (7, 8) and oak tree (11) species, there apparently have been “escapees” from concerted evolution, in which a few distinct rRNA types are maintained within the genome. Still, all copies of each distinct type are homogenized by concerted evolution in these exceptional cases. Apicomplexans present a rather interesting case of birth-and-death evolution of a small rRNA gene family (9). In these organisms, rRNA genes are dispersed throughout the genome and are not organized in a tandem array. This dispersed gene organization apparently facilitates birth-and-death evolution wherein rRNA genes diversify from one another, some of which are unique to a given species, whereas others are shared among species (9).

Unlike the other nuclear rRNA genes, the 5S genes of most eukaryotes are dispersed throughout the genome [e.g., as in Schizosaccharomyces pombe (12)] or found in their own distinct tandem arrays [e.g., as in soybeans (13)]. In some species, both types of gene arrangement are found [e.g., as in humans (14)]. This flexibility of 5S rRNA gene organization in different genomes is rather unusual and reminiscent of what is observed in some large multigene families that undergo birth-and-death evolution (Fig. 1), such as histones (15). Therefore, the purpose of this study was to investigate 5S rRNA multigene family evolution in filamentous fungi, because the genomes of four relatively closely related species from this group have been completely sequenced, are reasonably well assembled, and contain 5S rRNA gene families with 50–100 copies.

Materials and Methods

Genomic BLAST searches were used to mine 5S gene copies from the complete genomes of Aspergillus nidulans FGSC A4, Fusarium graminearum NRRL 31084 (also known as Gibberella zeae PH-1), Magnaporthe grisea 70–15, and Neurospora crassa OR74A (www.ncbi.nlm.nih.gov/ genomes/FUNGI/funtab.html). Initial searches in each genome were performed by using a 5S rRNA gene sequence from A. nidulans (GenBank accession no. K03162). Subsequent searches were performed by using genome-specific sequences obtained from the initial search. This process was repeated with different sequences until no new gene copies were found. Genomic BLAST search results occasionally included partial sequences of ~40–50% of the complete sequence length (119–120 nt). In all cases, the regions immediately flanking each partial sequence were directly examined to determine whether the sequence was truly incomplete or whether it merely contained large numbers of substitutions relative to the query sequence. Once all of the 5S genes were extracted from each of the above genomes, an alignment was constructed by using the computer program SEQUENCER 4.1.2 (Gene Codes, Ann Arbor, MI). The resulting alignment was subsequently inspected for errors but required virtually no editing because of the highly conserved nature of the sequences.

For each genome, we randomly chose several loci to screen for the presence of orthologous 5S genes in closely related species.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: rRNA, ribosomal RNA; LINE, long interspersed element; SINE, short interspersed element.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY924880–AY924946).

*To whom correspondence should be addressed. E-mail: rooney@ncuar.usda.gov.
Oligonucleotide primers in predicted protein-coding sequences flanking each selected 5S gene copy were used to amplify the corresponding region in related species. If a locus failed to amplify with primers set in protein-coding regions, we designed alternative primer sets in which one of the primers was located in the 5S rRNA gene itself. PCR amplifications were performed with Platinum TaqDNA polymerase high fidelity (Invitrogen) under standard reaction conditions. The oligonucleotide primer sequences and specific PCR reaction conditions are available upon request. Amplification products were purified by using Montage PCR cleanup filter plates (Millipore). Sequencing reactions were conducted as described in Ward et al. (16) by using BigDye 3.0 sequencing chemistry (Applied Biosystems). Reaction products were purified by ethanol precipitation and run on an ABI3100 or ABI3730 genetic analyzer (Applied Biosystems). DNA sequences were edited by using Sequencher 4.1.2 and added to the alignment containing the genomic sequences. The sequences have been deposited in GenBank under accession nos. AY924880–AY924946.

The computer program MEGA 2 (17) was used to conduct all phylogenetic and nucleotide sequence analyses. Kimura two-parameter distances (18) were computed and used to generate phylogenetic trees by using the neighbor-joining method (19). The statistical reliabilities of the internal branches were assessed for all trees by using 1,000 bootstrap pseudoreplicates.

Results

The 5S Gene Complement and Sequence Divergence Levels. The size of the 5S rRNA gene complement in the A. nidulans, F. graminearum, M. grisea, and N. crassa genomes is consistent with expectations based on genome size (Table 1). Accordingly, N. crassa, which has the largest genome size (43 Mb), possesses the greatest number of total 5S gene copies (92), and A. nidulans, which has the smallest genome size (31 Mb), possesses the least number of total gene copies (52). The F. graminearum and M. grisea genomes are approximately the same size (40 Mb) and have similar numbers of 5S gene copies (66 and 71, respectively). Of the total 5S gene copy number in each genome, a portion was comprised of truncated 5S gene copies (Table 1), but the majority consisted of intact gene copies. There are a total of 36 intact copies in the A. nidulans genome but only 16 distinct copy types, because 20 copies were identical to one or more of the others (Table 1). Similarly, the M. grisea genome contains 38 intact copies but only 17 distinct types. In contrast, the F. graminearum genome has 59 intact copies and 24 distinct types, whereas the N. crassa genome has 79 intact copies and 29 distinct types.

