



Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data

Austen R.D. Ganley and Takehiko Kobayashi

Genome Res. published online January 2, 2007
Access the most recent version at doi:[10.1101/gr.5457707](https://doi.org/10.1101/gr.5457707)

P<P Published online January 2, 2007 in advance of the print journal.

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2007, Cold Spring Harbor Laboratory Press

Highly efficient concerted evolution in the ribosomal DNA repeats: Total rDNA repeat variation revealed by whole-genome shotgun sequence data

Austen R.D. Ganley^{1,3,4,5} and Takehiko Kobayashi^{1,2,4}

¹National Institute for Basic Biology, Okazaki, 444-8585 Japan; ²SOKENDAI, Okazaki, 444-8585 Japan; ³Department of Biology, Duke University, Durham, North Carolina 27708, USA

Repeat families within genomes are often maintained with similar sequences. Traditionally, this has been explained by concerted evolution, where repeats in an array evolve “in concert” with the same sequence via continual turnover of repeats by recombination. Another form of evolution, birth-and-death evolution, can also explain this pattern, although in this case selection is the critical force maintaining the repeats. The level of intragenomic variation is the key difference between these two forms of evolution. The prohibitive size and repetitive nature of large repeat arrays have made determination of the absolute level of intragenomic repeat variability difficult, thus there is little evidence to support concerted evolution over birth-and-death evolution for many large repeat arrays. Here we use whole-genome shotgun sequence data from the genome projects of five fungal species to reveal absolute levels of sequence variation within the ribosomal RNA gene repeats (rDNA). The level of sequence variation is remarkably low. Furthermore, the polymorphisms that are detected are not functionally constrained and seem to exist beneath the level of selection. These results suggest the rDNA is evolving via concerted evolution. Comparisons with a repeat array undergoing birth-and-death evolution provide a clear contrast in the level of repeat array variation between these two forms of evolution, confirming that the rDNA indeed does evolve via concerted evolution. These low levels of intra-genomic variation are consistent with a model of concerted evolution in which homogenization is very rapid and efficiently maintains highly similar repeat arrays.

[Supplemental material is available online at www.genome.org.]

Repetitive elements are an abundant feature of genomes (Britten and Kohne 1968) and play critical roles in cell biology, genome structure, and adaptive evolution (Andersson et al. 1998; Shapiro and von Sternberg 2005). In many cases, repeats from a repeat family are highly similar to one another within a genome, a pattern that persists through evolutionary time even though sequence differences are apparent between species (Brown et al. 1972). Thus, the repeats seem to be maintained as a coherent family. This pattern is contrary to the expectations of conventional evolution, where repeats are expected to evolve independently (Fig. 1) and diverge with time. This unusual mode of evolution, where repeats within a genome are more similar to each other than they are to “orthologous” repeats in a related species, is defined as concerted evolution (Zimmer et al. 1980). The molecular process responsible for concerted evolution is known as homogenization (Dover 1982), and although not fully elucidated, is thought to involve continual turnover of repeat copies by unequal recombination (e.g., Smith 1973; Szostak and Wu 1980; Kobayashi et al. 1998). However, recent studies have shown that several repeat families previously thought to evolve by concerted evolution actually evolve via a different evolutionary process known as birth-and-death evolution (e.g., Nei et al. 1997, 2000; Rooney and Ward 2005). In birth-and-death evolution it is selection, not homogenization, that main-

tains the repeats as a coherent family, and occasional duplication/deletion is also thought to play a role. Therefore, relatively high levels of intragenomic repeat variation are expected in regions of low selective constraint (e.g., synonymous sites and noncoding regions) under birth-and-death evolution. To distinguish between concerted evolution and birth-and-death evolution, knowledge of the level of repeat variation is critical. Despite this, the level of intragenomic repeat variation has not been reported for large repeat arrays; therefore, evidence for concerted evolution of these arrays, as per the classical definition, is lacking.

The major problems in detecting variation within large tandem-repeat arrays are the prohibitive size and high levels of similarity of these arrays. Therefore, they represent uncharted territories in the genome, and are usually assumed to be comprised of repeats with identical sequences (e.g., Goffeau et al. 1996). However, it is possible that substantial cryptic variation exists within these repeat arrays, as the power to detect such cryptic variation is low (e.g., Williams et al. 1990; Copenhaver and Pikaard 1996; O'Donnell and Cigelnik 1997; Gonzalez and Sylvester 2001) and it is not likely to be noticed during routine applications (such as sequencing for phylogenetic analyses). The advent of whole-genome shotgun sequencing (WGSS) now gives us a way to determine the total level of variation in repeat arrays. In principle, WGSS methodology sequences the entire genome, with all parts receiving equal coverage (in contrast to map-based genome sequencing, where highly repetitive regions are usually not sequenced [Goffeau et al. 1996]). Therefore, information on the absolute level of sequence variation present within repeat arrays is contained within WGSS data.

⁴Present address: Division of Cytogenetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan.

⁵Corresponding author.

E-mail gausten@lab.nig.ac.jp; fax 81-055-981-6881.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5457707>.

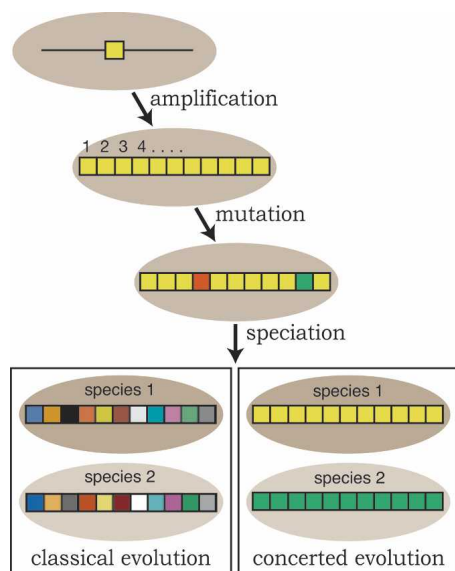


Figure 1. Classical evolution vs. concerted evolution. Repeats (individual boxes) in a repeat array are initially formed by a gene amplification event. The repeats accumulate mutations (alternatively colored boxes) through time. Under classical evolution these mutations persist and therefore, after speciation events, the orthologous relationships of the repeats remain (e.g., repeat #1 from species 1 will resemble repeat #1 from species 2 more closely than the other repeats in species 1; indicated schematically by different shades of the same color). However, under concerted evolution, homogenization continuously sweeps one repeat variant (at random) to fixation within the array. Therefore, the repeats within a genome are all expected to be similar, but differ in sequence from the repeats in a closely related species. Birth-and-death evolution is a more complex form of classical evolution mixed with some aspects of concerted evolution, and is not depicted.

The ribosomal RNA repeats (rDNA) are an extensively studied repetitive gene family. Each repeat unit contains three ribosomal RNA (rRNA) genes, the large subunit (LSU), small subunit (SSU), and 5.8S rRNA genes, as well as two transcribed spacers (the ITS1 and ITS2) and a large intergenic spacer (IGS) (Long and Dawid 1980). Variably, another rRNA gene, the 5S, may be present within the IGS. rDNA copy number varies widely; in most eukaryotes it is between 30 and 30,000 copies (Prokopowich et al. 2003) and the repeats are organized tandemly at one or more sites per haploid genome.

Fungi are an ideal group of eukaryotes to study at a genomic level due to their relative small genome sizes. We chose five fungi encompassing a diverse range of species for which WGSS data were available. These species were chosen as all have a simple rDNA organization, i.e., a single rDNA array. The first four are members of the phylum Ascomycota. *Saccharomyces cerevisiae* is a budding yeast from the family *Saccharomycetaceae*, and is a widely used model organism for biological research (the other model yeast, *Schizosaccharomyces pombe* was not included as it contains two rDNA arrays). The haploid genome sequence was completed by the Broad Institute (http://www.broad.mit.edu/annotation/genome/saccharomyces_cerevisiae/Home.html). *Saccharomyces paradoxus* is a close relative of *S. cerevisiae*. The diploid genome sequence was completed by the Broad Institute as part of a comparative study of *Saccharomycetaceae* genomes (Kellis et al. 2003) (http://www.broad.mit.edu/annotation/fungi/comp_yeasts/). *Ashbya gossypii* is a filamentous fungus from the *Saccharomycetaceae* family. It is a mild pathogen of cotton, and

has been used industrially to produce high vitamin B₂ levels. The haploid genome sequence was recently completed (Dietrich et al. 2004) (<http://agd.unibas.ch/>). *Aspergillus nidulans* is a common soil mold from the order Eurotiales. It has been a popular model organism for genetics for over 50 yr and is filamentous in growth form. The haploid genome sequence was completed by the Broad Institute (<http://www.broad.mit.edu/annotation/fungi/aspergillus/>). The final species, *Cryptococcus neoformans*, is a member of the phylum Basidiomycota, which includes the mushrooms. However, *C. neoformans* grows as a yeast and is a severe pathogen of immunocompromised humans. The haploid genome sequence was recently completed by the Stanford Genome Technology Center (Loftus et al. 2005) (<http://www-sequence.stanford.edu/group/C.neoformans/index.html>).

In this study we identified all of the rDNA reads from WGSS projects of these five fungal species and assembled these reads and scanned them for polymorphisms to identify the total level of variation in the rDNA. The results give us the first complete picture of the level of variation within a large, tandemly arrayed repeat family, and this variation is very low. We discuss the implications of these results for the evolution of the rDNA repeats.

Results

The rDNA reads from the five genome projects were each assembled into a rDNA unit alignment. The high similarity of the repeats in the rDNA means there is no way to determine which repeat in the array any given sequence read comes from (e.g., whether a given read comes from repeat #1 or repeat #100). Instead, the sequence data collapse down to a single rDNA unit alignment whose sequence coverage is a product of the genome coverage level and rDNA copy number. We performed automated searches, followed by manual corrections on these alignments to identify polymorphisms (see Methods for details). The polymorphisms that were identified were put into two classes, high-confidence and low-confidence polymorphisms, and results for both are presented. The low-confidence polymorphisms occur in areas of low-sequence quality, and thus, we believe that most of them are unlikely to be real, but we have included them for completeness. We found no evidence for any rearranged rDNA units, such as those found recently in humans (Caburet et al. 2005) in any of the five species.

Low-sequence variation within the rDNA repeats

Surprisingly, we found the number of polymorphisms within the rDNA arrays of all five species is very low. In *A. gossypii*, only three polymorphisms (six, including low-confidence polymorphisms) were found; *S. cerevisiae* contained four (or seven); *S. paradoxus* contained 13 (or 16); *A. nidulans* 11 (or 20); and the number in *C. neoformans* was slightly higher at 37 (or 43). The numbers of polymorphisms and their locations within the rDNA unit are presented in Figure 2. Most polymorphisms are present on only one or a few reads; however, three polymorphisms (two from *S. cerevisiae* and one from *S. paradoxus*) are present on a relatively high number of reads, indicating high copy number in the array (Fig. 2).

To make meaningful interpretations of these polymorphism levels, it is necessary to relate the polymorphism level to the copy number of the rDNA (as more copies are expected to harbor a greater absolute polymorphism level). Therefore, we determined