When the distances between the intact copies within each of the four genomes were analyzed, we found that 5S genes in the A. nidulans and M. grisea genomes are much more closely related than those of the F. graminearum and N. crassa genomes. For example, the average distance between the 16 distinct types in the A. nidulans genome is \( \bar{d} = 0.03 \). Similarly, the 17 distinct types in the M. grisea genome differed by \( \bar{d} = 0.02 \). In contrast, the F. graminearum genome contains 24 distinct types that show an average divergence of \( \bar{d} = 0.12 \), and the N. crassa genome contains 29 distinct types that show an average divergence of \( \bar{d} = 0.11 \). These average distances were compared to the estimates generated for all intact types, we found that they were either the same or highly similar (Table 1). In contrast, the estimates differed substantially between the N. crassa and F. graminearum genomes vs. the A. nidulans and M. grisea genomes (Table 1). Thus, 5S gene sequences are much more closely related in the A. nidulans and M. grisea genomes than in the F. graminearum and N. crassa genomes.

Pseudogenes and Birth-and-Death Evolution. As mentioned previously, in all four genomes, we found that a portion of the total gene number was comprised of truncated 5S gene sequences. These truncated sequences are \( \approx 40-50\% \) of the intact gene sequence length. We consider these truncated sequences to be pseudogenes because their lack of an intact coding sequence effectively destroys the secondary structure of the 5S rRNA molecule that they would have otherwise encoded. The presence of pseudogenes in a multigene family strongly suggests that the family evolves under a birth-and-death process (9, 15, 20). Under this process, a multigene family both expands because of gene duplication and contracts because of gene loss (e.g., as a result of unequal crossover). Thus, some family members exist in the genome for only short periods of evolutionary time, whereas others persist for much longer periods. Eventually, distinct gene copies accumulate differences, leading some of them to degenerate into pseudogenes. Birth-and-death evolution can be detected through two means: (i) a phylogenetic analysis of multigene family members, whereby a between-species gene clustering pattern will be evident; and (ii) through an examination of gene sequence divergence levels, in which case a relatively high proportion of changes between gene family members indicates a lack of homogenization (9, 15, 20) (Fig. 1). In contrast, sequences evolving in a concerted fashion will show evidence of extensive homogenization due to gene conversion or unequal crossover, resulting in a pattern of within-species gene clustering accom-

<table>
<thead>
<tr>
<th>Genome size, Mb</th>
<th>Total number of gene copies</th>
<th>Number of intact gene copies(^*)</th>
<th>Number of distinct intact copy types(^*)</th>
<th>Number of pseudogene copies</th>
<th>Proportion of intact copies/Mb</th>
<th>Proportion of pseudogene copies/Mb(^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. nidulans</td>
<td>31</td>
<td>52</td>
<td>36 (0.03)</td>
<td>16 (0.03)</td>
<td>16</td>
<td>1.16</td>
</tr>
<tr>
<td>M. grisea</td>
<td>40</td>
<td>71</td>
<td>38 (0.01)</td>
<td>17 (0.02)</td>
<td>33</td>
<td>0.95</td>
</tr>
<tr>
<td>F. graminearum</td>
<td>40</td>
<td>66</td>
<td>59 (0.09)</td>
<td>24 (0.12)</td>
<td>7</td>
<td>1.48</td>
</tr>
<tr>
<td>N. crassa</td>
<td>43</td>
<td>92</td>
<td>78 (0.07)</td>
<td>29 (0.11)</td>
<td>14</td>
<td>1.81</td>
</tr>
</tbody>
</table>

\(^*\)Average \( \bar{d} \) distance among sequences is given in parentheses.

\(^*\)Results for comparisons of pseudogene frequency by using a two-tailed \( G \) test were significant for A. nidulans vs. F. graminearum (\( P = 0.05 \)), M. grisea vs. F. graminearum (\( P = 0.007 \)), and M. grisea vs. N. crassa (\( P = 0.02 \)).
panied by very small distance estimates in comparisons between gene copies of the same organism (9, 15, 20).