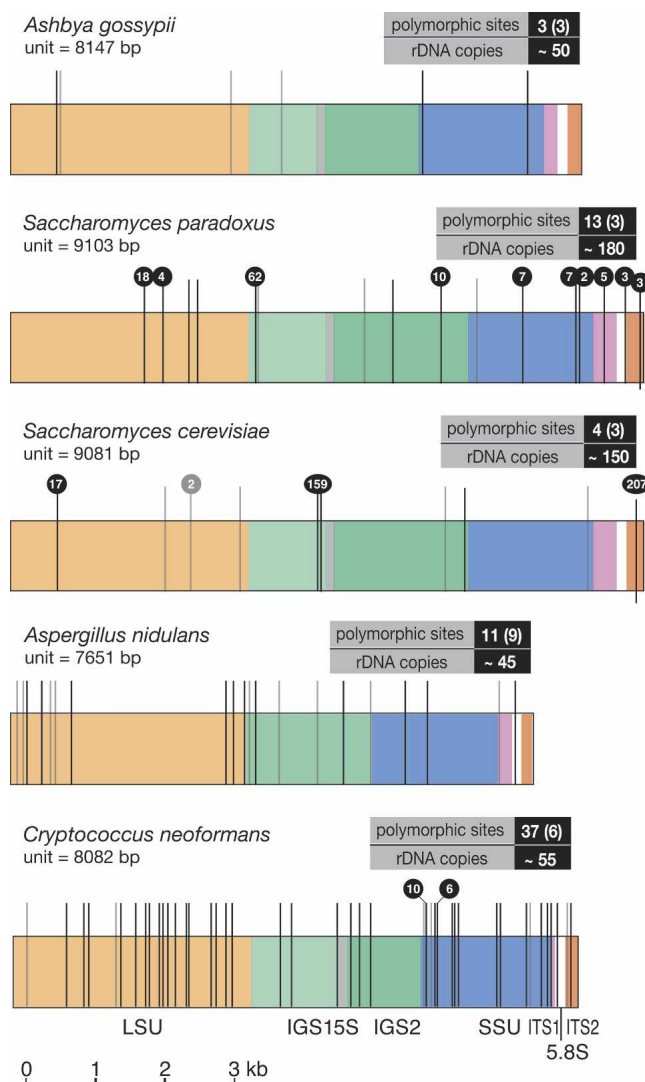


Figure 2. Positions of the polymorphisms in the rDNA unit. Polymorphic sites for each species are plotted onto a map of a single rDNA unit. rDNA features are color coded as indicated beneath the *bottom* rDNA unit. High-confidence polymorphisms are shown as black lines, low-confidence polymorphisms as gray lines, and the total numbers are boxed (low-confidence polymorphisms in parentheses). Although the polymorphisms are shown in a single repeat, in reality they are likely to be scattered throughout the array. Polymorphisms present in more than one sequence read are indicated by black balls above the line, with the number of reads shown in the ball. In many cases these are likely to result from the coverage level of the whole-genome shotgun sequence data. The rDNA unit length and copy number (from Fig. 3) are also indicated (*S. paradoxus* is the diploid copy number). Diagram is to scale.

rDNA copy number for these species. *A. gossypii* was previously reported to contain ~50 rDNA copies (Wendland et al. 1999) and *S. cerevisiae* is known to contain ~150 copies (Kobayashi et al. 1998). To determine rDNA copy number from the remaining three species, pulsed-field gel electrophoresis was performed on genomic DNA digested with restriction enzymes that do not cut within the rDNA (Fig. 3). The haploid rDNA copy number can then be calculated from the size of the resultant rDNA band. Using *S. cerevisiae* chromosomes as size markers, the sizes of the rDNA-containing bands are estimated as follows: *S. paradoxus* = ~850 kb; *A. nidulans* = ~360 kb; and *C. neoformans* = ~440 kb.

From these values, we calculated rDNA copy number: *S. paradoxus* contains ~90 rDNA copies (per haploid genome); *A. nidulans* ~45 copies; and *C. neoformans* ~55 copies. Two sharp bands are present in *S. paradoxus*, consistent with slight allelic variation in rDNA copy number in this diploid. The rDNA band is much more diffuse in *A. nidulans*, consistent with more copy-number variation. This is expected, as this fungus is filamentous; therefore, the time back to the last single nucleus is expected to be large, resulting in greater variation as a result of the continuous copy-number changes that occur during growth (Cowen et al. 2000).

To put these levels of variation in perspective, we compared the level of variation found here in the rDNA with the level of variation present in a repeat family undergoing birth-and-death evolution. We used the nucleotide diversity (π) statistic (Nei and Li 1979) to represent the level of repeat array variation, as it can be used to compare repeat families with different copy numbers. The polyubiquitin gene repeats (poly-u repeats) were chosen for comparison, as they were previously shown to be undergoing birth-and-death evolution rather than concerted evolution (Nei et al. 2000). Within a repeat array the amino acid sequences of the poly-u repeats are usually invariant, but synonymous sites in the nucleotide sequence are variable. DNA sequences of the poly-u repeats from the five fungal species used in the rDNA analyses were obtained from the relevant genome projects. These have four to six poly-u repeats in a single tandem array. The level of π within each array was calculated for each species for both the rDNA and poly-u repeats (Table 1). The contrast in variation between the rDNA and poly-u repeats is stark, with the poly-u repeats showing three to four orders of magnitude more diversity than the rDNA repeats. Many of the rDNA polymorphisms we observe are present on only a single sequence read, and these are probably sequence errors (see Discussion). If so, they artificially inflate the level of π , and therefore, we have also calculated π for the rDNA polymorphisms using only those polymorphisms present in more than one read (Table 1). To rule out drastic differences in evolutionary rates being responsible for these results (i.e., to rule out the possibility that the lower rate of intra-genomic variation in the rDNA is the result of sequence change intolerance), we measured the inter-specific divergences for these

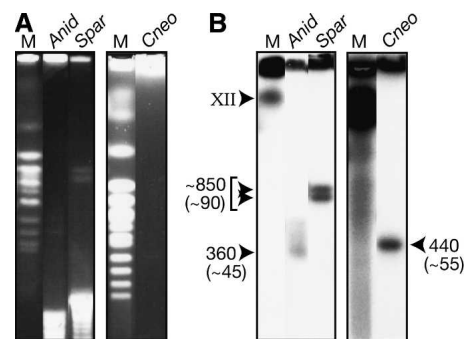


Figure 3. Determination of rDNA array sizes by pulsed field gel electrophoresis. (A) Ethidium bromide-stained gels showing the sizes of the rDNA arrays from *A. nidulans*, *S. paradoxus*, and *C. neoformans* after digestion of chromosomal plugs with *Hin*DIII, *Bam*HI, and *Age*I, respectively. (B) Southern blot of the gels from A probed with a conserved region of the LSU from *S. cerevisiae* to confirm the rDNA bands. Array sizes (kilobase) are indicated, as are rDNA copy numbers (in parentheses), and these were calculated using *S. cerevisiae* chromosomes as size markers (M).

Table 1. Intragenomic nucleotide diversity (π) between repeats in the array

Species	rDNA π^a	Poly-u repeats π	Ratio ^b	rDNA multi π^c
<i>Ashbya gossypii</i>	0.01×10^{-3}	88.6×10^{-3}	8860	0
<i>Saccharomyces cerevisiae</i>	0.08×10^{-3}	151×10^{-3}	1890	0.08×10^{-3}
<i>Saccharomyces paradoxus</i>	0.08×10^{-3}	141×10^{-3}	1760	0.06×10^{-3}
<i>Aspergillus nidulans</i>	0.07×10^{-3}	197×10^{-3}	2830	0
<i>Cryptococcus neoformans</i>	0.18×10^{-3}	157×10^{-3}	871	0.02×10^{-3}

^aThe total number of polymorphisms found in the rDNA is dependent on the coverage level. Therefore, observed polymorphisms were corrected to “real” polymorphism number using genome coverage level. All single-read polymorphisms were included as single-copy polymorphisms, and only high-confidence polymorphisms were used (e.g., we calculated a total of 40 polymorphic sites in *C. neoformans* among the 55 copies).

^bThe fold increase of poly-u repeat π relative to rDNA repeat π .

^cThe value of π when only polymorphisms present on more than one sequence read are counted.

two loci. These were calculated by comparing the consensus sequences between all five species separately for both loci. The levels of divergence are very similar (average pairwise nucleotide similarity between the five species is 79% for the polyubiquitin gene repeat vs. 76% for the rDNA repeat). Therefore, the overall level of sequence constraint in these two loci is very similar, but the intra-genomic variation is markedly different, demonstrating that the within-array evolutionary forces of these two loci differ greatly.

Polymorphisms exist beneath the radar of selection

We next looked to see how the polymorphisms we detected are distributed with respect to functional constraint. If selection acts directly on rDNA mutations, we would expect to preferentially find polymorphisms in rDNA residues that show low levels of constraint. Visual inspection of the positions of the polymorphisms (Fig. 2) reveals no obvious pattern of localization. Importantly, there is no bias of polymorphisms toward the IGS regions, which are the least selectively constrained regions of the rDNA. However, it is still possible that the polymorphisms found in the rDNA genes mostly fall on residues with low selective constraint. To test this, we used the LSU and SSU variability maps from the European Ribosomal RNA Database (Van de Peer et al. 1997; Ben Ali et al. 1999). These maps categorize each site in the rRNA gene-coding regions into one of six “bins” based on the level of conservation of that site across the eukaryotes. If the polymorphisms we observe are influenced by selection, we would expect a bias of these polymorphisms toward residues with low levels of conservation. We mapped every high-quality polymorphism from the rRNA gene-coding regions onto these variability maps. Pooling the results from all five species demonstrates that the numbers of polymorphisms found in each conservation “bin” are not significantly different from the expected distribution ($P > 0.95$; Table 2). This is also true ($P > 0.6$) if only polymorphisms present on more than one read are used, although the numbers are too low for statistical power (results not shown). Thus, there is no bias of these polymorphisms toward more variable sites in the rRNA genes, and we can conclude that at least most of the polymorphisms are not influenced by selective con-

straint. Finally, the most frequently found polymorphism in the *C. neoformans* data set (found in 10 reads) is a substitution in a site that is invariant across the fungi (using the data set of Berbee and Taylor 2001) (<http://www.treebase.org/treebase/>). Therefore, this polymorphism is likely to be highly deleterious, yet it seems to be present in more than one rDNA unit, indicating that polymorphisms present at relatively low frequencies are beneath the “radar” of selection, presumably due to redundancy of the rDNA (Hadjiolov 1984).

Mutational spectrum of the rDNA polymorphisms

If the polymorphisms we observe are appearing beneath the level of selection, they may give us insight into the spectrum of mutations offered by the mutation process. Therefore, we looked at the types of mutations that have arisen. Polymorphisms were divided into five classes of mutation (transitions, transversions, insertions, deletions, and complex mutations), and the proportions of these for each species (except *A. gossypii* and *S. cerevisiae*, as there are too few polymorphisms) are presented (for high-confidence polymorphisms) (Fig. 4). Each species has its own unique profile of mutation, although the polymorphism levels in *S. paradoxus* and *A. nidulans* are too low for strong conclusions. However, we find, as expected, that transitions are the most common form of substitution in all species. The most striking pattern is that of *C. neoformans*, which shows a very strong deletion bias (62% of all polymorphisms are deletions). If deletion/insertion bias predicts the direction of genome size evolution (Mira et al. 2001), then *C. neoformans* may be in the process of a genome size reduction, although current gene density is not especially high (Loftus et al. 2005).

C. neoformans is also characterized by several highly complex mutations. These mutations typically involve replacement of one large sequence tract (10–19 bp) with a similar-sized tract of unrelated sequence (10–28 bp). We speculated that these mutations were the result of sequence substitution mutations (Yoshiyama et al. 2001). However, complementary sequences near the inserted tracts were not found, ruling this explanation out. Attempts were made to detect two of these complex mutations by PCR, but the correct product was not detected (results not

Table 2. Variability level of the sites in which polymorphisms are found in rRNA genes

Variability “bin” ^a	Number of polymorphisms ^b	%	% of total bases ^c
Red	6	13.3%	11.1%
Orange	9	20.0%	20.0%
Yellow	7	15.6%	15.6%
Green	6	13.3%	13.3%
Blue	6	13.3%	17.8%
Purple	11	24.4%	22.2%

^aVariability “bins” are those from the LSU and SSU rRNA variability maps in Ben Ali et al. (1999) and Van de Peer et al. (1997), respectively. Red are the most variable, and purple the most highly conserved sites (Van de Peer et al. 1996).

^bIncludes all polymorphisms in the 18S, 28S, and 5.8S rRNA genes from all five species pooled together. Complex mutations (involving more than three nucleotides) and insertions were not included. Only high-confidence polymorphisms were used.

^cPercent of rRNA gene nucleotides expected for this sample size in each variability “bin,” calculated by summing all nucleotides found in each “bin” from the LSU and SSU gene variability maps of *Saccharomyces cerevisiae*. The observed numbers are not significantly different from the expected using the χ^2 test ($P > 0.95$).

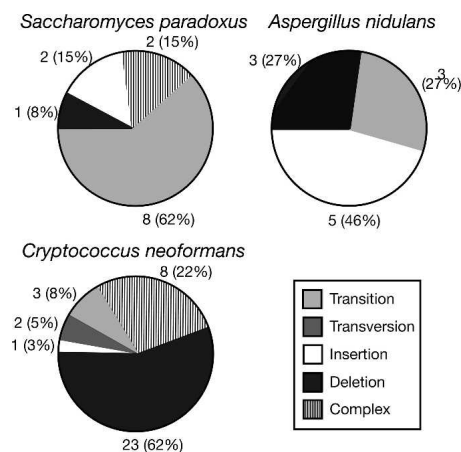


Figure 4. Mutational profile of the rDNA polymorphisms. The proportion of each of the five classes of mutation (listed in the boxed legend) that form the high-confidence polymorphisms are graphed for *S. paradoxus*, *A. nidulans*, and *C. neoformans*. Absolute numbers and percentages are given for each class. “Complex” mutations are defined as those involving more than 3 bp. For the full list of polymorphisms, see Supplemental Table 1.

shown). Therefore, these complex mutations are possibly artifacts of some kind.