We analyzed the phylogenetic relationships between the 5S genes from all four genomes along with sequences from closely related taxa that we obtained from our own laboratory work (Figs. 2 and 3). Comparison of the gene copies from the four complete genome sequences provided little evidence of interspecies clustering of 5S types. All 5S copies from the N. crassa, M. grisea, and A. nidulans genome sequences formed individual monophyletic groups, with 5S copies from the F. graminearum genome more broadly distributed in a weakly supported, but paraphyletic arrangement (Fig. 2). However, comparisons involving closely related species revealed extensive interspecies clustering and the presence of identical 5S types shared among species (Figs. 2 and 3). This is best seen in the case of F. graminearum (Fig. 2), for which we had samples of several closely related species. The 5S rRNA genes are split into several multispecies clades, some of which are well supported by relatively high bootstrap values. Because these clades represent orthologous loci (Fig. 2), the phylogeny is characterized by locus-specific clustering, as genes that were amplified and sequenced from a particular locus cluster together to the exclusion of genes from other loci regardless of the species from which they came (Fig. 2). The one exception involves a cluster in which the genes present at three loci were identical in all species examined (Fig. 2, G–I). However, this observation is not unexpected if the genes originated from a recent duplication event and/or if purifying selection is strong. In fact, the important observations that can be made from examining this phylogeny are that (i) 5S gene sequences form several clusters composed of different species and (ii) the gene sequences in these clusters are either identical to or display only one or two differences from each other. Taken together, these observations indicate that 5S nucleotide sequence homogeneity is maintained by purifying selection, not concerted evolution, and that the gene copies evolve under a birth-and-death process. The same patterns were found for Aspergillus and Neurospora (Fig. 3), although we screened fewer sister species and 5S gene loci in these cases.

Fig. 1. Expected gene relationships under the concerted evolution model (A) and the birth-and-death evolution model (B). Species are denoted by Greek letters, and genes are denoted by numbers. Under the concerted evolution model, genes cluster according to species; however, they do not under the birth-and-death model, except in cases of recent gene duplication (e.g., the a3 sequences in B). The figure was adapted from Rooney (9).

Fig. 2. Phylogeny of 5S rRNA genes from the filamentous fungal species examined in this study. Because of the large number of sequences analyzed, we show only the Fusarium gene sequences. The gene sequences from Aspergillus, Magnaporthe, and Neurospora are shown in Fig. 3. Each Fusarium species is shown in a different color. The clusters labeled A–I represent distinct loci.
Evidence for birth-and-death evolution was also found as a result of identifying instances in which specific 5S copies from the complete genomes were absent from the genomes of closely related species or strains. For example, at the top of Fig. 2, one clade represents a locus that contains sequences from only five Fusarium species; apparently, the other two (F. asiaticum and F. sporotrichioides) did not possess a 5S sequence at that locus. This suggests that the 5S gene was lost at that locus in the latter two taxa. We made similar observations for the Neurospora and Aspergillus species that we examined (Fig. 3). For instance, for one of the N. crassa loci that we examined, N. crassa and N. intermedia had a 5S gene at one locus but N. africana did not (Fig. 3). We also saw evidence of gene “birth” in our laboratory analyses. This was shown best in our analysis of M. grisea and its relatives. Here, we looked at one locus in three M. grisea genome strain relatives: (i) a strain of M. grisea that was not the genome strain, (ii) M. salvinii (a close relative of M. grisea), and (iii) Gaeumannomyces graminis var. tritici (a member of a genus

Fig. 3. Phylogeny of the 5S rRNA genes from Aspergillus (A), Magnaporthe (B), and Neurospora (C). Each species is shown in a different color.
closely related to Magnaporthe). None of these isolates possessed a 5S gene at that locus. The most parsimonious explanation is that the 5S gene was inserted in the M. grisea genome strain after it had diverged from these other genomes.

**Discussion**

For several decades now, researchers have been interested in understanding how rRNA genes and gene families diversify and evolve (e.g., see refs. 1, 3, 5, 7–9, and 21–27). Brown et al. (1) first proposed that rRNA genes evolve “horizontally” or in a concerted fashion on the basis of restriction digest patterns. Since then, rRNA multigene families have come to be viewed as a paradigm of concerted evolution (4, 28–31). Only rarely have any departures from this model been found, the most notable of which were mentioned at the beginning of this paper. Moreover, virtually all of these departures from the rRNA concerted evolution paradigm are limited in scope and usually involve the existence of one or a few divergent types (e.g., see refs. 7, 8, and 11). In addition, these departing types encompass many gene copies that are highly homogenized. So far, the only identified case of an rRNA multigene family undergoing birth-and-death evolution is the apicomplexan rRNA multigene family (9). This family encompasses a small number of copies (usually five to seven) that have a dispersed genomewide organization, which is quite unlike what is seen in most other eukaryotes. In this paper, we identify an example of a large dispersed rRNA multigene family undergoing birth-and-death evolution: the 5S rRNA genes of filamentous fungi within the ascomycete subphylum Pezizomycotina, as represented by the genera Aspergillus, Fusarium, Magnaporthe, and Neurospora.