Discussion

This study has given us the first quantitative picture of the level of polymorphism present within rDNA arrays. Combining the levels of polymorphism with rDNA copy number data, it is clear the level of variation across the rDNA arrays in all five fungal species is extremely low. Indeed, in each species many repeats must be identical in sequence across the entire unit (~7.5–9 kb of sequence), as the total number of polymorphisms is less than the rDNA copy number in every species. We also show that this level of variation is orders of magnitude lower than that of a repeat family undergoing birth-and-death evolution. Furthermore, there is no bias of the few observed polymorphisms to areas of low selective constraint, and polymorphisms present in a few copies seem to exist beneath the level of selection. Together these results demonstrate that the rDNA is evolving via concerted evolution, rather than birth-and-death evolution, and suggest that homogenization is highly efficient at maintaining the rDNA with near-identical repeats.

At first glance it may seem paradoxical that variation within the rDNA is very low when some regions of the rDNA (notably the IGS) evolve very rapidly. However, these features are easily reconciled under a model of rapid homogenization (Fig. 5). In this model there are three phases of homogenization in a hypothetical repeat array. In the first phase (mutation), mutations can occur stochastically anywhere within the repeat unit. The rDNA is highly redundant, so no selective pressure acts on these “unique” mutations and they can persist for some time. Indeed, most polymorphisms we observe seem to fall into this class: low-frequency polymorphisms that are located randomly throughout the repeat unit, irrespective of the level of constraint. In the second phase (transition), continual repeat turnover by homogenization (unequal recombination) results in mutated repeats being either deleted or duplicated (again stochastically). Deletion obviously removes that mutation from the array (homogenizes

the array). Duplication starts the mutated repeat on a process where it may increase in copy number through successive duplications. This is where natural selection comes in. A deleterious mutation will only be able to increase in copy number up to a certain threshold, above which the mutation will compromise fitness. Therefore, only mutations tolerated by natural selection can increase to high copy numbers in the array. Interestingly, the three high copy-number polymorphisms found in this study (two from *S. cerevisiae* and one from *S. paradoxus*) (Fig. 2) are present in the IGS and ITS, the regions of the rDNA with the lowest selective constraint, fully consistent with them being neutral variants that are being spread by homogenization. All other polymorphisms are present at low copy number, and therefore are likely to have arisen recently or are unable to spread to high copy number because of functional constraint. The final phase is fixation, where a “tolerated” mutant repeat completely replaces the previous repeats. Thus a new, variant sequence becomes homogenized in the array.

A rapid homogenization model of concerted evolution is consistent with the results from numerous previous studies (Liao et al. 1997; Ganley and Scott 1998, 2002; Skalicka et al. 2003; Averbeck and Eickbush 2005; Kovarik et al. 2005). Furthermore, rDNA polymorphisms between individuals in a population (e.g., Carbone and Kohn 2001; James et al. 2001; Ganley et al. 2005) are also evidence for rapid homogenization. Our results suggest that individuals are likely to have homogenized arrays, therefore, polymorphisms between individuals in a population represent cases where homogenization has spread a mutation to all the repeats in the array of one individual, creating a fixed polymorphism between individuals. This homogenization must have occurred in the evolutionary time separating the two polymorphic individuals. Other studies also support the existence of a threshold rDNA copy number, above which deleterious mutations are not tolerated. A notable case is the R1/R2 insertion elements in *Drosophila*, retrotransposons that insert into and inactivate the LSU. There is a dramatic level of turnover of these elements in the

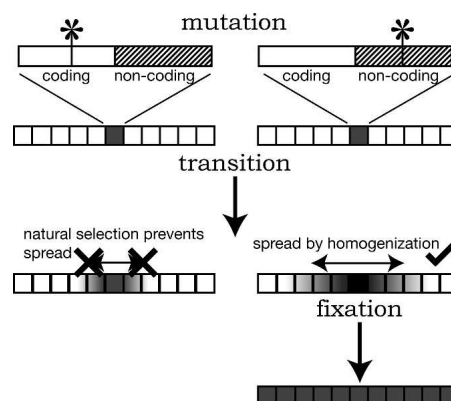


Figure 5. Three phases of repeat homogenization under a rapid homogenization model. First, a mutation occurs at either a selectively constrained (e.g., a coding part of the repeat), or a nonselectively constrained (e.g., a noncoding part of the repeat) site in a single unit from the stylized array. In the transition phase, only the unit with the nonselectively constrained mutation can increase to high copy number by homogenization. This mutation is able to sweep to fixation in the array. Thus, only mutations tolerated by selection can spread throughout the array, explaining why within the same repeat some regions are highly polymorphic while others are highly conserved, even though the entire repeat unit is subject to the identical homogenization process. See text for details.

rDNA, and this is consistent with inserted units being unable to achieve high copy number (because selection keeps the number of inactive units beneath a certain threshold), yet the elements are able to avoid extinction at the hand of homogenization through continual retrotransposition (e.g., Hollocher and Templeton 1994; Pérez-González and Eickbush 2002; Averbach and Eickbush 2005). Also, it has long been known that deletions of large numbers of rDNA units (a dramatic form of inactivation) are still viable (e.g., Ritossa et al. 1966; Russell and Rodland 1986; Takeuchi et al. 2003), and more minor copy-number variation in tandemly repeated rDNA arrays is routinely observed. Conversely, a repeat unit with a disrupted coding region was recently found at high frequencies in the 5S rRNA gene repeats in *Trypanosoma cruzi* (Westenberger et al. 2006); it is not clear why these seemingly deleterious repeats are maintained at high copy numbers.

One unusual feature of our data, as noted in the Results section, is that the majority of polymorphisms are found in only one read. Given that these genomes were sequenced to multiple levels of coverage, we expect “true” polymorphisms to be present in multiple reads at a frequency similar to the coverage level (i.e., for a threefold coverage level, we expect to find single-copy polymorphisms in approximately three reads). We suspect that these are the result of unknown sequencing errors. The possibility that these polymorphisms result from heavily methylated, inactive rDNA copies that are biased against in the genomic libraries, and which contain many mutations, was ruled out by digestion of *S. paradoxus*, *A. nidulans*, and *C. neoformans* genomic DNA with methylation sensitive/insensitive isoschizomers. No obvious rDNA methylation was revealed in any of these species (results not shown). Another possibility is that these polymorphisms are real and are the result of cell-to-cell variation due to ongoing mutation in the rDNA during cell growth, resulting in polymorphisms between cells in a colony. In either case, only counting polymorphisms present on more than one read will be a more realistic measure of total array variation.

No fixed (allelic) differences were found between the homologous arrays in *S. paradoxus*, even though fixed differences are expected in diploid organisms with low rates of sexual recombination (K. Klein, pers. comm.). We did observe one high-frequency IGS polymorphism in this species, but the polymorphism is present equally on both homologous rDNA arrays (see Supplemental Fig. 1). Although this is consistent with the idea of the cohesive spread of variants through populations by concerted evolution (Ohta and Dover 1984), we think it more likely results from a high frequency of intra-ascus mating and/or auto-diploidization (Johnson et al. 2004). High levels of inbreeding will severely limit heterozygous rDNA arrays.

Some differences between the five species were observed, particularly the high proportion of polymorphisms found on multiple reads in *S. paradoxus*. However, if single-read polymorphisms are excluded, the results for all five species are similar. All five species have a single rDNA array, and it will be interesting to see what pattern emerges from species with more complicated rDNA organizations. Also, sex in these species is thought to be rare or absent, and they all have a history of vegetative lab cultivation, features that will limit variation (K. Klein, pers. comm.). Finally, variation in the structural arrangement of repeat units has been detected within some species (Caburet et al. 2005), but we do not yet know how widespread this phenomenon is.

In summary, this study shows that the level of intragenomic variation in the rDNA arrays of these five fungal species is extremely low. This low level of variation provides a clear distinc-

tion between repeats evolving by concerted evolution and those evolving by birth-and-death evolution, and suggests that concerted evolution is very dynamic and efficient. The WGSS data methodology used here can be applied to other repeat arrays as well as to the rDNA of other species. Furthermore, the contrast between the variation observed in repeats undergoing concerted evolution versus repeats undergoing birth-and-death evolution provides a framework for determining the evolutionary behavior of other repeat families. These results are reassuring for researchers using the rDNA for phylogenetic purposes, as there is no evidence for major cryptic variation within the repeats, although the situation may be different in species with more complicated rDNA arrangements (e.g., multiple loci), and different life-history traits (e.g., higher rates of sexual recombination). Extending these analyses to more and varied species will help clarify this. Our results also provide critical data for the formulation of theoretical models of concerted evolution, as now both the rate of rDNA recombination (e.g., Szostak and Wu 1980; Gangloff et al. 1996) and the level of rDNA array heterogeneity are known. Finally, although our results seem to justify the decision of some genome projects not to sequence long repeat arrays such as the rDNA, such data are useful and can help us decipher the evolutionary dynamics of these intriguing regions of the genome.

Methods

Identifying polymorphisms

rDNA sequence reads were either obtained directly from the genome sequencing center (*A. gossypii* and *C. neoformans*) or for (*S. paradoxus* and *A. nidulans*) identified as follows. First, complete rDNA unit sequences were constructed by taking a portion of the rDNA (from GenBank), BLASTing (Altschul et al. 1997) this to the genome sequence using the NCBI Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>) and using the results to “walk” out in both directions until a complete unit was obtained. For *S. cerevisiae*, GenBank sequence U53879 was used. The complete unit was then used in a subsequent BLAST back to the genome sequence, with the searches performed conservatively to identify all rDNA reads, and these were downloaded from the NCBI Trace Archive. The consensus rDNA unit sequences for *S. paradoxus* and *C. neoformans* are deposited in the DNA Data Bank of Japan (accession nos. BR000309 and BR000310), and those of *S. cerevisiae* and *A. nidulans* are given in the Supplemental material. The *A. gossypii* sequence already exists (accession no. AF113137). Details for each species’ rDNA reads are as follows: *A. gossypii*, 2658 reads for ~3.3-fold coverage of the rDNA; *S. cerevisiae*, 10,764 reads for ~5.4-fold coverage; *S. paradoxus*, 5571 reads for ~1.7-fold coverage (diploid coverage level); *A. nidulans*, 4996 reads for ~7.7-fold coverage; and *C. neoformans*, 4206 reads for ~4.7-fold coverage. rDNA coverage level was calculated using total rDNA array length (from rDNA copy-number results) and the total rDNA read length (obtained by multiplying rDNA read number by average read length). Average read length was obtained either directly from the genome sequencing center or calculated from total read number, sequenced genome length, and coverage level. This was 500 bp for *A. gossypii*, *S. cerevisiae*, and *S. paradoxus*, and 534 bp for *A. nidulans*. The value for *C. neoformans* is not known, but we used 500 bp as an estimate. rDNA coverage level calculated in this way was consistently lower than overall genome coverage level, but this is not surprising as regions of unusual chromatin structure such as centromeres and telomeres are usually underrepresented in genomic libraries (e.g., Mais et al. 2005). The value is particularly low for *S. paradoxus* because this

is diploid rDNA coverage level, whereas haploid coverage level is normally quoted for whole genomes.

rDNA reads from each genome were assembled into a single unit with a high density of sequence reads using the Phred/Phrap/Consed software cluster (<http://www.phrap.org/phredphrapconsed.html>). Due to the large amount of data, this often required assemblies of subsets of the data. Any minor contigs formed were manually checked, and any reads that were real rDNA reads were subsequently incorporated into the main contig. Polymorphisms were then automatically identified using Consed, and every polymorphic chromatogram was checked by eye. From this, “true” polymorphisms were identified and put subjectively into two classes: high-confidence, and low-confidence polymorphisms. Reads containing polymorphisms are presented in Supplemental Table 1.

Molecular biology techniques

Strains used for molecular analyses in this study were *S. cerevisiae* RM11-1A, *S. paradoxus* NRRL Y-17217, *A. nidulans* FGSC-A4, and *C. neoformans* B-3501A, the sequenced strains obtained from the respective genome centers. Routine growth of *S. paradoxus* and *C. neoformans* was performed in YPD (1% yeast extract, 2% peptone, 2% glucose), while *A. nidulans* was grown on 5YEG (0.5% yeast extract; 1% glucose) plates. Genomic DNA extractions from *C. neoformans* were performed based on the method of Hoffman and Winston (1987). Briefly, cells suspended in Winston buffer with glass beads were first frozen at -20°C and then disrupted by shaking after addition of phenol/chloroform using a cell-disrupter (Multi-Bead Shocker, Yasui Kikai) with a 30-sec shake followed by a 60-sec rest for 12 cycles. The remainder of the procedure is identical.