Previous studies claimed to show that the 5S genes of A. nidulans (25) and N. crassa (26) undergo concerted evolution. However, we did not find this to be true for these species or for M. grisea or the Fusarium species on the basis of nucleotide sequence divergence levels, phylogenetic clustering patterns, and the existence of 5S pseudogenes. It is true that we observed multiple copies of identical gene sequences. However, the reasons for this are (i) strong purifying selection and (ii) frequent 5S gene copying and multiplication resulting in rapid gene turnover. When combined, these processes will curtail the chances of extensive divergence between genes. In a study on SS rRNA gene evolution in the filamentous fungi Epichloë and Neotyphodium, which are also members of the subphylum Pezizomycotina, Ganley and Scott (32) claimed that the 5S genes of these species evolve only partially in concert, with limited divergence between genes. After examining their results, we conclude that a more likely explanation is that the genes undergo birth-and-death evolution with strong purifying selection and rapid gene turnover.

Depending on how rapid the gene turnover rate is, the distinct gene types produced through birth-and-death evolution will show varying degrees of divergence from each other. In our study, we found that the gene types of A. nidulans and M. grisea were much more closely related than were the gene types of F. graminearum and N. crassa (Table 1), suggesting that rates of gene turnover are higher in the latter two genomes than in the former two. This conclusion is consistent with the finding that the genomes possessing very closely related gene types (A. nidulans and M. grisea) had a higher frequency of pseudogenes than the genomes whose gene types were more divergent (F. graminearum and N. crassa) (Table 1). Moreover, there was a strong, although not statistically significant ($r = -0.912$) between the number of pseudogenes per megabase and the average level of nucleotide sequence divergence among 5S gene types (Table 1). Overall, our results show that there are high levels of gene amplification and dispersal in all of the species studied, indicating that their genomes are highly dynamic. However, the rate of birth and death of 5S copies appears to be different in the four pezizomycete genera examined, with Magnaporthe having the fastest rate and Fusarium having the slowest rate. More generally, our results suggest that pseudogene formation is linked to rates of gene turnover in this large multigene family.

Given the prevalence of concerted evolution among tandemly arrayed rRNA gene families, some might hypothesize that the dispersed nature of the apicomplexan and ascomycete rRNA gene families is responsible for their exceptional patterns of evolution. Although this might be true of apicomplexan rRNA genes (9), the dispersed organization of pezizomycotine 5S genes is more a result of the birth-and-death process rather than a facilitator of it. For the most part, a dispersed genomic organization of multigene family members ordinarily arises through a process involving gene duplication followed by insertion of copies into other areas of the genome via unequal crossover, inversion and/or translocation. Mobile elements, on the other hand, use a process of transposition or retrotransposition to amplify, disperse, and subsequently integrate copies into the genome. Our findings appear to be more in line with this type of process for best explaining the pattern of 5S gene organization and dispersal observed in this study. Gene amplification involves the copying of a progenitor gene or genetic element and subsequent insertion of the daughter copies into other areas of the genome. The retrotransposition of long interspersed elements (LINEs) and short interspersed elements (SINEs) are classic examples of amplification (28–30, 33). The amplification of a LINE occurs when its primary transcript is reverse transcribed, copied, and then integrated into another area of the genome. SINEs operate essentially the same way but rely on LINEs, at least in part, for their own retroposition, amplification, and genomic integration (34–36).

We hypothesize that 5S genes are capable of multiplying and integrating into other areas of the genome through a process the same as, or akin to, retroposition. The 5S rRNA genes and SINEs have the same type of internal RNA polymerase III promoter (33, 37). Furthermore, Kapitonov and Jurka (38) have identified an unique class of SINEs that are derived from the fusion of a 5S rRNA gene and a LINE, showing that SS rRNA genes and retroelements can indeed interact with one another. A number of retroelements, including SINEs, LINEs, and related retrotransposons, have been identified in A. nidulans (39), F. graminearum (40), M. grisea (41, 42), N. crassa (43), and a number of other related fungal species (44–49). Interestingly, Essnault et al. (50) have shown that LINEs are capable of amplifying and integrating non-LINE mRNA transcripts by turning them into processed pseudogenes and then integrating them into other parts of the genome. Apparently, the mRNA transcripts redirect the LINE enzymatic machinery to accomplish this. As such, it is interesting to consider the possibility that the 5S rRNA gene of filamentous fungi might also be capable of redirecting the enzymatic machinery of LINEs, LINE-like elements, or even some other retroelement for their own amplification, dispersal, and genomic integration.

We thank J. R. Dettman (University of California, Berkeley) and S. B. Goodwin, K. O’Donnell, and S. W. Peterson (Agricultural Research Service, U.S. Department of Agriculture) for providing fungal strains. We are grateful to T. Usgaard and J. Robinson for laboratory support.