Determination of rDNA copy number

Preparation of chromosomal DNA in agarose plugs for CHEF analyses was performed as follows. For *S. paradoxus*, CHEF plugs were prepared as previously described using 2 mg/mL of Zymolase 20-T (Seikagaku) and a 6-h digestion (Birren et al. 1997). The method of Yelton et al. (1984) was used to prepare protoplasts from *A. nidulans*, with the following modifications: $\sim 5 \times 10^7$ conidia were inoculated into 100 mL of minimal medium (Scott and Kafer 1982) and grown at 37°C for 24 h; OM buffer contained 1.0 M MgSO_4 ; mycelia were digested using Yatalase (20 mg/mL; Takara) and Kitalase (5 mg/mL; Wako); and digestion was performed for 6 h at 30°C with shaking at 150 rpm. Preparation of CHEF plugs from these protoplasts then followed that of *S. paradoxus* (Birren et al. 1997). CHEF plugs from *C. neoformans* were prepared according to the method of Smith et al. (1988), except that 100 mg/mL of lysing enzyme (Sigma) for 1 h at 37°C were used for protoplast formation. For restriction enzyme in-plug digestion, CHEF plugs were first washed three times in TE (10/1) buffer for 20 min each, followed by three washes in 300 μL of restriction enzyme buffer solution using the appropriate manufacturer's buffer (for *Hin*DIII and *Bam*HI, Toyobo; for *Age*I, Promega) at 37°C for 20 min each. This was followed by incubation with 50–100 U of enzyme/half agarose plug and 0.01% BSA in 500 μL reaction volume for 3–16 h at 37°C , using either *Bam*HI (*S. paradoxus*), *Hin*DIII (*A. nidulans*), or *Age*I (*C. neoformans*). rDNA-containing fragments were separated on 1% PFC agarose (BioRad) gels using the CHEF Mapper (BioRad) with the following parameters: 6.0 V/cm; 0.2–204-sec ramped pulsed time; 15.2 h; $0.5 \times$ TBE buffer; and 14°C . Southern hybridizations were performed using standard procedures (Sambrook and Russell 2001) and rDNA bands were detected using a probe from a conserved region of the LSU derived from *S. cerevisiae*.

Nucleotide diversity

Sequences of the polyubiquitin repeat (poly-u repeats) arrays (after removal of introns; monomer ubiquitin sequences were not used; for *S. paradoxus* homologous diploid arrays were used) are deposited in the DNA Data Bank of Japan (accession nos. BR000305 to BR000308) or for *S. cerevisiae* can be found in the Supplemental material. Nucleotide diversity values (π) were calculated using the program DnaSP-4 (Rozas et al. 2003). rDNA arrays were constructed from rDNA copy number and high-confidence polymorphism number data after correcting for genome coverage level.

Acknowledgments

Particular thanks are owed to Fred Dietrich (Duke Univ. Medical Center) for the inspiration for this project, and for providing access to the *A. gossypii* genome sequence data. We also thank Eula Fung (Stanford Genome Tech. Center) for providing the *C. neoformans* rDNA reads and for helpful discussions; Alexey Egorov (NCBI Trace Archive) for retrieving the rDNA reads for *A. nidulans* and *S. paradoxus*; and Tim James and Rytas Vilgalys (Duke Univ.), and Takashi Horiuchi (Natl. Inst. Basic Biol.) for helpful discussions. Thanks are also due to K. Klein (Minnesota State Univ.) for helpful comments on the manuscript, and the communication of unpublished results. This work was supported by grants from the Clark Fellowship for Molecular Evolution and Comparative Genomics and the Japan Society for the Promotion of Science to ARDG, grants 17080010, 17370065, and 18207013 from the Ministry of Education, Science and Culture, Japan, and by a Human Frontier Science Program grant to T.K.

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Andersson, D.I., Slechta, E.S., and Roth, J.R. 1998. Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science* **282**: 1133–1135.
- Averbeck, K.T. and Eickbush, T.H. 2005. Monitoring the mode and tempo of concerted evolution in the *Drosophila melanogaster* rDNA locus. *Genetics* **171**: 1837–1846.
- Ben Ali, A., Wuyts, J., De Wachter, R., Meyer, A., and Van de Peer, Y. 1999. Construction of a variability map for eukaryotic large subunit ribosomal RNA. *Nucleic Acids Res.* **27**: 2825–2831.
- Berbee, M.L. and Taylor, J.W. 2001. Fungal molecular evolution: Gene trees and geologic time. In *The Mycota* (eds. D.J. McLaughlin, E. McLaughlin, and P.A. Lemke), pp. 229–246. Springer, Berlin.
- Birren, B., Green, E.D., Klapholz, S., Myers, R.M., and Roskams, J. 1997. Analyzing DNA. In *Genome analysis: A laboratory manual*. (eds. E.D. Green, B. Birren, S. Klapholz, R.M. Myers, and P. Hieter), pp. 125–127. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Britten, R.J. and Kohne, D.E. 1968. Repeated sequences in DNA. *Science* **161**: 529–540.
- Brown, D.D., Wensink, P.C., and Jordan, E. 1972. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: The evolution of tandem genes. *J. Mol. Biol.* **63**: 57–73.
- Caburet, S., Conti, C., Schurra, C., Lebofsky, R., Edelstein, S.J., and Bensimon, A. 2005. Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome Res.* **15**: 1079–1085.
- Carbone, I. and Kohn, L.M. 2001. A microbial population-species interface: Nested cladistic and coalescent inference with multilocus data. *Mol. Ecol.* **10**: 947–964.
- Copenhaver, G.P. and Pikaard, C.S. 1996. Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* **9**: 273–282.
- Cowen, L.E., Sanglard, D., Calabrese, D., Sirjusingh, C., Anderson, J.B.,

- and Kohn, L.M. 2000. Evolution of drug resistance in experimental populations of *Candida albicans*. *J. Bacteriol.* **182**: 1515–1522.
- Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pöhlmann, R., Luedi, P., Choi, S., et al. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**: 304–307.
- Dover, G.A. 1982. Molecular drive: A cohesive mode of species evolution. *Nature* **299**: 111–117.
- Gangloff, S., Zou, H., and 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, A.R.D. and Scott, B. 1998. Extraordinary ribosomal spacer length heterogeneity in a Neotyphodium endophyte hybrid: Implications for concerted evolution. *Genetics* **150**: 1625–1637.
- Ganley, A.R.D. and Scott, B. 2002. Concerted evolution in the ribosomal RNA genes of an *Epichloë* endophyte hybrid: Comparison between tandemly-arranged rDNA and dispersed 5S *rrn* genes. *Fungal Genet. Biol.* **35**: 39–51.
- Ganley, A.R.D., Hayashi, K., Horiuchi, T., and Kobayashi, T. 2005. Identifying gene-independent noncoding functional elements in the yeast ribosomal DNA by phylogenetic footprinting. *Proc. Natl. Acad. Sci.* **102**: 11787–11792.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274**: 546–567.
- Gonzalez, I.L. and Sylvestre, J.E. 2001. Human rDNA: Evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics* **73**: 255–263.
- Hadjiolov, A.A. 1984. The nucleolus and ribosome biogenesis. In *Cell biology monographs* (eds. M. Alfert et al.), Vol. 12. Springer-Verlag Wien, New York.
- Hoffman, C.S. and Winston, F. 1987. A ten-minute preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**: 267–272.
- Hollocher, H. and Templeton, A.R. 1994. The molecular through ecological genetics of *abnormal abdomen* in *Drosophila mercatorum*. VI. The non-neutrality of the Y chromosome rDNA polymorphism. *Genetics* **136**: 1373–1384.
- James, T.J., Moncalvo, J.-M., Li, S., and Vilgalys, R. 2001. Polymorphism at the ribosomal DNA spacers and its relation to breeding structure of the widespread mushroom *Schizophyllum commune*. *Genetics* **157**: 149–161.
- Johnson, L.J., Koufopanou, V., Goddard, M.R., Hetherington, R., Schäfer, S.M., and Burt, A. 2004. Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* **166**: 43–52.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- Kobayashi, T., Heck, D.J., Nomura, M., and Horiuchi, T. 1998. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: Requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes & Dev.* **12**: 3821–3830.
- Kovarik, A., Pires, J.C., Leitch, A.R., Lim, K.Y., Sherwood, A.M., Matyasek, R., Rocca, J., Soltis, D.E., and Soltis, P.S. 2005. Rapid concerted evolution of nuclear ribosomal DNA in two Tragopogon allopolyploids of recent and recurrent origin. *Genetics* **169**: 931–944.
- Liao, D., Pavelitz, T., Kidd, J.R., Kidd, K.K., and Weiner, A.M. 1997. Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J.* **16**: 588–598.
- Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A., et al. 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**: 1321–1324.
- Long, E.O. and Dawid, I.B. 1980. Repeated genes in eukaryotes. *Annu. Rev. Biochem.* **49**: 727–764.
- Mais, C., Wright, J.E., Prieto, J.-L., Raggett, S.L., and McStay, B. 2005. UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. *Genes & Dev.* **19**: 50–64.
- Mira, A., Ochman, H., and Moran, N.A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**: 589–596.
- Nei, M. and Li, W.-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* **76**: 5269–5273.
- Nei, M., Gu, X., and Sitnikova, T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci.* **94**: 7799–7806.
- Nei, M., Rogozin, I.B., and Piontkivska, H. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl. Acad. Sci.* **97**: 10866–10871.
- O'Donnell, K. and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. *Mol. Phylogenet. Evol.* **7**: 103–116.
- Ohta, T. and Dover, G.A. 1984. The cohesive population genetics of molecular drive. *Genetics* **108**: 501–521.
- Pérez-González, C.E. and Eickbush, T.H. 2002. Rates of R1 and R2 retrotransposition and elimination from the rDNA locus of *Drosophila melanogaster*. *Genetics* **162**: 799–811.
- Prokopowich, C.D., Gregory, T.R., and Crease, T.J. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46**: 48–50.
- Ritossa, F.M., Atwood, K.C., and Spiegelman, S. 1966. A molecular explanation of the bobbed mutants of *Drosophila* as partial deficiencies of “ribosomal DNA.” *Genetics* **54**: 819–834.
- Rooney, A.P. and Ward, T.J. 2005. Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. *Proc. Natl. Acad. Sci.* **102**: 5084–5089.
- Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Russell, P.J. and Rodland, K.D. 1986. Magnification of rRNA gene number in a *Neurospora crassa* strain with a partial deletion of the nucleolus organizer. *Chromosoma* **93**: 337–340.
- Sambrook, J. and Russell, D.W. 2001. *Molecular cloning: A laboratory manual*, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Scott, B.R. and Kafer, E. 1982. *Aspergillus nidulans*: An organism for detecting a range of genetic damage. In *Chemical mutagens* (eds. F.J. deSerres and A. Hollaender), Vol. 7, pp. 447–479. Plenum, New York.
- Shapiro, J.A. and von Sternberg, R. 2005. Why repetitive DNA is essential to genome function. *Biol. Rev. Camb. Philos. Soc.* **80**: 227–250.
- Skalicka, K., Lim, K.Y., Matyásek, R., Koukalová, B., Leitch, A.R., and Kovarik, A. 2003. Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. *Am. J. Bot.* **90**: 988–996.
- Smith, G.P. 1973. Unequal crossover and the evolution of multigene families. *Cold Spring Harb. Symp. Quant. Biol.* **38**: 507–513.
- Smith, C.L., Klcó, S.R., and Cantor, C.R. 1988. Pulsed field gel electrophoresis and the technology of large DNA molecules. In *Genome analysis: A practical approach* (ed. K.E. Davies), pp. 41–72. IRL Press, Oxford.
- Szostak, J.W. and Wu, R. 1980. Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* **284**: 426–430.
- Takeuchi, Y., Horiuchi, T., and Kobayashi, T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes & Dev.* **17**: 1497–1506.
- Van de Peer, Y., Chapelle, S., and De Wachter, R. 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res.* **24**: 3381–3391.
- Van de Peer, Y., Jansen, J., De Rijk, P., and De Wachter, R. 1997. Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Res.* **25**: 111–116.
- Wendland, J., Pöhlmann, R., Dietrich, F., Steiner, S., Mohr, C., and Philippsen, P. 1999. Compact organization of rRNA genes in the filamentous fungus *Ashbya gossypii*. *Curr. Genet.* **35**: 618–625.
- Westenberger, S.J., Sturm, N.R., and Campbell, D.A. 2006. *Trypanosoma cruzi* 5S rRNA arrays define five groups and indicate geographic origins of an ancestor of the heterozygous hybrids. *Int. J. Parasitol.* **36**: 337–346.
- Williams, S.M., Robbins, L.G., Cluster, P.D., Allard, R.W., and Strobeck, C. 1990. Superstructure of the *Drosophila* ribosomal gene family. *Proc. Natl. Acad. Sci.* **87**: 3156–3160.
- Yelton, M.M., Hamer, J.E., and Timberlake, W.E. 1984. Transformation of *Aspergillus nidulans* by using a *trpC* plasmid. *Proc. Natl. Acad. Sci.* **81**: 1470–1474.
- Yoshiyama, K., Higuchi, K., Matsumura, H., and Maki, H. 2001. Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in *Escherichia coli*. *J. Mol. Biol.* **307**: 1195–1206.
- Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W., and Wilson, A.C. 1980. Rapid duplication and loss of genes coding for the α chains of hemoglobin. *Proc. Natl. Acad. Sci.* **77**: 2158–2162.

Received May 1, 2006; accepted in revised form November 16, 2006.